

Einführung in die Faktorenanalyse

mit einer Einführung in die Matrixrechnung

Skriptum zu den Vorlesungen
Evaluation und Forschungsmethoden (Statistik III + IV)

U. Mortensen

Fachbereich Psychologie und Sportwissenschaften, Institut III
Westfälische Wilhelms-Universität
Fliegenerstr. 21

Letzte Änderung: 14. 06. 2013

Ich danke Herrn Dr. Hans Stocker für seine Hinweise auf (Tipp-)Fehler und unverständliche Sätze; die vorgenommenen Verbesserungen sind Resultate seiner Sorgfalt.

Inhaltsverzeichnis

1	Grundgedanken der Faktorenanalyse	4
1.1	Kurze Geschichte der Faktorenanalyse	4
1.2	Der allgemeine Ansatz	5
1.3	Messwerte als Funktion latenter Variablen	7
2	Vektoren, Matrizen, und latente Variablen	16
2.1	Vektoren	16
2.1.1	Latente Variablen, Vektoren, und Linearkombinationen	16
2.1.2	Skalarprodukte	20
2.1.3	Vektorlänge und Normierung	21
2.1.4	Das Skalarprodukt und der Winkel zwischen Vektoren .	22
2.1.5	Korrelation und Skalarprodukt	25
2.1.6	Vektorräume, Basisvektoren und latente Variablen . . .	26
2.2	Matrizen, Eigenvektoren und Ellipsoide	36
2.2.1	Definition einer Matrix	36
2.2.2	Multiplikation mit einem Skalar und Addition von Matrizen	37
2.2.3	Das Produkt von Matrizen	37
2.2.4	Zentrierung und Standardisierung	40
2.2.5	Der Rang einer Matrix	43
2.2.6	Symmetrische Matrizen	44
2.2.7	Die Einheitsmatrix und die inverse Matrix	46
2.2.8	Die Transformation von Vektoren	47
2.2.9	Die Rotation als Transformation	48
2.2.10	Eigenvektoren	49
2.2.11	Quadratische Formen und Ellipsoide	53
3	Hauptachsentransformation und Faktorenanalyse	56
3.1	Die Hauptachsentransformation	56
3.1.1	Die Singularwertzerlegung	61
3.1.2	Faktorladungen und Faktorwerte	64
3.1.3	Formale Aspekte der Interpretation	69
3.1.4	Faktorladungen als Korrelationen	71
3.1.5	Die Approximation von Z	72

3.1.6	Die Beziehung zur Hauptachsentransformation	73
3.1.7	Eine Anwendung auf die Regressionsrechnung	73
3.2	Faktorenanalyse	75
3.2.1	Vorbemerkung: zufällige Vektoren	75
3.2.2	Das Modell der Faktorenanalyse	75
3.2.3	Die Hauptkomponentenanalyse	78
3.2.4	Hauptkomponenten versus Faktoren	78
3.2.5	Der Biplot	81
3.2.6	Die Mahalanobis-Distanz	83
3.2.7	Die Mahalanobis-Distanz und die Distanz zwischen Personen	88
3.2.8	Die multivariate Normalverteilung	90
3.2.9	Beispiele	91
3.3	Die Hauptfaktorenanalyse	101
3.4	Die Schätzung der Kommunalitäten	103
3.5	Image-Analyse	104
3.6	Faktorentransformationen	105
4	Typen von Analysen	109
5	Dichotome Variable und nichtlineare Faktorenanalyse	111
5.1	Dichotome Variable I	111
5.2	Dichotome Variable II	116
5.3	Nichtlineare Modelle	118
5.4	Latent-Class-Modelle	121
6	Anhang	123
6.1	Eine alternative Herleitung	123
6.2	Lineare und statistische Unabhängigkeit	125
6.3	Koordinatenrotation	127
6.3.1	Ansatz I	127
6.3.2	Ansatz II	128
6.4	Extrema von Funktionen mit Nebenbedingungen	130
6.5	Herleitung der 2-dimensionalen Normalverteilung	133

Hinweis: Am Ende des Skripts findet man einen Index.

1 Grundgedanken der Faktorenanalyse

1.1 Kurze Geschichte der Faktorenanalyse

Die Ursprünge der Faktorenanalyse liegen in Versuchen, die Intelligenz zu messen, und scheinen auf Karl Pearson (1901) zurückzugehen, wenn auch Charles Spearman oft als Begründer dieser Methode genannt wird; den ebenso interessanten wie beklemmenden Hintergrund dieser Vermessungsbemühungen findet man in Stephen Jay Gould's Buch *Der falsch vermessene Mensch*, wo man auch eine ebenso nicht-mathematische wie informative Einführung in die Faktorenanalyse findet (p. 265). Es ging zunächst darum, die Existenz oder Nichtexistenz eines allgemeinen Fähigkeitsfaktors g nachzuweisen und Methoden zu seiner Berechnung zu finden. Hotelling (1933) schlug die Hauptachsen-Transformation, wie sie in diesem Skript im Zentrum der Betrachtungen steht, vor, auch wenn sie hier anders hergeleitet wird (vergl. aber den Ansatz, die Summe der Quadrate der Koordinaten auf der neuen Achse, wie sie in Gleichung (234) auf Seite 79) definiert wird, zu maximieren. Thurstone (1931) trug dann wesentlich zur Verbreitung der Theorie der Multiplen Faktoren bei, in der die Spearmansche g -Theorie kritisiert wird; weiter schuf er mit der Centroid-Methode als einer Annäherung an die Hauptachsenmethode eine Möglichkeit zur numerischen Approximation eben der Hauptachsenmethode, und mit dem Begriff der Einfachstruktur schlug er einen plausiblen Ansatz zur Interpretation vor. Ob man allerdings die Faktorenanalyse als Königsweg zu einer Theorie des menschlichen Intellekts und der Persönlichkeit ansehen kann, muß wohl eher bezweifelt werden. Kelly (1940) hat den Zweck der Faktorenanalyse mit der Bemerkung

”There is no search for timeless, spaceless, populationless truth in factor analysis; rather, it represents a simple, straightforward problem of description in several dimensions of a definite group functioning in definite manners, and he who assumes to read more remote verities into the factorial outcome is certainly doomed to disappointment.”

charakterisiert (vergl. auch Harman (1967)). Die Einstellung zur Faktorenanalyse und verwandten Verfahren sollte eher pragmatisch sein. Idealerweise sollte man eine Theorie über den betrachteten Gegenstandsbereich (Intelligenz, Persönlichkeitsstörungen, etc) haben; man kann dann diskutieren, in welchem Sinne faktorenanalytische Ansätze diese Theorie approximieren. In der psychologischen Forschungspraxis geht man allerdings oft umgekehrt vor, d.h. man versucht, über die Faktorenanalyse induktiv zu einer Theorie zu gelangen. Solche Versuchen liegt implizit die Annahme zugrunde, dass Theorien durch eine Art kanonischer Struktur gekennzeichnet werden können, die durch das Postulat der Existenz latenter, additiv wirkender ”Faktoren” definiert werden können (es gibt auch Ansätze für nichtlinear wirkende Faktoren, die aber in der Praxis

nur eine geringe Rolle zu spielen scheinen). Warum eine solche Struktur psychologischen Prozessen unterliegen soll, wird kaum jemals explizit diskutiert. Vielfach repräsentieren die mit den gängigen Standardverfahren gemessenen psychologischen Variablen nur Aspekte eines komplexen, nichtlinearen dynamischen Systems, und die Charakterisierung der Interaktion dieser Variablen anhand der der Faktorenanalyse zugrunde liegenden linearen Regression kann im Prinzip nur eine erste Näherung darstellen.

1.2 Der allgemeine Ansatz

Vielfach erhält man bei empirischen Untersuchungen große Datenmengen, die interpretiert werden müssen. So werden bei z.B. Meinungsumfragen 1000 oder 2000 Personen befragt, wobei der Fragebogen 20 oder 30 Fragen umfassen kann. Bei EEG-Untersuchungen, bei denen etwa an 15 Positionen am Kopf in Abständen von Millisekunden Potenziale gemessen werden, fallen ebenfalls sehr viele Messungen an. Selbst wenn nur 100 Personen einen Fragebogen mit 20 Fragen vorgelegt bekommen, liegt die Anzahl der zu interpretierenden Antworten bei 2000. So wichtig jede einzelne Antwort ist, so sehr ginge man in dieser Datenflut unter, wollte man jede einzelne Antwort hermeneutisch bewerten. Das Ziel wird also sein, Hypothesen über mögliche Strukturen in diesen Daten zu bilden und zu überprüfen. Zunächst geht es darum, die Korrelationen zwischen den gemessenen Variablen (Fragen in einem Fragebogen, Potenziale an verschiedenen Skalenpositionen, etc) in einer systematischen Weise zu erklären. Die Faktorenanalyse kann unter Umständen eine solche Klärung liefern. Es werden zuerst das Modell der Faktorenanalyse und die diesem Modell zugrundeliegenden Annahmen vorgestellt. Die Parameter des Modells - die Faktorladungen der Variablen und die Faktorwerte der Personen - sind unbekannt und müssen aus den Daten geschätzt werden. Die Schätzung der Parameter wird in Abschnitt 2 vorgestellt.

Man mißt also die Variablen V_1, \dots, V_n ; dies sind, wie schon angedeutet, Fragen eines Tests oder Fragebogens, physiologische Messungen, Potenziale an Positionen des Kopfes, etc. Es wird im Allgemeinen angenommen, dass die Messungen X_j dieser Variablen Intervallskalenniveau haben, so dass Produkt-Moment-Korrelationen zwischen den V_j berechnet werden können; Verallgemeinerungen für dichotome und nominale Daten können allerdings ebenfalls behandelt werden. Hier liegt allerdings der Fokus zuerst auf Daten mit Intervallskalenniveau. Bei n Variablen können $\binom{n}{2} = n(n-1)/2$ Korrelationen berechnet werden. Der Übersichtlichkeit halber fasst man sie in einer Matrix R zusammen:

$$R = \begin{pmatrix} r_{11} & & & & & \\ r_{21} & r_{22} & & & & \\ r_{31} & r_{32} & r_{33} & & & \\ \vdots & \vdots & \vdots & & & \\ r_{n1} & r_{n2} & r_{n3} & \cdots & r_{nn} & \end{pmatrix}. \quad (1)$$

Da $r_{12} = r_{21}$, $r_{13} = r_{31}$, etc, ist nur die untere Hälfte der Matrix angeschrieben worden, da diese Hälfte bereits alle Informationen über die Korrelationen enthält. Allgemein gilt $r_{jk} = r_{kj}$; man sagt, die Matrix R sei symmetrisch. Korreliert man eine Messwertreihe mit sich selbst, d.h. bestimmt man r_{jj} , so wird man $r_{jj} = 1$ finden, so dass man in R auch $r_{11} = r_{22} = \dots = r_{nn} = 1$ setzen kann. In den tatsächlichen Berechnungen wird man i.a. auch von diesen Werten ausgehen, es sei aber darauf hingewiesen, dass die Korrelation auch Werte kleiner als eins sein kann. Denn wie aus der Konstruktion von Tests bekannt ist kann die Korrelation zwischen ein und derselben Variablen, die aber zu verschiedenen Zeitpunkten (an den selben Personen) gemessen wird, wegen des jeweilig auftretenden Meßfehlers kleiner als eins sein.

Die Anordnung der Korrelationen in einer Matrix ist zwar übersichtlich, aber wenn es um die Interpretation der Korrelationen geht, hilft auch eine übersichtliche Anordnung oft nur wenig. Bei $n = 3$ Variablen hat man $3 \cdot 2/2 = 3$ Korrelationen zu analysieren, bei $n = 4$ sind es bereits $4 \cdot 3/2 = 6$, bei $n = 10$ sind es schon $10 \cdot 9/2 = 45$. Der Wunsch nach einer systematischen "Erklärung" der beobachteten Korrelationen wird dann verständlich. Die Korrelation zwischen zwei Variablen V_j und V_k läßt sich u. U. durch die Annahme erklären, dass beide Variablen mindestens eine gemeinsame Variable erfassen. Die berühmte Korrelation zwischen der Anzahl der Alkoholiker und der Anzahl der Priester in den USA läßt sich durch die gesamtwirtschaftliche Lage erklären: ist sie schlecht, erhöht sich die Zahl der Arbeitslosen, von denen sich ein Teil in den Alkohol flüchtet und ein anderer Teil in die Priesterschaft. Partialisiert man diese "latente" Variable aus, so geht die Korrelation auf Null zurück. Die Korrelation zwischen den Lösungen verschiedener Denksportaufgaben läßt sich durch die Annahme erklären, dass die Aufgaben nur gelöst werden können, wenn bestimmte Fähigkeiten vorhanden sind: ein wenig Umgang mit Zahlen, die Fähigkeit, Informationen im Gedächtnis halten zu können, die Fähigkeit, Sachverhalte sprachlich formulieren zu können. Diese Fähigkeiten repräsentieren gewissermassen "latente" Variablen, die von den tatsächlich gemessenen Variablen erfasst werden und die den Korrelationen zugrunde liegen. Partialisiert man sie der Reihe nach aus den gemessenen Variablen heraus, werden die Korrelationen gegen Null gehen.

Man kann sagen, dass dies der Grundgedanke der Faktorenanalyse ist. Die Frage ist nun, wie die möglichen latenten Variablen bestimmt werden können. Die Möglichkeit, nach Maßgabe von Hypothesen Variable explizit zu messen und dann aus den V_j heraus zu partialisieren ist denkbar, aber natürlich völlig unpraktikabel. Die Idee ist also, die latenten Variablen aus den Korrelationen zwischen den V_j *herauszurechnen*, oder, wie auch gesagt wird, zu *extrahieren*. Dazu muß ein formales Modell aufgestellt werden. In einem solchen Modell werden die Annahmen über den Zusammenhang von gemessenen Variablen - den V_j - und den latenten Variablen spezifiziert. Aus diesen Annahmen ergeben sich dann die Möglichkeiten, die latenten Variablen explizit zu bestimmen.

1.3 Messwerte als Funktion latenter Variablen

Dazu werde die Korrelation $r(V_j, V_k) = r_{jk}$ betrachtet. Es werde angenommen, dass sie auf nur eine latente Variable zurückzuführen sei. Dann kann man für die Messwerte X_j und X_k die Gleichungen

$$x_{ij} = \alpha_{j1}F_{i1} + e_{ij} \quad (2)$$

$$x_{ik} = \alpha_{k1}F_{i1} + e_{ik} \quad (3)$$

anschreiben. Der erste Index i steht dabei für die i -te Person, an der sowohl der Wert von V_j als auch der Wert von V_k gemessen wird. α_{j1} ist als Regressionskoeffizient aufzufassen, der den Anteil bestimmt, mit dem die latente Variable F_1 in die Variable V_j eingeht, und α_{k1} ist der entsprechende Anteil, mit dem F_1 in V_k eingeht. Man bemerke, dass diese Regressionsgewichte, also α_{j1} und α_{k1} , spezifisch für die Variablen V_j und V_k sind, nicht aber für die i -te Person. Die Messwerte x_{ij} und x_{ik} für die i -te Person ergeben sich aus dem Wert F_{i1} , d.h. aus dem Wert für F_1 , den die i -te Person hat. Repräsentiert also F_1 die Fähigkeit zu logischem Denken, so ist F_{i1} der Messwert, den die i -te Person auf einer Skala hat, die die Fähigkeit zu logischem Denken erfasst. e_{ij} und e_{ik} sind die üblichen Messfehler.

Korrelationen und Standardisierung: Man kann nun die Korrelation r_{jk} aus den Gleichungen (2) und (3) voraussagen. Dazu wird man die X_j - und X_k -Werte standardisieren, d.h. man wird zu den z -Werten

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (4)$$

$$z_{ik} = \frac{x_{ik} - \bar{x}_k}{s_k} \quad (5)$$

übergehen. \bar{x}_j ist das arithmetische Mittel der X_j -Werte und s_j ist die zugehörige Standardabweichung, und \bar{x}_k und s_k sind die analogen Werte für X_k . Sicherlich ist¹

$$\bar{x}_j = \alpha_{j1}\bar{F}_1 + \bar{e}_j \quad (6)$$

$$\bar{x}_k = \alpha_{k1}\bar{F}_1 + \bar{e}_k \quad (7)$$

\bar{F}_1 ist der Mittelwert der F_{i1} -Werte. Dann ist z.B.

$$x_{ij} - \bar{x}_j = \alpha_{j1}(F_{ij} - \bar{F}_1) + (e_{ij} - \bar{e}_j). \quad (8)$$

Dividiert man durch s_j , so erhält man einen Ausdruck für z_{ij} :

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} = \frac{\alpha_{j1}}{s_j}(F_{ij} - \bar{F}_1) + \frac{e_{ij} - \bar{e}_j}{s_j}.$$

¹ $\bar{x}_j = \sum_i x_{ij}/m = \sum_i (\alpha_{j1}F_{i1} + e_{ij})/m = \alpha_{j1} \sum_i F_{i1}/m + \sum_i e_{ij}/m = \alpha_{j1}\bar{F}_1 + \bar{e}_j.$

Es wird sich als vorteilhaft erweisen, wenn man statt der Differenzen $F_{i1} - \bar{F}_1$ ebenfalls standardisierte Werte in den Gleichungen hat. Ist s_1^1 die Standardabweichung der F_{i1} -Werte, so kann man den folgenden "Trick" anwenden: man dividiert $F_{i1} - \bar{F}_1$ durch s_1^1 , so dass man einen standardisierten F_{i1} -Wert erhält, und multipliziert gleichzeitig mit s_j^1 , damit die Gleichung korrekt bleibt:

$$z_{ij} = \alpha_{j1} \frac{s_1^1 (F_{i1} - \bar{F}_1)}{s_j^1} + \frac{e_{ij} - \bar{e}_j}{s_j}.$$

Setzt man nun zur Abkürzung

$$a_{j1} = \alpha_{j1} \frac{s_1^1}{s_j^1}, \quad q_{i1} = \frac{F_{i1} - \bar{F}_1}{s_1^1}, \quad \varepsilon_{ij} = \frac{e_{ij} - \bar{e}_j}{s_j}, \quad (9)$$

so erhält man für z_{ij} die wesentlich übersichtlichere Gleichung

$$z_{ij} = a_{j1} q_{i1} + \varepsilon_{ij}. \quad (10)$$

In der gleichen Weise verfährt man bei der Standardisierung der X_k -Werte und erhält

$$z_{ik} = a_{k1} q_{i1} + \varepsilon_{ik}. \quad (11)$$

Man beachte, dass die q_{i1} -Werte ebenfalls standardisierte Werte sind, d.h. sie haben einen Mittelwert gleich Null und eine Varianz gleich 1. Die Korrelation zwischen den X_j - und den X_k -Werten läßt sich nun wie folgt anschreiben:

$$\begin{aligned} r_{jk} &= \frac{1}{m} \sum_{i=1}^m z_{ij} z_{ik} = \frac{1}{m} \sum_{i=1}^m (a_{j1} q_{i1} + \varepsilon_{ij})(a_{k1} q_{i1} + \varepsilon_{ik}) \\ &= \frac{1}{m} \sum_{i=1}^m a_{j1} a_{k1} q_{i1}^2 + \frac{1}{m} \sum_{i=1}^m a_{j1} q_{i1} \varepsilon_{ik} + \frac{1}{m} \sum_{i=1}^m a_{k1} q_{i1} \varepsilon_{ij} \\ &= a_{j1} a_{k1} \frac{1}{m} \sum_{i=1}^m q_{i1}^2 + a_{j1} \frac{1}{m} \sum_{i=1}^m q_{i1} \varepsilon_{ik} + a_{k1} \frac{1}{m} \sum_{i=1}^m q_{i1} \varepsilon_{ij} \end{aligned} \quad (12)$$

Da die q_{i1} -Werte standardisierte Werte sind, folgt sofort

$$\frac{1}{m} \sum_{i=1}^m q_{i1}^2 = 1. \quad (13)$$

Aber auch die ε_{ij} - und ε_{ik} -Werte sind standardisierte Werte, so dass die beiden übrigen Summen in (12) als Korrelationen zwischen den q_{i1} -Werten und den standardisierten Fehlern aufgefasst werden können. Fasst man die q_{i1} -Werte als "wahre" Werte im Sinne der Klassischen Testtheorie auf, so sind sie nur in dem Sinne zufällige Variable, als sie von Person zu Person variieren und also zufällig nur deswegen sind, weil die Person zufällig in die Stichprobe gelangt ist. Die ε_{ij} und ε_{ik} -Werte sind aber auch für jede Person zufällig, weil ihr Wert auch bei der selben Person von Messung zu Messung zufällig variiert. Deswegen kann

man die Annahme machen, dass die Korrelation zwischen den q_{i1} -Werten und den ε_i -Werten gleich Null ist:

Annahme:

$$\frac{1}{m} \sum_{i=1}^m q_{i1} \varepsilon_{ik} = \frac{1}{m} \sum_{i=1}^m q_{i1} \varepsilon_{ij} = 0. \quad (14)$$

Berücksichtigt man nun (13) und (14), so vereinfacht sich der Ausdruck (12) für r_{jk} zu

$$r_{jk} = a_{j1} a_{k1}. \quad (15)$$

Die Korrelation zwischen den Variablen V_j und V_k ist also gerade gleich dem Produkt der Regressionsgewichte a_{j1} und a_{k1} ! Auf diese Weise hat man die beobachtete Korrelation zwischen den beiden Variablen durch Rückführung auf eine gemeinsam gemessene "latente" Variable "erklärt".

Die Diskussion der Frage, wie man die a_{j1} und a_{k1} tatsächlich findet, wird zunächst zurückgestellt; es kommt hier nur darauf an, das Modell der Faktorenanalyse zu entwickeln. Es kann ja sein, dass die beobachteten Korrelationen nicht durch eine einzelne latente Variable erklärt werden können, denn es können ja verschiedene solche Variablen durch die gemessenen Variablen erfasst werden. Können also die Korrelationen r_{jk} nicht durch den Ansatz (15) erfasst werden, wird man eine zweite latente Variable F_2 annehmen. Statt der Gleichungen (2) und (3) kommt man dann zu dem erweiterten Modell

$$x_{ij} = \alpha_{j1} F_{i1} + \alpha_{j2} F_{i2} + e_{ij} \quad (16)$$

$$x_{ik} = \alpha_{j1} F_{i1} + \alpha_{j2} F_{i2} + e_{ik} \quad (17)$$

Natürlich wird man versuchen, F_2 so zu bestimmen, dass man zu einer möglichst ökonomischen Erklärung der Korrelationen kommt. Die generelle Idee dazu ist, dass die F_{i1} - und F_{i2} -Werte keine redundanten Größen sein sollen, d.h. sie sollen keine Merkmale repräsentieren, die wechselseitig auseinander vorhergesagt werden können. Diese Idee wird weiter unten noch genauer spezifiziert werden.

Man kann nun ebenfalls die Standardisierung der x_{ij} und x_{ik} vornehmen und kommt dann unter der Annahme, dass auch die Korrelationen der latenten Variablen F_1 und F_2 mit den Fehlern e_{ij} und e_{ik} gleich Null sind, zu der Gleichung

$$r_{jk} = a_{j1} a_{k1} + a_{j2} a_{k2}. \quad (18)$$

Gelingt es nicht, die r_{jk} in dieser Form darzustellen, wird man eine weitere latente Variable F_3 annehmen, und

$$x_{ij} = \alpha_{j1} F_{i1} + \alpha_{j2} F_{i2} + \alpha_{j3} F_{i3} + e_{ij}, \quad j = 1, \dots, n \quad (19)$$

schreiben. Geht man von der speziellen (aber nicht notwendig zu machenden) Annahme aus, dass die latenten Variablen F_1 , F_2 und F_3 paarweise unkorreliert sind, so erhält man für die Korrelationen den Ausdruck

$$r_{jk} = a_{j1} a_{k1} + a_{j2} a_{k2} + a_{j3} a_{k3}. \quad (20)$$

Allgemein kann man davon ausgehen, dass man $r < n$ latente Variable benötigt, die dann zu dem allgemeinen Ausdruck

$$x_{ij} = \alpha_{j1}F_{i1} + \alpha_{j2}F_{i2} + \cdots + \alpha_{js}F_{is} + e_{ij}, \quad j = 1, \dots, n \quad (21)$$

führen, der standardisiert die Form

$$z_{ij} = a_{j1}q_{i1} + a_{j2}q_{i2} + \cdots + a_{js}q_{is} + \varepsilon_{ij} \quad (22)$$

annimmt. Für die meisten Untersuchungen in der Psychologie ist diese Gleichung der Ausgangspunkt der Analyse:

1. **Datenreduktion** Im Fall $s < n$ lassen sie die n gemessenen Variablen durch weniger, eben s , latente Variable beschreiben. In diesem Sinne führt die Faktorenanalyse zu einer *Datenreduktion*. Die Datenreduktion erleichtert die Analyse der gefundenen Zusammenhänge zwischen den Variablen oft erheblich.

2. **Ladungen** Die a_{j1}, \dots, a_{js} heißen die *Ladungen* der j -ten, gemessenen Variablen auf den latenten Dimensionen.

Die Ladungen der Variablen sind dieselben für alle Personen bzw. für alle Einheiten, an denen sie gemessen wurden. Eine Ladung repräsentiert den Anteil, mit dem eine gemessene Variable eine latente Variable erfasst. a_{jk} ist also der Anteil, mit dem die j -te gemessene Variable (das j -te Item) die k -te latente Variable erfasst.

3. **Faktorwerte** Die q_{i1}, \dots, q_{is} heißen die *Faktorwerte* der Personen $i = 1, \dots, m$ auf den gleichen latenten Dimensionen. Die Definition (9) der q_{i1}, \dots, q_{is} als standardisierte Werte impliziert, dass

$$\bar{q}_1 = \frac{1}{m} \sum_{i=1}^m q_{i1} = \bar{q}_2 = \frac{1}{m} \sum_{i=1}^m q_{i2} = \cdots = \bar{q}_n = \frac{1}{m} \sum_{i=1}^m q_{in} = 0. \quad (23)$$

Der Faktorwert q_{ik} repräsentiert das Ausmaß, in dem die k -te latente Variable bei der i -ten Person ausgeprägt ist. Die Faktorwerte einer Person sind dieselben für alle Variablen. Sie entsprechen den "wahren" Werte der i -ten Person - im Sinne der Klassischen Testtheorie - auf den verschiedenen latenten Dimensionen.

4. **Unabhängigkeit von Ladungen und Faktorwerten** Vom Modell der Faktorenanalyse her gesehen sind die Ladungen a_{jk} der gemessenen Variablen und die Faktorwerte q_{ik} der Personen unabhängig voneinander: eine Person verfügt über das durch F_k repräsentierte Merkmal im Ausmaß q_{ik} unabhängig davon, mit welchem Verfahren oder Test F_k gemessen wird, und die gemessene Variable bzw. das Item X_j "benötigen" das Ausmaß a_{jk} der Eigenschaft F_k , damit X_j "positiv" beantwortet

wird, oder X_j erfasst das Merkmal F_K stets zu einem Anteil a_{jk} , unabhängig davon, bei welcher Person oder bei welchem Objekt gemessen wird. Andererseits sind die a_{jk} und die q_{ik} unbekannte Parameter, die aus den Daten, also den Messungen für X_1, \dots, X_n , geschätzt werden müssen. Es wird sich zeigen, dass die q_{ik} als Funktion der X_j und der a_{jk} ausgedrückt werden können, und die a_{jk} als Funktion der X_j und der q_{ik} ; in die Schätzung der Parameter aus den Daten gehen also auf implizite Weise bestimmte Abhängigkeiten zwischen den beiden Klassen von Parametern ein.

5. **Linearität** Eine Funktion $f(x_1, \dots, x_n)$ der n Variablen x_1, \dots, x_n heißt *linear*, wenn

$$f(x_1, \dots, x_n) = a_1x_1 + a_2x_2 + \dots + a_nx_n$$

In (21) ist f durch x_{ij} gegeben, und die x_1, \dots, x_n stehen für die F_1, \dots, F_r . Offenbar repräsentiert (21) ein lineares Modell der Wirkung latenter Variablen. Die Restriktion auf lineare Modelle ist keine triviale Einschränkung, denn viele psychologische Modelle sind als nichtlineare Modelle konzipiert worden: So ist etwa die Arbeitsmotivation den Befunden von Vroom (1964) zufolge durch das Produkt von Erwartung und Valenz (expectency \times valence) definiert, und die Performanz einer Person durch das Produkt von Fähigkeit und Motivation (ability \times motivation) (vergl. Busemeyer und Jones (1983))². Somit kann man vermuten, dass Performanz einer Wechselwirkung der Form Fähigkeit \times Erwartung \times Valenz, also einem Term $F_1F_2F_3$ entspricht. Generell kann man sagen, dass Wechselwirkungen zwischen Variablen oft eine nichtvernachlässigbare Komponente in der Wirkungsweise von Variablen sind, und es gibt keinen Grund, sie nicht auch für latente Variablen anzunehmen.

Eine Möglichkeit, einen rein linearen Ansatz zu rechtfertigen, besteht im Hinweis auf die Tatsache, dass man in den meisten Fällen Funktionen in eine Reihe entwickeln kann, d.h. man kann nichtlineare Funktionen durch geeignet gewählte Polynome im Prinzip beliebig genau approximieren. Ist also $f(x)$ irgendeine Funktion, so kann man eine Approximation der Form

$$f(x) \approx a_0 + a_1x + a_2x^2 + \dots + a_px^p$$

finden, die durch geeignete Wahl von p im Prinzip beliebig genau gemacht werden kann. Existieren die Ableitungen

$$f'(x) = \frac{df(x)}{dx}, \quad f''(x) = \frac{d^2(x)}{dx^2}, \quad \text{etc,}$$

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

so kann z.B. $f(x + \Delta x)$ durch eine Taylor-Reihe³

$$f(x + \Delta x) = f(x) + \Delta x f'(x) + \frac{\Delta x^2}{2} f''(x) + \frac{\Delta x^3}{2 \cdot 3} f'''(x) + \dots$$

approximiert werden; für $x = 0$ und Umbenennung von Δx in x erhält man auch eine Approximation für $f(x)$ (der Umweg über $f(x + \Delta x)$ für $x \neq 0$ wird nötig, wenn $f(0)$ nicht existiert, wie etwa für $f(x) = \log x$), bzw. einen nicht berechenbaren "Wert" wie $-\infty$ annimmt). Die k -ten Ableitungen $f^{(k)}/k!$, entsprechen dann den Koeffizienten im approximierenden Polynom. Solche Approximationen lassen sich auch für Funktionen von mehr als einer Variablen herleiten.

Die ersten Terme der Reihenentwicklung sind oft lineare Terme, gefolgt von nichtlinearen Termen. So ist für hinreichend kleine Werte von x die Exponentialfunktion e^x durch die Reihe

$$e^{ax} = e^0 + xae^0 + \frac{x^2a^2}{2}e^0 + \frac{x^3a^3}{6}e^0 \dots = \sum_{k=0}^{\infty} \frac{(ax)^k}{k!}$$

definiert und für hinreichend kleine Werte von x erhält man $e^x \approx 1 + x$. Allgemein erhält man oft für hinreichend kleine Werte der unabhängigen Variablen eine Approximation durch lineare Funktionen. Insofern kann man das lineare Faktorenmodell stets als Approximation auffassen, die für "hinreichend kleine" Werte der latenten Variablen F_1, \dots, F_r gilt.

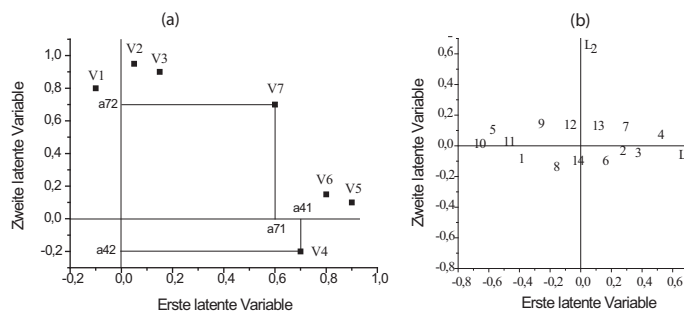
Der Wert s der Anzahl der benötigten latenten Dimensionen oder latenten Variablen ist ebenfalls unbekannt und muß aus den Daten bestimmt werden. Wegen der oben gemachten Unabhängigkeitsannahmen findet man, dass für die Korrelation r_{jk} zwischen zwei beliebigen Variablen V_j und V_k die Beziehung

$$r_{jk} = \frac{1}{m} \sum_{i=1}^m z_{ij}z_{ik} = \sum_{t=1}^s a_{jt}a_{kt}, \quad s \leq n \quad (24)$$

gilt. Die Gleichung setzt allerdings stillschweigend voraus, dass man den Wert für die Anzahl der Dimensionen, s , kennt. Eine perfekte Übereinstimmung zwischen den Korrelationen r_{jk} und der "Vorhersage", wie sie auf der rechten Seite der Gleichung durch die Summe $\sum_{t=1}^s a_{jt}a_{kt}$ gemacht wird, wird man dann bekommen, wenn man $s = n$ setzt, wenn also die Anzahl der latenten Dimensionen gleich der Anzahl der untersuchten Variablen ist. Die Annahme $r = n$ ist aber nicht besonders interessant, denn man möchte die Daten, also die Messwerte x_{ij} bzw. die standardisierten Werte z_{ij} so ökonomisch wie möglich beschreiben, d.h. man sucht den kleinstmöglichen Wert für s , für den noch eine hinreichende *Approximation* der Korrelationen r_{jk} gemäß (24) gelingt. Die Suche nach dem kleinstmöglichen Wert für s ist der *datenreduzierende Aspekt* der Faktorenanalyse.

³Brook Taylor (1685 – 1731), englischer Mathematiker, zeigte als erster die Möglichkeit einer solchen Entwicklung, die für viele mathematische Analysen von grundlegender Bedeutung ist.

Abbildung 1: (a) Repräsentation von Variablen (V_1, \dots, V_7) und (b) Personen in Koordinatensystemen, die latente Dimensionen repräsentieren; man beachte, dass die Koordinaten der Personen wegen (23) den Mittelwert 0 haben. Aus den Variablen V_4, V_5 und V_6 kann u. U. die Bedeutung der ersten latenten Dimension erschlossen werden, und aus den Variablen V_1, V_2 und V_3 die der zweiten latenten Dimension. (b) zeigt, dass die erste Dimension am meisten zwischen den Personen differenziert, die zweite Dimension differenziert weniger.



Anmerkung: Die Gleichung (22) zeigt, dass der der Faktorenanalyse zugrundeliegende Ansatz immer nur als Approximation, nicht aber sinnvoll als Theorie "an sich" verstanden werden kann. Denn zunächst einmal soll der Ansatz (22) für jeden Gegenstandsbereich, auf den die Analyse angewendet wird, gelten, er soll also für EEG-Daten wie für politische Meinungsumfragen gleichermaßen gültig sein. Warum sich aber Messwerte stets als Summe von Produkten darstellen lassen sollen, in denen ein Faktor (der Faktorwert q_{ik}) das i -te gemessene Objekt auf einer k -ten "latenten" Variablen charakterisiert und der andere Faktor (die Ladung a_{jk}) die j -te gemessene Variable auf der gleichen latenten Dimension abbildet, ist völlig unklar, so lange man (22) zum nicht weiter hinterfragten Ausgangspunkt der Analyse macht. (22) kann als Ansatz zu einer möglichst ökonomischen Beschreibung gewählt werden, allerdings ist dann die *Reifikation*, d.h. die verdinglichende Annahme der Existenz der Dimensionen, die durch die q_{ik} und p_{jk} repräsentiert werden, noch lange nicht gerechtfertigt.

Einen Spezialfall für (24) erhält man, wenn man $j = k$ setzt. Dann ist

$$r_{jj} = \frac{1}{m} \sum_{i=1}^m z_{ij}^2 = 1 = \sum_{t=1}^s a_{jt}^2. \quad (25)$$

Es gilt wieder die Anmerkung, dass diese Gleichung im allgemeinen nur dann exakt gilt, wenn $s = n$. Für $s < n$ wird die Beziehung nur angenähert gelten. Man nennt

$$h_j^2 = \sum_{t=1}^s a_{jt}^2, \quad (26)$$

also die Summe der a_{jt}^2 für den gewählten Wert für s , die *Kommunalität* der j -ten Variablen, und es wird eben im allgemeinen $h_j^2 < 1$ sein. r_{jj} ist ja die Varianz der standardisierten j -ten Variablen (da $\bar{z}_j = 0$ ist $\sum_i z_{ij}^2/m$ ein Ausdruck, der dem für eine Varianz entspricht), und h_j^2 gibt an, in welchem Ausmaß die Varianz einer Variablen durch die latenten Dimensionen erklärt wird, von denen ja angenommen wird, dass sie *gemeinsam* (daher der Ausdruck "Kommunalität") in allen Variablen enthalten sind.

Die Gleichung (25) erlaubt es zumindest im Prinzip, die Anzahl der latenten Dimensionen, die "hinter" den gemessenen Variablen wirken, abzuschätzen. Denn für gegebene Daten müssen die Gleichungen $r_{jj} = 1$ für alle j , $1 \leq j \leq n$ gelten, da ja jede Variable mit sich selbst korreliert wird. Andererseits muß $r_{jj} = \sum_t a_{jt}^2$ sein, wenn die Anzahl s der latenten Dimensionen korrekt gewählt wurde. Es zeigt sich, dass diese Beziehungen i.a. nur für $s = n$ gelten, - aber dann hat man keinen "datenreduzierenden" Effekt. Man wird $s < n$ wählen, und dann gilt nach (26)

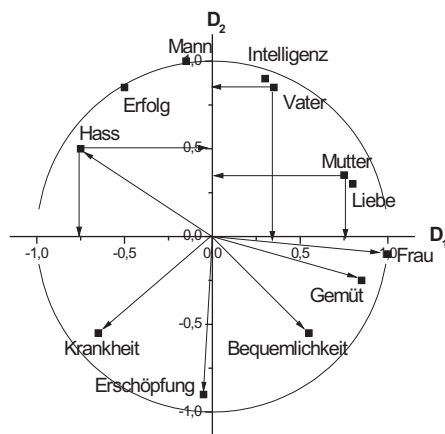
$$r_{jj} = 1 = h_j^2 + \varepsilon_j^2, \quad (27)$$

und ε_j^2 repräsentiert den Effekt von "spezifischen" latenten Dimensionen, also latenten Variablen, die nur in der Variable V_j enthalten sind, und "(Mess-)Fehlern". Da die a_{j1}, \dots, a_{js} die Koordinaten des Punktes sind, der die Variable V_j im "Variablenraum" repräsentiert, gibt nach dem Satz des Pythagoras h_j^2 das Quadrat der Distanz dieses Punktes vom Koordinatenursprung an. Ist also die Anzahl s zu berücksichtigender latenter Dimension gut gewählt worden, so wird $h_j^2 \approx 1$ für alle j gelten, d.h. die repräsentierenden Punkte liegen dicht an der Oberfläche einer s -dimensionalen Kugel mit dem Radius 1. Für den Fall $s = 2$ ist diese Kugel gerade ein (Einheits-)Kreis. Abb. 2 illustriert diesen Sachverhalt.

Der Gleichung (22) entsprechend wird z_{ij} im faktorenanalytischen Zusammenhang als eine Summe von Produkten, d.h. als ein sogenanntes *Skalarprodukt*, dargestellt. Dies gilt auch für die Korrelation r_{jk} , vergl. (12). In einem gewissen Sinn kann dann auch z_{ij} als Korrelation aufgefasst werden, nämlich als Korrelation zwischen den q_{i1}, \dots, q_{is} -Werten einerseits und den a_{j1}, \dots, a_{js} -Werten andererseits. Die q_{i1}, \dots, q_{is} geben an, sie die i -te Person mit den latenten Dimensionen "ausgestattet" ist, und die a_{j1}, \dots, a_{js} geben an, in welchem Ausmaß die Variablen V_1, \dots, V_n die latenten Dimensionen überhaupt erfassen. Dieser Sachverhalt soll noch einmal hervorgehoben werden:

In einem allgemeinen Sinne kann man sagen, dass das faktorenanalytische Modell impliziert, dass die Messwerte, insbesondere die standardisierten Messwerte, als Korrelationen zwischen der Ausprägung der latenten Variablen bei den Personen - repräsentiert durch die q_{it} - und den Anteilen a_{jt} , mit denen die latenten Variablen durch die V_j erfasst werden, aufgefasst werden können. Allerdings sind nur die q_{it} standardisierte Variablen, die a_{jt} nicht - so ist es möglich, dass $a_{jt} > 0$ für alle j und t gilt, was für standardi-

Abbildung 2: Begriffliche Stereotypen in den 50-er Jahren nach P. R. Hofstätter. Die Punkte, die die Begriffe repräsentieren, liegen nahe beim Einheitskreis, so dass die begriffliche Struktur gut durch eine 2-dimensionale "Lösung" beschrieben werden kann. D_1 repräsentiert das "weibliche Prinzip", D_2 das "männliche Prinzip". Entgegen geisteswissenschaftlichen Vorstellungen (Wellek, 1977) sind diese "Prinzipien" nicht polar, also als Gegensätze auf *einer* Dimension, angeordnet, sondern es handelt sich um voneinander unabhängige Prinzipien. In einer Person, gleich ob weiblich oder männlich, können also beide Prinzipien gleichermaßen vorhanden oder nicht vorhanden sein.



sierte Variablen nicht möglich ist. Der Begriff des Skalarprodukts liefert eine genauere Charakterisierung der z_{ij} -Werte.

Da die a_{jt} nur für die Variablen und die q_{it} nur für die Personen charakteristisch sind, könnte man auf die Idee kommen, dass diese Parameter völlig unabhängig voneinander sind. Dies ist nicht der Fall; es wird in den folgenden Abschnitten deutlich werden, dass es hier wechselseitige Abhängigkeiten gibt. Die Daten zusammen mit den Personenparametern q_{it} bestimmen die Variablenparameter a_{jt} , und die Daten zusammen mit den Variablenparametern bestimmen die Personenparameter q_{it} .

Zur Natur der latenten Variablen Bisher sind die latenten Variablen einfach als Variablen eingeführt worden, aus denen die gemessenen Variablen additiv zusammengesetzt sind. Die Frage ist nun, ob die latenten Variablen deswegen notwendig auch "atomare", also nicht weiter zerlegbare Größen repräsentieren. Man betrachte dazu das Beispiel in Abbildung 2 auf Seite 15. Eine latente Dimension (oder Variable) repräsentiert das "weibliche Prinzip", die andere das "männliche Prinzip". Es ist nicht klar, warum derartige "Prinzipien" in sich homogene, nicht weiter aufspaltbare Merkmale sein sollen. Alles, was man sagen kann ist, dass diese Merkmale oder Prinzipien relativ zu den anderen Größen ('Mutter', 'Hass', 'Krankheit' etc) als konstant zusammengesetzte Größen erscheinen, die in in verschiedener Ausprägung in andere Größen

eingehen. Anhand des Vektorbegriffs läßt sich dieser Sachverhalt klarer formulieren; in Abschnitt 2 wird deshalb auf diese Eigenschaft latenter Variablen noch einmal näher eingegangen.

Bisher ist nur das Modell vorgestellt worden. Es muß nun geklärt werden, wie die latenten Variablen gefunden werden können, d.h. wie die *Parameter* q_{it} und die a_{jt} des Modells tatsächlich berechnet werden können.

2 Vektoren, Matrizen, und latente Variablen

In den Gleichungen (21) und (22) sind die Grundgleichungen für x_{ij} bzw. für die standardisierten Werte z_{ij} gegeben worden; sie werden hier noch einmal zur Erinnerung aufgeführt:

$$x_{ij} = \alpha_{j1}F_{i1} + \alpha_{j2}F_{i2} + \cdots + \alpha_{js}F_{is} + e_{ij}, \quad j = 1, \dots, n \quad (28)$$

$$z_{ij} = a_{j1}q_{i1} + a_{j2}q_{i2} + \cdots + a_{js}q_{is} + \varepsilon_{ij} \quad (29)$$

Es sind also $n \times s$ Parameter α_{jt} ($j = 1, \dots, n$ und $t = 1, \dots, s$) bzw. a_{jt} und $m \times s$ Parameter F_{it} bzw. q_{it} ($i = 1, \dots, m$) zu schätzen. Die Schätzung setzt die Kenntnis einiger Begriffe und Ergebnisse der linearen Algebra voraus, die im Folgenden kurz vorgestellt werden sollen.

Es sei daraufhin gewiesen, dass die Gleichungen für x_{ij} und z_{ij} als Regressionsgleichungen aufgefasst werden können. In einer linearen multiplen Regression betrachtet man ja Beziehungen der Form

$$Y = b_0 + b_1X_1 + \cdots + b_sX_s + e, \quad (30)$$

wobei X_1, \dots, X_s sind die Prädiktorvariablen und b_0, b_1, \dots, b_s die zu schätzenden Regressionsgewichte sind. In (28) entsprechen den X_1, \dots, X_s die F_{i1}, \dots, F_{is} , und den b_0, b_1, \dots, b_s die $\alpha_{j1}, \dots, \alpha_{js}$. Y entspricht den x_{ij} . Für (29) gelten die analogen Beziehungen. Der einzige Unterschied zwischen der Regressionsgleichung (30) und den Gleichungen (28) und (29) ist, dass die Prädiktorvariablen F_{i1}, \dots, F_{is} nicht explizit gegeben sind, sondern zusammen mit den Regressionsgewichten aus den Daten geschätzt werden müssen.

2.1 Vektoren

2.1.1 Latente Variablen, Vektoren, und Linearkombinationen

Um die folgenden Begriffsbildungen zu motivieren, werde die Gleichung (28) - im Prinzip - für alle i angeschrieben, wobei der Einfachheit halber $s = 2$ angenommen wird; es geht ja nur um die Einführung des Vektor- und des Matrixbegriffs.

$$x_{i1} = \alpha_{11}F_{i1} + \alpha_{12}F_{i2} + e_{i1}$$

$$\begin{aligned}
x_{21} &= \alpha_{11}F_{21} + \alpha_{12}F_{22} + e_{21} \\
x_{31} &= \alpha_{11}F_{31} + \alpha_{12}F_{32} + e_{31} \\
&\vdots \\
x_{m1} &= \alpha_{11}F_{m1} + \alpha_{12}F_{m2} + e_{m1}
\end{aligned}
\tag{31}$$

Diese Gleichungen können auch in der folgenden Form geschrieben werden:

$$\begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \\ \vdots \\ x_{m1} \end{pmatrix} = \alpha_{11} \begin{pmatrix} F_{11} \\ F_{21} \\ F_{31} \\ \vdots \\ F_{m1} \end{pmatrix} + \alpha_{12} \begin{pmatrix} F_{12} \\ F_{22} \\ F_{32} \\ \vdots \\ F_{m2} \end{pmatrix} + \begin{pmatrix} e_{11} \\ e_{21} \\ e_{31} \\ \vdots \\ e_{m1} \end{pmatrix}
\tag{32}$$

Es handelt sich zunächst tatsächlich nur um eine andere Schreibweise für die Gleichungen (32). Die x_{11}, \dots, x_{m1} können als Koordinaten eines Punktes aufgefasst werden, der die Variable V_1 in einem m -dimensionalen Raum repräsentiert (für $m \leq 3$ gibt es eine anschauliche Vorstellung von dieser Repräsentation). Ebenso können die $F_{1k}, F_{2k}, \dots, F_{mk}$ als Koordinaten eines Punktes aufgefasst werden, der die k -te latente Variable in einem m -dimensionalen Raum abbildet. Aber auch die Koeffizienten $(\alpha_{j1}, \alpha_{j2})$ können als Koordinaten eines Punktes in einem 2-dimensionalen Raum interpretiert werden. Dieser Punkt bildet die Variable V_j im Raum der latenten Variablen ab. Es zeigt sich aber, dass die Interpretation der x_{11}, \dots, x_{m1} , der $F_{1k}, F_{2k}, \dots, F_{mk}$ und der $(\alpha_{j1}, \alpha_{j2})$ als Komponenten von *Vektoren* der geeigneter ist als die Punktinterpretation, u.a. weil dann die Korrelation zwischen den Variablen ebenfalls geometrisch repräsentiert wird: sie steht in einer bestimmten Beziehung zu den Winkeln zwischen den repräsentierenden Vektoren. Demnach wäre also ein Vektor die Anordnung von Zahlen in einer Spalte, – wobei es aber auf die Reihenfolge der Zahlen ankommt. Denn wenn man die Reihenfolge zweier Zahlen miteinander vertauscht, vertauscht man die Messwerte zweier Objekte oder Personen, und das darf natürlich nicht geschehen. Diese Charakterisierung eines Vektors ist allerdings unvollständig, und eine vollständige Definition wird bald gegeben werden. Jedenfalls heißen die Zahlen, die zu einem Vektor zusammengefasst werden, die *Komponenten* des Vektors. Weiter zeigt der Vergleich von (32) mit (33), dass Faktoren wie α_{11} und α_{12} , die mit allen Komponenten eines Vektors multipliziert werden, vor die Klammer gezogen werden können, die einen Vektor kennzeichnen. Die Vereinfachung besteht zunächst nur darin, dass diese Faktoren jetzt nur noch einmal geschrieben werden müssen. Anders gesehen soll diese Schreibweise zeigen, dass eben alle Komponenten eines Vektors mit dem davor stehenden Faktor multipliziert werden sollen. Die Schreibweise ist also als eine Art Handlungsanweisung zu verstehen.

Die Schreibweise (33) ist immer noch zu ausladend, um von allgemeinem Nutzen zu sein. Man bezeichnet einen Vektor deshalb oft durch einen einzelnen Buchstaben mit einem Pfeil darüber, etwa \vec{x} (es gibt andere Schreibweisen, z.B.

einfach ein Buchstabe in Fettschrift, \mathbf{x}). Dementsprechend kann man für die Vektoren in (33) auch die kürzeren Schreibweisen

$$\vec{X}_j = \begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \\ \vdots \\ x_{m1} \end{pmatrix}, \quad \vec{F}_1 = \begin{pmatrix} F_{11} \\ F_{21} \\ F_{31} \\ \vdots \\ F_{m1} \end{pmatrix}, \quad \vec{F}_2 = \begin{pmatrix} F_{12} \\ F_{22} \\ F_{32} \\ \vdots \\ F_{m2} \end{pmatrix}, \quad \vec{e}_1 = \begin{pmatrix} e_{11} \\ e_{21} \\ e_{31} \\ \vdots \\ e_{m1} \end{pmatrix} \quad (34)$$

eingeführen und erhält dann statt (33) den einfacheren Ausdruck

$$\vec{X}_j = \alpha_{j1}\vec{F}_1 + \alpha_{j2}\vec{F}_2 + \vec{e}_1 \quad (35)$$

Ein Pfeil über einem Buchstaben bezeichnet also einen Vektor, d.h. eine ganze Spalte von Zahlen, eben den Komponenten des Vektors; gelegentlich werden Vektoren auch durch Fettschrift gekennzeichnet, etwa \mathbf{x} , \mathbf{a} , etc. $\alpha_{j1}\vec{F}_1$ bedeutet, wie oben schon gesagt, dass alle Komponenten des Vektors mit dem Faktor α_{j1} multipliziert werden sollen; eine analoge Aussage gilt natürlich für $\alpha_{j2}\vec{F}_2$. Natürlich sind $\alpha_{j1}\vec{F}_1$ und $\alpha_{j2}\vec{F}_2$ auch wieder Vektoren, denn diese Ausdrücke stehen ja wieder für Spalten von Zahlen. Nach (35) ist der Vektor \vec{X}_j eine Summe von Vektoren, und eine Summe von Vektoren ist offenbar als ein Vektor definiert, dessen Komponenten die Summen der Komponenten der Vektoren $\alpha_{j1}\vec{F}_1$ und $\alpha_{j2}\vec{F}_2$ ist. Dies muß so sein, wie der Vergleich mit (32) zeigt. Dieser Vergleich zeigt auch, dass mit (35) eine sehr vereinfachte Schreibweise erreicht worden ist.

Anmerkung: Bei der Beschreibung des Modells der Faktorenanalyse sind F_1, \dots, F_s als latente Variablen eingeführt worden, d.h. als Variablen, die jeweils Werte aus einer Menge möglicher Werte annehmen können. Häufig wird implizit angenommen, dass die Menge der möglichen Werte durch die Menge \mathbb{R} der reellen Zahlen gegeben ist, so dass ein Kontinuum möglicher Werte zur Verfügung steht. Die Vektoren $\vec{F}_1, \dots, \vec{F}_s$ bezeichnen aber stets Stichproben von etwa m Werten aus der Menge der möglichen Werte. Damit kann eine konkrete Stichprobe, als ein konkret vorliegender Datensatz gemeint sein. Man kann aber auch die Menge der möglichen m -dimensionalen Vektoren betrachten, wobei jede Komponente einen Wert aus \mathbb{R} annehmen kann. \square

Es wird noch eine allgemeine Definition von Vektoren gegeben:

Definition 2.1 Ein n -dimensionaler Vektor \vec{x} ist ein geordnetes n -Tupel von Zahlen x_1, \dots, x_n :

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}. \quad (36)$$

Die x_1, \dots, x_n heißen die Komponenten des Vektors.

Hat man mehrere Vektoren \vec{x}_j , $j = 1, \dots, m$, so kann man die Komponenten durch Doppelindizierung kennzeichnen: x_{1j}, x_{2j}, \dots

Anmerkungen:

1. Der Begriff *geordnetes n -Tupel* besagt, dass es auf die Anordnung der Komponenten des Vektors ankommt. Verändert man die Reihenfolge der Komponenten in (36), so erhält man einen anderen Vektor.
2. Ein Vektor wird wie in (36) immer als eine "Spalte" von Komponenten aufgefasst. Schreibt man die Komponenten in einer Zeile an, so erhält man einen "gestürzten" oder "transponierten" Vektor, der mit \vec{x}' bezeichnet und gelegentlich als Zeilenvektor bezeichnet wird. Um Platz zu sparen, kann man dann

$$\vec{x} = (x_1, \dots, x_n)'$$

schreiben; der Strich an der letzten Klammer zeigt dann an, dass der Zeilenvektor wieder gestürzt werden soll, - wodurch er wieder zu einem Spaltenvektor wird.

3. Graphisch wird ein Vektor durch einen Pfeil dargestellt. Die Komponenten des Vektors sind die Differenzen zwischen den Koordinaten des Endpunktes und des Anfangspunktes des Pfeils. Damit legen die Komponenten sowohl die Orientierung, d.h. die Richtung, in die der Pfeil zeigt, wie auch seine Länge fest.
4. Eine einzelne Zahl λ kann als Spezialfall eines Vektors aufgefasst werden, nämlich eines Vektors, der nur eine Komponente hat. Man spricht dann von einem *Skalar*, im Unterschied zu einem Vektor, der dann mindestens zwei Komponenten haben sollte.

Die folgenden Rechenregeln sind schon in Gleichung (34) angewandt worden:

1. **Multiplikation mit einem Skalar λ :**

$$\lambda\vec{x} = (\lambda x_1, \lambda x_2, \dots, \lambda x_n)'. \quad (37)$$

2. **Addition von Vektoren:** Sind \vec{x} und \vec{y} zwei n -dimensionale Vektoren, so ist

$$\vec{x} + \vec{y} = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)'. \quad (38)$$

3. **Linearkombinationen:** Sind $\lambda_1, \dots, \lambda_k$ Skalare und $\vec{x}_1, \dots, \vec{x}_k$ n -dimensionale Vektoren, so heißt der durch

$$\vec{y} = \lambda_1\vec{x}_1 + \dots + \lambda_k\vec{x}_k \quad (39)$$

definierte Vektor \vec{y} eine *Linearkombination* der Vektoren $\vec{x}_1, \dots, \vec{x}_k$. Für $k = 2$, $\lambda_1 = 1$ und $\lambda_2 = -1$ ergibt sich insbesondere die Differenz zweier Vektoren als Spezialfall:

$$\vec{z} = (z_1, \dots, z_n)' = \vec{x} - \vec{y} = (x_1 - y_1, \dots, x_n - y_n)'. \quad (40)$$

Wie noch verdeutlicht werden wird werden im Modell der Faktorenanalyse die Vektoren, die gemessene Variablen repräsentieren, als Linearkombinationen von Vektoren dargestellt, die latente Variablen repräsentieren. Es ist diese Repräsentation, die die Schätzung der latenten Variablen ermöglicht.

2.1.2 Skalarprodukte

(35) enthält noch mehr Informationen über mögliche Verknüpfungen von Vektoren. Betrachtet man die i -te Komponente von \vec{X}_j , also die Komponente x_{ij} . Nach (32), aber natürlich auch nach (35), muß dann

$$x_{ij} = \alpha_{j1}F_{i1} + \alpha_{j2}F_{i2}. \quad (41)$$

gelten. Hier treten ausser den Komponenten von $\vec{\alpha}_j$ noch je eine Komponente der Vektoren \vec{F}_1 und \vec{F}_2 auf, nämlich F_{i1} und F_{i2} . Diese beiden Zahlen definieren wiederum einen Vektor, nämlich den Vektor

$$\vec{Y}_i = \begin{pmatrix} F_{i1} \\ F_{i2} \end{pmatrix}. \quad (42)$$

Der Index i in \vec{Y}_i soll dabei andeuten, dass die i -te Person gemeint ist. Nach (41) ist x_{ij} eine Summe von Produkten der Komponenten von $\vec{\alpha}_j$ und \vec{Y}_i . Man schreibt dafür

$$\vec{\alpha}_j' \vec{Y}_i = \langle \vec{\alpha}_j, \vec{Y}_i \rangle = \alpha_{j1}F_{i1} + \alpha_{j2}F_{i2}. \quad (43)$$

Sowohl die Schreibweise $\vec{\alpha}_j' \vec{Y}_i$ wie die alternative Schreibweise $\langle \vec{\alpha}_j, \vec{Y}_i \rangle$ sind üblich; letztere ist gelegentlich deutlicher. Die Summe von Produkten auf der rechten Seite dieser Gleichung definiert, was mit der Schreibweise $\vec{\alpha}_j' \vec{Y}_i$ gemeint ist: man bildet das Produkt der beiden ersten Komponenten von $\vec{\alpha}_j$ und \vec{Y}_i und addiert sie zu dem Produkt der beiden zweiten Komponenten dieser beiden Vektoren. Dass die beiden Vektoren hier nur zwei Komponenten haben, ist dabei unwesentlich, sie können irgendeine Anzahl n von Komponenten haben. Natürlich impliziert (43), dass $\vec{\alpha}_j' \vec{Y}_i = x_{ij}$. Man nennt die Summe der Produkte der zueinander korrespondierenden Komponenten zweier Vektoren das *Skalarprodukt* der beiden Vektoren, hier also von $\vec{\alpha}_j$ und \vec{Y}_i . Da der Vektor $\vec{\alpha}_j$ spezifisch für die j -te Variable X_j ist und der Vektor \vec{Y}_i spezifisch für die i -te Person ist, bedeutet (43), dass sich wegen $x_{ij} = \vec{\alpha}_j' \vec{Y}_i$ der Messwert x_{ij} der i -ten Person für die j -te Variable dem bis jetzt verfolgten Ansatz entsprechend als Skalarprodukt eines variablenspezifischen Vektors und eines personenspezifischen Vektors repräsentieren läßt. Dieser Befund ist ein zentraler Bestandteil

des faktorenanalytischen Ansatzes. Natürlich muß der Index i nicht für eine Person stehen, denn man ja z.B. auch Messungen bei einer Person zu verschiedenen Zeitpunkten t_i durchführen; auf die Interpretation dieses Falles wird später explizit eingegangen werden.

Ein Skalarprodukt ist also im wesentlichen eine Summe von Produkten. Dementsprechend ist der bekannte Produkt-Moment-Korrelationskoeffizient

$$\begin{aligned} r_{jk} &= \frac{\frac{1}{m} \sum_{i=1}^m (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{s_j s_k} \\ &= \frac{1}{m} \sum_{i=1}^m \left(\frac{x_{ij} - \bar{x}_j}{s_j} \right) \left(\frac{x_{ik} - \bar{x}_k}{s_k} \right) = \frac{1}{m} \sum_{i=1}^m z_{ij} z_{ik} \end{aligned} \quad (44)$$

(bis auf den Faktor $1/m$) ein Skalarprodukt. Die allgemeine Definition eines Skalarprodukts ist:

Definition 2.2 *Es seien $\vec{x} = (x_1, \dots, x_n)'$ und $\vec{y} = (y_1, \dots, y_n)'$ irgend zwei n -dimensionale Vektoren. Dann heißt die Summe der Produkte ihrer Komponenten*

$$\vec{x}'\vec{y} = \langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^n x_i y_i \quad (45)$$

das Skalarprodukt von \vec{x} und \vec{y} .

Anmerkung: Die Schreibweisen $\vec{x}'\vec{y}$ und $\langle \vec{x}, \vec{y} \rangle$ für das Skalarprodukt werden gleichermaßen verwendet.

2.1.3 Vektorlänge und Normierung

Die Vektoren $\vec{\alpha}_j$ und \vec{Y}_i haben eine bestimmte Länge. Sie ergibt sich direkt aus dem Skalarprodukt eines Vektors mit sich selbst, denn es ist etwa

$$\vec{\alpha}_j' \vec{\alpha}_j = \langle \vec{\alpha}_j, \vec{\alpha}_j \rangle = \alpha_{j1}^2 + \alpha_{j2}^2 = \|\vec{\alpha}_j\|^2. \quad (46)$$

Nach dem Satz des Pythagoras ist dies aber das Quadrat der Länge von $\vec{\alpha}_j$. $\|\vec{\alpha}_j\|^2$ ist zunächst wieder nur ein Symbol für $\vec{\alpha}_j' \vec{\alpha}_j$; gelegentlich wird auch $|\vec{\alpha}_j|^2$ dafür geschrieben. Die Länge des Vektors $\vec{\alpha}_j$ ist dann durch

$$\|\vec{\alpha}_j\| = \sqrt{\vec{\alpha}_j' \vec{\alpha}_j} = \sqrt{\langle \vec{\alpha}_j, \vec{\alpha}_j \rangle} \quad (47)$$

gegeben; hier zeigt sich, dass die Einführung des Symbols $\|\cdot\|$ Sinn macht, denn sie zeigt kürzer als $\sqrt{\vec{\alpha}_j' \vec{\alpha}_j}$ an, dass die Länge eines Vektors gemeint ist. Die Länge des Vektors \vec{Y}_i ist dann natürlich durch

$$\|\vec{Y}_i\| = \sqrt{\vec{Y}_i' \vec{Y}_i} \quad (48)$$

gegeben.

Ist die Länge eines Vektors gleich 1, so heißt der Vektor *normiert*. Man kann einen Vektor, der nicht die Länge 1 hat, stets *normieren*. Dazu multipliziert man ihn mit einem Faktor λ , der so gewählt wird, dass

$$\lambda \|\vec{\alpha}\| = 1 \quad (49)$$

gilt. Es folgt sofort, dass

$$\lambda = \frac{1}{\|\vec{\alpha}\|}. \quad (50)$$

Dies bedeutet, dass man einen Vektor normiert, indem man jede seiner Komponenten mit dem Reziprokwert seiner Länge multipliziert. In der Tat ist ja

$$\frac{1}{\|\vec{\alpha}_j\|} \sqrt{\alpha_{j1}^2 + \alpha_{j2}^2} = \sqrt{\left(\frac{\alpha_{j1}}{\|\vec{\alpha}_j\|}\right)^2 + \left(\frac{\alpha_{j2}}{\|\vec{\alpha}_j\|}\right)^2} = 1.$$

Setzt man nun

$$\beta_{j1} = \frac{\alpha_{j1}}{\|\vec{\alpha}_j\|}, \quad \beta_{j2} = \frac{\alpha_{j2}}{\|\vec{\alpha}_j\|}, \quad (51)$$

so erhält man den normierten Vektor $\vec{\beta}_j = (\beta_{j1}, \beta_{j2})'$. Man normiert also einen Vektor, indem man jede seiner Komponenten durch die Länge des Vektors teilt.

Allgemein gilt: ist $\vec{x} = (x_1, \dots, x_n)'$ ein n -dimensionaler Vektor, so ist (nach dem Satz des Pythagoras) seine Länge durch

$$\|\vec{x}\| = \sqrt{\sum_{i=1}^n x_i^2} \quad (52)$$

gegeben. Der Vektor wird *normiert*, indem man seine Komponenten durch $\|\vec{x}\|$ dividiert:

$$\vec{x}_0 = \left(\frac{x_1}{\|\vec{x}\|}, \frac{x_2}{\|\vec{x}\|}, \dots, \frac{x_n}{\|\vec{x}\|} \right)'. \quad (53)$$

Man rechnet leicht nach, dass nun $\|\vec{x}_0\| = 1$ gilt.

2.1.4 Das Skalarprodukt und der Winkel zwischen Vektoren

Wenn sich Vektoren durch Pfeile darstellen lassen, so macht es sicher Sinn, den Winkel, den die beiden Vektoren bilden, zu betrachten. Hierzu zieht man den bekannten Kosinussatz heran. Ein Dreieck habe die Seiten a , b und c . Man kann diese Seiten als Vektoren auffassen, deren Längen durch $\|\vec{\alpha}\|$, $\|\vec{\beta}\|$ und $\|\vec{\gamma}\|$ gegeben seien. Ist ϕ der Winkel zwischen den Seiten, d.h. Vektoren $\vec{\alpha}$ und $\vec{\beta}$, so gilt der

Kosinussatz:

$$\|\vec{\gamma}\|^2 = \|\vec{\alpha}\|^2 + \|\vec{\beta}\|^2 - 2\|\vec{\alpha}\|\|\vec{\beta}\| \cos \phi. \quad (54)$$

Nun ist aber $\vec{\gamma} = \vec{\alpha} - \vec{\beta}$, und bestimmt man $\|\vec{\gamma}\|^2 = \|\vec{\alpha} - \vec{\beta}\|^2$ und vereinfacht dann (54), so erhält man die Beziehung

$$\cos \phi = \frac{\langle \vec{\alpha}, \vec{\beta} \rangle}{\|\vec{\alpha}\| \|\vec{\beta}\|} \quad (55)$$

bzw.

$$\langle \vec{\alpha}, \vec{\beta} \rangle = \|\vec{\alpha}\| \|\vec{\beta}\| \cos \phi. \quad (56)$$

Aus dieser Beziehung erhält man sofort eine Aussage über eine geometrische Beziehung zwischen den Vektoren $\vec{\alpha}$ und $\vec{\beta}$. Denn der Wert des Skalarprodukts für zwei gegebene Vektoren hängt offenbar vom Winkel ϕ zwischen ihnen ab. Der maximal mögliche Wert von $\cos \phi$ ist 1, und der wird für den Winkel $\phi = 0$ angenommen. Dann hat man jedenfalls

$$\max[\langle \vec{\alpha}, \vec{\beta} \rangle] = \|\vec{\alpha}\| \|\vec{\beta}\|. \quad (57)$$

Das max steht für Maximum, und das Maximum des Skalarprodukts $\langle \vec{\alpha}, \vec{\beta} \rangle$ ist nicht das absolute Maximum, - das wäre unendlich und würde erreicht, wenn mindestens eine der beiden Vektorlängen $\|\vec{\alpha}\|$ oder $\|\vec{\beta}\|$ unendlich wäre und die andere nicht gleich Null ist. Das hier gemeinte Maximum ist das Maximum des Skalarprodukts für *gegebene* Vektorlängen $\|\vec{\alpha}\|$ und $\|\vec{\beta}\|$.

Der minimale Wert von $\cos \phi$ ist -1, und dieser Wert wird für den Winkel $3\pi/2$ angenommen, und man hat

$$\min \langle \vec{\alpha}, \vec{\beta} \rangle = -\|\vec{\alpha}\| \|\vec{\beta}\|. \quad (58)$$

Für $\phi = \pi/2$ (d.h für einen rechten Winkel von 90^0) ist $\cos \phi = 0$, so dass

$$\langle \vec{\alpha}, \vec{\beta} \rangle = 0 \text{ genau dann, wenn } \phi = \pi/2. \quad (59)$$

Dieser Fall tritt also genau dann ein, wenn die Vektoren senkrecht aufeinander stehen, also einen *rechten Winkel* (von 90^0) bilden. Die Vektoren $\vec{\alpha}$ und $\vec{\beta}$ heißen dann *orthogonal* (zueinander). Man schreibt dafür auch kurz $\vec{\alpha} \perp \vec{\beta}$.

Anwendung: Diese Resultate können auf die Charakterisierung der Messwerte x_{ij} als Skalarprodukte - vergl. (41) und (43) - angewendet werden. Denn $x_{ij} = \langle \vec{\alpha}_j, \vec{Y}_i \rangle$ und nach (56) kann man dann

$$x_{ij} = \|\vec{\alpha}_j\| \|\vec{Y}_i\| \cos \phi \quad (60)$$

schreiben; ϕ ist der Winkel zwischen den Vektoren $\vec{\alpha}_j$ und \vec{Y}_i . x_{ij} nimmt dann nach (57) den maximal bzw. minimal möglichen Wert

$$\max[x_{ij}] = \|\vec{\alpha}_j\| \|\vec{Y}_i\|, \quad \min[x_{ij}] = -\|\vec{\alpha}_j\| \|\vec{Y}_i\| \quad (61)$$

an. Diese "extremen" Werte für x_{ij} hängen also einerseits davon ab, wie lang der Vektor $\vec{\alpha}_j$ ist, d.h. ab er von den Anteilen α_{j1} und α_{j2} , mit denen die

Variable X_j überhaupt die zugrundeliegenden latenten Dimensionen erfasst, und von der Länge des Vektors \vec{Y}_i , der die i -te Person bezüglich der latenten Dimension charakterisiert, d.h. von den Anteilen F_{i1} und F_{i2} , die die Ausstattung der i -ten Person mit den latenten Dimensionen. Insbesondere gilt

$$x_{ij} = 0 \text{ genau dann, wenn } \vec{\alpha}_j \perp \vec{Y}_i. \quad (62)$$

Die i -te Person liefert also den Messwert $x_{ij} = 0$ in der j -ten Variable, wenn der Variablenvektor $\vec{\alpha}_j$ und der Personvektor \vec{Y}_i orthogonal sind. Ein Spezialfall dieses Falles ist, wenn entweder $\alpha_{j1} = \alpha_{j2} = 0$, die Variable X_j die latenten Dimensionen also gar nicht erfasst, oder $F_{i1} = F_{i2} = 0$, die i -te Person also die Merkmale, die die latenten Dimensionen repräsentieren, gar nicht hat, oder wenn $\vec{\alpha}_j = \vec{Y}_i = \vec{0}$. Das eigentlich Bemerkenswerte an dem Resultat (62) ist aber, dass weder $\vec{\alpha}_j = \vec{0}$ noch $\vec{Y}_i = \vec{0}$ zutreffen müssen, damit der Fall $x_{ij} = 0$ eintritt. Es genügt, wenn die i -te Person und die j -te Variable gewissermaßen inkompatibel miteinander sind. Diese Inkompatibilität wird deutlicher, wenn man sich daran erinnert, dass x_{ij} ja auch als Korrelation gedeutet werden kann.

Häufig werden aber nicht die Messwerte selbst, sondern ihre standardisierten Werte betrachtet. Statt x_{ij} wird also der Wert

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (63)$$

für die Datenanalyse zugrunde gelegt, wobei s_j die Standardabweichung der j -ten Variablen ist. Auch hier gilt dann die Beziehung

$$z_{ij} = q_{i1}a_{j1} + q_{i2}a_{j2} = \langle \vec{q}_i, \vec{a}_j \rangle, \quad (64)$$

vergl. (22). Ist nun $\langle \vec{q}_i, \vec{a}_j \rangle = z_{ij}$, so heißt dies natürlich wieder, dass $\vec{q}_i \perp \vec{a}_j$, aber es bedeutet nicht, dass $x_{ij} = 0$, sondern dass $x_{ij} = \bar{x}_j$.

Zwischenbetrachtung: die Cauchy-Schwarzsche Ungleichung Das Ergebnis (61) führt zu einer Herleitung der *Cauchy-Schwarzschen Ungleichung*, die für viele Betrachtungen nützlich ist. Es seien (a_1, \dots, a_n) und (b_1, \dots, b_n) irgendwelche reellen Zahlen. Dann gilt die folgende Ungleichung:

$$\left| \sum_{i=1}^n a_i b_i \right|^2 \leq \sum_{i=1}^n |a_i|^2 \sum_{i=1}^n |b_i|^2, \quad (65)$$

bzw.

$$\left| \sum_{i=1}^n a_i b_i \right| \leq \sqrt{\sum_{i=1}^n |a_i|^2} \sqrt{\sum_{i=1}^n |b_i|^2}, \quad (66)$$

wobei das Gleichheitszeichen genau dann gilt, wenn $a_i = \alpha b_i$ für alle i .

Denn es sei $(a_1, \dots, a_n) = \vec{a}$ und $(b_1, \dots, b_n) = \vec{b}$. Dann ist $\sum_{i=1}^n a_i b_i$ gerade das Skalarprodukt von \vec{a} und \vec{b} , und

$$\sum_i a_i^2 = \|\vec{a}\|^2, \quad \sum_i b_i^2 = \|\vec{b}\|^2,$$

und nach (56) gilt

$$\langle \vec{a}, \vec{b} \rangle = \sum_{i=1}^n a_i b_i = \|\vec{a}\| \|\vec{b}\| \cos \phi \leq \|\vec{a}\| \|\vec{b}\|,$$

wenn ϕ der Winkel zwischen \vec{a} und \vec{b} ist, und $\cos \phi \leq 1$; dies ist aber gerade (66), und (65) folgt ebenfalls. $|\langle \vec{a}, \vec{b} \rangle|$ wird maximal für $\phi = 0$, wenn also \vec{a} und \vec{b} die gleiche Orientierung haben. Die beiden Vektoren unterscheiden sich dann höchstens durch ihre Länge, d.h. es muß in diesem Fall

$$a_i = \alpha b_i \quad (67)$$

gelten. \square

2.1.5 Korrelation und Skalarprodukt

Es ist

$$r_{jk} = \frac{1}{m} \sum_{i=1}^m z_{ij} z_{ik} = \frac{1}{m} \langle \vec{z}_j, \vec{z}_k \rangle \quad (68)$$

schreiben kann. \vec{z}_j ist der Vektor mit den Komponenten $(x_{ij} - \bar{x}_j)/s_j$, und \vec{z}_k hat die Komponenten $(x_{ik} - \bar{x}_k)/s_k$. Wendet man (56) auf r_{jk} an, so erhält man

$$r_{jk} = \frac{1}{m} \|\vec{z}_j\| \|\vec{z}_k\| \cos \phi_{jk}, \quad (69)$$

und ϕ ist der Winkel zwischen den m -dimensionalen Vektoren \vec{z}_j und \vec{z}_k . Nun ist bekanntlich

$$\|\vec{z}_j\|^2 = \sum_{i=1}^m z_{ij}^2 = \sum_{i=1}^m \frac{(x_{ij} - \bar{x}_j)^2}{s_x^2} = \frac{1}{s_x^2} \sum_{i=1}^m (x_{ij} - \bar{x}_j)^2 = \frac{1}{s_x^2} m s_x^2 = m.$$

Analog dazu ist

$$\|\vec{z}_k\|^2 = \sum_{i=1}^m z_{ik}^2 = m,$$

so dass $\|\vec{z}_j\| = \|\vec{z}_k\| = \sqrt{m}$, und damit hat man

$$r_{jk} = \frac{1}{m} m \cos \phi_{jk} = \cos \phi_{jk}. \quad (70)$$

Der Korrelationskoeffizient ist also gerade gleich dem Kosinus des Winkels zwischen den Vektoren, deren Komponenten die standardisierten Messwerte sind! Man findet sofort $\max r_{jk} = 1$ und $\min r_{jk} = -1$, und insbesondere $r_{jk} = 0$ genau dann, wenn \vec{z}_j und \vec{z}_k orthogonal zueinander sind. Alternativ kann man von (66) ausgehen:

$$|r_{jk}| \leq \frac{1}{m} \left| \sum_{i=1}^m |z_{ij}|^2 \right|^{1/2} \left| \sum_{i=1}^m |z_{ik}|^2 \right|^{1/2} = \frac{1}{m} \sqrt{m} \sqrt{m} = 1,$$

d.h. es muß $-1 \leq r_{jk} \leq 1$ gelten.

Nach (21) und (22) sind x_{ij} bzw. z_{ij} Skalarprodukte, wenn man die Fehler e_{ij} und ε_{ij} vernachlässigt. Setzt man $s = n$, so erhält man

$$\begin{aligned} x_{ij} &= \alpha_{j1}F_{i1} + \alpha_{j2}F_{i2} + \cdots + \alpha_{jn}F_{in} + e_{ij}, \quad j = 1, \dots, n \\ z_{ij} &= a_{j1}q_{i1} + a_{j2}q_{i2} + \cdots + a_{jn}q_{in} + \varepsilon_{ij}. \end{aligned}$$

Es folgt dann, dass $x_{ij} = 0$ bzw. $z_{ij} = 0$ genau dann, wenn die Vektoren $(\alpha_{j1}, \dots, \alpha_{jn})'$ und $(F_{i1}, \dots, F_{in})'$ bzw. $(a_{j1}, \dots, a_{jn})'$ und $(q_{i1}, \dots, q_{in})'$ orthogonal sind. Salopp gesprochen ist dies dann der Fall, wenn die persönliche Ausstattung einer Person mit den latenten Variablen einerseits und die Erfassung eben dieser latenten Variablen durch einen "Test" andererseits nicht miteinander korrelieren. Repräsentiert man die Variable X_j also durch einen Punkt mit den Koordinaten $\{\alpha_{j1}, \alpha_{j2}\}$ und nimmt man diesen Punkt als Endpunkt des Personenvektors \vec{Y}_i , so ist $x_{ij} = 0$ immer dann, wenn $\vec{Y}_i \perp \vec{x}_j$, und $x_{ij} = \max$ genau dann, wenn der Winkel zwischen \vec{Y}_i und \vec{x}_j gleich Null ist. Der Wert für x_{ij} wird minimal genau dann, wenn die beiden Vektoren genau in zueinander entgegengesetzte Richtungen zeigen. Die Betrachtungen für z_{ij} sind analog.

2.1.6 Vektorräume, Basisvektoren und latente Variablen

Vorbetrachtungen Um die Einführung der in diesem Abschnitt vorgestellten Begriffe zu motivieren, sollen die Grundgedanken des faktoranalytischen Modells noch einmal in hzusammengefasster Form vorgestellt werden:

1. Gesucht ist die kleinste Anzahl $r \leq \min(m, n)$ latenter Variablen, mit denen sich die Daten "erklären" lassen. m ist die Anzahl von Personen (Probanden, Patienten, Befragte) oder Objekten (Scherben bei einer archäologischen Grabung) etc., n ist die Anzahl der gemessenen Merkmale (Fragen in einem Test oder Fragebogen, Positionen von Elektroden bei einer EEG-Untersuchung, Eigenschaften der gefundenen Scherben, etc).
2. Die latenten Variablen sollen durch m -dimensionale Vektoren $\vec{F}_1, \vec{F}_2, \dots, \vec{F}_r$ repräsentiert werden. Die latenten Variablen sollen nicht redundant sein, d.h. sie sollen Merkmale repräsentieren, die nicht auseinander vorhergesagt werden können.

Die Nichtredundanz der latenten Variablen kann durch Eigenschaften der Vektoren $\vec{F}_1, \vec{F}_2, \dots, \vec{F}_r$ ausgedrückt werden. Speziell ist damit gemeint, dass die entsprechenden Eigenschaften nicht auseinander vorhergesagt werden können. Diese mangelnde Vorhersagbarkeit kann durch eine Art Regressionsbedingung ausgedrückt werden: es soll *nicht* möglich sein, einen der Vektoren $\vec{F}_k, 1 \leq k \leq r$ als Kombination (gewogene Summe) der anderen auszudrücken, also *es soll*

nicht gelten, dass für irgendeinen der Vektoren \vec{F}_k eine Aussage der Art

$$\vec{F}_k = \lambda_1 \vec{F}_1 + \cdots + \lambda_{k-1} \vec{F}_{k-1} + \lambda_{k+1} \vec{F}_{k+1} \cdots + \lambda_r \vec{F}_r \quad (71)$$

gilt.

Anmerkung: Es sei gleich angemerkt, dass diese Bedingung *nicht* bedeutet, dass die F_1, \dots, F_r auch unkorreliert sind (s. unten). \square

Bringt man in (71) \vec{F}_k auf die rechte Seite, so erhält man die Vektorgleichung

$$\vec{0} = \lambda_1 \vec{F}_1 + \lambda_2 \vec{F}_2 + \cdots + \lambda_r \vec{F}_r, \quad (72)$$

wobei $\vec{0}$ der Nullvektor ist, d.h. ein Vektor ist, dessen Komponenten alle gleich Null sind. Ließe sich \vec{F}_k als gewogene Summe der übrigen Vektoren darstellen, wäre $\lambda_k = -1$ in (72). Um die Bedeutung der Forderung (71) zu sehen werde für den Moment angenommen, dass es möglich ist, \vec{F}_k wie in (71) auszudrücken, - es gelte also

$$\vec{F}_k = \lambda_1 \vec{F}_1 + \cdots + \lambda_{k-1} \vec{F}_{k-1} + \lambda_{k+1} \vec{F}_{k+1} \cdots + \lambda_r \vec{F}_r.$$

\vec{F}_k wird dann als "gewogene" Summe der übrigen Vektoren, die alle latente Variablen repräsentieren, dargestellt. Repräsentieren die \vec{F}_k bestimmte Merkmale M_1, \dots, M_r , so bedeutet dieser Ausdruck, dass die Ausprägungen von etwa M_k bei den Personen oder Objekten durch die Ausprägungen der übrigen Merkmale berechnet und insofern vorhergesagt werden können. Die Information über M_k ist dann in den übrigen Merkmalen enthalten. Damit ist aber M_k kein Merkmal mehr, das zur "Erklärung" der gemessenen Variablen, also der Vektoren \vec{X}_j , benötigt wird, denn die Information, die in M_k enthalten ist, ist ja bereits in den anderen Merkmalen $M_{k'}$, $k' \neq k$, enthalten. Damit also M_k einen Beitrag zur Erklärung der \vec{X}_j leisten kann, sollte \vec{F}_k eben nicht durch die übrigen Vektoren vorhergesagt werden können. Die Nichtvorhersagbarkeit bedeutet, dass es keine Koeffizienten $\lambda_{k'}$, $k' \neq k$ gibt derart, dass \vec{F}_k als gewogene Summe der übrigen Vektoren darstellbar ist. Mit Bezug auf die Gleichung (72) bedeutet die Nichtredundanz, dass alle Koeffizienten λ_k gleich Null sind, wie weiter unten noch erläutert werden wird.

Die Standardisierung der \vec{X}_j bedeutet, dass auch die \vec{F}_k standardisiert sind, d.h. die Mittelwerte der Komponenten sind jeweils gleich Null und die Varianz der Komponenten ist jeweils gleich 1. Die durch m dividierten Skalarprodukte $\vec{F}_k' \vec{F}_{k'}/m$ entsprechen dann Korrelationen zwischen den latenten Variablen. Die Forderung (71) bedeutet noch *nicht*, dass die Korrelationen zwischen den latenten Variablen alle gleich Null sind! Denn angenommen, es gilt (72) mit $\lambda_1 = \lambda_2 = \cdots = \lambda_r = 0$, d.h. keiner der Vektoren ist aus den anderen vorhersagbar. Man multipliziere die Gleichung mit \vec{F}_k' :

$$\vec{F}_k' \vec{0} = \lambda_1 \vec{F}_k' \vec{F}_1 + \cdots + \lambda_r \vec{F}_k' \vec{F}_r.$$

Sicherlich ist $\vec{F}_k' \vec{0} = 0$ (d.h. jeder Vektor \vec{F}_k ist orthogonal zu $\vec{0}$). Aber da $\lambda_k = 0$ für alle k gilt $\vec{F}_k' \vec{0} = 0$ unabhängig davon, ob die Skalarprodukte und damit die Korrelationen $\vec{F}_k' \vec{F}_{k'}$ von Null verschieden sind oder nicht.

Man kann natürlich postulieren, dass $\vec{F}'_k \vec{F}_{k'} = 0$ für alle $k \neq k'$, d.h. dass die $\{\vec{F}_k\}$ paarweise orthogonal sind. Es läßt sich dann zeigen (s. unten), dass dann auch stets (71) erfüllt ist. Die Unkorreliertheit, d.h. die Orthogonalität der \vec{F}_k ist also eine schärfere Forderung; (71) ist auch mit "obliquen" und damit korrelierten Lösungen für die \vec{F}_k kompatibel.

Vektorräume: Die Betrachtungen zur minimalen Anzahl r von latenten Variablen und zur Redundanz bzw. Nichtredundanz der \vec{F}_k lassen sich durch bestimmte Begriffsbildungen der linearen Algebra (Vektoralgebra) leicht formulieren. Dazu gehören die Begriffe der linearen Abhängigkeit bzw. Unabhängigkeit von Vektoren sowie des Vektorraumes und des Teilraumes eines Vektorraumes. Über diese Begriffe lassen sich dann Wege aufzeigen, die zur Bestimmung der latenten Variablen und ihrer "Gewichte" führen.

Auf Seite 19 wurden einige Rechenregeln für Vektoren eingeführt: die Multiplikation eines Vektors mit einem Skalar, die Addition von Vektoren und die Linearkombination von Vektoren. Zunächst wird der Begriff des *Vektorraumes* eingeführt. Ein Vektorraum ist eine Menge von Vektoren, üblicherweise von gleicher Dimensionalität, d.h. von Vektoren mit gleicher Anzahl von Komponenten, wobei alle Linearkombinationen wieder Element dieser Menge sein sollen. Diese Einschränkung ist wesentlich. Eine beliebig definierte Menge von Vektoren bildet noch nicht notwendig einen Vektorraum: So betrachte man einen Kreis, dessen Mittelpunkt mit dem Nullpunkt eines 2-dimensionalen Koordinatensystems zusammenfällt. Weiter betrachte man die Menge der Vektoren, deren Anfangspunkt ebenfalls im Nullpunkt des Koordinatensystems liegt und deren Endpunkte auf dem Kreis liegen. Diese Menge ist kein Vektorraum. Denn wenn man eine Linearkombination aus irgendzwei dieser Vektoren bildet, wird ihr Endpunkt nicht notwendig ebenfalls auf dem Kreis liegen, dh die Linearkombination ist kein Element der Menge der Vektoren mit Endpunkt auf dem Kreisumfang (wenn man ihren Anfangspunkt in den Nullpunkt des Koordinatensystems verschiebt). Die gleiche Argumentation zeigt, dass diese Menge von Vektoren auch kein Teilraum eines Vektorraumes ist, für den ja Geschlossenheit in Bezug auf die Verknüpfung von Vektoren gelten soll, d.h. Linearkombinationen von Vektoren aus dem Teilraum sollen wieder Elemente des Teilraums sein.

Man kann dann bestimmte Vektoren wählen derart, dass alle Vektoren des Vektorraumes sich als Linearkombination dieser *Basisvektoren* darstellen lassen. Die die latenten Variablen repräsentierenden Vektoren $\vec{F}_1, \dots, \vec{F}_r$ erweisen sich als solche Basisvektoren. Da die Vektoren \vec{X}_j m -dimensionale Vektoren sind, aber $r < m$ sein soll, bilden die $\vec{F}_1, \dots, \vec{F}_r$ aber nur die Basis eines Teilraumes. Wie gezeigt werden wird, lassen sich nicht *alle* m -dimensionalen Vektoren als Linearkombination der $\vec{F}_1, \dots, \vec{F}_r$ darstellen, wenn $r < m$ ist. Der Eindeutigkeit der Begriffsbildung wegen wird noch eine formale Definition eines Vektorraumes und eines Teilraums gegeben:

Definition 2.3 Ein Vektorraum ist eine Menge V von Vektoren \vec{x} , für die mit $\lambda, \mu \in \mathbb{R}$ gilt

- (i) $\vec{x} \in V$, so auch $(\lambda + \mu)\vec{x} = \lambda\vec{x} + \mu\vec{x} \in V$,
- (ii) $\vec{x} \in V$, dann auch $(\lambda\mu)\vec{x} = \lambda(\mu\vec{x})$,
- (iii) $\vec{x}_1, \vec{x}_2 \in V$, dann auch $\lambda(\vec{x}_1 + \vec{x}_2) = \lambda\vec{x}_1 + \lambda\vec{x}_2 \in V$.

Es sei $V_0 \subset V$; V_0 ist ein Teilvektorraum von V , wenn für alle Vektoren aus V_0 wieder die Bedingungen (i) bis (iii) gelten, wobei V durch V_0 ersetzt wird.

Anmerkungen:

1. **n -dimensionaler Vektorraum** Nach Definition 2.3 sind alle Linearkombinationen von Vektoren aus einem Vektorraum ebenfalls Elemente eben dieses Vektorraums. Auf den ersten Blick mag diese Definition ein wenig leer erscheinen. Man muß aber bedenken, dass die Addition von Vektoren nur Sinn macht, wenn die zu addierenden Vektoren die gleiche Anzahl von Komponenten haben. Ist diese Anzahl gleich n , n irgendeine natürliche Zahl, so spricht man dementsprechend von einem n -dimensionalen Vektorraum V_n . Eine Linearkombination von n -dimensionalen Vektoren ist demnach wieder ein n -dimensionaler Vektor, – und nicht etwa ein $(n - r)$ -dimensionaler oder $(n + s)$ -dimensionaler Vektor, $r < n$, $s > 0$. Eine Menge von Vektoren zusammen mit einer Regel, derzufolge Kombinationen von Vektoren zwar wieder zu Vektoren führt, deren Dimensionalität aber eine andere als die der kombinierten Vektoren ist, ist also *kein* Vektorraum im Sinne der Definition 2.3.

Betrachtet man den Vektor \vec{X}_j , dessen Komponenten die Messwerte von m Personen für ein Merkmal M_j ($j = 1, \dots, n$) sind, so ist \vec{X}_j ein m -dimensionaler Vektor, der ein Element eines m -dimensionalen Vektorraums ist. Jede Komponente repräsentiert eine der insgesamt m Dimensionen, d.h. jede Person steht für eine Dimension. Die latenten Merkmale werden dann ebenfalls durch m -dimensionale Vektoren \vec{F}_k repräsentiert, denn jede Person hat eine bestimmte Ausprägung des Merkmals M_k . Ebenso kann man den Vektor \vec{Y}_i betrachten, dessen Komponenten die Messwerte der i -ten Person auf den verschiedenen Merkmalen M_1, \dots, M_n sind. \vec{Y}_i ist ein n -dimensionaler Vektor, also ein Element eines n -dimensionalen Vektorraumes. Die einzelnen Komponenten repräsentieren dann die *gemessenen* Merkmale; jedes Merkmal definiert eine Dimension. Es wird noch gezeigt werden, dass die Probandendimensionen und die Merkmalsdimensionen in einer bestimmten Abhängigkeitsbeziehung zueinander stehen.

2. **Basis eines Vektorraumes** Es seien⁴ $\mathbf{b}_1, \dots, \mathbf{b}_n$, $r \leq n$ n -dimensionale

⁴Vektoren werden entweder durch einen Pfeil über einem Buchstaben bezeichnet, wie in \vec{X} , oder durch fette Schreibweise, zB \mathbf{x} .

Vektoren. Lassen sich aus diesen Vektoren *alle* n -dimensionalen Vektoren als Linearkombination darstellen, so heißt $\mathbf{b}_1, \dots, \mathbf{b}_n$ eine *Basis* des n -dimensionalen Vektorraums. Damit die $\mathbf{b}_1, \dots, \mathbf{b}_n$ eine Basis bilden, müssen sie eine bestimmte Eigenschaft haben: sie müssen *linear unabhängig* sein, worauf weiter unten noch explizit eingegangen wird (s. Def. 2.4).

3. **Teilbasis eines Vektorraums** Es sei $\mathbf{b}_1, \dots, \mathbf{b}_n$ eine Basis des n -dimensionalen Vektorraums. Dann ist $\mathbf{b}_1, \dots, \mathbf{b}_r$, $r < n$, eine *Teilbasis* des V_n , die einen Teilraum V_0^r des V_n im Sinne der Definition (2.3) definiert. Der Punkt bei dieser Definition ist, dass keine der Linearkombinationen – also der Vektoren, die sich als Linearkombination der $\mathbf{b}_1, \dots, \mathbf{b}_r$ darstellen lassen – in dem Teilraum des V_n liegt, der zu V_0^r komplementär ist; durch Linearkombination von Vektoren aus V_0^r kommt man gewissermaßen nicht aus dem V_0^r heraus. Dieser Sachverhalt wird weiter unten noch elaboriert.

Der Begriff des Teilraums ist für die Faktorenanalyse wesentlich: wie weiter unten noch elaboriert werden wird, impliziert dieser Versuch, die n m -dimensionalen Vektoren $\vec{X}_1, \dots, \vec{X}_n$ durch latente Variablen F_1, \dots, F_r , d.h. durch Vektoren $\vec{F}_1, \dots, \vec{F}_r$ mit $r < n$ "erklären", dass die \vec{X}_j , $j = 1, \dots, n$, in einem r -dimensionalen Teilraum des m -dimensionalen Vektorraumes liegen, und dass die $\vec{F}_1, \dots, \vec{F}_r$ eine Basis dieses Teilraumes bilden. Dazu müssen die $\vec{F}_1, \dots, \vec{F}_r$ linear unabhängig sein.

Es ist oben angemerkt worden, dass Vektoren linear unabhängig sein müssen, damit sie die Basis oder Teilbasis eines Vektorraums bilden können. Vektoren sind linear unabhängig, wenn keiner von ihnen als Linearkombination der übrigen dargestellt werden kann; lineare Unabhängigkeit ist also ein wesentliches Merkmal für Vektoren $\vec{F}_1, \dots, \vec{F}_r$, die latente Merkmale repräsentieren sollen. Formal wird der Begriff der linearen Unabhängigkeit in der folgenden Definition charakterisiert:

Definition 2.4 Gegeben seien r n -dimensionale Vektoren⁵ $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_r$. Gilt

$$\vec{0} = \lambda_1 \vec{x}_1 + \dots + \lambda_r \vec{x}_r, \quad (73)$$

dann und nur dann⁶, wenn $\lambda_1 = \lambda_2 = \dots = \lambda_r = 0$, so heißen die $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_r$ linear unabhängig, andernfalls heißen sie linear abhängig.

Erläuterung: Zunächst werde der Begriff der linearen Abhängigkeit erläutert. Dazu werde angenommen, der der Vektoren \vec{x}_j sei als Linearkombination der übrigen Vektoren darstellbar, etwa \vec{x}_1 . Dann soll also gelten

$$\vec{x}_1 = \lambda_2 \vec{x}_2 + \dots + \lambda_r \vec{x}_r.$$

⁵ n ist hier eine allgemeine Dimensionalitätsbezeichnung, – man kann ebensogut m - oder r -dimensionale Vektoren betrachten.

⁶ $\vec{0}$ ist der Nullvektor, seine Komponenten sind alle gleich Null.

Eine solche Beziehung ist aus der multiplen Regression bekannt: man sagt $\vec{y} = \vec{x}_1$ anhand der "Prädiktoren" $\vec{x}_2, \dots, \vec{x}_r$ "voraus". Die $\lambda_2, \dots, \lambda_r$ sind dann Regressionskoeffizienten. Das Faktorenmodell

$$\vec{x}_j = \lambda_{j2}\vec{F}_1 + \dots + \lambda_{jr}\vec{F}_r + \varepsilon_j$$

ist ebenfalls von dieser Struktur. Die "Voraussage" von \vec{x}_1 oder allgemein von \vec{x}_j anhand anderer Vektoren besagt ja gerade, dass \vec{x}_1 oder \vec{x}_j als eine von den übrigen Vektoren abhängige Größe betrachtet wird. Damit man auf die Form (73) kommt, muß man \vec{x}_1 oder \vec{x}_j nur auf die rechte Seite bringen:

$$\vec{0} = -\vec{x}_1 + \lambda_2\vec{x}_2 + \dots + \lambda_r\vec{x}_r;$$

mit $\lambda_1 = -1$ hat man gerade (73). Damit die Voraussage gelingt, dürfen natürlich die Koeffizienten λ_j nicht alle gleich Null sein. Wegen $\lambda_1 = -1$ sind in der Tat nicht alle Koeffizienten gleich Null. Der Sachverhalt impliziert natürlich sofort, dass noch mindestens ein anderer λ -Wert von Null verschieden sein muß, damit \vec{x}_1 als Linearkombination der übrigen Vektoren darstellbar ist.

Nun werde der Fall betrachtet, dass \vec{x}_1 *nicht* als Linearkombination der übrigen \vec{x}_j darstellbar und damit linear unabhängig von der übrigen Vektoren ist. Wenn von Null verschiedene λ_j -Werte lineare Abhängigkeit bedeuten bzw implizieren, dann folgt⁷ aus der linearen Unabhängigkeit der Vektoren, dass $\lambda_j = 0$ für alle j gilt. Keiner der Vektoren \vec{x}_j ist dann als Linearkombination der übrigen Vektoren darstellbar. Zum Beispiel betrachte man die Variablen IQ (= Intelligenzquotient, wie er in einem Intelligenztest erfasst wird), Stirnhöhe, Augenabstand, Kopfumfang, Länge der Nase, Ausgeprägtheit des Kinns. Wer ein rechter Phrenologe sein will versucht, den IQ anhand dieser Variablen vorherzusagen. Er mißt also bei m Personen den IQ und fasst die Werte in einem Vektor \vec{x}_1 zusammen. Ebenso mißt er bei diesen Personen die Stirnhöhe (\vec{x}_2), den Augenabstand (\vec{x}_2), den Kopfumfang (\vec{x}_3), die Länge der Nase (\vec{x}_4) und die Ausgeprägtheit des Kinns (\vec{x}_5) und sucht nun die Koeffizienten λ_2 bis λ_5 , die die Voraussage

$$\vec{x}_1 = \lambda_2\vec{x}_2 + \dots + \lambda_5\vec{x}_5$$

erlauben. Er wird lange suchen und keine Werte für die λ_j finden, denn die anatomischen Maße enthalten keine Information über die Intelligenz, wie man heute weiß. \vec{x}_1 ist von den übrigen Vektoren linear unabhängig.

Die Forderung, dass die latenten Variablen repräsentierenden Vektoren $\vec{F}_1, \dots, \vec{F}_r$ linear unabhängig sind, ist also verständlich: wenn keiner dieser Vektoren als Linearkombination der übrigen dargestellt werden kann, so bedeutet dies, dass jeder dieser Vektoren Information repräsentiert, die nicht auch schon von den anderen Vektoren repräsentiert wird. Die Definition der linearen Abhängigkeit bzw. Unabhängigkeit hat weitere Implikationen:

⁷Dieser Schluß ist eine Anwendung des *modus tollens*: wenn p und q irgendzwei Aussagen sind und "Wenn p , dann auch q " gilt, so folgt aus nicht- q ($\neg q$), dass nicht- p ($\neg p$) gilt, d.h. $p \rightarrow q \Rightarrow \neg q \rightarrow \neg p$. Wenn es regnet (p), dann ist die Strasse nass (q). Nun ist aber die Strasse trocken ($\neg q$), ergo kann es nicht regnen ($\neg p$).

1. Sind die Vektoren $\vec{x}_1, \dots, \vec{x}_n$ paarweise orthogonal, so sind sie auch l.u.; sind sie l.u., so sind sie nicht notwendig auch orthogonal.

Beweis: Denn es gelte

$$\vec{0} = \lambda_1 \vec{x}_1 + \lambda_2 \vec{x}_2 + \dots + \lambda_n \vec{x}_n,$$

und die Vektoren seien paarweise orthogonal. Dann kann man das Skalarprodukt von $\vec{0}$ mit z.B. \vec{x}_1 bilden:

$$\vec{x}_1' \vec{0} = \lambda_1 \vec{x}_1' \vec{x}_1 + \lambda_2 \vec{x}_1' \vec{x}_2 + \dots + \lambda_n \vec{x}_1' \vec{x}_n = 0,$$

denn notwendig $\vec{x}_1' \vec{0} = 0$. Außerdem muß $\vec{x}_1' \vec{x}_2 = \dots = \vec{x}_1' \vec{x}_n = 0$ gelten, wegen der postulierten Orthogonalität. Dann folgt $0 = \lambda_1 \|\vec{x}_1\|^2$, und $\|\vec{x}_1\|^2 \neq 0$ nach Voraussetzung, so dass $\lambda_1 = 0$ gelten muß. In dieser Weise fährt man fort mit \vec{x}_2, \vec{x}_3 , etc, und findet so, dass alle $\lambda_j = 0$ sein müssen; also sind die \vec{x}_j auch l.u. □

2. Die n Einheitsvektoren $\vec{e}_j = (0, \dots, 0, 1, 0, \dots, 0)'$, die 1 an der j -ten Stelle, $j = 1, \dots, n$, sind l.u. Das sieht man sofort, denn aus $\vec{0} = \lambda_1 \vec{e}_1 + \dots + \lambda_n \vec{e}_n$ folgt sofort für die j -te Komponente von $\vec{0}$ die Gleichung $0 = \lambda_j 1$ und damit $\lambda_j = 0$, für alle j . Darüber hinaus sind die \vec{e}_j orthogonal. Z.B. ist

$$\vec{e}_1' \vec{e}_2 = 1 \cdot 0 + 0 \cdot 1 + 0 \cdot \dots + 0 = 0,$$

also sind \vec{e}_1 und \vec{e}_2 orthogonal. Analog zeigt man, das \vec{e}_j und \vec{e}_k für $j \neq k$ orthogonal sind.

3. Die Vektoren \vec{u}_j und \vec{u}_k haben verschiedene Orientierungen. Dann sind die beiden Vektoren linear unabhängig.

Beweis: Es gelte $\vec{0} = \lambda_j \vec{u}_j + \lambda_k \vec{u}_k$. Angenommen, es gelte $\lambda_j \neq 0$, $\lambda_k \neq 0$. Dann folgt $\vec{u}_j = -(\lambda_k/\lambda_j) \vec{u}_k$; aber dieses Resultat bedeutet, daß \vec{u}_j und \vec{u}_k die selbe Richtung haben, also entgegen der Voraussetzung, daß sie die verschiedene Richtungen haben. Also muß $\lambda_j = \lambda_k = 0$ gelten, d.h. die beiden Vektoren sind linear unabhängig. □

4. Es sei V eine Menge von Vektoren, und die l.u. Vektoren $\vec{x}_1, \dots, \vec{x}_n$ seien ein Teil dieser Menge. Können *alle* Vektoren in V als Linearkombination der $\vec{x}_1, \dots, \vec{x}_n$ dargestellt werden, so bilden die $\vec{x}_1, \dots, \vec{x}_n$ eine Basis von V .

Die Vektoren $\vec{y}_1, \dots, \vec{y}_p$ seien ebenfalls l.u. und mögen ebenfalls eine Basis von V sein. Dann gilt $p = n$.

Beweis: V enthalte mehr Vektoren als nur den Nullvektor, und es gelte $n < p$. Alle Vektoren aus V können aus $\vec{x}_1, \dots, \vec{x}_n$ als Linearkombination erzeugt werden, also auch die Vektoren $\vec{y}_1, \dots, \vec{y}_p$. Die Vektoren $\vec{x}_1, \dots, \vec{x}_n, \vec{y}_1$ sind dann jedenfalls linear abhängig; man

kann dann etwa \vec{x}_1 als Linearkombination der Vektoren $\vec{x}_2, \dots, \vec{x}_n, \vec{y}_1$ darstellen. Weiter kann man mit den Vektoren $\vec{x}_2, \dots, \vec{x}_n, \vec{y}_1, \vec{y}_2$ alle Vektoren von V erzeugen; dabei sind diese Vektoren linear abhängig, da ja \vec{y}_2 bereits eine Linearkombination der Vektoren $\vec{x}_2, \dots, \vec{x}_n, \vec{y}_1$ ist. Also kann etwa \vec{x}_2 als Linearkombination der Vektoren

$$\vec{x}_3, \dots, \vec{x}_n, \vec{y}_1, \vec{y}_2$$

dargestellt werden. So verfährt man weiter, bis man schließlich

$$\vec{y}_1, \dots, \vec{y}_n$$

als Basis erhält, aus der alle Vektoren aus V erzeugt werden können. Da $n < p$, kann \vec{y}_{n+1} als Linearkombination der $\vec{y}_1, \dots, \vec{y}_n$ berechnet werden, entgegen der Annahme, daß $\vec{y}_1, \dots, \vec{y}_p$ eine Basis ist, diese Vektoren also linear unabhängig sind. Also kann $n < p$ nicht gelten; es muß also $n = p$ sein. \square

5. Es sei V_n die Menge aller n -dimensionalen Vektoren. Jede Basis von V_n enthält genau n linear unabhängige Vektoren $\vec{u}_1, \dots, \vec{u}_n$.

Beweis: Es existieren genau n n -dimensionale Einheitsvektoren \vec{e}_j , $j = 1, \dots, n$. Ein beliebiger Vektor $\vec{u} \in V_n$ kann dann als Linearkombination der \vec{e}_j dargestellt werden,

$$\vec{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = u_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + u_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \dots + u_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \quad (74)$$

Also bilden die \vec{e}_j , $j = 1, \dots, n$ eine Basis. Nach 4 haben alle Basen die gleiche Anzahl von Vektoren, d.h. jede Basis des V_n enthält genau n Vektoren. \square

Hat man also $r < n$ linear unabhängige n -dimensionale Vektoren, so kann man damit nur eine Teilmenge der Vektoren in V_n erzeugen. Hat man dagegen $r > n$ n -dimensionale Vektoren $\vec{x}_1, \dots, \vec{x}_r$, so können sie nicht linear unabhängig sein. Denn es läßt sich ja stets eine Menge von l.u. Vektoren, d.h. eine Basis, $\vec{u}_1, \dots, \vec{u}_n$ finden, mit der sich *alle* Vektoren von V_n darstellen lassen. Da alle Basen von V_n genau n Vektoren enthalten, kann $\vec{x}_1, \dots, \vec{x}_r$ keine Basis sein, diese Vektoren sind l.a.; - man kann stets nur Mengen von höchstens n l.u. Vektoren finden.

Es ist wichtig, zwischen Orientierung und Dimension zu unterscheiden. n -dimensionale Vektoren, die sich alle nur hinsichtlich der Länge unterscheiden, nicht aber hinsichtlich ihrer Orientierung, liegen alle in einem 1-dimensionalen Teilraum des n -dimensionalen Raums. Sie sind sicherlich linear abhängig.

Wählt man irgend 2 Vektoren mit unterschiedlicher Orientierung, so definieren sie eine Ebene, also einen 2-dimensionalen Teilraum des n -dimensionalen

Raums. Es sei etwa ein 3-dimensionaler Vektorraum gegeben, dh eine Menge von 3-dimensionalen Vektoren, die alle als Linearkombination einer Basis von drei 3-dimensionalen Vektoren dargestellt werden können. In diesem Raum ist jede Ebene ein Teilvektorraum. Um dies zu sehen, Dazu werden zwei 3-dimensionale Vektoren \vec{x}_1 und \vec{x}_2 mit verschiedener Orientierung herausgegriffen; sie sind dann linear unabhängig und bilden eine Teilbasis des 3-dimensionalen Vektorraums. Die Linearkombinationen

$$\vec{y} = a_1\vec{x}_1 + a_2\vec{x}_2$$

(erzeugt, indem man a_1 und a_2 verschiedene Wert annehmen läßt) liegen in einer Ebene des 3-dimensionalen Raumes. Die Orientierung dieser Ebene ist durch die Orientierung eines Vektor \vec{n} charakterisierbar, der senkrecht auf dieser Ebene steht, – dies ist der *Normalenvektor* dieser Ebene. Dies bedeutet, dass \vec{n} notwendig orthogonal zu \vec{x}_1 und \vec{x}_2 ist, so dass

$$\vec{x}_1' \vec{n} = 0, \quad \vec{x}_2' \vec{n} = 0.$$

(Der Normalenvektor kann auch für einen 1-dimensionalen Teilraum definiert werden, er steht dann senkrecht auf der Geraden, die diesen Teilraum definiert.) Damit man sieht, dass die Ebene ein Teilraum des 3-dimensionalen Vektorraumes ist, muß man nur zeigen, dass jede Linearkombination von \vec{y} in eben dieser Ebene liegt und damit orthogonal zu \vec{n} ist. Es sei also $\vec{y} = a_1\vec{x}_1 + a_2\vec{x}_2$. Dann ist

$$\vec{y}' \vec{n} = a_1\vec{x}_1' \vec{n} + a_2\vec{x}_2' \vec{n} = 0, \tag{75}$$

denn $\vec{x}_1' \vec{n} = \vec{x}_2' \vec{n} = 0$ nach Voraussetzung. Mithin liegt \vec{y} in der durch \vec{x}_1 und \vec{x}_2 aufgespannten Ebene, und die Ebene ist ein Teilraum des 3-dimensionalen Vektorraumes. Man bemerke, dass die Vektoren im 2-dimensionalen Teilraum alle *3-dimensionale* Vektoren und nicht 2-dimensionale Vektoren sind! Ebenso zeigt man, dass eine Gerade ein Teilraum des 3-dimensionalen Teilraums ist.

Das in (75) repräsentierte Argument ist aber offenbar ganz unabhängig von der Anzahl der Komponenten der Vektoren \vec{n} , \vec{y} , \vec{x}_1 und \vec{x}_2 , so dass das Argument nicht auf 3-dimensionale Vektoren beschränkt ist, – es gilt für beliebige n -dimensionale Vektorräume. Es läßt sich überdies zeigen⁸, dass ein beliebiger n -dimensionaler Vektor \vec{x} stets als Summe zweier Vektoren \vec{x}_r und \vec{x}_{n-r} darstellbar ist, von denen der eine aus einem Teilraum V_r ($r < n$) stammt und der zweite aus dem dazu komplementären Teilraum V_{n-r} gewählt werden kann und für den gilt, dass er orthogonal zu \vec{x}_r ist. Man sieht daran, dass man durch Linearkombinationen von Vektoren aus einem Teilraum nicht aus diesem Teilraum hinausgelangen kann, denn hat man etwa im 2-dimensionalen Fall $\vec{x} = a\vec{u} + b\vec{v}$, wobei \vec{u} und \vec{v} bestimmte 2-dimensionale Vektoren sind, und ist $\vec{n} \in V_{n-r}$ der zu \vec{u} und \vec{v} orthogonale Vektor, der in die dritte Dimension weist, so folgt $\vec{n}' \vec{x} = a\vec{n}' \vec{u} + b\vec{n}' \vec{v} = 0$, da ja $\vec{n}' \vec{u} = \vec{n}' \vec{v} = 0$ ist, d.h. \vec{n} steht auch senkrecht auf der Linearkombination \vec{x} , so dass \vec{x} notwendig wieder ein Vektor im

⁸vergl. Skript *Vektoren und Matrizen*.

2-dimensionalen Raum ist. Die inhaltliche Bedeutung dieses Sachverhalts ist, dass die lineare Kombination von Merkmalen niemals eine Merkmalsmischung erzeugen kann, in der neue Qualitäten enthalten sind.

Im faktorenanalytischen Zusammenhang repräsentiert jede Dimension gewissermaßen eine Qualität, die sich nicht als Kombination der übrigen Dimensionen bzw. Qualitäten darstellen läßt. Dies ist der für die Faktorenanalyse relevante Kern des Begriffs der linearen Unabhängigkeit. Linear unabhängige Vektoren können, müssen aber nicht orthogonal sein. Wählt man orthogonale Basisvektoren als Faktoren, so sind die paarweisen Skalarprodukte dieser Vektoren gleich Null; in diesem Sinne sind die Faktoren unkorreliert. Wählt man nicht-orthogonale ("oblique") Basisvektoren, so sind sie zwar linear unabhängig, aber die Skalarprodukte zwischen ihnen sind nicht gleich Null und in diesem Sinne sind sie "korreliert". Die Korrelation zwischen ihnen bedeutet aber noch nicht, dass einer der Basisvektoren durch die anderen "vorausgesagt" werden kann, ihre lineare Unabhängigkeit verhindert diese Voraussagbarkeit. Die Qualitäten, die durch die obliquen Basisvektoren repräsentiert werden, treten dann in gewissen Kopplungen auf, auch wenn sie sich nicht aus den jeweilig anderen Qualitäten erklären lassen.

Anmerkung Für die psychologische Diagnostik wird von bestimmten Testtheorien die strikte Eindimensionalität eines gemessenen Merkmals gefordert, etwa wenn das Rasch-Modell angewendet werden soll. Dieses Modell erlaubt die Schätzung von Personenparametern und Itemparametern (die "Schwierigkeit" der Testaufgaben (= Items), die wiederum unabhängig von der Stichprobe der Probanden sein sollen (spezifische Objektivität)). Die Forderung nach Eindimensionalität ist allerdings keineswegs trivial. Man denke an die Fähigkeit, arithmetische Berechnungen im Kopf durchführen zu können. Aufgaben aus dem Kleinen Einmaleins können von den meisten Probanden nachgerade automatisch aus dem Gedächtnis beantwortet werden – kaum einer wird lange nachdenken müssen, um die Frage wieviel $3 \cdot 4$ oder $8 \cdot 7$ ist zu beantworten. Bei Aufgaben der Art $12 \cdot 27$ wird es schon schwieriger, denn nur noch wenige werden sie aus dem Gedächtnis heraus beantworten. Statt dessen wird man sich an einen Algorithmus erinnern, nach dem die Aufgabe in Teilaufgaben zerlegt wird, deren Resultate dann im Kurzzeitgedächtnis gespeichert werden müssen, um dann zu einer Antwort kombiniert zu werden. Die Fähigkeit zum Kopfrechnen erfordert gewissermaßen als rechenunabhängige Komponente eine gewisse Kapazität des Kurzzeitgedächtnisses. Wird die Aufgabe gestellt, den Wert von $\pi/e^{-1.75}$ zu berechnen, muß nicht nur eine erhebliche Kurzzeitspeicherkapazität zur Verfügung stehen, sondern noch Wissen über den Wert von π , der Zahl e und von einem Algorithmus, nach dem $e^{-1.75}$ berechnet werden kann. Mit der Schwierigkeit der Aufgaben verändert sich hier auch die Menge der Merkmale, die für das Lösen der Aufgaben vorhanden sein müssen. Eindimensionalität setzt voraus, dass die verschiedenen Merkmale, die für das Lösen der Aufgaben notwendig sind, stets zu gleichen Anteilen vorhanden sind; schwierigere Aufgaben verlangen, dass alle diese Merkmale in größerem

Ausmaß vorhanden sind, die Verhältnisse der Ausmaße aber konstant bleiben. Es ist klar, dass diese Konstanz der Anteile eine sehr restriktive Forderung ist.

Variieren die Verhältnisse der Ausprägungen der Merkmale von Aufgabe zu Aufgabe, so wird die Forderung der Eindimensionalität verletzt. Je nach Art der Variation wird eine zweite, möglicherweise noch eine dritte oder gar eine vierte etc zur Beschreibung der Daten erforderlich. Nimmt man zwei Grunddimensionen an, so wird damit postuliert, dass sich alle Aufgaben (und Personen) durch die Linearkombination zweier Basisvektoren darstellen lassen. Wie oben gezeigt, wird keine dieser Linearkombinationen in eine dritte Dimension weisen.

Die latenten Variablen entsprechen einer Basis des n -dimensionalen Vektorraumes, wenn n Variablen betrachtet werden. Hat man m Objekte ω_i mit Messungen x_{i1}, \dots, x_{in} für die n Variablen, so hat man m n -dimensionale Vektoren, die in einem n -dimensionalen Vektorraum liegen. Die Hoffnung des Faktorenanalytikers ist, dass sie alle in einem Teilraum mit der Dimension $s < n$ liegen; dann werden nur s latente Variable zu ihrer Deutung benötigt, dh es werden nur s Basisvektoren benötigt. Der folgende Ausflug in die Theorie der Matrizen zeigt, wie man eine solche Basis finden kann.

Dass empirisch gemessene Vektoren in einem echten Teilraum des n -dimensionalen Vektorraumes liegen ist äußerst unwahrscheinlich, denn wegen der unvermeidlichen Messfehler wird man sie nicht exakt anhand einer Teilbasis des n -dimensionalen Vektorraumes berechnen können. Die Annahme einer echten Teilbasis als Repräsentanten der latenten Variablen ist also stets eine Hypothese, es sei denn, man findet eine Möglichkeit, messfehlererzeugte latente Variablen (die also keine wirklichen latenten Variablen sind) und echte latente Variablen zu erzeugen. Die folgenden Betrachtungen erlauben ist, zumindest vernünftige Approximationen zu ermöglichen.

2.2 Matrizen, Eigenvektoren und Ellipsoide

2.2.1 Definition einer Matrix

Schreibt man die n m -dimensionalen Spaltenvektoren $\vec{X}_j, j = 1, \dots, n$ nebeneinander an, so entsteht eine $(m \times n)$ -Matrix, vergl. Tabelle 1. Schreibt man m n -dimensionale Zeilenvektoren $\vec{Y}_i, i = 1, \dots, m$, untereinander an, so entsteht ebenfalls eine $(m \times n)$ -Matrix. m und n heißen auch die Dimensionen einer Matrix, um anzuzeigen, dass sie eben n m -dimensionale Spaltenvektoren und m n -dimensionale Zeilenvektoren enthält. Die übliche Schreibweise ist allerdings

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \quad (76)$$

Tabelle 1: Eine $(m \times n)$ -Matrix

	\vec{X}_1	\vec{X}_2	\cdots	\vec{X}_n
\vec{Y}_1	x_{11}	x_{12}	\cdots	x_{1n}
\vec{Y}_2	x_{21}	x_{22}	\cdots	x_{2n}
			\vdots	
\vec{Y}_m	x_{m1}	x_{m2}	\cdots	x_{mn}

Gelegentlich schreibt man für eine Matrix auch $X = (x_{ij})$, $i = 1, \dots, m$, $j = 1, \dots, n$, um anzuzeigen, wie man die Elemente einer Matrix, hier als der Matrix X , bezeichnen will und wieviele Zeilen (m) und Spalten (n) die Matrix hat.

2.2.2 Multiplikation mit einem Skalar und Addition von Matrizen

Es sei X eine $(m \times n)$ -Matrix, d.h. X habe m Zeilen und n Spalten. $\lambda \in \mathbb{R}$ sei eine reelle Zahl, d.h. ein Skalar. Dann bedeutet λX , dass jedes Element x_{ij} von X (d.h. das Element in der i -ten Zeile und j -ten Spalte) mit λ multipliziert werden soll.

Es sei Y ebenfalls eine $(m \times n)$ -Matrix. Dann ist die Summe $X + Y$ der beiden Matrix durch die Elemente $x_{ij} + y_{ij}$ definiert, d.h. man summiert zwei Matrizen, indem man die zueinander korrespondierenden Elemente addiert. Diese Regel impliziert, dass X und Y notwendig die gleiche Anzahl von Zeilen und die gleiche Anzahl von Spalten haben.

2.2.3 Das Produkt von Matrizen

Auf Seite 17 wurde in den Gleichungen (32) die Messwerte x_{i1} durch latente Variable F_{ij} ausgedrückt; in der Gleichung (33) wurde der gesamte Spaltenvektor \vec{X}_1 als Linearkombination der Vektoren \vec{F}_1 und \vec{F}_2 dargestellt. Man kann nun die beiden Vektoren \vec{F}_1 und \vec{F}_2 zu einer Matrix F zusammenfassen und die Koeffizienten α_{j1} und α_{j2} zu einem Vektor \mathbf{a}_j :

$$F = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \\ \vdots & \vdots \\ F_{i1} & F_{i2} \\ \vdots & \vdots \\ F_{m1} & F_{m2} \end{pmatrix}, \quad \mathbf{a}_j = \begin{pmatrix} \alpha_{j1} \\ \alpha_{j2} \end{pmatrix} \quad (77)$$

Man kann nun das Produkt der Matrix F mit dem Vektor \mathbf{a}_j erklären als die der Reihe nach berechneten Skalarprodukte der *Zeilenvektoren* (F_{i1}, F_{i2}) mit

dem Vektor \mathbf{a}_j :

$$x_{ij} = \alpha_{j1}F_{i1} + \alpha_{j2}F_{i2},$$

die, untereinander angeschrieben, die Komponenten eines neuen Vektors ergeben, und zwar des Spaltenvektors \vec{X}_j :

$$\vec{X}_j = F\mathbf{a}_j = \begin{pmatrix} F_{11}\alpha_{1j} + F_{12}\alpha_{2j} \\ F_{21}\alpha_{1j} + F_{22}\alpha_{2j} \\ \vdots \\ F_{m1}\alpha_{1j} + F_{m2}\alpha_{2j} \end{pmatrix} = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{mj} \end{pmatrix} \quad (78)$$

Das Matrixprodukt von F mit \mathbf{a}_j entsteht also einfach durch Bildung der Skalarprodukte der Zeilenvektoren von F mit dem Spaltenvektor \mathbf{a}_j . Inspiziert man den Spaltenvektor in der Mitte, so sieht man, dass beim ersten Summanden stets der Faktor α_{1j} auftaucht, und beim zweiten Summanden tritt stets der Faktor α_{2j} auf, – diese Faktoren sind die Komponenten von \mathbf{a}_j . Man kann den Vektor in der Mitte also als Summe bzw. als Linearkombination der Spaltenvektoren von F auffassen:

$$\begin{pmatrix} F_{11}\alpha_{1j} + F_{12}\alpha_{2j} \\ F_{21}\alpha_{1j} + F_{22}\alpha_{2j} \\ \vdots \\ F_{m1}\alpha_{1j} + F_{m2}\alpha_{2j} \end{pmatrix} = \alpha_{1j} \begin{pmatrix} F_{11} \\ F_{21} \\ \vdots \\ F_{m1} \end{pmatrix} + \alpha_{2j} \begin{pmatrix} F_{12} \\ F_{22} \\ \vdots \\ F_{m2} \end{pmatrix} = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{mj} \end{pmatrix} \quad (79)$$

Man kann also die Regel aufstellen:

Regel 1: Multipliziert man eine Matrix F von rechts mit einem Spaltenvektor \mathbf{a} , so entsteht ein Spaltenvektor \vec{X} , der eine Linearkombination der Spaltenvektoren \vec{F}_k von F ist.

Dabei sind die Komponenten von \mathbf{a} die Koeffizienten der \vec{F}_k , $k = 1, \dots, n$, n die Anzahl der Spalten von F . Der Vektor \vec{X} hat so viele Komponenten wie die Vektoren \vec{F}_k .

Will man das Produkt $F\mathbf{a}_j$ für eine Reihe von Vektoren \mathbf{a}_j bilden, etwa für $j = 1, \dots, r$, so kann man die \mathbf{a}_j als Spalten einer $(n \times r)$ -Matrix A zusammenfassen, wobei jetzt a_{ij} statt α_{ij} geschrieben wird:

$$A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_r] = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ & & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nr} \end{pmatrix}. \quad (80)$$

Das Produkt FA liefert nun eine Matrix X , deren Spalten gerade die Vektoren $F\mathbf{a}_j = \vec{X}_j$ sind, und jeder Spaltenvektor \vec{X}_j ist eine Linearkombination der Spaltenvektoren $\vec{F}_1, \dots, \vec{F}_n$ von F :

$$FA = X = [\vec{X}_1, \dots, \vec{X}_r], \quad F\mathbf{a}_j = \vec{X}_j. \quad (81)$$

Statt eine Matrix F von rechts mit einem *Spaltenvektor* \mathbf{a} zu multiplizieren, kann man sie auch von links mit einem *Zeilenvektor* $\mathbf{b} = (b_1, \dots, b_m)'$ multiplizieren

$$\begin{aligned} \mathbf{b}'F &= (b_1, b_2, \dots, b_m) \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \\ \vdots & \vdots \\ F_{i1} & F_{i2} \\ \vdots & \vdots \\ F_{m1} & F_{m2} \end{pmatrix} \\ &= (b_1F_{11} + b_2F_{21} + \dots + b_mF_{m1}, b_1F_{12} + b_2F_{22} + \dots + b_mF_{m2}) \quad (82) \end{aligned}$$

Die Multiplikation von links mit \mathbf{b}' liefert offenbar einen *Zeilenvektor*. Man beachte, dass der Vektor \mathbf{b} so viele Komponenten hat wie die Matrix F Zeilen hat. Die Grundregel ist wieder die gleiche wie bei der Bildung des Produkts von F von rechts mit einem Spaltenvektor: die Komponenten des Vektors $\mathbf{b}'F = \mathbf{c}'$ sind Skalarprodukte eines Zeilenvektors mit einem Spaltenvektor. \mathbf{c} hat so viele Komponenten wie die Matrix F Spalten hat. Schreibt man \mathbf{c} als Spaltenvektor an, so erhält man

$$\mathbf{c} = \begin{pmatrix} b_1F_{11} + b_2F_{21} + \dots + b_mF_{m1} \\ b_1F_{12} + b_2F_{22} + \dots + b_mF_{m2} \end{pmatrix} = b_1 \begin{pmatrix} F_{11} \\ F_{12} \end{pmatrix} + \dots + b_m \begin{pmatrix} F_{m1} \\ F_{m2} \end{pmatrix} \quad (83)$$

Die Spaltenvektoren auf der rechten Seite sind aber gerade die *Zeilenvektoren* von F . Man hat dementsprechend die

Regel 2: Die Multiplikation einer Matrix $(m \times n)$ -Matrix F von links mit einem m -dimensionalen Zeilenvektor \mathbf{b}' liefert einen Zeilenvektor \mathbf{c}' . \mathbf{c} ist eine Linearkombination der Zeilenvektoren von F .

F sei eine $(m \times n)$ -Matrix. Hat man verschiedene m -dimensionale Zeilenvektoren, die von links mit der Matrix F multipliziert werden sollen, so kann man sie untereinander schreiben, erhält damit eine Matrix B und hat das Matrixprodukt

$$BF = C. \quad (84)$$

Anmerkung: Voraussetzung für die Möglichkeit, das Produkt BF zu bilden ist, dass die Matrix B so viele Spalten (m) wie die Matrix F Zeilen hat. \square

Man kann die Definition des Matrixprodukts zusammen mit den Regeln 1 und 2 so zusammenfassen:

Matrixprodukt: Gegeben sei die $(m \times n)$ -Matrix $A = A_{m,n}$ und die $(n \times r)$ -Matrix $B = B_{n,r}$. Dann kann das Produkt $C = C_{m,r} = A_{m,n}B_{n,r}$ gebildet werden; C ist eine $(m \times r)$ -Matrix (die Indizierung der Matrizen mit ihren jeweiligen Zeilen- und Spaltenzahlen

kann helfen, die *Dimensionen* (eben die Zahl der Zeilen und Spalten) im Auge zu behalten). Dann gilt stets

1. Die Spaltenvektoren von C sind Linearkombinationen der Spaltenvektoren von A ,
2. Die Zeilenvektoren von C sind Linearkombinationen der Zeilenvektoren von B .

Assoziativität Die Matrizen A und B seien wie oben definiert, und zusätzlich sei D eine $(r \times s)$ -Matrix. Dann kann man das Produkt $E = CD$ bilden; E ist eine $(m \times s)$ -Matrix. Da $C = AB$ ist dann $E = CD = ABD$. Man kann nun fragen, ob es bei der Bildung dieses Produkts auf die Reihenfolge ankommt, d.h. ob man auch A mit dem Produkt BD multiplizieren kann. Dies ist möglich, denn es gilt das *Assoziativgesetz* der Matrixmultiplikation:

$$E = (AB)D = A(BD). \quad (85)$$

Man überzeugt sich von der Richtigkeit dieser Aussage durch Nachrechnen.

Kommutativität Sind A und B wie oben definiert, so zeigt sich, dass im Allgemeinen das Produkt AB *nicht* gleich dem Produkt BA ist. Zum einen ist es eine Voraussetzung für die Bildung des Produkts BA , dass B so viele Spalten hat wie A Zeilen hat, d.h. es muß notwendig $m = n$ gelten, aber diese Bedingung ist nicht notwendig erfüllt. Es sei $C = AB$. Die Spalten von C sind dann Linearkombinationen der Spalten von A . Ist ebenfalls $C = BA$, so müssen die Spaltenvektoren von C ebenfalls Linearkombinationen der Spalten von B sein. Dies kann in Spezialfällen gelten, muß aber nicht gelten, so dass *im Allgemeinen* die Aussage

$$AB \neq BA \quad (86)$$

gilt. Man sagt, die Matrixmultiplikation sei im Allgemeinen *nicht kommutativ*. Ein Spezialfall, bei dem Kommutativität gegeben ist, ist die Multiplikation von Diagonalmatrizen, vorausgesetzt, sie haben die gleiche Anzahl von Zeilen und Spalten.

Transponierte eines Produkts Es sei $C = AB$. Dann gilt für die Transponierte C' von C

$$C' = (AB)' = B'A'. \quad (87)$$

Denn die Zeilenvektoren von C sind ja Linearkombinationen der Zeilenvektoren von B und müssen deshalb gleich den Spaltenvektoren von C' sein. Die Zeilenvektoren von B sind aber die Spaltenvektoren von B' und müssen also im Produkt $(AB)'$ als linker Faktor auftreten, also folgt $(AB)' = B'A'$.

2.2.4 Zentrierung und Standardisierung

Die hier hergeleitete Darstellung von Kovarianz- und Korrelationsmatrizen dient mehr der Einübung des Umgangs mit Matrizen und kann übersprungen werden.

Es sei X eine zufällige Veränderliche, für die die Messungen x_1, \dots, x_m vorliegen mögen. Das arithmetische Mittel und die Varianz der Messungen sind durch

$$\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i, \quad s^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})^2 = \frac{1}{m} \sum_{i=1}^m x_i^2 - \bar{x}^2, \quad (88)$$

wobei statt s^2 im Allgemeinen $\hat{s}^2 = \sum_i (x_i - \bar{x})^2 / (m - 1)$ berechnet wird; die Division durch $m - 1$ statt durch m gleicht eine Verzerrung (Bias) aus, mit dem die Schätzung s^2 für kleinere Werte von m behaftet ist. Die in (88) angegebene Formel für s^2 zeigt an, dass die Varianz eben als Mittelwert der Abweichungsquadrate $(x_i - \bar{x})^2$ definiert ist. Die Messwerte x_i werden standardisiert, wenn sie (i) *zentriert* werden, d.h. wenn \bar{x} subtrahiert wird, und wenn die Differenzen $x_i - \bar{x}$ durch die Standardabweichungen s (bzw. \hat{s}) dividiert werden:

$$z_i = \frac{x_i - \bar{x}}{s} \quad (89)$$

Der Mittelwert der z_i ist stets gleich Null und die Varianz ist stets gleich 1.

Die Spalten einer gegebenen Matrix X mögen die Messwerte von Variablen V_j enthalten, – x_{ij} ist der i -te Messwert der Variablen X_j . Wenn die Maßeinheiten der verschiedenen X_j verschieden sind, ist die Kovarianz etwa der Variablen X_j und X_k oft schwer zu interpretieren. Der Übergang zu standardisierten und damit von der Kovarianz zur Korrelation eliminiert der Effekt der verschiedenen Maßeinheiten. Die Standardisierung der X_j läßt sich dann in Matrixform anschreiben.

Dazu sei $\mathbf{1}_m = (1, 1, \dots, 1)'$ ein m -dimensionaler Vektor, dessen Komponenten alle gleich 1 sind. Dann ist

$$X' \mathbf{1}_m = \begin{pmatrix} \sum_{i=1}^m x_{i1} \\ \sum_{i=1}^m x_{i2} \\ \vdots \\ \sum_{i=1}^m x_{in} \end{pmatrix} \quad (90)$$

der Vektor der Spaltensummen von X , und folglich ist

$$\bar{X} = \frac{1}{m} X' \mathbf{1}_m = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_n \end{pmatrix} \quad (91)$$

der Vektor der arithmetischen Mittel $\bar{x}_1, \dots, \bar{x}_n$. Weiter werde das Vektorprodukt $\bar{X} \bar{X}'$ betrachtet:

$$\bar{X} \bar{X}' = \begin{pmatrix} \bar{x}_1 \bar{x}_1 & \bar{x}_1 \bar{x}_2 & \cdots & \bar{x}_1 \bar{x}_n \\ \bar{x}_2 \bar{x}_1 & \bar{x}_2 \bar{x}_2 & \cdots & \bar{x}_2 \bar{x}_n \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_n \bar{x}_1 & \bar{x}_n \bar{x}_2 & \cdots & \bar{x}_n \bar{x}_n \end{pmatrix} \quad (92)$$

Es werde nun die Kovarianz c_{jk} zwischen X_j und X_k betrachtet (für $j = k$ ist $c_{jj} = s_j^2$). Es ist

$$c_{jk} = \frac{1}{m} \sum_{i=1}^m (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) = \frac{1}{m} \sum_{i=1}^m x_{ij}x_{ik} - \bar{x}_j\bar{x}_k. \quad (93)$$

Das Produkt $\bar{x}_j\bar{x}_k$ ist gerade das Element in der j -ten Zeile und k -ten Spalte der in (92) definierten Matrix $\mathbf{X}\bar{X}'$. Die Matrix C der Kovarianzen c_{jk} ergibt sich nun gemäß

$$C = \frac{1}{m} X'X - \bar{X}\bar{X}' = \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})'. \quad (94)$$

Substituiert man den Ausdruck (91) für \bar{X} , so erhält man

$$C = \frac{1}{m} X'X - \bar{X}\bar{X}' = \frac{1}{m} \left(X'X - \frac{1}{m} X'1_m 1_m' X \right). \quad (95)$$

Definiert man nun die *Zentrierungsmatrix*

$$H = I - \frac{1}{m} 1_m 1_m', \quad (96)$$

I die Einheitsmatrix, so ergibt sich die Darstellung

$$C = \frac{1}{m} X'HX. \quad (97)$$

Die Matrix H ist symmetrisch und *idempotent*. Die Symmetrie folgt sofort aus aus der Symmetrie von $1_m 1_m'$, und die Idempotenz zeigt sich wegen

$$H^2 = \left(I - \frac{1}{m} 1_m 1_m' \right) \left(I - \frac{1}{m} 1_m 1_m' \right) = I - \frac{1}{m} 1_m 1_m' - \frac{1}{m} 1_m 1_m' + \frac{1}{m} 1_m 1_m' \frac{1}{m} 1_m 1_m',$$

und wegen

$$\frac{1}{m} 1_m \frac{1}{m} 1_m' = \frac{1}{m^2} 1_m m 1_m' = \frac{1}{m} 1_m 1_m'$$

folgt $H^2 = H$.

Der Korrelationskoeffizient ist durch

$$r_{jk} = \frac{c_{jk}}{s_j s_k} \quad (98)$$

definiert. Definiert man die Diagonalmatrix

$$S^{-1} = \begin{pmatrix} 1/s_1 & 0 & \cdots & 0 \\ 0 & 1/s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/s_n \end{pmatrix}, \quad (99)$$

so findet man für die Matrix $R = (r_{jk})$ der Korrelationen

$$R = S^{-1}CS^{-1}. \quad (100)$$

2.2.5 Der Rang einer Matrix

Es sei X eine $(m \times n)$ -Matrix und die Spaltenvektoren seien Linearkombinationen von $r \leq \min(m, n)$ linear unabhängigen Vektoren $\vec{u}_1, \dots, \vec{u}_r$. Die Anzahl r heißt *Spaltenrang* von X . Die m n -dimensionalen Zeilenvektoren können ebenfalls als Linearkombinationen von s linear unabhängigen Vektoren dargestellt werden; s heißt der *Zeilenrang*. Die Frage ist, ob $r = s$ – in diesem Fall muß nur von einem *Rang* der Matrix X gesprochen werden – oder ob im allgemeinen Fall $r \neq s$ gilt.

Die Frage ist nicht nur von formalem Interesse. In der Persönlichkeitstheorie wurde die Frage diskutiert, ob für eine gegebene Datenmatrix X die Anzahl der Personenfaktoren gleich der Anzahl der Testfaktoren sei oder nicht. Man könnte der Begriff eines Persönlichkeitstyps mit dem eines Persönlichkeitsfaktors identifizieren, und es wäre doch denkbar, dass es etwa drei Persönlichkeitstypen gäbe, die Tests aber insgesamt fünf Merkmalsdimensionen erfassen. In diesem Fall wäre $r = 5$ und $s = 3$. Es gilt aber der allgemeine Satz

Satz 2.1 *Es sei X eine beliebige $m \times n$ -Matrix. Dann ist stets der Zeilenrang s gleich dem Spaltenrang r . r ist höchstens gleich der kleineren der Zahlen m, n , so dass gilt*

$$r = s \leq \min(m, n). \quad (101)$$

Beweis: Der Spaltenrang von X sei r ; dann existiert eine $(m \times r)$ -Matrix A mit r l.u. Spaltenvektoren und eine $(r \times n)$ -Matrix B derart, dass

$$X = AB; \quad (102)$$

es gilt demnach $\mathbf{x}_j = A\mathbf{b}_j = b_{1j}\mathbf{a}_1 + \dots + b_{rj}\mathbf{a}_r$. Gleichzeitig bedeutet aber (102) auch, dass die Zeilenvektoren von X als Linearkombinationen der Zeilenvektoren von B dargestellt werden. Es gibt aber nur r Zeilenvektoren in B , von denen (noch) nicht klar ist, dass sie alle l.u. sind, so dass der Zeilenrang s von X höchstens gleich r sein kann, – $s \leq r$.

Umgekehrt sei s der Zeilenrang von X . Dann existiert eine $(s \times n)$ -Matrix D mit s l.u. Zeilenvektoren sowie eine $(m \times s)$ -Matrix C derart, dass

$$X = CD; \quad (103)$$

C enthält die zur Darstellung der Zeilenvektoren als Linearkombinationen der Zeilen von D notwendigen Koeffizienten. (103) bedeutet aber wiederum, dass die Spalten von X als Linearkombination der Spalten von C darstellbar sind, und diese Matrix hat s Spalten, so dass der Spaltenrang r nicht größer als s sein kann, mithin muß $r \leq s$ gelten. Somit muß sowohl $r \leq s$ also auch $s \leq r$ gelten, woraus folgt, dass $r = s$ gelten muß. \square

Bemerkung: Die Matrizen A in (102) und C in (103) müssen nicht identisch sein, da die Wahl der l.u. Vektoren nicht eindeutig ist; das gilt entsprechend für die Matrizen B und D . \square

2.2.6 Symmetrische Matrizen

Definition 2.5 Es sei $M = (m_{ij})$ eine quadratische Matrix, also $i = 1, \dots, n$ und $j = 1, \dots, n$. Es gelte

$$M' = M, \quad (104)$$

d.h. $m_{ij} = m_{ji}$ für alle i und alle j . Dann heißt M symmetrisch.

Symmetrische Matrizen spielen in der multivariaten Analyse eine besondere Rolle. So ist eine Matrix $R = (r_{ij})$, deren Elemente Korrelationen sind, stets symmetrisch, denn r_{ij} ist stets gleich r_{ji} – ob man die Variable V_i mit der Variablen V_j korreliert oder V_j mit V_i ist egal. Es sind Eigenschaften symmetrischer Matrizen, die ausgenutzt werden, um die latenten Dimensionen zu bestimmen.

Ein Spezialfall symmetrischer Matrizen sind *Diagonalmatrizen*. Eine Matrix $M = (m_{ij})$ heißt *diagonal*, wenn

$$m_{ij} = \begin{cases} \lambda_i \neq 0, & i = j \\ 0, & i \neq j \end{cases} \quad (105)$$

Diagonalmatrizen sind natürlich symmetrisch, weil eben $m_{ij} = m_{ji} = 0$. Ein Beispiel für eine Diagonalmatrix wird im Anschluß an die folgenden Betrachtungen gegeben.

Hat man zwei $(m \times n)$ -Matrizen X und Y mit den Elementen x_{ij} und y_{ij} , so kann man die Summe und die Differenz dieser beiden Matrizen definieren: die Summe besteht aus den Elementen $x_{ij} + y_{ij}$, und die Differenz aus den Elementen $x_{ij} - y_{ij}$. Wichtig ist darüber hinaus die Multiplikation von Matrizen; sie wird jetzt eingeführt.

Beispiel 2.1 Es sei A eine $(m \times n)$ -Matrix, \mathbf{x} sei ein n -dimensionaler Vektor und \mathbf{y} sei ein m -dimensionaler Vektor. Dann ist

$$A\mathbf{x} = \mathbf{y} \quad (106)$$

ein System von m Gleichungen, wie man sich durch Ausschreiben sofort klar macht. Ist $\mathbf{y} = \vec{0}$ der Nullvektor, so heißt das Gleichungssystem *homogen*, andernfalls heißt es *inhomogen*. Es sei $\mathbf{y} \neq \vec{0}$. Ist der Vektor \mathbf{x} nicht bekannt, so sind die Komponenten von \mathbf{x} die n Unbekannten. Man kann die Gleichung $A\mathbf{x} = \mathbf{y}$ dann als eine Hypothese auffassen, derzufolge eine Lösung – also der Vektor \mathbf{x} – überhaupt existiert. Um die Hypothese zu überprüfen, muß man sich daran erinnern, dass \mathbf{y} ja eine Linearkombination der Spaltenvektoren von A sein muß, d.h. \mathbf{y} muß ein Element in dem von den Spaltenvektoren von A aufgespannten Vektorraum $\mathcal{C}(A)$ sein. Für $m > n$ können aber nicht alle m -dimensionalen Vektoren durch die Spaltenvektoren erzeugt werden, d.h. es ist möglich, dass \mathbf{y} nicht in $\mathcal{C}(A)$ liegt. Da man durch Bildung von Linearkombinationen der Spalten von A nicht aus $\mathcal{C}(A)$ hinauskommt, heißt dies, dass kein

Vektor \mathbf{x} existieren *muß*, der der Gleichung (106) genügt, ein solcher Vektor *kann* existieren. Ist $m = n$, d.h. ist A eine quadratische Matrix, so existiert unter bestimmten Bedingungen eine Matrix A^{-1} derart, dass $A^{-1}A = I$ die Einheitsmatrix ist – eine hinreichende Bedingung ist, dass die Spaltenvektoren paarweise orthogonal zueinander sind. Dann besteht A^{-1} aus Spaltenvektoren, die orthogonal zu den Spaltenvektoren in A sind. Dann ist

$$\mathbf{x} = A^{-1}\mathbf{y} \quad (107)$$

und die Lösung \mathbf{x} des inhomogenen Gleichungssystems ist eine Linearkombination der Spaltenvektoren von A^{-1} .

Insbesondere sei $\mathbf{y} = \vec{0}$ der Nullvektor. Der Nullvektor soll dann als Linearkombination der Spalten von A dargestellt werden. Sind die Spaltenvektoren linear unabhängig, so ist dies nach Definition der linearen Unabhängigkeit nur möglich, wenn \mathbf{x} ebenfalls der Nullvektor ist, – es gibt dann nur diese eine Lösung $\mathbf{x} = \vec{0}$. Eine Lösung $\mathbf{x} \neq \vec{0}$ existiert also nur, wenn die Spaltenvektoren von A linear abhängig sind. \square

Die Korrelationsmatrix: Die Matrixmultiplikation liefert eine in vielerlei Hinsicht nützliche Möglichkeit, Die Korrelationen zwischen allen Paaren von n Variablen darzustellen. Dazu geht man von der Matrix X der Rohwerte x_{ij} zunächst zur Matrix der standardisierten Werte z_{ij} über, wobei $z_{ij} = (x_{ij} - \bar{x}_j)/s_j$, \bar{x}_j der Mittelwert der Werte der j -ten Variablen (des j -ten Spaltenvektors \vec{X}_j) und s_j die Standardabweichung der Komponenten (der Messwerte) von \vec{X}_j . Man spricht von *Spaltenstandardisierung*. Die Korrelation zwischen der j -ten und der k -ten Variable ist dann

$$r_{jk} = \frac{1}{m} \sum_{i=1}^m z_{ij}z_{ik} = \frac{1}{m} \vec{Z}'_j \vec{Z}_k. \quad (108)$$

Da $\vec{Z}'_j \vec{Z}_k = \vec{Z}'_k \vec{Z}_j$ folgt $r_{jk} = r_{kj}$. Fasst man nun alle Vektoren \vec{Z}_j zu einer Matrix Z zusammen,

$$Z = [\vec{Z}_1, \vec{Z}_2, \dots, \vec{Z}_n],$$

so enthält die gestürzte Matrix Z' diese Spaltenvektoren als Zeilenvektoren. Bildet man nun die Skalarprodukte von all diesen Zeilenvektoren mit all den Spaltenvektoren, so erhält man gerade die Matrix R der Korrelationen:

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} & \dots & r_{1n} \\ r_{21} & 1 & r_{23} & \dots & r_{2n} \\ r_{31} & r_{32} & 1 & \dots & r_{3n} \\ & & & \ddots & \\ r_{n1} & r_{n2} & r_{n3} & \dots & 1 \end{pmatrix} = \frac{1}{m} Z'Z. \quad (109)$$

Hier ist sicherlich $r_{12} = r_{21}$, $r_{13} = r_{31}$ etc, allgemein $r_{jk} = r_{kj}$. Die Matrix R ist offenbar symmetrisch. Die Einsen in den Diagonalzellen von R sind die

Korrelationen $r_{11}, r_{22}, \dots, r_{nn}$. Da R durch das Produkt $Z'Z$ definiert ist, ist die Symmetrie eine Folge dieser speziellen Art der Produktbildung, die gelegentlich auch als *Kreuzprodukt* bezeichnet wird. Kreuzprodukte liefern stets symmetrische Matrizen.

Sind alle Variablen unkorreliert, d.h. gilt $r_{ij} = 0$ für alle $i \neq j$, so ist R eine Diagonalmatrix. Für empirische Messungen wird man Korrelationen, die exakt gleich Null sind, sehr selten finden. Aber für Variablen, von denen aus theoretischen Gründen gefordert wird, dass sie unabhängig im Sinne von unkorreliert sein sollen, müssen die entsprechenden Skalarprodukte gleich Null sein. So betrachte man die Gleichung (??), wo bezüglich der Matrix L postuliert wird, dass die Spaltenvektoren, deren Komponenten Ausprägungen auf unabhängigen Dimensionen repräsentieren, unkorreliert sind, d.h. es wird postuliert, dass die Skalarprodukte verschiedener Spaltenvektoren von L verschwinden (verschiedene Spaltenvektoren sind demnach *orthogonal*). Aus (??) folgt dann

$$X'X = AL'LA' = A\Lambda A' \quad (110)$$

mit

$$L'L = \Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \vdots & 0 \\ 0 & 0 & \cdots & \lambda_N \end{pmatrix} \quad (111)$$

Offenbar ist $\lambda_j = \|\vec{F}_j\|^2$, d.h. die Diagonalelemente λ_j sind gleich den Quadraten der Längen der \vec{F}_j . Die Matrix $X'X$ ist symmetrisch, denn nach (??) muß $(X'X)' = X'X$, also (104) gelten. Wie in Abschnitt 2.2.10 gezeigt wird, liefert die Gleichung (110) die Lösung für das Problem, die latenten Variablen zu bestimmen.

2.2.7 Die Einheitsmatrix und die inverse Matrix

Die Einheitsmatrix spielt im sogenannten *Matrixkalkül* (dies ist die Menge der Regeln für das Rechnen mit Matrizen) die Rolle der 1 beim Rechnen mit reellen Zahlen (Skalaren). Die Einheitsmatrix ist eine Diagonalmatrix mit $\lambda_j = 1$ für alle j , d.h.

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \quad (112)$$

I steht für Identität. Die Elemente dieser Matrix sind alle gleich 0, bis auf die Elemente in den Diagonalzellen; diese Elemente sind alle gleich 1. Die i -te Zeile einer Einheitsmatrix heißt auch i -ter (Zeilen-)Einheitsvektor; analog dazu ist der j -te Spaltenvektor der j -te (Spalten-)Einheitsvektor. Multipliziert

man eine Matrix A mit I , so erhält man wieder A :

$$AI = A, \quad IA = A. \quad (113)$$

Um AI zu berechnen, muß I so viele Zeilen haben, wie A Spalten hat, und um IA zu berechnen, muß I so viele Spalten haben, wie A Zeilen hat.

Es sei A eine quadratische Matrix, d.h. A habe so viele Zeilen wie Spalten. Unter bestimmten Umständen existiert nun eine Matrix A^{-1} derart, dass

$$A^{-1}A = AA^{-1} = I. \quad (114)$$

A^{-1} heißt dann die zu A *inverse* Matrix, oder auch die *Kehrmatrix*. Sie ist in Gleichung (107), Seite 45, bereits aufgetreten. Für die Matrizen A und A^{-1} gilt also die Kommutativität der Multiplikation.

2.2.8 Die Transformation von Vektoren

Es sei \vec{x} ein n -dimensionaler Vektor, und T sei eine $(m \times n)$ -Matrix, d.h. T habe m Zeilen und n Spalten. Bildet man das Produkt $T\vec{x}$, bildet man also alle Skalarprodukte der m n -dimensionalen Zeilenvektoren von T mit \vec{x} , so entsteht ein neuer Vektor, \vec{y} . Da T m Zeilen hat, gibt es m Skalarprodukte, also ist \vec{y} ein m -dimensionaler Vektor:

$$T\vec{x} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{m1} & t_{m2} & \cdots & t_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_j t_{1j}x_j \\ \sum_j t_{2j}x_j \\ \vdots \\ \sum_j t_{mj}x_j \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \quad (115)$$

Man sagt, der Spaltenvektor \vec{x} wird durch T in den Spaltenvektor \vec{y} transformiert. Man beachte, dass $m = n$ sein kann, aber nicht sein muß; wichtig ist aber, dass die Anzahl von Spalten von T mit der Anzahl der Komponenten von \vec{x} übereinstimmen muß.

Man kann ebenfalls den Fall betrachten, dass der gestürzte oder *transponierte* Vektor \vec{x}' transformiert wird; dazu bildet man das Produkt $\vec{x}'T$. Jetzt müssen alle Skalarprodukte von \vec{x}' mit den Spaltenvektoren von T gebildet werden, - was nur möglich ist, wenn die Anzahl der *Zeilen* von T mit der Anzahl n der Komponenten von \vec{x} übereinstimmt. T muß jetzt also eine $(n \times m)$ -Matrix sein, wobei $n = m$ sein kann, aber nicht sein muß:

$$\begin{aligned} \vec{x}'T &= (x_1, x_2, \dots, x_n) \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1m} \\ t_{21} & t_{22} & \cdots & t_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{nm} \end{pmatrix} \\ &= \left(\sum_i x_i t_{i1}, \sum_i x_i t_{i2}, \dots, \sum_i x_i t_{im} \right) \end{aligned} \quad (116)$$

d.h.

$$\vec{x}'T = (y_1, y_2, \dots, y_n) = \vec{y}' \quad (117)$$

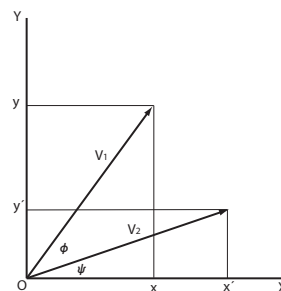
Zusammenfassend kann man sagen:

1. Ein Spaltenvektor \vec{x} wird in einen anderen Spaltenvektor \vec{y} transformiert durch Multiplikation *von links* mit einer Matrix T , die so viele Spalten hat wie \vec{x} Komponenten hat. Hat T m Zeilen, so ist \vec{y} ein m -dimensionaler Vektor. Ist $m = n$, so unterscheidet sich \vec{y} von \vec{x} hinsichtlich (i) der Länge und (ii) der Orientierung. Unterscheiden sich \vec{x} und \vec{y} nur hinsichtlich der Orientierung, so heißt die Transformationsmatrix T auch *Rotationsmatrix*. Da $\|\vec{x}\| = \|\vec{y}\|$ gelten soll, folgt $\|\vec{x}\|^2 = \vec{y}'\vec{y} = \vec{x}'T'T\vec{x}$ genau dann, wenn $T'T = I$ die Einheitsmatrix. Dies bedeutet, dass die Spaltenvektoren von T orthogonal sind und die Länge 1 haben, d.h. T ist orthonormal.
2. Ein n -dimensionaler Zeilenvektor \vec{x}' wird in einen Zeilenvektor \vec{y}' transformiert durch Multiplikation von rechts mit einer Matrix T , die n Zeilen hat und m Spalten, wobei $m \neq n$ möglich ist. \vec{y}' ist dann m -dimensional. Für den Fall $m = n$ ist die Transformation wieder eine Rotation, wenn $\|\vec{x}\| = \|\vec{y}\|$; $\vec{y}'\vec{y} = \vec{x}'TT'\vec{x} = \vec{x}'\vec{x}$ gilt genau dann, wenn $TT' = I$ die Einheitsmatrix, wodurch wieder die Orthonormalität von T impliziert wird.

2.2.9 Die Rotation als Transformation

Es gelte $T\vec{x} = \vec{y}$ und es sei $\|\vec{x}\| = \|\vec{y}\|$, d.h. die Transformation ist eine Rotation: nur die Orientierung des Vektors \vec{x} wird verändert; in anderen Worten, die Vektoren \vec{x} und \vec{y} unterscheiden sich nur hinsichtlich ihrer Orientierung. Dieser

Abbildung 3: Rotation eines Vektors



Fall ist im Zusammenhang mit faktorenanalytischen Untersuchungen von beträchtlichem Interesse. Denn hier sind die Rotationen von Koordinatenachsen von Bedeutung, da bestimmte Orientierungen der Achsen u.U. bessere Interpretationen der latenten Dimensionen erlauben als andere Orientierungen. Die

Koordinatenrotation kann nun mit der Betrachtung von Vektoren, die sich nur hinsichtlich ihrer Orientierung unterscheiden, in Zusammenhang gebracht werden. Denn betrachtet man den Vektor \vec{v}_1 , so habe er in den (X, Y) -Koordinaten die Komponenten (x, y) , also $\vec{x} = (x, y)'$. Rotiert man die Koordinatenachsen, so dass ein System (X', Y') entsteht, so hat der gleiche Vektor in diesem Koordinatensystem die Komponenten (x', y') . Die Veränderung der Orientierung, die in einem gegebenen Koordinatensystem den Übergang vom Vektor \vec{v}_1 zum Vektor \vec{v}_2 bedeutet, entspricht einer Rotation des Koordinatensystems. Die Transformationsmatrix T , die die Rotation von \vec{v}_1 in den Vektore \vec{v}_2 bewirkt, entspricht der Rotation des Koordinatensystems in entgegengesetzter Richtung.

In Abb. 3 unterscheiden sich die Orientierungen von \vec{v}_1 und \vec{v}_2 durch den Winkel ϕ ; die Transformation

$$\vec{v}_1 \xrightarrow{T} \vec{v}_2, \text{ d.h. } T\vec{v}_1 = \vec{v}_2 \quad (118)$$

bedeutet, dass die Elemente von T in irgendeiner Weise durch den Winkel ϕ bestimmt sein müssen. Es lässt sich zeigen (vergl. Abschnitt 6.3, insbesondere Abschn. 6.3.1), dass T durch

$$T = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} \quad (119)$$

gegeben ist, und die dazu inverse Transformation ist

$$T^{-1} = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}. \quad (120)$$

Offensichtlich ist $T^{-1} = T'$, d.h. T ist orthonormal (man überprüfe, dass tatsächlich $TT' = T'T = I$ gilt). In Abschnitt 6.3.2 wird gezeigt, dass eine Transformation der Koordinaten zur gleichen Matrix T führt, allerdings ist der Nachweis aufwendiger.

2.2.10 Eigenvektoren

Die Gleichung (110) auf Seite 46 wurde bereits als Schlüssel für die Lösung des Problems, die latenten Variablen zu bestimmen, genannt. In diesem Abschnitt wird gezeigt, warum diese Aussage gilt. Dazu wird ein für die Schätzung der latenten Dimensionen wichtiger Spezialfall einer Transformation vorgestellt.

Definition 2.6 Es sei M eine $(n \times n)$ -Matrix, \vec{t} sei ein n -dimensionaler Vektor, λ sei ein Skalar, und es gelte

$$M\vec{t} = \lambda\vec{t}. \quad (121)$$

Dann heißt \vec{t} *Eigenvektor* von M und λ heißt der zugehörige *Eigenwert*.

Anmerkungen:

1. **Orientierung:** Generell ergibt das Produkt $M\vec{t}$ stets einen Vektor, etwa \vec{s} , also $M\vec{t} = \vec{s}$. Man kann sagen, dass M den Vektor \vec{t} in den Vektor \vec{s} transformiert. Die Komponenten von \vec{s} sind die Skalarprodukte der Zeilenvektoren von M mit \vec{t} , und \vec{s} wird i.a. eine andere Orientierung und eine andere Länge als \vec{t} haben. Die Gleichung (121) bedeutet aber, dass Eigenvektoren ein Spezialfall sind: $M\vec{t} = \vec{s} = \lambda\vec{t}$ impliziert ja, dass \vec{s} nun die gleiche Orientierung wie \vec{t} hat, und dass sich \vec{s} und \vec{t} nur durch einen Faktor, den Eigenwert λ , und mithin nur in ihrer Länge unterscheiden.
2. **Normiertheit:** Die Definitionsgleichung (121) besagt, dass sich die Vektoren \vec{t} und $\vec{s} = \lambda\vec{t}$ nur durch einen Faktor, eben λ , unterscheiden; hat \vec{t} die Länge $\|\vec{t}\|$, so hat \vec{s} die Länge $\|\vec{s}\| = \lambda\|\vec{t}\|$. Über die Länge von \vec{t} und damit über die von \vec{s} ist nichts ausgesagt; sind s_i und t_i die zueinander korrespondierenden Komponenten von \vec{t} und \vec{s} , so folgt nur, dass $s_i/t_i = \lambda$ für alle i . Berechnet man nun die Eigenvektoren einer Matrix, so legt man mit ihren Komponenten auch die Länge dieser Vektoren fest. Da aber die die Definition der Eigenvektoren eine Spezifizierung der Längen nicht beinhaltet, muß man sich hinsichtlich der Länge schon aus Gründen der tatsächlichen Berechnung irgendwie festlegen. Deswegen legt man für die Berechnung fest, dass ihre Länge gleich 1 ist, dass die Eigenvektoren also normiert sind. Dies bedeutet nicht, dass man, wenn es für irgendwelche Zwecke nützlich ist, nicht auch eine andere Länge wählen kann.

Eine etwas andere Betrachtungsweise geht wie folgt: Angenommen, \vec{t}_j sei nicht normiert und habe also eine Länge $\|\vec{t}_j\| = \tau$. Dann kann man \vec{t}_j normieren, indem man seine Komponenten durch die Länge τ teilt. Es entsteht ein Vektor $\vec{t}_{0j} = (1/\tau)\vec{t}_j$, oder $\vec{t}_j = \tau\vec{t}_{0j}$. Setzt man diesen Ausdruck in (121) ein, so erhält man

$$M\tau\vec{t}_{0j} = \tau\lambda\vec{t}_{0j}. \quad (122)$$

Der Skalar τ kürzt sich heraus und es bleibt nur noch $M\vec{t}_{0j} = \lambda\vec{t}_{0j}$ übrig. Man kann dann einfach \vec{t}_j statt \vec{t}_{0j} schreiben und die Normiertheit voraussetzen.

□

Die Gleichung (110) ist offenbar von der Form

Definition 2.7 Der Ausdruck $\vec{t}'M\vec{t} = k_0$ heißt *quadratische Form*; k_0 ist ein Skalar. Die quadratische Form heißt *positiv definit* bzw. *positiv semi-definit*, wenn für alle \vec{t} , $k_0 > 0$ bzw. $k_0 \geq 0$ gilt.

Bemerkung: k_0 ist ein Skalar, denn $\vec{t}'M$ ergibt einen Zeilenvektor, und k_0 ist das Skalarprodukt dieses Zeilenvektors mit \vec{t} . Eine Begründung für den Ausdruck 'quadratische Form' wird in Abschnitt 2.2.11 gegeben. □

Satz 2.2 Die Matrix M ist positiv semi-definit genau dann, wenn eine Matrix G existiert derart, dass $M = G'G$.

Beweis: Existiert G derart, dass $G'G = M$, so gilt für einen beliebigen reellen Vektor \vec{t} die Beziehung $\vec{t}'M\vec{t} = \vec{t}'G'G\vec{t}$. Setzt man $\vec{y} = G\vec{t}$, so erhält man $\vec{t}'G'G\vec{t} = \vec{y}'\vec{y} = y_1^2 + \dots + y_n^2 \geq 0$. Also ist M positiv-semidefinit. \square

Anmerkungen:

1. Da $R = Z'Z$, Z die (spalten-)standardisierte Datenmatrix, ist die Korrelationsmatrix R im allgemeinen positiv definit. Gäbe es fehlerfreie Daten, so wäre allerdings der Fall denkbar, dass sie auch positiv-semidefinit sein kann; es läßt sich zeigen, dass dieser Fall eintritt, wenn es weniger latente Variablen gibt als tatsächlich gemessene Variablen. Diesen Fall nimmt man zwar bei der Anwendung der Faktorenanalyse an, allerdings bewirken Messfehler sowie numerische Ungenauigkeiten bei der Berechnung der Korrelationen, dass R von einem numerischen Standpunkt aus positiv-definit ist.

Schreibt man $R = Z'Z$, so hat man stillschweigend den Faktor $1/m$ in die Matrix Z absorbiert.

2. Ist M positiv (semi-)definit, so ist sie notwendig auch symmetrisch, denn Matrizen, die als Produkt $G'G$ dargestellt werden können, sind notwendig symmetrisch. Aber eine symmetrische Matrix ist nicht notwendig auch positiv (semi-)definit. Ist M symmetrisch, aber nicht positiv (semi-)definit, so definiert sie eine Menge von Hyperbeln; ist M dagegen positiv definit, so definiert sie eine Menge von Ellipsen bzw. Ellipsoiden (vergl. Abschnitt 2.2.11). Da die Korrelationsmatrix R notwendig positiv definit ist, definiert R notwendig Ellipsen bzw. Ellipsoide. \square

Satz 2.3 Es seien \vec{t}_j und \vec{t}_k Eigenvektoren einer symmetrischen Matrix mit zugehörigen Eigenwerten $\lambda_j \neq 0$, $\lambda_k \neq 0$ und $\lambda_j \neq \lambda_k$. Dann sind \vec{t}_j und \vec{t}_k orthogonal.

Beweis: Es seien \vec{t}_j und \vec{t}_k zwei verschiedene Eigenvektoren von M , d.h. es gelte

$$M\vec{t}_j = \lambda_j\vec{t}_j \quad (123)$$

$$M\vec{t}_k = \lambda_k\vec{t}_k, \quad (124)$$

wobei $\lambda_j \neq \lambda_k$. Um zu sehen, dass die beiden Vektoren orthogonal bzw. orthonormiert sind, multipliziert man (123) von links mit \vec{t}_k' , und (124) von links mit \vec{t}_j' . Man erhält

$$\vec{t}_k' M \vec{t}_j = \lambda_j \vec{t}_k' \vec{t}_j \quad (125)$$

$$\vec{t}_j' M \vec{t}_k = \lambda_k \vec{t}_j' \vec{t}_k \quad (126)$$

Es ist sicherlich $\vec{t}_k' \vec{t}_j = \vec{t}_j' \vec{t}_k$, denn diese Größen sind ja Skalarprodukte und damit einfache reelle Zahlen. Weiter ist $(\vec{t}_j' M \vec{t}_k)' = \vec{t}_k' M' \vec{t}_j$, und wegen der vorausgesetzten Symmetrie gilt $M = M'$, so dass $\vec{t}_k' M \vec{t}_j = \vec{t}_j' M \vec{t}_k$ bzw. $\vec{t}_k' M \vec{t}_j - \vec{t}_j' M \vec{t}_k = 0$ folgt. Subtrahiert man nun die zweite Gleichung von der ersten, so erhält man die Differenz

$$\vec{t}_k' M \vec{t}_j - \vec{t}_j' M \vec{t}_k = \vec{t}_j' \vec{t}_k (\lambda_j - \lambda_k).$$

Aber $\vec{t}_k' M \vec{t}_j - \vec{t}_j' M \vec{t}_k = 0$, wie eben gezeigt wurde, also muß $\vec{t}_j' \vec{t}_k (\lambda_j - \lambda_k) = 0$ gelten. Da aber $\lambda_j - \lambda_k \neq 0$ vorausgesetzt wurde, muß $\vec{t}_j' \vec{t}_k = 0$ folgen, und dies bedeutet ja die Orthogonalität von \vec{t}_j und \vec{t}_k . \square

Anmerkung: Generell kann für einen Eigenvektor $\|\vec{t}\| = 1$ angenommen werden, s. oben. Die Eigenvektoren einer Matrix können zu einer Matrix $T = [\vec{t}_1, \dots, \vec{t}_n]$ zusammengefasst werden, und die Gleichungen $M \vec{t}_j = \lambda_j \vec{t}_j$ können zu der Matrixgleichung

$$MT = T\Lambda, \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) \quad (127)$$

zusammengefasst werden. Da die Eigenvektoren orthonormal sind, ist T orthonormal, d.h. es gilt $T'T = TT' = I$, I wieder die Einheitsmatrix. Multiplikation von links mit T' liefert dann

$$T'MT = \Lambda. \quad (128)$$

Multiplikation der Gleichung (127) von rechts mit T' liefert

$$M = T\Lambda T'. \quad (129)$$

1. Die Gleichung (128) zeigt, wie durch "Prämultiplikation" von M mit T' und "Postmultiplikation" von M mit T die Matrix M "auf Diagonalfom" gebracht wird.
2. Die Gleichung (129) zeigt, wie Λ durch Prämultiplikation von Λ mit T und Postmultiplikation mit T' in die Nichtdiagonalfom M überführt wird.
3. Die Gleichungen (128) und (129) charakterisieren die Matrix T als Matrix der Eigenvektoren der symmetrischen Matrix M und Λ als Matrix der zugehörigen Eigenwerte.
4. In Gleichung (110), Seite 46 wurde die Gleichung

$$X'X = \Lambda\Lambda'$$

vorge stellt. Sie ergab sich aus dem Ansatz $X = LA'$, wobei die Spaltenvektoren von L als paarweise orthogonal angenommen wurden, so dass $\Lambda = L'L$ eine Diagonalmatrix ist. Da $X'X$ eine symmetrische Matrix ist, legt ein Vergleich mit (129) nahe, dass die Eigenvektoren von $M = X'X$

eine mögliche Lösung für die Matrix A sind; Λ ist dann die Diagonalmatrix der Eigenwerte von $X'X$. In der Tat wird sich zeigen, dass sich die Lösung über die Anwendung der Eigenvektoren und Eigenwerte von $X'X$ und XX' ergibt.

Der folgende Satz erweist sich als nützlich:

Satz 2.4 *Es sei M eine symmetrische $(n \times n)$ -Matrix mit der zugehörigen Matrix T von Eigenvektoren und der Diagonalmatrix Λ von Eigenwerten. Alle n Eigenwerte seien von Null verschieden. Dann gilt*

$$M^{-1} = T\Lambda^{-1}T', \quad (130)$$

wobei $\Lambda^{-1} = \text{diag}(1/\lambda_1, \dots, 1/\lambda_n)$ ist.

Beweis: Es ist

$$M^{-1} = (T\Lambda T')^{-1} = (T')^{-1}\Lambda^{-1}T^{-1}.$$

Aber $T' = T^{-1}$, so dass $(T')^{-1} = (T^{-1})^{-1} = T$, und damit hat man schon $M^{-1} = T\Lambda^{-1}T'$. \square

2.2.11 Quadratische Formen und Ellipsoide

Es sei insbesondere

$$M = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \quad (131)$$

Es sei $\vec{x} = (x_1, x_2)'$ und es werde die Quadratische Form $\vec{x}'M\vec{x} = k_0$, also

$$(x_1, x_2) \begin{pmatrix} a & c \\ c & b \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = k_0 \quad (132)$$

betrachtet, wobei k_0 eine Konstante ist. Multipliziert man die Gleichung aus, so erhält man

$$\vec{x}'M\vec{x} = ax_1^2 + bx_2^2 + 2cx_1x_2 = k_0. \quad (133)$$

Dies ist die Gleichung einer Ellipse. Die Endpunkte der Vektoren \vec{x} , die der Gleichung (132) genügen, liegen auf einer Ellipse. Für $c \neq 0$ ist die Ellipse nicht achsenparallel.

Gesucht ist nun die Länge der Halbachsen dieser Ellipse. Es wird angenommen, dass der Mittelpunkt der Ellipse mit dem Ursprung des Koordinatensystem zusammenfällt. \vec{x} sei ein Vektor, der der Gleichung (133) genügt. Das Quadrat der Länge r (r für Radius) von \vec{x} ist $r^2 = \vec{x}'\vec{x}$. Wenn \vec{x} mit der ersten Halbachse zusammenfällt, ist r maximal, und wenn \vec{x} mit der zweiten Halbachse zusammenfällt, ist r minimal. Die Längen der Halbachsen ergeben sich also als Extrema für $r = \|\vec{x}\|$ bzw. für $r^2 = \|\vec{x}\|^2$. Es gibt zwei Möglichkeiten, diese Extrema zu bestimmen. Die erste geht davon aus, dass die Längen

der Halbachsen nicht von der Lage bzw. der Rotation der Ellipse abhängen, so dass man die Längen auch für die rotierte Ellipse, deren Achsen parallel zu den Koordinatenachsen sind, bestimmen kann. Es sei nun T' eine Transformationsmatrix, insbesondere eine Rotationsmatrix, die die Vektoren \vec{x} in die Vektoren \vec{y} überführt, die die rotierte, achsenparallele Ellipse definieren. Dann folgt

$$\vec{y} = T' \vec{x}. \quad (134)$$

Dann folgt $\vec{x}' = \vec{y}' T'$ und (133) kann in der Form

$$\vec{y}' T' M T \vec{y} = k_0 \quad (135)$$

geschrieben werden. Da die \vec{y} eine achsenparallele Ellipse beschreiben sollen, muß $T' M T$ eine Diagonalmatrix sein, d.h. es existiert eine Matrix $\Lambda = \text{diag}(\lambda_1, \lambda_2)$ derart, dass

$$T' M T = \Lambda. \quad (136)$$

Dies ist aber gerade die Gleichung (128), d.h. die Matrix T muß die Eigenvektoren von M enthalten, und Λ enthält die zugehörigen Eigenwerte von M . T' überführt die nicht achsenparallele Ellipse (133) in die achsenparallele Ellipse

$$\vec{y}' \Lambda \vec{y} = k_0, \quad (137)$$

und T transformiert die achsenparallele Ellipse in die rotierte Ellipse (133).

Ist nun $\vec{y} = (y_1, y_2)'$, so ist (137) äquivalent zu

$$y_1^2 \lambda_1 + y_2^2 \lambda_2 = \lambda_1 y_1^2 + \lambda_2 y_2^2 = k_0. \quad (138)$$

Der spezielle Vektor $\vec{y}_0 = (a_1, 0)$ definiert die erste Halbachse der achsenparallelen Ellipse. (138) impliziert dann

$$a_1^2 \lambda_1 = k_0. \quad (139)$$

Daraus folgt für die Länge a_1 der ersten Halbachse die Beziehung

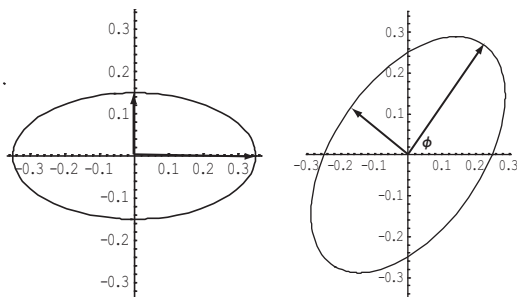
$$a_1 = \sqrt{\frac{k_0}{\lambda_1}}. \quad (140)$$

Analog folgt für die Länge a_2 der zweiten Halbachse

$$a_2 = \sqrt{\frac{k_0}{\lambda_2}}. \quad (141)$$

Da bei der Rotation die Länge nicht verändert wird, kann man sagen, dass die Längen der Halbachsen zu $1/\sqrt{\lambda_j}$, $j = 1, 2$, also zu den Reziprokwerten der Wurzeln aus den Eigenwerten proportional sind. Gilt $\lambda_1 \geq \lambda_2$ (Programme zur Berechnung der Eigenwerte geben diese gewöhnlich der Größe nach geordnet aus), so ist die "erste" Hauptachse also die *kürzeste*. Allgemein ist für ein n -dimensionales Ellipsoid die "erste" Hauptachse die kürzeste, die "zweite" die zweitkürzeste, etc.

Abbildung 4: Hauptachsen als Eigenvektoren



Die Abb. 4 zeigt Ellipsen in verschiedenen Orientierungen und die jeweiligen - nicht normierten - Eigenvektoren. Normiert man diese Eigenvektoren, so sind sie durch die Spalten der in (119) definierten Matrix T gegeben.

Die zweite Möglichkeit, die Längen der Halbachsen zu bestimmen, besteht darin, die Lagrangsche Multiplikatorenregel (vergl. Anhang, Abschnitt 6.4) anzuwenden. Das Quadrat der Länge von \vec{x} werde wieder durch $r^2 = x_1^2 + x_2^2$ angegeben. Betrachtet werden muß nun

$$r^2 = x_1^2 + x_2^2 - \mu(ax_1^2 + bx_2^2 + 2cx_1x_2 - k_0). \quad (142)$$

Die partiellen Ableitungen nach x_1 bzw. x_2 werden gleich Null gesetzt und ergeben das Gleichungssystem

$$\begin{aligned} 0 &= 2x_1 - (2a\mu x_1 + 2c\mu x_2) \\ 0 &= 2x_2 - (2b\mu x_2 + 2c\mu x_1), \end{aligned}$$

wobei sich der Faktor 2 herauskürzt, d.h. man erhält

$$0 = x_1 - \mu(ax_1 + cx_2) \quad (143)$$

$$0 = x_2 - \mu(bx_2 + cx_1). \quad (144)$$

Nun ist aber

$$M\vec{x} = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 \\ bx_1 + cx_2 \end{pmatrix}.$$

Damit kann man die Gleichungen (143) und (144) vektoriell schreiben:

$$\vec{0} = \vec{x} - \mu M\vec{x}, \quad (145)$$

woraus $\mu M\vec{x} = \vec{x}$ oder

$$M\vec{x} = \frac{1}{\mu} \vec{x} \quad (146)$$

folgt. Damit hat man gezeigt, dass der Vektor mit maximaler Länge ein Eigenvektor von M ist, mit $1/\mu$ als zugehörigem Eigenwert. Multipliziert man von links mit \vec{x} , so erhält man

$$\vec{x}'M\vec{x} = \frac{1}{\mu}\vec{x}'\vec{x} = \frac{1}{\mu}r^2,$$

woraus sofort

$$\frac{1}{\mu}r^2 = k_0$$

folgt, denn die Nebenbedingung $\vec{x}'M\vec{x} = k_0$ muß ja erfüllt sein. Daraus folgt, dass die Länge des Vektors durch

$$r = \|\vec{x}\| = \sqrt{k_0\mu} = \sqrt{\frac{k_0}{\lambda}} \quad (147)$$

ist, wenn $\lambda = 1/\mu$ wieder den zu \vec{x} gehörenden Eigenwert bezeichnet. Dieses Ergebnis korrespondiert zu (140) und wurde gefunden, ohne auf den Fall $c = 0$ zu rekurren.

3 Hauptachsentransformation und Faktorenanalyse

3.1 Die Hauptachsentransformation

Gegeben sei eine quadratische Form $\vec{x}'M\vec{x} = k_0$, und M sei symmetrisch. Die Menge der Vektoren \vec{t} , die dieser Gleichung genügen, beschreiben entweder ein Ellipsoid oder ein Hyperboloid (für den Fall $n = 2$ also entweder eine Ellipse oder eine Hyperbel). Ist M eine Diagonalmatrix, d.h. sind die Elemente $m_{ij} = 0$ für alle i und j mit $i \neq j$, so ist z.B. das Ellipsoid achsenparallel.

Satz 3.1 *Es sei M eine symmetrische, positiv-semidefinite Matrix, so dass $\vec{x}'M\vec{x} = k_0$ ein Ellipsoid charakterisiert. Weiter sei T die Matrix der Eigenvektoren von M . Dann beschreiben die Vektoren \vec{y} , die der Bedingung $\vec{y}'\Lambda\vec{y} = k_0$ mit $\Lambda = TMT'$ genügen, ein achsenparalleles Ellipsoid.*

Beweis: Es sei Λ eine Diagonalmatrix und es gelte $\vec{y}'\Lambda\vec{y} = k_0$. Weiter existiere eine Matrix T derart, dass $\vec{x} = T\vec{y}$. Dann folgt

$$\vec{x}'M\vec{x} = \vec{y}'T'MT\vec{y} = k_0.$$

Aber es soll gleichzeitig $\vec{y}'\Lambda\vec{y} = k_0$ gelten, und deshalb ergibt sich

$$T'MT = \Lambda. \quad (148)$$

Nach (128) bedeutet diese Gleichung aber, dass T die Matrix der Eigenvektoren von M sein muß, und Λ ist die Matrix der zugehörigen Eigenwerte. Die Matrix T der Eigenvektoren von M transformiert also die Vektoren \vec{y} so, dass sich ein orientiertes Ellipsoid ergibt. \square

Anmerkung: Für eine beliebige quadratische Matrix T definiert die Gleichung $T\vec{y} = \vec{x}$ eine Transformation der Vektoren \vec{y} in die Vektoren \vec{x} , wobei die \vec{x} sich hinsichtlich ihrer Länge wie auch ihrer Orientierung unterscheiden. Ist T aber orthonormal, d.h. sind die Spaltenvektoren von T orthonormal, so gilt $TT' = T'T = I$, I die Einheitsmatrix. In diesem Fall verändert sich bei der Transformation nur die Orientierung, nicht aber die Länge, wie man sofort sieht:

$$\vec{x}'\vec{x} = \|\vec{x}\|^2 = \vec{y}'T'T\vec{y} = \vec{y}'\vec{y} = \|\vec{y}\|^2. \quad (149)$$

In (119), Seite 49, wird ein Beispiel für eine derartige Transformationsmatrix gegeben. Weiter impliziert die Orthonormalität von T , dass $T'T\vec{y} = \vec{y} = T'\vec{x}$, d.h. die Matrix T' transformiert (rotiert) die Vektoren \vec{x} in die Vektoren \vec{y} der achsenparallelen Ellipse.

□

Die Transformation $T\vec{y}$ bzw. $T'\vec{x}$ heißt deshalb *Hauptachsentransformation*.

Satz 3.2 Die Matrizen M und T seien wie in Satz 3.1 definiert. Die Eigenvektoren $\vec{t}_1, \dots, \vec{t}_n$, d.h. die Spalten von T , haben die Orientierung der Hauptachsen des zu M gehörigen Ellipsoids.

Beweis: Die Aussage läßt sich auf verschiedene Weise beweisen. Bei der ersten geht man einfach davon aus, dass die erste Hauptachse des achsenparallelen Ellipsoids durch $by_1 = y_1\vec{e}_1 = y_1(1, 0, \dots, 0)'$ gegeben ist, wobei \vec{e}_1 der erste Einheitsvektor ist und $y_1 = \|\vec{y}_1\|$ die Länge des Vektors ist, der die erste Hauptachse definiert. Durch T wird \vec{y}_1 in \vec{x}_1 rotiert; \vec{x}_1 definiert die erste Hauptachse des durch $\vec{x}'M\vec{x} = k$ definierten Ellipsoids; $\vec{x}_1 = T\vec{y}_1$, und

$$\vec{x}_1 = \vec{t}_1y_1 + 0\vec{t}_2 + \dots + 0\vec{t}_n = y_1\vec{t}_1,$$

d.h. \vec{x}_1 hat die Orientierung von \vec{t}_1 und die Länge y_1 . Hier wird von der Tatsache Gebrauch gemacht, dass die Gleichung $\vec{x} = T\vec{y}$ bedeutet, dass \vec{x} eine Linearkombination der Spaltenvektoren von T ist.

Ein anderer Beweis macht von der Technik der Bestimmung eines Extremwerts einer Funktion unter Nebenbedingungen Gebrauch. Es gelte $\vec{x}'M\vec{x} = k_0$, k_0 eine Konstante. Die Endpunkte der \vec{x} mögen auf einem Ellipsoid liegen. Insbesondere sei \vec{y} der Vektor, der der ersten Halbachse entspricht; \vec{y} genügt also ebenfalls der Bedingung $\vec{y}'M\vec{y} = k_0$. Aus der Definition der ersten Hauptachse folgt, dass $\|\vec{y}\| = \max\|\vec{x}\|$ gelten muß. Um \vec{y} zu bestimmen, muß also $\vec{x}'\vec{x} = \|\vec{x}\|^2$ maximiert werden unter den Nebenbedingung, dass $\vec{x}'M\vec{x} = k_0$ gilt. Dazu definiert man

$$Q(\vec{x}) = \vec{x}'\vec{x} - \lambda(\vec{x}'M\vec{x} - k_0), \quad (150)$$

wobei λ ein Lagrange-Multiplikator ist. Man leitet nun Q nach \vec{x} ab und setzt die Ableitung gleich Null; die Lösung der entstehenden Gleichung liefert den gesuchten Vektor \vec{y} . Man findet

$$\left. \frac{dQ}{d\vec{x}} \right|_{\vec{x}=\vec{y}} = 2M\vec{y} - 2\lambda\vec{y} = 0,$$

woraus sich die Gleichung

$$(M - \lambda I)\vec{y} = \vec{0} \quad (151)$$

ergibt, I die Einheitsmatrix. Aber dieses Resultat bedeutet

$$M\vec{y} = \lambda\vec{y}, \quad (152)$$

woraus folgt, dass \vec{y} proportional zu einem Eigenvektor von M sein muß. Da \vec{y} derjenige Vektor aus der Menge der Vektoren \vec{x} ist, der maximale Länge hat, korrespondiert er zu dem Eigenvektor \vec{t}_1 mit maximalem Eigenwert λ_1 . Es existiert demnach eine Zahl q derart, dass $\vec{y} = q\vec{t}_1$, und \vec{t}_1 zeigt in die Richtung der ersten Hauptachse. Der zweite Eigenvektor \vec{t}_2 ist orthogonal zu \vec{t}_1 und zeigt demnach in Richtung der zweiten Hauptachse, etc. \square

Eigenvektoren und Rotation: In (119) wurde T als Rotationsmatrix eingeführt. Also kann man schließen, dass im 2-dimensionalen Fall die Matrix der Eigenvektoren die Form

$$T = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}$$

hat. Der erste Eigenvektor $\vec{t}_1 = (t_{11}, t_{21})'$ ist demnach durch den ersten Spaltenvektor von T gegeben, also $t_{11} = \cos \phi$, $t_{21} = -\sin \phi$. Daraus läßt sich der Winkel ϕ berechnen:

$$\cos^{-1}(t_{11}) = \phi, \quad \sin^{-1}(t_{21}) = -\phi. \quad (153)$$

Satz 3.3 Die Länge a einer Halbachse ist umgekehrt proportional zum Eigenwert λ des entsprechenden Eigenvektors \vec{t} , also $a \propto \sqrt{1/\lambda}$.

Beweis: Die Länge einer Halbachse sei a . Nach Satz 3.2 haben die Eigenvektoren die Orientierung der Halbachsen, so dass der Vektor \vec{y} , der eine bestimmte Hauptachse definiert, der Bedingung $M\vec{y} = \lambda\vec{y}$ genügt. Die Länge des Eigenvektors kann beliebig gewählt werden, also kann man insbesondere $\|\vec{y}\| = a$ setzen. Dann ist $\vec{y}'\vec{y} = \|\vec{y}\|^2 = a^2$, und man erhält aus $M\vec{y} = \lambda\vec{y}$ durch Multiplikation von links mit \vec{y}' die Beziehung

$$\vec{y}'M\vec{y} = k_0 = \lambda\|\vec{y}\|^2 = \lambda a^2,$$

woraus

$$a = \sqrt{k_0/\lambda} \quad (154)$$

folgt. \square

Anmerkung über die Transformation von Spalten- und Zeilenvektoren: Allgemein wird ein Vektor transformiert, indem man ihn mit einer Matrix multipliziert. Da Vektoren standardmäßig als Spaltenvektoren angeschrieben

werden, ergibt sich die Gleichung (134), d.h. $T\vec{y} = \vec{x}$, - hier wird der Vektor von links mit einer Matrix multipliziert. Gelegentlich ist es nützlich oder nötig, einen Zeilenvektor in einen anderen Zeilenvektor zu überführen. Dazu muß der zu transformierende Zeilenvektor von rechts mit einer geeigneten Matrix multipliziert werden (denn Zeilenvektor mal Matrix ergibt einen Zeilenvektor). Man hat also

$$\vec{y}'P = \vec{x}'. \quad (155)$$

Gilt für die Spaltenvektoren die Gleichung $T\vec{y} = \vec{x}$, so folgt durch Transponieren oder "Stürzen" die Gleichung

$$(T\vec{y})' = \vec{y}'T' = \vec{x}',$$

und damit $P = T'$. Für die Hauptachsentransformation sind beide Ansätze gleichwertig, denn aus $\vec{x}'M\vec{x} = k_0$ wird $\vec{y}'T'MT\vec{y} = \vec{y}'\Lambda\vec{y} = k_0$. Bei der Schätzung der Parameter a_{jk} und q_{ik} geht man in der Tat von der Transformation von Zeilenvektoren aus.

Die multivariate Normalverteilung: Die multivariate Normalverteilung ist durch

$$f(\vec{x}) = c_0 \exp \left[-\frac{1}{2}(\vec{x} - \vec{\mu})'S^{-1}(\vec{x} - \vec{\mu}) \right] \quad (156)$$

definiert, wobei c_0 ein Normierungsfaktor ist, und S^{-1} ist die Inverse der Varianz-Kovarianz-Matrix S der n Komponenten von \vec{x} . S hat die allgemeine Form

$$S = \begin{pmatrix} s_1^2 & s_1s_2r_{12} & \cdots & s_1s_nr_{1n} \\ s_2s_1r_{21} & s_2^2 & \cdots & s_2s_nr_{2n} \\ & & \vdots & \\ s_ns_1r_{n1} & s_ns_2r_{n2} & \cdots & s_n^2 \end{pmatrix}. \quad (157)$$

Die allgemeine Form der Inversen S^{-1} wird hier nicht gegeben, aber auf die Spezialfälle $n = 2$ und $n = 3$ wird weiter unten zurückgekommen.

S ist eine symmetrische, positiv-definite Matrix, und damit ist auch S^{-1} symmetrisch und positiv definit. $(\vec{x} - \vec{\mu})'S^{-1}(\vec{x} - \vec{\mu}) = k_0$ definiert also ein Ellipsoid. Es gelte $ST = T\Lambda$, d.h. T ist die Matrix der Eigenvektoren von S und Λ ist die Diagonalmatrix der zugehörigen Eigenwerte. Da S symmetrisch ist, sind die Eigenvektoren orthonormal und es gilt $T'ST = \Lambda$. Für die Inverse Λ^{-1} folgt dann

$$\Lambda^{-1} = \begin{pmatrix} 1/\lambda_1 & 0 & \cdots & 0 \\ 0 & 1/\lambda_2 & \cdots & 0 \\ 0 & & \vdots & \\ 0 & 0 & \cdots & 1/\lambda_n \end{pmatrix} = (T'ST)^{-1} = T^{-1}S^{-1}(T')^{-1}. \quad (158)$$

Aber für orthonormale Matrizen gilt $T^{-1} = T'$ und damit $(T')^{-1} = T$, also hat man

$$T'S^{-1}T = \Lambda^{-1}; \quad (159)$$

diese Aussage ist auch dem Satz 2.4, Gleichung (130), Seite 53, für $S = M$ zu entnehmen.

Es werde zur Abkürzung $\vec{x} - \vec{\mu} = \vec{u}$ geschrieben, so dass die Gleichung $(\vec{x} - \vec{\mu})' S^{-1} (\vec{x} - \vec{\mu}) = k_0$ in der Form $\vec{u}' S^{-1} \vec{u} = k_0$ erscheint. Dann impliziert (159)

$$T\vec{u} = \vec{w}, \quad \vec{w}' \Lambda^{-1} \vec{w} = k_0, \quad (160)$$

d.h. die Vektoren \vec{w} definieren ein achsenparalleles Ellipsoid.

Anmerkung: Die Länge der Hauptachsen dieses Ellipsoids ist identisch mit der der Längen des ursprünglichen Ellipsoids. Schreibt man $\Lambda^{-1} = \text{diag}(\mu_1, \dots, \mu_n)$, so ist die Länge der Hauptachsen nach (154) durch $a = \sqrt{k_0/\mu}$ gegeben. Da aber $\mu = 1/\lambda$, λ der entsprechende Eigenwert von S , so erhält man für die Länge der Achse

$$a = \sqrt{k_0 \lambda}, \quad (161)$$

d.h. die Länge der Halbachse ist proportional zum entsprechenden Eigenwert von S .

Der Spezialfall $n = 2$: Man kann $r_{12} = r_{21} = r$ setzen und erhält insbesondere für S

$$S = \begin{pmatrix} s_1^2 & s_1 s_2 r \\ s_1 s_2 r & s_2^2 \end{pmatrix}, \quad (162)$$

und für die Inverse hat man

$$S^{-1} = \begin{pmatrix} \frac{1}{s_1^2(1-r^2)} & -\frac{r}{s_1 s_2(1-r^2)} \\ -\frac{r}{s_1 s_2(1-r^2)} & \frac{1}{s_2^2(1-r^2)} \end{pmatrix}, \quad -1 < r < 1. \quad (163)$$

Man bemerke die Bedingung $-1 < r < 1$ in (163). Setzte man nämlich $r = 1$ oder $r = -1$, wo würden die Elemente in S^{-1} unendlich, da dann $1 - r^2 = 0$. Der Fall $r = 0$ muß in (162) angenommen werden; für die Inverse findet man dann

$$S^{-1} = \begin{pmatrix} 1/s_1^2 & 0 \\ 0 & 1/s_2^2 \end{pmatrix}. \quad (164)$$

(Vergl. Satz 3.3, p. 3.3, insbesondere (154).)

Schreibt man die Dichtefunktion für $n = 2$ explizit aus, so erhält man

$$f(x_1, x_2) = \frac{1}{2\pi s_1 s_2 \sqrt{1-r^2}} \exp \left[-\frac{1}{2(1-r^2)} \left(\left(\frac{x_1 - \mu_1}{s_1} \right)^2 + \left(\frac{x_2 - \mu_2}{s_2} \right)^2 - \frac{2r(x_1 - \mu_1)(x_2 - \mu_2)}{s_1 s_2} \right) \right] \quad (165)$$

Für konstantes $k_0 = \vec{u}' S^{-1} \vec{u}$, $\vec{u} = \vec{x} - \vec{\mu}$, liegen die Endpunkte der Vektoren \vec{u} auf einer durch S^{-1} definierten Ellipse. Für die Eigenwerte von S^{-1} erhält man

$$\lambda_{1,2} = \frac{\frac{1}{2}[s_1^2 + s_2^2 \mp \sqrt{s_1^4 + s_2^4 - 2(s_1 s_2)^2(1-r^2)}]}{(s_1 s_2)^2(1-r^2)} \quad (166)$$

Die Gleichung (161) kann nun angewendet werden, um die Länge der Halbachsen in Abhängigkeit von s_1 , s_2 und r zu bestimmen.

Für $r = 0$ insbesondere erhält man eine achsenparallele Ellipse, mit den zugehörigen Eigenwerten

$$\begin{aligned}\lambda_{1,2} &= \frac{\frac{1}{2}[s_1^2 + s_2^2 \mp \sqrt{s_1^4 + s_2^4 - 2(s_1 s_2)^2}]}{(s_1 s_2)^2} \\ &= \frac{\frac{1}{2}[s_1^2 + s_2^2 \mp \sqrt{(s_1^2 - s_2^2)^2}]}{(s_1 s_2)^2} \\ &= \frac{\frac{1}{2}[s_1^2 + s_2^2 \mp (s_1^2 - s_2^2)]}{(s_1 s_2)^2},\end{aligned}$$

d.h.

$$\lambda_1 = 1/s_1^2, \quad \lambda_2 = 1/s_2^2. \quad (167)$$

Nach Satz 3.3, Gleichung (154) (Seite 58) sind dann die Längen der Halbachsen proportional zu den Varianzen s_1^2 und s_2^2 .

3.1.1 Die Singularwertzerlegung

Es wird jetzt eine Möglichkeit vorgestellt, die die latenten Variablen repräsentierenden Vektoren zu schätzen. Man erinnere sich: diese Vektoren sollen linear unabhängig sein, müssen aber nicht orthogonal sein. Orthogonale Basisvektoren sind ein Spezialfall, der lineare Unabhängigkeit impliziert.

Die einfachste Methode, eine Menge von Basisvektoren zu finden, beruht allerdings auf der Annahme orthogonaler Basisvektoren. Das ist zunächst einmal keinerlei Einschränkung, denn die Basis eines Vektorraumes oder Teilraumes ist nicht eindeutig bestimmt. Es kommt ja nur darauf an, irgendeine Menge von linear unabhängigen Vektoren als (Teil-)Basis zu finden. Es ist stets möglich, durch Wahl einer geeigneten Transformationsmatrix von einer Basis zu einer anderen überzugehen. Es ist also auch möglich, von einer orthogonalen Basis oder Teilbasis zu einer nicht-orthogonalen ("obliquen") Basis oder Teilbasis überzugehen. Einen solchen Übergang wird man in Betracht ziehen, wenn sich für eine oblique Basis eine bessere Interpretierbarkeit als für eine orthogonale Basis andeutet.

Im Folgenden wird eine Zerlegung der Datenmatrix X (oder $\S Z$) hergeleitet, die stets auf orthogonale (Teil-)Basen für die Zeilen- und Spaltenvektoren von X oder Z führt. Diese Zerlegung kann für *jede* Matrix X oder Z gefunden werden, und insofern findet man stets eine "Lösung" für das Problem, solche Basen zu finden. Ob sie tatsächlich eine sinnvolle, inhaltliche Interpretation als latente Variablen für die \vec{X}_j oder \vec{Y}_i zulassen ist eine Frage, die nur anhand des im Folgenden beschriebenen Resultats nicht beantwortet werden kann.

Die Singularwertzerlegung (SVD): Es sei Z eine beliebige⁹ $(m \times n)$ -Matrix. Ohne Beschränkung der Allgemeinheit kann man sagen, dass die Spaltenvektoren Linearkombinationen von n linear unabhängigen, m -dimensionalen Vektoren sind (auf den Fall $r < n$ l.u. Vektoren wird noch zurückgekommen). Für den j -ten Spaltenvektor \vec{Z}_j von Z hat man die Darstellung

$$\vec{Z}_j = u_{1j}\vec{v}_1 + \cdots + u_{nj}\vec{v}_n,$$

wobei die \vec{u}_k in \vec{u}_k und die \vec{v}_k in \vec{v}_k umbenannt wurden. Diese Gleichung entspricht (29), wobei wegen der Standardisierung $\|\vec{v}_k\| = 1$ für alle k . Hier tritt allerdings kein Fehlerterm ε_{ij} auf, was weiter unten noch erläutert werden wird.

Die Vektoren \vec{v}_k , $k = 1, \dots, n$ sind unbekannt, ebenso die n -dimensionalen Vektoren $\vec{u}_1, \dots, \vec{u}_m$. Diese Vektoren müssen aus den Daten ausgerechnet werden. Es sei $V = [\vec{v}_1, \dots, \vec{v}_n]$ die Matrix, die entsteht, wenn man die Vektoren \vec{v}_k nebeneinander schreibt, dh die Spalten von V seien die \vec{v}_k . Die Berechenbarkeit von V bedeutet nun, dass eine Matrix U existieren muß derart, dass

$$ZU = V. \quad (168)$$

Den Regeln der Matrixmultiplikation entsprechend sind also die Spaltenvektoren von V Linearkombinationen der Spaltenvektoren von X ; ebenso kann man sagen, dass die Spaltenvektoren von V sind Transformationen der Spaltenvektoren von X sind. Es werde nun gefordert, dass die Matrix V orthogonal ist, d.h. die Spaltenvektoren von V sollen paarweise orthogonal zueinander sein. Dann impliziert (168), dass

$$V'V = U'Z'ZU = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) \quad (169)$$

denn V enthält ja nach Voraussetzung orthogonale Spaltenvektoren, $V'V$ enthält die Skalarprodukte dieser Vektoren, die wegen der Orthogonalität alle gleich Null sind bis auf die Skalarprodukte der \vec{v}_k mit sich selbst, d.h.

$$\lambda_k = \vec{v}_k' \vec{v}_k = \|\vec{v}_k\|^2, \quad k = 1, \dots, n \quad (170)$$

Die λ_k in den Diagonalzellen von Λ bedeuten hier also die Quadrate der Längen den \vec{v}_k . Wenn aber $U'Z'ZU$ eine Diagonalmatrix ist, so bedeutet dies, dass U gerade die Matrix der orthonormalen Eigenvektoren von $Z'Z$ ist, und die lassen sich mit Hilfe numerischer Methoden berechnen. Damit ist die Frage, wie die Vektoren \vec{v}_k gefunden werden können, auch schon beantwortet! Die Transformation der Spaltenvektoren von Z in die von V bedeutet einfach eine Hauptachsentransformation.

⁹ Z beliebig heißt, dass die Matrix keine standardisierten Werte enthalten muß. Es wird hier Z geschrieben, damit man ohne weitere Umbenennung den Falls standardisierter Werte betrachten kann.

Wegen der Orthonormalität von U erhält man aus (168) sofort $ZUU' = VU' = Z$, oder

$$Z = VU'. \quad (171)$$

Während die Spaltenvektoren der Matrix U auf die Länge 1 normiert sind (als Eigenvektoren von $Z'Z$ werden sie schon als normierte Vektoren berechnet), sind die von V nicht notwendig normiert. Aber man kann sie normieren, indem man ihre Komponenten jeweils durch die Längen der Vektoren dividiert. Die Längen sind aber gerade die Wurzeln $\sqrt{\lambda_k}$. Man erhält die Matrix Q der normierten Vektoren aus V , indem man V mit

$$\Lambda^{-1/2} = \text{diag}(1/\sqrt{\lambda_1}, \dots, 1/\sqrt{\lambda_n})$$

multipliziert, d.h.

$$Q = V\Lambda^{-1/2}. \quad (172)$$

Damit erhält man die Darstellung

$$V = Q\Lambda^{1/2}. \quad (173)$$

In (171) eingesetzt erhält man die

Singularwertzerlegung (SVD)¹⁰

$$Z = Q\Lambda^{1/2}U' \quad (174)$$

Der Sachverhalt, dass U eine orthonormale Matrix ist, ergab sich bereits aus der Tatsache, dass U die Matrix der Eigenvektoren von $Z'Z$ ist. Bestimmt man nun die Matrix ZZ' , so erhält man

$$ZZ' = Q\Lambda^{1/2}U'U\Lambda^{1/2}Q' = Q\Lambda Q'. \quad (175)$$

Der Ausdruck $Q\Lambda Q'$ bedeutet aber, dass Q die Matrix der (normierten) Eigenvektoren von ZZ' ist, woraus folgt, dass Q ebenfalls eine orthonormale Matrix ist, d.h. die Spaltenvektoren von Q sind orthonormal. Die Diagonalelemente von Λ enthalten die von Null verschiedenen Eigenwerte von ZZ' , die offenbar gleich den von Null verschiedenen Eigenwerten von $Z'Z$ sind.

Anmerkungen:

1. Es ist $Z'Z/m = R$, R die Matrix der Korrelationen zwischen den gemessenen Variablen V_1, \dots, V_n . Formal entspricht die Matrix ZZ'/n einer Matrix der Korrelationen zwischen den Personen, weil ihre Elemente die Skalarprodukte der Zeilenvektoren von Z sind. Tatsächlich handelt es sich aber nicht um Korrelationen zwischen den Personen, weil die Matrix Z durch *Spaltennormierung* der Rohdatenmatrix X entstanden ist,

¹⁰Singular Value Decomposition

d.h. man hat für jede Variable V_j den Mittelwert \bar{x}_j und die entsprechende Streuung s_j bestimmt und daraus die $z_{ij} = (x_{ij} - \bar{x}_j)/s_j$ berechnet. Damit ZZ' Korrelationen zwischen den Personen enthält, hätte man eine *Zeilenstandardisierung* vornehmen müssen, d.h. man hätte für jede Person den Mittelwert \bar{x}_i über die Messwerte in den verschiedenen Variablen und die jeweilige Streuung s_i dieser Werte bestimmen müssen und die Standardisierung in bezug auf diese Werte vornehmen müssen. Ein solches Vorgehen macht aber nur Sinn, wenn die Variablen nicht inhaltlich verschiedene Größen sind: man kann ja nicht gut Äpfel und Birnen mitteln. Voraussetzung für eine Zeilenstandardisierung ist also, dass die verschiedenen Variablen tatsächlich nicht verschieden sind, sondern stets die gleiche Variable sind, - etwa gemessen zu verschiedenen Zeitpunkten.

2. Die Singularwertzerlegung ist ein Theorem der linearen Algebra und ist deswegen nicht an die psychologische Theorie latenter Variablen gebunden. Jede reelle Matrix X läßt sich in der Form $X = Q\Lambda^{1/2}P'$ darstellen, mit Q die Matrix der Eigenvektoren von XX' und P die Matrix der Eigenvektoren von $X'X$, Λ die Diagonalmatrix der von Null verschiedenen Eigenwerte (von $X'X$ sowohl wie von XX'). Die Berechnung der "Faktoren" über die Singularwertzerlegung bedeutet also noch nicht, dass entsprechende psychologische Dimensionen auch existieren. Die Existenz solcher Dimensionen ist eine Hypothese, die noch überprüft werden muß.

3.1.2 Faktorladungen und Faktorwerte

In der SVD

$$Z = Q\Lambda^{1/2}U'$$

sind die Spaltenvektoren von Q und U normiert. Q ist eine $(m \times n)$ -Matrix, und U ist eine $(n \times n)$ -Matrix. Schon die Dimensionen (m und n) der Matrizen zeigen, dass Q die Personen, U dagegen die Variablen V_1, \dots, V_n repräsentiert. Die Frage ist, welche Rolle $\Lambda^{1/2}$ in diesem Zusammenhang spielt.

Eine Möglichkeit besteht darin, die Matrix $\Lambda^{1/4}$ zu definieren; ihre Diagonalelemente enthalten die Wurzeln aus den Wurzeln der Eigenwerte. Man kann dann die Matrizen $\tilde{Q} = Q\Lambda^{1/4}$ und $\tilde{U} = U\Lambda^{1/4}$ definieren und erhält

$$Z = \tilde{Q}\tilde{U}' = Q\Lambda^{1/4}\Lambda^{1/4}U' = Q\Lambda^{1/2}U' \quad (176)$$

Wieder repräsentieren \tilde{Q} und \tilde{U} die Personen bzw. die Variablen. Die Multiplikation mit $\Lambda^{1/4}$ bedeutet eine Skalierung der Koordinaten in Q und U . Bei bestimmten Anwendungen der SVD wird tatsächlich diese Skalierung gewählt, etwa in der Korrespondenzanalyse, die zur Analyse von Häufigkeitstabellen verwendet wird, wo man eine identische Skalierung der Zeilen- und Spaltenkategorien wünscht. Bei faktorenanalytischen Untersuchungen ist man aber oft nur an der Skalierung der Variablen interessiert. In einem solchen Fall wird

nur die Matrix U skaliert. Dazu definiert man die Matrix

$$A = U\Lambda^{1/2}, \quad (177)$$

so wird (174) zu

$$Z = QA'. \quad (178)$$

Für die Spaltenvektoren \vec{Z}_j von Z erhält man daraus die Darstellung

$$\vec{Z}_j = a_{1j}\vec{q}_1 + \dots + a_{nj}\vec{q}_n \quad (179)$$

wobei a_{kj} das Element in der k -ten Zeile und j -ten Spalte von A' ist und \vec{q}_k ist der k -te Spaltenvektor von Q . Da die \vec{q}_k paarweise orthonormal sind, folgt, dass sie auch linear unabhängig sind. Dies bedeutet, dass man mit der SVD eine Lösung für die Aufgabe, die \vec{Z}_j als Linearkombination von linear unabhängigen Vektoren darzustellen, gefunden hat.

Definition 3.1 *Die Koeffizienten*

$$a_{1j}, a_{2j}, \dots, a_{nj}$$

heißen die Faktorladungen der j -ten gemessenen Variablen auf den latenten Variablen (oder Dimensionen) F_1, \dots, F_n . Die

$$q_{1i}, \dots, q_{ni}$$

heißen Faktorwerte (Faktor-Scores) der Personen auf den latenten Variablen.

Umgekehrt kann man natürlich auch Q mit $\Lambda^{1/2}$ skalieren; darauf wird weiter unten noch zurückgekommen. Zuerst wird die Bedeutung der Skalierung $A = U\Lambda^{1/2}$ diskutiert. Die Matrix $R = (r_{jk})$ der Korrelationen zwischen den Variablen V_j und V_k ist wegen (177) durch

$$R = \frac{1}{m} Z'Z = A Q' Q A' = A A' \quad (180)$$

gegeben, d.h. die Korrelationen zwischen den Variablen lassen sich aus den Faktorladungen gemäß

$$r_{jk} = \frac{1}{m} \sum_{i=1}^n a_{ij} a_{ik} \quad (181)$$

errechnen. Insbesondere hat man

$$r_{jj} = \frac{1}{m} \sum_{i=1}^n a_{ij}^2 = 1. \quad (182)$$

Für $n = 2$ bedeutet dies, dass $a_{1j}^2 + a_{2j}^2 = 1$, d.h. die Variablen V_j liegen alle auf dem Umfang eines Kreises mit dem Radius 1. Für $n = 3$ liegen sie alle auf der Oberfläche einer Kugel mit dem Radius 1, und für $n > 3$ liegen sie

auf der Oberfläche einer Hyperkugel mit dem Radius 1. Dieser Sachverhalt ist von Bedeutung, wenn man nur $r < n$ Dimensionen berücksichtigen will; in diesem Fall betrachtet man die $n - r$ vernachlässigten Dimensionen als Fehler. Die Gleichungen (180) und (182) gelten dann i. A. nur approximativ, erlauben aber eine Abschätzung der Güte der Approximation.

Anmerkung: Nach Gleichung (180), $R = AA'$, hängen die Korrelationen zwischen den Variablen nur von $A = U\Lambda^{1/2}$ ab, d.h. die Größen q_{ij} , die die Personen auf den latenten Variablen abbilden (d.h. die Ausprägungen der latenten Variablen bei den einzelnen Personen) *gehen nicht* in die Korrelationen ein. Die Korrelationen hängen nur von den Ausmaßen, mit denen die gemessenen Variablen die latenten Variablen erfassen ab. \square

Die Definition von A , $A = U\Lambda^{1/2}$, impliziert

$$a_{jk} = u_{jk}\sqrt{\lambda_k} \quad (183)$$

Daraus folgt

$$\sum_{j=1}^n a_{jk}^2 = \lambda_k \sum_{k=1}^n u_{jk}^2 = \lambda_k \quad (184)$$

denn $\sum_k u_{jk}^2 = 1$ wegen der Normiertheit der Vektoren in U . Der k -te Eigenwert λ_k ist also gleich der Summe der Quadrate der Ladungen der Variablen auf der k -ten Dimension. Für große λ_k -Werte werden also auch die Werte der $|a_{jk}|$ groß sein, d.h. große λ_k -Werte deuten größere Faktorladungen der Variablen an. Eine große Faktorladung indiziert aber, dass die entsprechende Dimension ausgeprägt in einer gemessenen Variablen enthalten ist. Auf diese Weise zeigen die λ_k -Werte die Bedeutung einer Dimension. Dieser Aspekt wird noch einmal erläutert, wenn die Skalierung der Faktor-Scores mit den $\sqrt{\lambda_k}$ diskutiert wird.

Nun werde die Matrix $V = Q\Lambda^{1/2}$ betrachtet, so dass $Z = VU'$. In diesem Fall erhält man

$$R = \frac{1}{m}Z'Z = \frac{1}{m} = \frac{1}{m}UV'VU' = \frac{1}{m}U\Lambda U', \quad (185)$$

wegen

$$V'V = \Lambda^{1/2}Q'Q\Lambda^{1/2} = \Lambda.$$

Für den k -ten Vektor \vec{V}_k ergibt sich also

$$\vec{V}_k'\vec{V}_k = \lambda_k \sum_{i=1}^m q_{ik}^2 = \lambda_k, \quad (186)$$

denn $\sum_i q_{ik}^2 = 1$ wegen der Normiertheit der \vec{q}_k . Die q_{ik} sind aber standardisierte Werte, dh

$$\sum_{i=1}^m \lambda_j q_{ik} = \lambda_k \sum_{i=1}^m q_{ik} = 0,$$

wegen $\sum_i q_{ik} = 0$. Daraus folgt, dass λ_k/m die Varianz der v_{ik} -Werte ist, d.h. λ_k ist die *Varianz der Koordinaten der Personen auf der k-ten Dimension*. Da die Dimensionen unabhängig voneinander sind folgt, dass

$$\sigma_{total}^2 = \frac{1}{m} \sum_{k=1}^n \lambda_k \quad (187)$$

als Gesamtvarianz aufgefaßt werden kann, und dementsprechend

$$\pi_k = \frac{\lambda_k}{\sum_j \lambda_j} \quad (188)$$

als Anteil der Gesamtvarianz, der durch die k -te latente Variable (Dimension) erklärt wird.

Abschätzung der Anzahl der Dimensionen: Der Rang r einer $(m \times n)$ -Matrix ist stets kleiner oder höchstens gleich der kleineren der beiden Zahlen m und n , $r \leq \min(m, n)$. Da im Allgemeinen $m > n$ ist bisher angenommen worden, dass für den Rang von Z $r = n$ gilt. Die SVD "erklärt" mit dieser Annahme die Datenmatrix stets vollständig. Tatsächlich wird die wahre Anzahl r der latenten Dimensionen aber i. A. kleiner als n sein. Aufschluß darüber kann die Betrachtung der λ_k bzw. der π_k geben. Sind nur einige der π_k , etwa r , "groß" und die restlichen "klein", so kann man vermuten, dass die "kleinen" Eigenwerte bzw. π_k -Werte nur "Rauschen" und keine latenten Variablen repräsentieren. Es sind dann die Messfehler, die bewirken, dass alle Eigenwerte von Null verschieden sind. Eine zu dieser Überlegung korrespondierende graphische Methode ist der *Scree-Test*. Dabei werden die Anteile π_k gegen ihren Rangplatz aufgetragen; das Ergebnis sieht dann wie ein *scree*, d.h. wie eine Geröllhalde aus, wie man sie an steilen Felswänden im Gebirge beobachten kann; am Fuß der Wand sammelt sich Geröll, das sanfter ausschwingt als die Felswand - vergl. Abb. 6, p. 81. Sind die ersten s Anteile groß und kommt es dann zu einem Absturz wie bei der Felswand, so dass sich restlichen Anteile wie das Geröll am Fuß der Felswand zeigen, so ist s oft eine gute Schätzung der Anzahl zu berücksichtigenden latenten Dimensionen.

Eigenwerte und Korrelationen Wegen der SVD gilt für die spaltenstandardisierten Daten Z die Beziehung

$$Z = Q\Lambda^{1/2}P',$$

und damit für die Matrix der Korrelationen

$$R = \frac{1}{m} Z'Z = \frac{1}{m} P\Lambda P'. \quad (189)$$

Es werde zuerst der Spezialfall $R = I$, I die $(n \times n)$ -Einheitsmatrix, betrachtet, d.h. der Fall $r_{jk} = 0$ für alle $j \neq k$. Multiplikation von rechts mit P liefert

$$IP = P = P\Lambda. \quad (190)$$

Schreibt man P als Anordnung der Spaltenvektoren $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_n$, so hat man also

$$[\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_n] = [\lambda_1 \mathbf{P}_1, \lambda_2 \mathbf{P}_2, \dots, \lambda_n \mathbf{P}_n],$$

d.h.

$$\mathbf{P}_j = \lambda_j \mathbf{P}_j, \quad j = 1, \dots, n$$

woraus sofort

$$\lambda_j = 1 \text{ für alle } j$$

folgt, d.h. alle Eigenwerte in Λ haben den Wert 1.

Nun gelte $r_{jk} = 1$ für alle j, k , d.h. alle Variablen korrelieren perfekt miteinander. Dies heißt, dass $\mathbf{Z}'_j \mathbf{Z}_k = m$ für alle j, k , d.h. $\mathbf{Z}_j = \mathbf{Z}_k$ für alle j, k . Die Spaltenvektoren von Z sind aber Linearkombinationen der Spalten von Q , wenn $Z = Q\Lambda^{1/2}P'$, oder von L , wenn man $L = Q\Lambda^{1/2}$ setzt. Mithin muß

$$\mathbf{Z}_j = p_{j1}\mathbf{L}_1 + p_{j2}\mathbf{L}_2 + \dots + p_{jn}\mathbf{L}_n$$

gelten, wobei $(p_{j1}, p_{j2}, \dots, p_{jn})'$ die j -te Spalte von P' , d.h. die j -te Zeile von P ist. Analog dazu gilt für \mathbf{Z}_k

$$\mathbf{Z}_k = p_{k1}\mathbf{L}_1 + p_{k2}\mathbf{L}_2 + \dots + p_{kn}\mathbf{L}_n$$

Da nach Voraussetzung $\mathbf{Z}_j - \mathbf{Z}_k = \vec{0}$ sein muß, folgt

$$\vec{0} = (p_{j1} - p_{k1})\mathbf{L}_1 + (p_{j2} - p_{k2})\mathbf{L}_2 + \dots + (p_{jn} - p_{kn})\mathbf{L}_n.$$

Da aber die \mathbf{L}_j linear unabhängig sind, folgt $p_{ji} - p_{ki} = 0$ für alle i , d.h.

$$p_{ji} = p_{ki}, \quad i = 1, \dots, n$$

Dies bedeutet aber, dass die Matrix identische Zeilen hat, d.h. die Spalten enthalten jeweils identische Werte. Sind diese Werte ungleich Null so verschwinden die Skalarprodukte verschiedener Spaltenvektoren nicht, entgegen der Voraussetzung, dass verschiedene Spaltenvektoren orthogonal sein müssen. Also müssen die Elemente der Spalten alle gleich Null sein bis auf die Elemente einer Spalte, die alle gleich 1 sein müssen, und dies bedeutet, dass $\mathbf{Z}_j = \mathbf{L}_1$ für alle j , d.h. man benötigt nur eine latente Variable (was intuitiv ein triviales Resultat ist, denn wenn die \mathbf{Z}_j alle identisch sind, definieren sie eben nur eine Orientierung im \mathbf{R}^m).

Sind also die gemessenen Variablen paarweise unkorreliert, so folgt, dass die Eigenwerte alle gleich 1 sind, und identische Eigenwerte, falls sie ungleich Null sind, implizieren paarweise unkorrelierte Korrelationen zwischen den Variablen. In diesem Fall sind die Varianzen der Projektionen der Fälle auf die latenten Dimensionen alle gleich groß, d.h. die Punktekongfiguration ist eine Hyperkugel im n -dimensionalen Raum. Korrelieren die Variablen andererseits perfekt miteinander, so wird ein 1-dimensionaler Teilraum im \mathbf{R}^m definiert, – man benötigt nur eine "latente" Variable.

Die beiden Fälle werden in der Praxis nicht in "Reinkultur" auftreten, aber man kann den Fall praktisch unkorrelierter Variablen haben, der so gut wie keine Reduktion auf einen Teilraum zuläßt, oder den Fall hochkorrelierender Variablen, der eine Reduktion auf einen 1-dimensionalen Teilraum erlaubt. Verteilen sich die Korrelationen auf den Bereich $-1 < r_{jk} < 1$, so nehmen die λ_j verschieden große Werte an, wobei besonders kleine λ_j -Werte häufig als Effekt von "Fehlern" gedeutet werden können. \square

3.1.3 Formale Aspekte der Interpretation

Die Zerlegung von Z in der Form $Z = Q\Lambda^{1/2}U'$ legt zunächst nicht fest, ob eher auf die Personen ("Typen") fokussiert werden soll oder eher auf die Variablen, - die Personenstruktur steckt in der normierten Matrix Q bzw. in $V = Q\Lambda^{1/2}$, und die Variablenstruktur findet sich in der Matrix U bzw. in $U\Lambda^{1/2}$. Da sich die Elemente von Q oder V auf Personen (oder Objekte, die vermessen werden) beziehen, muß etwa q_{ik} oder v_{ik} etwas wie das Ausmaß, in dem die i -te Person oder das i -te Objekt über das k -te latente Merkmal verfügt repräsentieren. Analog dazu repräsentiert a_{jk} oder u_{jk} das Ausmaß oder den Anteil, in dem die j -te gemessene Variable die k -te latente Variable erfaßt.

Nun hat eine Person das k -te latente Merkmal in einem bestimmten Ausmaß unabhängig davon, in welchem Ausmaß eine Variable V_j dieses k -te Merkmal erfaßt. In diesem Sinne kann man sagen, dass die q_{ik} und die u_{jk} voneinander unabhängige Größen sind. Aber schon bei der Betrachtung der Zeilen- und Spaltenvektoren \vec{Y}_i und \vec{q}_k in den Gleichungen(??) und (??) (Seite ??) zeigte sich eine wechselseitige Abhängigkeit dieser Parameter: es war

$$\begin{aligned} \vec{Y}_i &= q_{i1}\vec{b}_1 + \dots + q_{ir}\vec{b}_r, \quad \vec{b}_k = \begin{pmatrix} b_{1k} \\ \vdots \\ b_{nk} \end{pmatrix}, \\ \vec{X}_j &= b_{1j}\vec{q}_1 + \dots + b_{rj}\vec{q}_r, \quad \vec{q}_k = \begin{pmatrix} q_{1k} \\ \vdots \\ q_{mk} \end{pmatrix}, \quad k = 1, \dots, r \end{aligned}$$

Die Koeffizienten ("Gewichte") der \vec{b}_k sind die Komponenten der \vec{q}_k , und die Koeffizienten der \vec{q}_k sind die Komponenten der \vec{b}_k . Anhand der SVD läßt sich diese Abhängigkeit ebenfalls sehr direkt ausdrücken:

$$Z = Q\Lambda^{1/2}U' = QA', \quad A = U\Lambda^{1/2} \quad (191)$$

bzw.

$$z_{ij} = q_{i1}a_{1j} + q_{i2}a_{2j} + \dots + q_{in}a_{nj} \quad (192)$$

wird zunächst so interpretiert, dass sich der (standardisierte) Messwert z_{ij} der i -ten Person im j -ten Test als Skalarprodukt personspezifischer "Scores" q_{ik} und testspezifischer "Ladungen" a_{jk} , $k = 1, \dots, r \leq n$ anschreiben läßt. Aus

(191) folgt $A'A = \Lambda$ und nach Multiplikation von links mit Q bzw. von rechts mit $A\Lambda^{-1}$

$$A = Z'Q \quad (193)$$

$$Q = ZAA^{-1}. \quad (194)$$

Schreibt man die Gleichungen für Z , A und Q für individuelle Elemente aus, so erhält man

$$z_{ij} = q_{i1}a_{1j} + q_{i2}a_{2j} + \dots + q_{in}a_{nj} \quad (195)$$

$$\begin{aligned} q_{ik} &= z_{i1}a_{1k}/\lambda_k + z_{i2}a_{2k}/\lambda_k + \dots + z_{in}a_{nk}/\lambda_k \\ &= z_{i1}u_{1k}/\sqrt{\lambda_k} + z_{i2}u_{2k}/\sqrt{\lambda_k} + \dots + z_{in}u_{nk}/\sqrt{\lambda_k} \end{aligned} \quad (196)$$

$$a_{jk} = z_{1j}q_{1k} + \dots + z_{mj}q_{mk}. \quad (197)$$

Die letzten beiden Gleichungen, also (196) und (197) sind hier von besonderem Interesse (die erste Gleichung (195) ist ja der Ausgangsbetrachtungen). (196) zeigt, dass der Faktorwert q_{ik} der i -ten Person auf der k -ten latenten Dimension sich als gewogene Summe der Messwerte der i -ten Person in den n "Tests" darstellen läßt, wobei die "Gewichte" durch die Ladungen a_{jk} der Tests $j = 1, \dots, n$, wiederum gewichtet mit dem Reziprokwert des k -ten Eigenwerts, als dem Varianzanteil der k -ten Dimension, gegeben sind. In der klassischen Testtheorie kann man diese Beziehung benutzen, um den Score einer Person auf einer latenten Dimension oder latenten Variablen zu berechnen. Die Gleichung (197) wiederum drückt die Ladungen als gewogene Summe der Scores der Probanden im j -ten Test aus; Gewichte sind nun die q_{ik} , also die Faktorscores der Personen auf der k -ten latenten Dimension. a_{jk} ist eine Art gewogener Mittelwert der Scores der Probanden in Bezug auf die k -te latente Dimension, während q_{ik} eine Art gewogenes Mittel der Antworten einer Person in Bezug auf die latente Variable ist.

Man kann die Beziehung zwischen den z_{ij} -Werten, den Faktorscores und den Ladungen für die k -te latente Variable in einer Tabelle aufzeigen:

	Messwerte						Σ	
	z_{11}	z_{12}	\dots	z_{1j}	\dots	z_{1n}	q_{1k}	
	z_{21}	z_{22}	\dots	z_{2j}	\dots	z_{2n}	q_{2k}	
	z_{31}	z_{32}	\dots	z_{3j}	\dots	z_{3n}	q_{3k}	
				\vdots			\vdots	.
	z_{i1}	z_{i2}	\dots	z_{ij}	\dots	z_{in}	q_{ik}	
				\vdots			\vdots	
	z_{m1}	z_{m2}	\dots	z_{mj}	\dots	z_{mn}	q_{mk}	
Σ	a_{1k}	a_{2k}	\dots	a_{jk}	\dots	a_{nk}		

Die q_{ik} sind Skalenwerte für die i -te Zeilenkategorie (hier also der i -ten Person), und die a_{jk} entsprechen Skalenwerten für die j -te Spaltenkategorie (hier

dem j -ten "Test"). Nun sind Personen mit Merkmalen ausgestattet unabhängig von Items, mit denen man diese Merkmale erfassen kann, so dass man die a_{jk} und q_{ik} für voneinander unabhängige Größen halten könnte. Andererseits handelt es sich bei den a_{jk} und q_{ik} um Schätzungen der Merkmalsausprägungen, und die Beziehungen (196) und (197) zeigen, dass sie nicht unabhängig voneinander sind. Diese Größen sind Lösungen für ein Skalierungsproblem, die in einem etwas anderem Zusammenhang als *Dual Scaling* bekannt sind (die Korrespondenzanalyse ist ebenfalls ein Spezialfall des Dual Scaling-Prinzips, vgl. "Einführung in die Korrespondenzanalyse"¹¹).

3.1.4 Faktorladungen als Korrelationen

Eine eher inhaltliche Deutung der Faktorladungen ergibt sich aus der Tatsache, dass sie als Korrelationen zwischen der j -ten Variablen und der k -ten latenten Variablen gedeutet werden können. Denn multipliziert man die Gleichung $Z = QA'$ von links mit Q' , so ergibt sich $Q'Z = A'$, denn es ist ja $Q'Q = I$, und also $A = (Q'Z)' = Z'Q$. Speziell für die Ladung a_{jk} der j -ten Variablen auf der k -ten latenten Variablen erhält man

$$a_{jk} = u_{jk}\sqrt{\lambda_k} = z_{1j}q_{1k} + z_{2j}q_{2k} + \dots + z_{mj}q_{mk} \quad (199)$$

$$= \|\vec{z}_j\| \|\vec{q}_k\| \cos \phi_{jk} = \rho_{jk}, \quad (200)$$

wobei ϕ_{jk} der Winkel zwischen dem Vektor \vec{z}_j und dem Vektor \vec{q}_k ist. Demnach ist a_{jk}

- nach (199) die gewogene Summe (das gewogene Mittel) der z -Werte für die j -te Variable über die Personen, wobei die q_{ik} die Gewichte sind, und
- gleichzeitig die Korrelation ρ_{jk} zwischen den standardisierten Messwerten z_{ij} und den Ausprägungen der Personen auf den latenten Dimensionen, denn auch die q_{ik} sind ja standardisierte Variable.

Sind die z_{ij} durch $z_{ij} = (x_{ij} - \bar{x}_j)/s_j\sqrt{m}$ definiert, folgt aus dem Sachverhalt, dass a_{jk} eine Korrelation ist, sofort

$$-1 \leq a_{jk} \leq 1, \quad \text{wenn} \quad z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j\sqrt{m}}. \quad (201)$$

Es gilt

1. Aus (200) folgt, dass – vorausgesetzt $\vec{z}_j \neq \vec{0}$ und $\vec{q}_k \neq \vec{0}$ – $a_{jk} = 0$ genau dann ist, wenn \vec{z}_j und \vec{q}_k orthogonal sind, wenn also $\phi_{jk} = \pi/2$, d.h. wenn der Winkel zwischen den Vektoren \vec{z}_j und \vec{q}_k gleich 90° ist. Die Messwerte in der j -ten Variablen sind dann unabhängig von den Ausstattungen der Personen mit der k -ten latenten Dimension, d.h. diese Dimension geht in die Messwerte für die j -te Variable nicht in systematischer Weise ein.

¹¹<http://www.uwe-mortensen.de/skripten.html>

2. Generell gilt $-1 \leq \cos \phi_{jk} \leq 1$. Nach (200) ist $|a_{jk}| = \max$ dann, wenn $\cos \phi_{jk} = 1$, wenn also \vec{z}_j und \vec{q}_k die gleiche Orientierung haben; dann ist $\phi_{jk} = 0$. Da $\|\vec{q}_k\| = 1$ und $\|\vec{z}_j\| = 1$ (wenn $z_{ij} = (x_{ij} - \bar{x}_j)/s_j\sqrt{m}$) folgt $\max[a_{jk}] = 1$. Weiter folgt sofort, dass der Minimalwert von $a_{jk} = -1$ ist, wenn $\cos \phi_{jk} = -1$ ist, wenn also $\phi_{jk} = 3\pi/2$ ist. Bei geeigneter Definition der z_{ij} tritt dieser Fall also ein, wenn $z_{ij} = \pm q_{ik}$ (vergl. (67), p. 25), wenn also die Messwerte vollständig durch die Ausprägungen der Personen auf der k -ten latenten Dimension bestimmt sind, wenn also die j -te Variable ausschließlich die k -te latente Variable erfasst. Da aber die q_{ik} der Annahme nach Konstante sind, die z_{ij} aber standardisierte Werte fehlerbehafteter Messungen x_{ij} sind, wird die Beziehung $z_{ij} = q_{ik}$ auch dann nicht in strenger Form gelten, wenn die j -te Variable tatsächlich nur die k -te Dimension erfasst; dann gilt $z_{ij} = \pm q_{ik} + \varepsilon_{ij}$, ε_{ij} ein "Fehler", der durch die in den Messungen x_{ij} enthaltenen Messfehler erzeugt wird, so dass $z_{ij} = \pm q_{ik}$ in strenger Form selten erfüllt sein wird.

Im allgemeinen wird aber die j -te Variable mehr als nur eine latente Dimension erfassen, so dass es kaum vorkommen wird, dass \vec{z}_j und \vec{q}_k identische oder entgegengesetzte Orientierungen haben.

3.1.5 Die Approximation von Z

Die Matrix $(m \times n)$ -Matrix Z wird vollständig dargestellt durch die SVD

$$Z = Q\Lambda^{1/2}P',$$

wenn Q eine $(m \times n)$ -Matrix und P eine $(n \times n)$ -Matrix ist. Λ ist dann ebenfalls eine $(n \times n)$ -Matrix, d.h. alle Eigenwerte sind größer als Null. Dies ist der allgm eine Fall, auch wenn es tatsächlich nur $r < \min(m, n)$ latente Variable gibt, denn der unvermeidliche Messfehler impliziert, dass rechnerisch n "latente" Variablen benötigt werden, um Z exakt vorherzusagen. Aber das Ziel ist ja üblicherweise, nur die wichtigen ersten r latenten Variablen zu berücksichtigen. Man berechnet also

$$\hat{Z}_r = Q_r \Lambda_r^{1/2} P_r', \quad r < \min(m, n)$$

wobei \hat{Z}_r eine Approximation von Z ist; Λ_r ist nun eine $(r \times r)$ -Matrix, weil nur die größten r Eigenwerte und damit nur die zu diesen Eigenwerten korrespondierenden Eigenvektoren in Q_r und P_r berücksichtigt werden. Es kann nun gezeigt werden, dass \hat{Z} eine Approximation von Z im Sinne der Methode der Kleinsten Quadrate ist.

Dazu sei \hat{z}_{ij} das Element in der i -ten Zeile und j -ten Spalte von \hat{Z}_r ; \hat{z}_{ij} ist eine Approximation für das entsprechende Element z_{ij} in Z . Es ist $Z - \hat{Z}_r$ eine Matrix, deren Elemente durch $z_{ij} - \hat{z}_{ij}$ gegeben sind. Dann ist

$$\phi(r) = \|Z - \hat{Z}_r\|^2 = \sum_{i=1}^m \sum_{j=1}^n (z_{ij} - \hat{z}_{ij})^2. \quad (202)$$

Es kann gezeigt werden, dass für einen vorgegebenen Wert von r die Funktion $\phi(r)$ minimal wird, wenn $\hat{Z}_r = Q_r \Lambda_r^{1/2} P_r'$ gegeben ist. Der Beweis ist ein wenig länglich; er wurde von Gabriel (1978) geliefert.

3.1.6 Die Beziehung zur Hauptachsentransformation

Die Transformation der Vektoren von Z gemäß $ZU = V$, wobei V eine Matrix mit orthogonalen Spaltenvektoren ist, stellt eine Hauptachsentransformation dar. Wie in den Gleichungen (134) bis (136) gezeigt wurde, transformiert die Matrix U aber gerade das Ellipsoid, das durch $Z'Z$ definiert wird, in ein achsenparalleles Ellipsoid, d.h. die Transformationsmatrix U definiert eine Hauptachsentransformation. Die Punktwolke der Personen mit den Koordinaten x_{i1}, \dots, x_{in} repräsentiert dieses Ellipsoid. Die Punkte mit den Koordinaten v_{i1}, \dots, v_{in} bilden das gleiche Ellipsoid in rotierten Achsen, so dass es nun achsenparallel erscheint. Die v_{1k}, \dots, v_{mk} sind die Koordinaten der m Personen auf der k -ten Achse. \bar{v}_k ist der Mittelwert dieser Koordinaten, und es ist $\bar{v}_k = 0$. Wie bereits gezeigt wurde ist $\sum_i v_{ik}^2 = s_k^2$ die Varianz dieser Koordinaten. Andererseits ist $\sum_i v_{ik}^2 = \|\vec{V}_k\|^2 = \lambda_k$ das Quadrat der Länge von \vec{F}_k . Damit ist gezeigt, dass

$$\lambda_k = s_k^2, \quad (203)$$

d.h. die Eigenwerte sind gleich der Varianzen der Koordinaten der Personen auf der k -ten latenten Dimension. Man vergleiche dieses Ergebnis mit (161), - diese Gleichung wurde für die multivariate Normalverteilung hergeleitet.

Die Zerlegung (174) liefert die neuen latenten Variablen $V = Q\Lambda^{1/2}$. Die Eigenvektoren $\vec{q}_1, \vec{q}_2, \dots, \vec{q}_n$ korrespondieren zu den zugehörigen Eigenwerten $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Nach (203) bedeutet dies, dass die Eigenvektoren nach der Größe der Varianzen s_k^2 , $k = 1, \dots, n$ der Koordinaten auf der jeweiligen Achse angeordnet sind.

3.1.7 Eine Anwendung auf die Regressionsrechnung

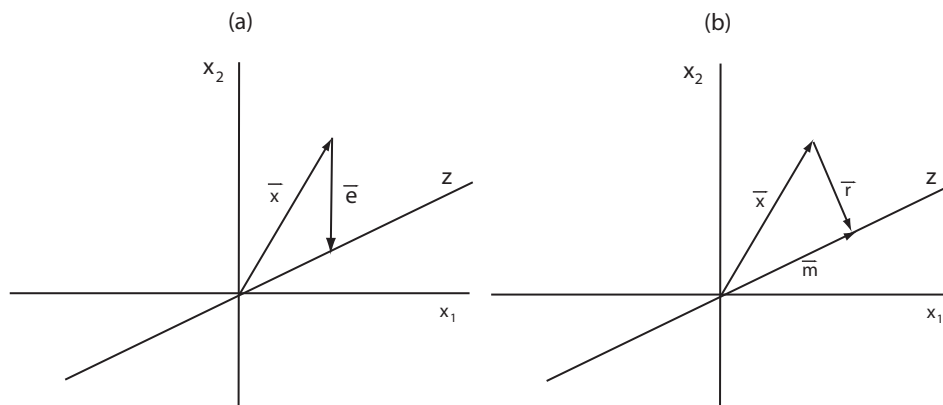
Die Gleichung (30) auf Seite 16 definiert die klassische multiple Regression; eine Variable Y wird anhand von Prädiktoren X_1, \dots, X_s "vorhergesagt". Dazu werden die Regressionsgewichte b_0, b_1, \dots, b_s so bestimmt, dass die Summe der "Abweichungsquadrate" $\sum_i (y_i - \hat{y}_i)^2 = \sum_i e_i^2$ ein Minimum wird, wobei

$$\hat{y}_i = b_0 + b_1 x_{i1} + \dots + b_s x_{is},$$

und \hat{y}_i ist die Vorhersage für den Messwert y_i von Y , und x_{i1}, \dots, x_{is} sind die i -ten Messungen für die X_1, \dots, X_s . Es ist $e_i = y_i - \hat{y}_i$, und

$$\sum_i (y_i - \hat{y}_i)^2 = \sum_i e_i^2 = Q(b_0, b_1, \dots, b_s). \quad (204)$$

Abbildung 5: Zur Minimalisierung des Fehlers \vec{e} in der Multiplen Regression (a), und des Lotes \vec{r} , (b)



e_i entspricht dem Vektor \vec{e} in Abbildung 5 (a); die Gerade z soll so bestimmt werden, dass $Q(b_0, b_1, \dots, b_s)$ minimal wird. Statt die Vektoren \vec{e}_i zu minimalisieren, kann man auch die Lote \vec{r}_i (vergl. Abb. 5, (b)) minimalisieren. Aus dem Satz des Pythagoras folgt sofort

$$|\vec{r}_i|^2 + |\vec{m}_i|^2 = |\vec{z}_i|^2, \quad i = 1, \dots, n \quad (205)$$

also

$$\sum_{i=1}^n |\vec{r}_i|^2 = \sum_{i=1}^n |\vec{z}_i|^2 - \sum_{i=1}^n |\vec{m}_i|^2. \quad (206)$$

Das Skalarprodukt $\vec{x}_i' \vec{x}_i = |\vec{x}_i|^2$ liegt fest (= Daten); die Minimalisierung von $\sum_{i=1}^n |\vec{r}_i|^2$ geschieht also durch Maximierung von $\sum_{i=1}^n |\vec{m}_i|^2$. Man sieht aber leicht, dass eine direkte Anwendung der Methode der Kleinsten Quadrate wie in (204) nicht möglich ist. Nun kann die Gerade z aber als Resultat einer Rotation des Koordinatensystems (x_1, x_2) gesehen werden, wobei die Achse x_1 gerade in z übergeht (und die neue Achse für x_2 senkrecht auf z steht). Die Vektoren \vec{m}_i sind dann die Koordinaten der Endpunkte der \vec{x}_i auf der neuen Achse z . Es sei nun X eine $(n \times 2)$ -Matrix der Komponenten der Vektoren \vec{x}_i , d.h. die i -te Zeile von X enthalte gerade die Komponenten x_{i1} und x_{i2} von \vec{x}_i . Die Matrixgleichung

$$XP = L \quad (207)$$

wobei P eine orthonormale Matrix ist, definiert nun eine Rotation des ursprünglichen Koordinatensystems, oder, damit gleichwertig, der Vektoren \vec{x}_i um einen bestimmten Winkel ϕ , und die Zeilen von L enthalten die neuen Komponenten dieser Vektoren, etwa (F_{i1}, F_{i2}) . Offenbar ist dann $F_{i1} = \vec{m}_i$ und $F_{i2} = \vec{r}_i$. Die Maximierung von $\sum_i |\vec{m}_i|^2$ bedeutet dann die Maximierung der Varianz $\sum_i F_{i1}^2$ der F_{i1} -Werte. Nach Satz 6.1, Seite 125, wird diese Varianz aber maximiert, wenn P die Matrix der Eigenvektoren von $X'X$ ist, wobei die

Messungen in X als zentriert vorausgesetzt werden (andernfalls ist der Mittelwert der F_{i1} -Werte nicht gleich Null und $\sum_i F_{i1}^2$ ist keine Varianz). Weiter ist dann

$$\sum_i |\vec{m}_i|^2 = \sum_i F_{i1}^2 = \lambda_1, \quad (208)$$

und λ_1 ist der größte Eigenwert von P .

3.2 Faktorenanalyse

3.2.1 Vorbemerkung: zufällige Vektoren

Der Vektor $\vec{X} = (X_1, \dots, X_n)'$ heißt *zufälliger Vektor* oder auch einfach *Zufallsvektor*, wenn die Komponenten zufällige Variablen sind. Ist $E(X_j) = \mu_j$ der Erwartungswert von X_j , so ist

$$E(\vec{X}) = (E(X_1), \dots, E(X_n))' = (\mu_1, \dots, \mu_n)' \quad (209)$$

Mit $x_j = X_j - \mu_j$ ist

$$\vec{x} = (x_1, \dots, x_n)' \quad (210)$$

der Vektor der Abweichungen vom jeweiligen Erwartungswert.

Das Skalarprodukt zweier Vektoren ist als Summe der Produkte der korrespondierenden Elemente erklärt worden, $\vec{x}'\vec{y} = \sum_i x_i y_i$. Ein andere Art von Vektorprodukt ergibt sich, wenn man

$$\vec{x}\vec{y}' = \begin{pmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ & & \vdots & \\ x_n y_1 & x_n y_2 & \cdots & x_n y_n \end{pmatrix} \quad (211)$$

eingführt. Dann ist

$$\text{Kov}(\vec{X}) = E(\vec{x}\vec{x}') = \begin{pmatrix} E(x_1^2) & E(x_1 x_2) & \cdots & E(x_1 x_n) \\ E(x_2 x_1) & E(x_2^2) & \cdots & E(x_2 x_n) \\ & & \vdots & \\ E(x_n x_1) & E(x_n x_2) & \cdots & E(x_n^2) \end{pmatrix}. \quad (212)$$

die Matrix der Varianzen $E(x_j^2)$ und Kovarianzen $E(x_i x_j)$ der Komponenten von \vec{X} .

3.2.2 Das Modell der Faktorenanalyse

Auf Seite 10 wurde die Gleichung (21), also

$$x_{ij} = \alpha_{j1} F_{i1} + \alpha_{j2} F_{i2} + \cdots + \alpha_{js} F_{is} + e_{ij}, \quad j = 1, \dots, n \quad (213)$$

als Modell der Erklärung der x_{ij} anhand latenter Variablen eingeführt, wobei die x_{ij} jetzt die Abweichungen $X_{ij} - \mu_j$ bedeuten sollen. Die Koeffizienten $\alpha_{j1}, \dots, \alpha_{js}$ sind formal Regressionskoeffizienten und selbst keine zufällige Variablen, - allerdings sind ihre Schätzungen zufällige Variablen. Die F_{i1}, \dots, F_{is} dagegen sind zufällige Variablen insofern, als die Personen "zufällig" aus einer Population ausgewählt werden. Fasst man sie zu Vektoren

$$\vec{F}_1 = (F_{11}, \dots, F_{m1})', \quad \vec{F}_2 = (F_{12}, \dots, F_{m2})', \quad \text{etc}$$

zusammen, so sind die $\vec{F}_1, \dots, \vec{F}_s$ zufällige Vektoren.

Es werden die folgenden Annahmen gemacht:

$$E(\vec{F}) = \vec{0} \quad (214)$$

$$E(\vec{e}) = \vec{0} \quad (215)$$

$$\text{Kov}(\vec{F}) = E(\vec{F}\vec{F}') \begin{cases} = I, & \text{orthogonales Modell} \\ \neq I, & \text{obliques Modell} \end{cases} \quad (216)$$

$$\text{Kov}(\vec{e}) = E(\vec{e}\vec{e}') = \Psi = \text{diag}(\psi_1^2, \dots, \psi_n^2) \quad (217)$$

$$\text{Kov}(\vec{e}, \vec{F}) = E(\vec{e}\vec{F}') = 0, \quad (218)$$

wobei I wieder die Einheitsmatrix ist. (217) bedeutet dann die Forderung, dass die Komponenten von \vec{e} , also die Fehler (spezifischen Faktoren) in den einzelnen X_j unkorreliert sein sollen, und (218) bedeutet, dass die latenten Variablen und die Fehler jeweils unkorreliert sein sollen (vergl. dazu (14) auf Seite 9). Die Diagonalelemente von $E(\vec{e}\vec{e}')$ sind allerdings nicht gleich Null, vielmehr gilt

$$E(\vec{e}\vec{e}') = \Psi = \begin{pmatrix} \psi_1^2 & 0 & \cdots & 0 \\ 0 & \psi_2^2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \psi_n^2 \end{pmatrix} \quad (219)$$

Die Varianzen ψ_k^2 werden auch *Restvarianzen* genannt.

Die α_{jk} können in einer Matrix \tilde{A} zusammengefasst werden. Für die Komponenten von $\vec{x} = \vec{X} - \mu$ gilt dann

$$\vec{x} = \vec{X} - \mu = \tilde{A}\vec{F} + \vec{e}, \quad (220)$$

und für die Varianz-Kovarianz-Struktur von \vec{X} ergibt sich

$$\begin{aligned} (\vec{X} - \vec{\mu})(\vec{X} - \vec{\mu})' &= (\tilde{A}\vec{F} + \vec{e})(\tilde{A}\vec{F} + \vec{e})' \\ &= (\tilde{A}\vec{F} + \vec{e})(\tilde{A}\vec{F})' + \vec{e}\vec{e}' \\ &= \tilde{A}\vec{F}\vec{F}'\tilde{A}' + \vec{e}\tilde{A}'\vec{F}' + \tilde{A}\vec{F}\vec{e}' + \vec{e}\vec{e}'. \end{aligned} \quad (221)$$

Für die Varianz-Kovarianz-Matrix von \vec{X} ergibt sich dann, wegen $\vec{x} = \vec{X} - \vec{\mu}$,

$$\begin{aligned} \text{Kov}(\vec{x}) = \Sigma &= E((\vec{X} - \vec{\mu})(\vec{X} - \vec{\mu})') \\ &= \tilde{A}E(\vec{F}\vec{F}')\tilde{A}' + E(\vec{e}\vec{e}')\tilde{A}' + \tilde{A}E(\vec{F}\vec{e}') + E(\vec{e}\vec{e}') \end{aligned} \quad (222)$$

Nimmt man an, dass die Komponenten von \vec{F} unkorreliert sind, so gilt nach (216) $E(\vec{F}\vec{F}') = I$, so dass $\tilde{A}E(\vec{F}\vec{F}')\tilde{A}' = \tilde{A}\tilde{A}'$, und nach (214), (215) und (218) folgt $E(\vec{e}\vec{L}')\tilde{A}' = \tilde{A}E(\vec{F}\vec{e}') = 0$ und $E(\vec{e}\vec{e}') = \Psi$. Demnach gilt das

Fundamentaltheorem der Faktorenanalyse:

$$\Sigma = \tilde{A}\tilde{A}' + \Psi. \quad (223)$$

(Thurstone, 1935). Für die Kovarianz c_{jk} zwischen der j -ten und der k -ten Variablen V_j und V_k findet man demnach

$$c_{jk} = \sum_{s=1}^t a_{js}a_{ks} + \begin{cases} \psi_j^2 & j = k \\ 0 & j \neq k \end{cases}, \quad t \leq n \quad (224)$$

Die folgende Beziehung erleichtert die Interpretation der Parameter in \tilde{A} . Aus (220) folgt

$$\begin{aligned} \text{Kov}(\vec{X}, \vec{F}) &= E[(\vec{X} - \vec{\mu})L'] \\ &= E[(\tilde{A}\vec{F} + \vec{e})L'] \\ &= E[\tilde{A}\vec{F}\vec{F}' + \vec{e}\vec{F}'] \\ &= \tilde{A}E[\vec{F}\vec{F}'] + E[\vec{e}\vec{F}'], \\ &= \tilde{A}. \end{aligned} \quad (225)$$

Die Gleichung (225) besagt, dass die Ladungen in \tilde{A} als Kovarianzen zwischen den gemessenen Variablen V_j und den latenten Faktoren interpretiert werden können.

Die Diagonalelemente der Matrix $\text{Kov}(\vec{x})$ enthalten die Varianzen der X_j . Nach (223) und (224) muß dann für die j -te Diagonalelemente

$$\sigma_j^2 = c_{jj} = a_{j1}^2 + \cdots + a_{jt}^2 + \psi_j^2 \quad (226)$$

gelten.

Definition 3.2 *Die Summe*

$$h_j^2 = a_{j1}^2 + \cdots + a_{jt}^2 \quad (227)$$

heißt Kommunalität, und ψ_j^2 ist die Rest- oder spezifische Varianz.

Die Kommunalität h_j^2 ist der Teil der Varianz σ_j^2 , der durch die *auf alle* gemessenen latenten Variablen erklärt wird. Sie wird offenbar durch die Quadrate der Ladungen der j -ten Variablen auf den r latenten Variablen, den Faktoren, definiert.

Faktorenmuster und -strukturen Geht man von standardisierten Variablen aus, so kann man die Werte der Variablen V_j durch die Vektoren \vec{z}_j repräsentieren; man hat die Gleichungen

$$\vec{z}_j = a_{j1}\vec{Q}_1 + a_{j2}\vec{Q}_2 + \cdots + a_{jt}\vec{Q}_t + d_j\vec{U}_j, \quad j = 1, 2, \dots, n \quad (228)$$

wobei $\vec{U}_j = \vec{Q}_j^s + \vec{\varepsilon}_j$ ist, \vec{Q}_j^s ein für die j -te Variable spezifischer Faktor, und $\vec{\varepsilon}_j$ ein Vektor, der Messfehler repräsentiert. Die Gleichungen (228) heißen auch *Faktorenmuster* (*factor pattern*).

Man kann darüber hinaus auch die Korrelationen zwischen den Variablen und den latenten Dimensionen (den "Faktoren") betrachten; diese Korrelationen definieren die *Faktorenstruktur* (*factor structure*). Bildet man das Skalarprodukt von \vec{Q}_k mit \vec{z}_j , so erhält man aus (228) die Gleichungen

$$\langle \vec{Q}_k, \vec{z}_j \rangle = \vec{Q}'_k \vec{z}_j = a_{j1} \vec{Q}'_k \vec{Q}_1 + \dots + a_{jt} \vec{Q}'_k \vec{Q}_t + d_j \vec{Q}'_k \vec{U}_j, \quad j, k = 1, \dots, t \quad (229)$$

die also die Faktorenstruktur definieren. Sind nun die $\vec{Q}_1, \dots, \vec{Q}_t, \vec{U}_j$ paarweise orthogonal, so verschwinden die Skalarprodukte für verschiedene latente Variable und man erhält als Spezialfall einer Faktorenstruktur wieder die Beziehungen

$$\langle \vec{Q}_k, \vec{z}_j \rangle = \vec{Q}'_k \vec{z}_j = a_{jk}, \quad j = 1, \dots, n; \quad k = 1, \dots, t, \quad (230)$$

d.h. die Ladungen sind gerade die Korrelationen zwischen den gemessenen Variablen und den latenten Dimensionen.

3.2.3 Die Hauptkomponentenanalyse

3.2.4 Hauptkomponenten versus Faktoren

Diese Analyse geht auf Pearson (1901) und Hotelling (1933) zurück. Es sollen dabei die beobachteten Variablen linear durch einige wenige sogenannte Hauptkomponenten erklärt werden. Die Hauptkomponenten sollen unkorreliert sein und nach fallendem Anteil der Gesamtvarianz der gemessenen Variablen, den sie jeweils erklären, geordnet sein. Im Unterschied zur Faktorenanalyse, bei der die Kovarianzen erklärt werden sollen, zielt die Hauptkomponentenanalyse auf eine Erklärung der Varianz. Die Hauptkomponentenanalyse besteht im wesentlichen in einer Anwendung der Hauptachsentransformation.

Die Hauptachsentransformation: Es ist sinnvoll, nicht die $X_j - \mu_j$, sondern die standardisierten Variablen $(X_j - \mu_j)/\sigma_j$ zu betrachten, da sich dann die üblicherweise verschiedenen Maßeinheiten der X_j nicht störend auswirken. Die Korrelation zwischen zwei Variablen V_j und V_k durch

$$r_{jk} = \frac{1}{m} \sum_{i=1}^m z_{ij} z_{ik}$$

gegeben. Um den Faktor $1/m$ nicht immer explizit mit angeben zu müssen, ist es sinnvoll, die z -Werte etwas anders als üblich zu definieren, nämlich gemäß

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j \sqrt{m}}, \quad \text{bzw.} \quad z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j \sqrt{m-1}}; \quad (231)$$

man hat dann

$$r_{jk} = \sum_{i=1}^m z_{ij} z_{ik} = \vec{Z}_j' \vec{Z}_k, \quad (232)$$

wobei $\vec{Z}_j = (z_{1j}, \dots, z_{mj})'$ und $\vec{Z}_k = (z_{1k}, \dots, z_{mk})'$.

Bekanntlich gilt für eine beliebige Matrix Z stets die Singularwertzerlegung

$$Z = Q\Lambda^{1/2}P', \quad (233)$$

(vergl. (174), Seite 63), wobei Q eine Matrix ist, deren Spalten die Eigenvektoren von ZZ' sind, und P ist eine Matrix, deren Spaltenvektoren die Eigenvektoren von $R = Z'Z$ sind. $\Lambda^{1/2}$ ist eine Diagonalmatrix, deren Diagonalelemente die Wurzeln aus den von Null verschiedenen Eigenwerten von ZZ' bzw. $Z'Z$ sind; die von Null verschiedenen Eigenwerte dieser beiden Matrizen sind identisch. Im allgemeinen findet man, dass die Anzahl der von Null verschiedenen Eigenwerte gleich $\min(m, n)$ ist. Im letzten Teil von Abschnitt 3.1.1 wurde gezeigt, dass die Matrix P eine Hauptachsentransformation der Spaltenvektoren von Z bewirkt. Ebenso kann man sagen, dass $ZP = L = Q\Lambda^{1/2}$ die Transformation der Zeilenvektoren von Z in die Zeilenvektoren von L bedeutet, und umgekehrt bedeutet $Z = Q\Lambda^{1/2}P'$ die Transformation der Vektoren in L in die Vektoren in Z .

Faktorladungen und Faktorwerte: Bei der "klassischen" Herleitung der Hauptfaktorlösung konzentrierte man sich auf die Faktorladungen. Ausgehend von dem Ansatz

$$z_{ij} = q_{i1}a_{j1} + \dots + q_{in}a_{jn}$$

sollen zunächst die Ladungen $a_{11}, a_{21}, \dots, a_{n1}$ der Variablen ("Tests") auf dem ersten Faktor gefunden werden. Die Idee ist, die a_{j1} so zu bestimmen, dass die Summe

$$T_1 = a_{11}^2 + a_{21}^2 + \dots + a_{n1}^2 \quad (234)$$

maximal wird. Ohne weitere Nebenbedingung ist aber das Maximum $T_1 = \infty$, - und diese Lösung nicht interessant, denn da n endlich ist, bedeutet $T_1 = \infty$, dass mindestens eine der Ladungen a_{j1} unendlich wird; außerdem kann man dann auch keine Rangordnung verschiedener Faktoren mehr herstellen. Eine geeignete Nebenbedingung ergibt sich aber, wenn man berücksichtigt, dass ja

$$r_{jk} = \begin{cases} a_{j1}a_{k1} + \dots + a_{jn}a_{kn}, & j \neq k \\ h_j^2, & j = k \end{cases}$$

gelten soll, wobei $r_{jj} = h_j^2$ die Kommunalität der j -ten Variablen ist. Über diese Beziehung werden die unbekanntenen Ladungen an die empirisch vorgegebenen r_{jk} gekoppelt. Man findet auf diese Weise, dass die Ladungen durch

$$a_{j1} = p_{j1}\sqrt{\lambda_1}$$

gegeben ist, wobei p_{j1} das j -te Element des 1-ten Eigenvektors \vec{p}_1 der Korrelationsmatrix R ist, und λ_1 ist der zugehörige Eigenwert, - vergl. etwa Harman (1967), p. 138. Die Ladungen für die möglicherweise existierende zweite latente Dimension werden auf analoge Weise aus den Residuen ${}_1r_{jk}$ der Korrelationen gewonnen, die sich aus der Beziehung $r_{jk} = a_{j1}a_{k1} + \dots + a_{jn}a_{kn}$ gemäß

$${}_1t_{jk} = r_{jk} - a_{j1}a_{k1} = a_{j2}a_{k2} + \dots + a_{jn}a_{kn}$$

ergeben. Das Ziel ist, über die a_{jk} die Korrelationen r_{jk} zu erklären; die Faktorwerte $q_{i\nu}$ werden gar nicht weiter betrachtet; dass eine Analyse von Personentypen gleichzeitig möglich ist, wird bei diesem Ansatz nicht deutlich.

Die Hauptkomponentenmethode: Da i.a. die Anzahl der von Null verschiedenen Eigenwerte gleich n ist (wegen zufälliger Fehler und numerischen Ungenauigkeiten bei der Berechnung der r_{jk}), enthält A auch n Eigenvektoren, so dass A eine $(n \times n)$ -Matrix ist. Die Hauptkomponentenmethode besteht darin, die Vektoren in Z durch weniger als n "latente" Vektoren zu erklären, etwa durch nur $s < n$. Es seien Q_s und A_s die aus Q und A gewonnenen Teilmatrizen, in denen nur die zu den ersten s Eigenwerten - die als der Größe nach geordnet angenommen werden - korrespondierenden Eigenvektoren enthalten sind. Dann ist

$$\hat{Z} = Q_s A_s' \quad (235)$$

eine Approximation von Z . Diese Approximation wird um so besser sein, je mehr Varianz durch die ersten s Vektoren in Q bzw. A erklärt wird. Exakt wird dann

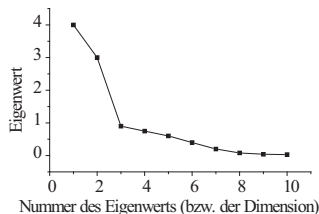
$$Z = Q_s A_s' + E \quad (236)$$

gelten, wobei E jetzt sozusagen den "fehlenden Rest" bezeichnet.

Der Scree-Test für die Anzahl zur berücksichtigenden Dimensionen:

Die Anteile der Gesamtvarianz der Variablen, die durch die einzelnen latenten Dimensionen "erklärt" werden, sind bereits in Abschnitt ?? hergeleitet worden. Um die Anzahl der latenten Dimensionen abzuschätzen, kann man von (??) ausgehen. Dann ist $\phi_s \times 100$ der Prozentsatz der Gesamtvarianz, der durch die ersten s Dimensionen erklärt wird. Eine Möglichkeit, die Anzahl der Dimensionen abzuschätzen, ist, den Wert von ϕ_s festzulegen, und dann s so zu bestimmen, dass dieser Wert gerade erreicht wird, etwa $\phi_s \times 100 = 80\%$. Ein solches Vorgehen ist aber ein wenig willkürlich. Eine andere Möglichkeit besteht darin, einen *Scree-Plot* anzufertigen. Dabei werden die λ_k -Werte gegen die k -Werte aufgetragen, vergl. Abb. 6. Man kann den Wert von s nach Maßgabe einer möglichen abrupten Änderung im Graphen wählen. In Abb. 6 findet man eine solche Veränderung nach dem zweiten Eigenwert, denn $\lambda_2 = 3$ und $\lambda_3 = .9$.

Abbildung 6: Scree-Test



Ein übliches Kriterium zur Bestimmung des Wertes von s wurde von Kaiser (1959) vorgeschlagen. Diesem Kriterium zufolge werden alle Dimensionen berücksichtigt, für die der zugehörige Eigenwert ≥ 1 ist. Dieses Kriterium ist allerdings kritisiert worden, vergl. Cliff (1988).

3.2.5 Der Biplot

Es werden also sowohl die Personen wie auch die Variablen jeweils durch Punkte in einem Koordinatensystem dargestellt, dass dieselben latenten Variablen repräsentiert. Allerdings haben die beiden Koordinatensysteme verschiedene Maßeinheiten. Denn zwar enthalten sowohl Q wie P auf die Länge 1 normierte Vektoren, aber die Koordinaten der Variablen für die k -te latente Variable sind mit dem Faktor $\sqrt{\lambda_k}$ multipliziert worden, während die \vec{q}_k noch alle die Länge 1 haben. Eine alternative Gewichtung ergibt sich, wenn man zu den $\vec{F}_k = \vec{q}_k \sqrt{\lambda_k}$ zurückkehrt. Jetzt werden also die Variablen durch die normierten Koordinaten p_{jk} dargestellt. Die folgende Betrachtung wird vereinfacht, wenn man $\tilde{\Lambda} = \Lambda^{1/2}$ setzt; $\tilde{\Lambda}$ ist also eine Diagonalmatrix, in deren Diagonalzellen die Wurzeln $\sqrt{\lambda_j}$ stehen. Dann gilt jedenfalls

$$Z = Q\Lambda^{1/2}P' = Q\tilde{\Lambda}P'. \quad (237)$$

Weiter ist sicherlich

$$\tilde{\Lambda} = \tilde{\Lambda}^\alpha \tilde{\Lambda}^{1-\alpha},$$

und man erhält

$$Z = Q\tilde{\Lambda}^\alpha \tilde{\Lambda}^{1-\alpha} P'. \quad (238)$$

Setzt man

$$L^* = Q\tilde{\Lambda}^\alpha, \quad A^* = \tilde{\Lambda}^{1-\alpha} P', \quad (239)$$

so erhält man mit L^* und A^* skalierte Koordinaten für die Personen einerseits und die "Tests" andererseits. Für α kann man nun viele Werte annehmen, allerdings konzentriert man sich i.a. auf die folgenden Fälle:

1. **Der Fall $\alpha = 0$:** Dies ist der übliche Fall; man hat $L^* = Q$ und $A^* = A = P\Lambda^{1/2}$. Der Fall entspricht der klassischen R -Analyse, die

auf eine Diskussion der Variablen - der "Tests" - zielt und weniger auf die Diskussion des durch Q definierten Personenraumes.

Man spricht auch von diesem Fall als dem CMP-Fall (CMP steht für *column metric preserving*), wobei mit dem Spalten (columns) hier die Spalten von Z gemeint sind: das Ziel ist, die Varianzen und Kovarianzen der Variablen, also der Spalten von Z , zu analysieren.

Da $L^* = Q$ ergibt sich eine weitere Interpretation hinsichtlich der Personen, allgemein der "Beobachtungen" (gemeint sind Gelegenheiten, an denen die Werte der Variablen gemessen werden). Die Euklidische Distanz zwischen zwei Personen entspricht der Mahalanobis-Distanz zwischen den gleichen Personen oder Beobachtungen, wenn man letztere in Koordinaten der ursprünglichen Variablen ausdrückt. Die Mahalanobis-Distanz wird wie die Euklidische Distanz in Abschnitt 3.2.6 definiert und interpretiert.

2. **DerFall $\alpha = 1$:** Die ist der Fall $L^* = L = Q\Lambda^{1/2}$, $A^* = P$. Man fokussiert auf die Struktur des Personenraums und damit auf die klassische Q-Analyse, die auf eine Identifizierung von Personentypen zielt.

Man spricht auch vom RMP-Fall (RMP steht für *row metric preserving*), weil die Struktur der Distanzen zwischen Paaren von Reihen - die hier i.a. Personen repräsentieren, bewahrt bleibt.

In diesem Fall ist die Euklidische Distanz zwischen den Personen, gemessen in L^* -Koordinaten, gleich der Euklidische Distanz zwischen den Personen, gemessen in den x - bzw. z -Werten; vergl. Abschnitt 3.2.6.

3. **DerFall $\alpha = 1/2$:** Hier wird den Zeilen und den Spalten gleiches Gewicht gegeben. Dieser Fall wird insbesondere bei der Analyse von Kontingenztabelle (Korrespondenzanalyse) betrachtet, bei der die Residuen $n_{ij} - \hat{n}_i + \hat{n}_{+j}/N$ nach geeigneter Gewichtung einer SVD unterzogen werden.

Der Fall $\alpha = 1/2$ erweist sich als nützlich für die Interpretation von Wechselwirkungen in 2-Faktor-Experimenten (Gower und Hand, 1996).

Die Berechnung einer Q-Analyse wurde ursprünglich durch explizite Berechnung der Korrelationen zwischen Personen vorgenommen, wobei notwendig über die Tests oder Variablen gemittelt wurde. Es ergab sich eine Diskussion der Frage, ob die Anzahl der Typenfaktoren gleich der Anzahl der Testfaktoren sei oder nicht. Erst die Herleitung der latenten Dimensionen über die SVD zeigt auf einfache Weise, dass zwischen Typen- und Testfaktoren ein enger Zusammenhang besteht¹². Auf die SVD scheinen zuerst Eckart und Young (1936) hingewiesen zu haben, die die Approximation der Datenmatrix Z durch eine Anzahl $s < n$ von latenten Variablen, d.h. den datenreduzierenden Aspekt

¹²Die SVD ist seit Mitte des 19-ten Jahrhunderts bekannt und wurde nicht im Rahmen faktorenanalytischer Betrachtungen hergeleitet.

der Faktorenanalyse diskutiert haben. In diesem Zusammenhang wurde von der SVD auch lange als vom *Eckart-Young-Theorem* gesprochen, was insofern inkorrekt ist, als bei der SVD die Approximation durch weniger als n latente Variablen noch nicht betrachtet wird.

Man kann die Personen und die Variablen also im Prinzip in einem gemeinsamen Koordinatensystem darstellen. Eine solche Darstellung wird nach Gabriel (1971) *Biplot* genannt, wobei diese Bezeichnung von Gabriel für den Fall $s = 2$ reserviert wird. Für den Fall $s \geq 3$ wird der Ausdruck *Bimodell* eingeführt. Die ungleiche Gewichtung von Q und A einerseits oder L und P andererseits muß dabei aber berücksichtigt werden. Die Produktdarstellung von $\tilde{\Lambda}$ geht auf Gabriel (1971) zurück.

3.2.6 Die Mahalanobis-Distanz

Es sei eine spaltenstandardisierte Datenmatrix Z gegeben; die Spalten repräsentieren Variablen ("Tests"), die Zeilen Personen. Eine Zeile der Matrix Q der Faktorwerte der Personen enthält die Koordinaten der entsprechenden Person auf den latenten Dimensionen: q_{i1} ist der Faktorwert, also die Koordinate der i -ten Person auf der ersten latenten Dimension, q_{i2} die Koordinate dieser Person bezüglich der zweiten latenten Dimension, etc. Die Ähnlichkeit der i -ten und der k -ten Person läßt sich durch die Euklidische Distanz zwischen den Punkten, die jeweils eine Person repräsentieren, angeben. Die Euklidische Distanz ist dabei wie folgt definiert:

Definition 3.3 *Es seien x_1, \dots, x_n und y_1, \dots, y_n die Koordinaten zweier Punkte x und y in einem n -dimensionalen Raum. Dann ist die Euklidische Distanz durch*

$$d(x, y) = \left(\sum_{j=1}^n (x_j - y_j)^2 \right)^{1/2} = \sqrt{\sum_{j=1}^n (x_j - y_j)^2} = ((\vec{x} - \vec{y})'(\vec{x} - \vec{y}))^{1/2}. \quad (240)$$

gegeben. Dabei ist $\vec{x} = (x_1, \dots, x_n)'$, $\vec{y} = (y_1, \dots, y_n)'$.

Die Begriff der Euklidischen Distanz kann als Verallgemeinerung des aus der Schule bekannten Satzes des Pythagoras gesehen werden.

Sind also q_{i1}, \dots, q_{in} und q_{k1}, \dots, q_{kn} die Koordinaten der i -ten bzw. k -ten Person im Personenraum, so ist die Euklidische Distanz zwischen ihnen durch

$$d_{ik} = \sqrt{\sum_{j=1}^n (q_{ij} - q_{kj})^2} = ((\vec{q}_i - \vec{q}_k)'(\vec{q}_i - \vec{q}_k))^{1/2}. \quad (241)$$

Mit \vec{q}_i und \vec{q}_k sind demnach die Vektoren mit den Komponenten (= Koordinaten der Endpunkte) q_{ij} bzw. q_{kj} , $j = 1, \dots, n$ gemeint.

Die Euklidische Distanz ist die bekannte "kürzeste Verbindung zwischen zwei Punkten", - die nach Euklid eben durch eine Gerade definiert ist. Die explizite Benennung dieser kürzesten Verbindung als "euklidisch" legt aber nahe, dass sich Distanzen auch anders definieren lassen, - generell eben als "nicht-euklidisch". So ist in einer Stadt die kürzeste Verbindung nicht notwendig durch eine Gerade gegeben, sondern durch bestimmte Straßen, die einen gekrümmten Verlauf nehmen können, und gelegentlich wird man rechts oder links abbiegen müssen. Im Rahmen der multidimensionalen Skalierung wird man dementsprechend einen verallgemeinerten Distanzbegriff einführen, der den euklidischen als Spezialfall enthält. Hier soll ein Distanzbegriff definiert werden, der auf den indischen Mathematiker Mahalanobis¹³ zurückgeht:

Definition 3.4 *Es seien \vec{x} und \vec{y} zwei n -dimensionale Vektoren, und es sei S eine positiv-definite, symmetrische Matrix. Dann heißt*

$$\delta_{xy} = [(\vec{x} - \vec{y})' S^{-1} (\vec{x} - \vec{y})]^{1/2} \quad (242)$$

die Mahalanobis-Distanz zwischen den Endpunkten der Vektoren \vec{x} und \vec{y} . Der Spezialfall $S = I$, I die Einheitsmatrix, liefert die euklidische Distanz

$$d_{xy} = [(\vec{x} - \vec{y})' (\vec{x} - \vec{y})]^{1/2} = \left[\sum_{i=1}^n (x_i - y_i)^2 \right]^{1/2}. \quad (243)$$

Die Mahalanobis-Distanz findet man u.a. bei der Definition der multivariaten Normalverteilung: sind die Komponenten x_j des Vektors \vec{x} normalverteilte Variablen mit dem Erwartungs- oder Mittelwert μ_j (oder der Schätzung \bar{x}_j von μ_j), und setzt man $\vec{y} = \vec{\mu} = (\mu_1, \dots, \mu_n)'$, so ist die multivariate Normalverteilung durch

$$f(\vec{x}) = A \exp \left[-\frac{1}{2} (\vec{x} - \vec{\mu})' S^{-1} (\vec{x} - \vec{\mu}) \right] \quad (244)$$

definiert, wobei S die Varianz-Kovarianz-Matrix der Komponenten von \vec{x} ist, und A ist eine Konstante derart, dass $\int_X f(\vec{x}) d\vec{x} = 1$ ist. Variiert man den Vektor \vec{x} so, dass $f(\vec{x}) = k_0$ eine Konstante, so muß auch $(\vec{x} - \vec{\mu})' S^{-1} (\vec{x} - \vec{\mu})$ gleich einer Konstanten sein, etwa $(\vec{x} - \vec{\mu})' S^{-1} (\vec{x} - \vec{\mu}) = \delta$. δ ist dann gleich der Mahalanobis-Distanz zwischen den Endpunkten von \vec{x} und $\vec{\mu}$, und einem bestimmten Wert von δ entspricht ein bestimmter Wert der Dichte $f(\vec{x})$.

Die Betrachtungen in diesem Abschnitt sind nicht an die Annahme gekoppelt, die Daten seien multivariat verteilt. Es geht hier einfach um den Sachverhalt, dass eine symmetrische Matrix stets ein Ellipsoid definiert; eine symmetrische 2×2 -Matrix definiert speziell eine Ellipse. Dass deshalb auch die Daten multivariat normalverteilt sind, folgt daraus nicht. Auf die multivariate Normalverteilung wird in Abschnitt 3.2.8 eingegangen.

¹³Prasanta Chandra Mahalanobis (29. Juni 1893 (Kalkutta) - 28. Juni 1972 (Kalkutta)) war ein indischer Physiker und Statistiker.

Zur Bedeutung der Mahalanobis-Distanz: Die Längen der Vektoren $\vec{u} = \vec{x} - \vec{\mu}$ variieren zwischen a_2 , der kleinsten Länge, und a_1 , der maximalen Länge. Diese Längen sind euklidische Distanzen zwischen dem Nullpunkt $\|\vec{u}\| = 0$ (für $\vec{x} = \vec{\mu}$) und dem Endpunkt eines Vektors \vec{u} auf der Ellipse. Allen diesen euklidischen Distanzen entspricht dieselbe Mahalanobis-Distanz. Nun ist $f(\vec{x})d\vec{x}$ ein Wahrscheinlichkeitselement, d.h. $f(\vec{x})d\vec{x}$ ist die Wahrscheinlichkeit, einen Vektor zwischen \vec{x} und $\vec{x} + d\vec{x}$ zu beobachten. Für alle Vektoren $\vec{u} = \vec{x} - \vec{\mu}$ ist diese Wahrscheinlichkeit gleich groß, sofern sie der gleichen Mahalanobis-Distanz entsprechen. Es ist gewissermaßen am "leichtesten", eine bestimmte euklidische Distanz zurückzulegen, wenn sie in Richtung der Hauptachse mit der maximalen Länge liegt, und am "schwierigsten", wenn sie in Richtung der Hauptachse mit der minimalen Länge liegt; je mehr die zurückzulegende euklidische Distanz in die Richtung dieser Hauptachse weist, desto mehr muß ein durch die Wahrscheinlichkeitsverteilung gegebener "Widerstand" überwunden werden. Die Mahalanobis-Distanz repräsentiert gewissermaßen den durch die Wahrscheinlichkeitsverteilung gegebenen "Widerstand", der der Überwindung einer euklidischen Distanz in Abhängigkeit von ihrer Richtung entgegengesetzt wird. Aus der gewöhnlichen Regressionsrechnung ist bekannt, dass Messwertpaare (u_1, u_2) – also die Komponenten des Vektors \vec{u} –, die der Regressionsgleichung $\hat{u}_2 = \alpha u_1 + \beta$ entsprechen, bei denen also u_2 wenig von \hat{u}_2 abweicht, mit größerer Wahrscheinlichkeit auftreten als solche, bei denen u_2 Werte in der zu $\hat{u}_2 = \alpha u_1 + \beta$ orthogonalen Richtung abweicht. Eine gegebene Mahalanobis-Distanz δ charakterisiert eine Menge A_δ von Abweichungen, die alle mit gleicher, von δ abhängenden Wahrscheinlichkeit auftreten; die mit dieser Wahrscheinlichkeit auftretenden Abweichungen sind um so kleiner, je mehr sie von der Regressionsgeraden abweichen.

Wie unterschiedlich die euklidischen Distanzen in A_δ sind, hängt natürlich von den Parametern der Ellipse ab. Es sei etwa

$$S = \begin{pmatrix} s_1^2 & s_1 s_2 r \\ s_2 s_1 r & s_2^2 \end{pmatrix}. \quad (245)$$

Dann ist¹⁴

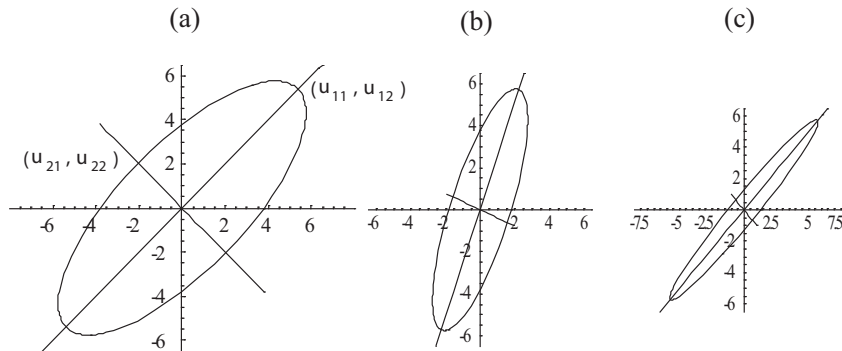
$$S^{-1} = \begin{pmatrix} \frac{s_2^2}{s_1^2 s_2^2 (1-r^2)} & -\frac{r}{s_1 s_2 (1-r^2)} \\ -\frac{r}{s_1 s_2 (1-r^2)} & \frac{s_1^2}{s_1^2 s_2^2 (1-r^2)} \end{pmatrix}. \quad (246)$$

Man sieht, dass S^{-1} eine Diagonalmatrix ist, wenn $r = 0$ ist. Es sei $\vec{u} = \vec{x} - \vec{\mu} = (u_1, u_2)'$, so dass

$$\begin{aligned} \delta^2 = (\vec{x} - \vec{\mu})' S^{-1} (\vec{x} - \vec{\mu}) &= \vec{u}' S^{-1} \vec{u} \\ &= \frac{s_2^2 u_1^2}{s_1^2 s_2^2 (1-r^2)} + \frac{s_1^2 u_2^2}{s_1^2 s_2^2 (1-r^2)} - \frac{2s_1 s_2 r u_1 u_2}{s_1^2 s_2^2 (1-r^2)} \\ &= \frac{u_1^2}{s_1^2 (1-r^2)} + \frac{u_2^2}{s_2^2 (1-r^2)} - \frac{2r u_1 u_2}{s_1 s_2 (1-r^2)}. \end{aligned} \quad (247)$$

¹⁴Wie S^{-1} bestimmt wird, wird hier nicht hergeleitet.

Abbildung 7: Ellipsen als Orte von Punkten mit gleicher Mahalanobis-Distanz.
 (a) $s_1 = s_2 = 5.75$, $r = .75$, (b) $s_1 = 2.75$, $s_2 = 5.75$, $r = .75$, (c) $s_1 = s_2 = 5.75$, $r = .975$



Die Parameter der Ellipse sind werden also durch die Werte von s_1 , s_2 und r , der Korrelation, festgelegt.

Es gibt verschiedene Fälle:

1. **Der Fall $S = I$:** In diesem Fall gilt

$$\delta^2 = d^2 = \vec{u}'\vec{u} = \sum_{i=1}^n u_i^2,$$

d.h. die Mahalanobis-Distanz ist gleich der euklidischen Distanz.

2. **Der Fall $S \neq I$:** Es gibt wieder zwei Fälle:

- (a) **Der Fall $r = 0$:** In diesem Fall gilt für $n = 2$

$$\delta^2 = \frac{u_1^2}{s_1^2} + \frac{u_2^2}{s_2^2}; \quad (248)$$

hier werden offenbar die Abweichungen $u_i = x_i - \mu_i$ in Standardabweichungen ausgedrückt und damit wie bei der z -Standardisierung auf einer gemeinsamen Skala repräsentiert. Nach der Tchebyscheffschen Ungleichung $P(|x_i - \mu_i| \geq k\sigma_i) \leq 1/k^2$ gilt ja insbesondere

$$P(|x_i - \mu_i| \geq 3\sigma_i) \leq \frac{1}{9} = .111,$$

d.h. $P(|u_i| \leq 3\sigma_i) \leq .89$: in ca 90 % der Fälle sind die u_i kleiner als 3 σ -Einheiten. Der Übergang von u_i zu $y_i = u_i/\sigma_i$ bedeutet also nicht nur den Übergang zu Werten y_i für die Abweichungen vom Mittel- bzw. Erwartungswert, die frei von Maßeinheiten sind, sondern auch den Übergang zu Werten, die eine Bewertung von

Abbildung 8: Mahalanobis-Distanzen als Funktion von r bei einer 2-dimensionalen Normalverteilung, $s_1 = 1.275$, $s_2 = 2.175$. (a) $\vec{x}_1 = (.375, .475)'$, (b) $\vec{x}_2 = (3.75, 4.75)'$. Die Form der Beziehung ist in beiden Fällen gleich, aber der längere Vektor \vec{x}_2 impliziert größere Werte von δ .

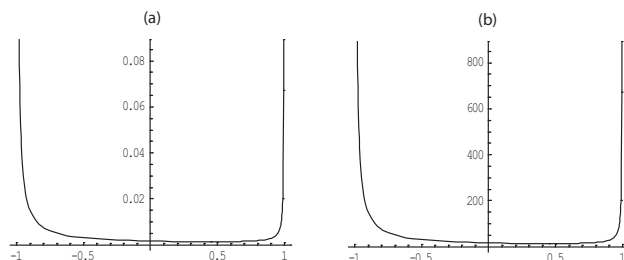
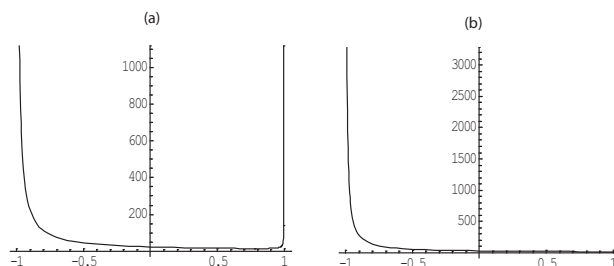


Abbildung 9: Mahalanobis-Distanzen als Funktion von r bei einer 2-dimensionalen Normalverteilung, $s_1 = s_2 = 1.275$. (a) $\vec{x}_1 = (3.75, 4.75)'$, (b) $\vec{x}_2 = (4.75, 4.75)'$. Gleiche Varianzen und/oder gleiche Vektorkomponenten implizieren keinen symmetrischen Verlauf von $\delta(r)$.



”klein” und ”groß” durch Bezug auf eben die σ -Einheiten erlauben: ein Wert $|u_i|$ ist etwa dann richtig ”groß”, wenn er größer als 3σ ist.

- (b) **Der Fall $r \neq 0$:** Setzt man wieder $y_i = u_i/s_i$, so nimmt (247) die Form

$$\delta^2 = \frac{y_1^2 + y_2^2 - 2ry_1y_2}{1 - r^2} \quad (249)$$

an. Man kann nun die Fälle $|r| \rightarrow 0$ und $|r| \rightarrow 1$ betrachten. Der Fall $r \rightarrow 0$ führt auf den schon behandelten Fall (248) zurück. Für $r \rightarrow 1$ dagegen folgt für $y_1 \neq y_2$ aus (249)

$$\delta^2 \rightarrow \frac{y_1^2 + y_2^2 - 2ry_1y_2}{1 - r^2} \rightarrow \infty \text{ für } r \rightarrow 1;$$

je größer der Wert von r , desto größer ist, für gegebene Werte von y_1 und y_2 , die Mahalanobis-Distanz δ . Eine Ausnahme bildet der

Fall $y_1 = y_2$; dann erhält man

$$\delta^2 = \frac{y_1^2 + y_2^2 - 2ry_1y_2}{1 - r^2} = \frac{2y_1^2(1 - r)}{1 - r^2} = \frac{2y_1^2(1 - r)}{(1 - r)(1 + r)} = \frac{2y_1^2}{1 + r} \rightarrow y_1^2$$

für $r \rightarrow 1$; dieser Fall ist möglich, hat aber die Wahrscheinlichkeit Null¹⁵ und ist deshalb von keinem praktischem Interesse.

Für $r \rightarrow -1$ folgt analog

$$\delta^2 \rightarrow \frac{y_1^2 + y_2^2 + 2ry_1y_2}{1 - r^2} \rightarrow \infty \text{ für } r \rightarrow -1.$$

Die Betrachtungen für den Fall $y_1 = y_2$ übertragen sich allerdings nicht auf den Fall $r \rightarrow -1$; in diesem Fall gilt ebenfalls $\delta \rightarrow \infty$.

Zusammenfassend kann man also sagen, dass $\delta^2 \rightarrow \infty$ für $|r| \rightarrow 1$; je größer also der Betrag $|r|$ der Korrelation ist. In den Abbildungen 8 und 9 wird die Beziehung zwischen r und δ illustriert.

3.2.7 Die Mahalanobis-Distanz und die Distanz zwischen Personen

Die Personen unterscheiden sich hinsichtlich der Messwerte, die sie für die Variablen oder Tests V_1, \dots, V_n erhalten haben. Es kann interessant sein, die Ähnlichkeit einer Person zu einer anderen Person zu bestimmen, die etwa einen bestimmten Typ repräsentiert. Eine solche Ähnlichkeitsbestimmung kann etwa dazu dienen, zu entscheiden, welchem Typus eine Person am ehesten entspricht (Berufsberatung, klinische Diagnose, etc.). Die Unterschiede zwischen der i -ten und der k -ten Person lassen sich dann durch die Differenzen

$$z_{i1} - z_{k1}, z_{i2} - z_{k2}, \dots, z_{in} - z_{kn} \quad (250)$$

angeben. Man könnte diese Differenzen zu einem Distanzmaß zusammenfassen, das die Gesamtähnlichkeit bzw. Gesamtunähnlichkeit angibt, z.B. die euklidische Distanz

$$d_{ik}^2 = \sum_{j=1}^n (z_{ij} - z_{kj})^2.$$

Da die Variablen aber im allgemeinen miteinander korrelieren, ist d_{ik} schwer zu interpretieren; die Differenzen $z_{ij} - z_{kj}$, $j = 1, 2, \dots, n$ werden dann ja auch korreliert sein und demnach nicht unabhängig voneinander in das Distanzmaß d_{ik} eingehen.

In Abschnitt 3.2.5 sind verschiedene Skalierungen der Personen- und Variablenkoordinaten diskutiert worden. Der gängige Fall ist der Fall $\alpha = 0$:

¹⁵Die Wahrscheinlichkeit, dass eine stetige zufällige Veränderliche einen bestimmten Wert annimmt, hat stets die Wahrscheinlichkeit Null. Nur das Ereignis, dass eine solche Variable einen Wert in einem bestimmten, endlichen Intervall annimmt, kann eine Wahrscheinlichkeit ungleich Null haben.

Der Fall $\alpha = 0$: Hier sind die Personen durch die Koordinaten q_{i1}, \dots, q_{in} charakterisiert, und die q_{ij} für $j = 1, \dots, n$ sind unkorreliert. Also macht es Sinn, die Distanz oder Unähnlichkeit der Personen i und k durch die euklidische Distanz im Q -Raum auszudrücken:

$$d_{ik}^2 = \sum_{j=1}^n (q_{ij} - q_{kj})^2 = (\vec{q}_i - \vec{q}_k)'(\vec{q}_i - \vec{q}_k), \quad (251)$$

wobei $\vec{q}_i = (q_{i1}, \dots, q_{in})'$, und \vec{q}_k ist analog definiert. Da die q_{ij} und q_{kj} bereits standardisierte Werte sind, müssen sie nicht mehr in s -Einheiten (s für Standardabweichung) ausgedrückt werden, um sie vergleichbar zu machen, und die Unkorreliertheit der q -Werte führt dann, würde man eine allgemeine Mahalanobis-Distanz wählen, sowieso auf die euklidische Distanz. Es soll gezeigt werden, in welcher Beziehung d_{ik}^2 zu einer Distanz steht, die über die Differenzen $z_{ij} - z_{kj}$ berechnet wird.

Satz 3.4 *Es gelte die SVD $Z = Q\Lambda^{1/2}P'$ und die i -te und die k -te Person seien durch die i -te und die k -te Zeile von Q charakterisiert. Dann gilt*

$$d_{ik}^2 = (\vec{q}_i - \vec{q}_k)'(\vec{q}_i - \vec{q}_k) = (\vec{z}_i - \vec{z}_k)'S^{-1}(\vec{z}_i - \vec{z}_k) = \delta_{ik}^2, \quad (252)$$

wobei $\vec{z}_i = (z_{i1}, \dots, z_{in})'$, und \vec{z}_k ist analog definiert; die euklidische Distanz zwischen der i -ten und der k -ten Person im Q -Raum ist also gleich der Mahalanobis-Distanz zwischen den gleichen Personen im Variablenraum.

Beweis: Aus $Z = Q\Lambda^{1/2}P'$ folgt

$$ZP\Lambda^{-1/2} = Q.$$

Sicherlich ist

$$\vec{q}_i' = (q_{i1}, \dots, q_{in}) = (z_{i1}, \dots, z_{in})P\Lambda^{-1/2} = \vec{z}_i'P\Lambda^{-1/2}.$$

Dann folgt

$$(\vec{q}_i - \vec{q}_k)'(\vec{q}_i - \vec{q}_k) = (\vec{z}_i - \vec{z}_k)'P\Lambda^{-1/2}\Lambda^{-1/2}P'(\vec{z}_i - \vec{z}_k),$$

oder

$$d_{ik}^2 = (\vec{z}_i - \vec{z}_k)'P\Lambda^{-1}P'(\vec{z}_i - \vec{z}_k).$$

Aber

$$P\Lambda^{-1}P' = S^{-1},$$

so dass man

$$d_{ik}^2 = (\vec{z}_i - \vec{z}_k)'S^{-1}(\vec{z}_i - \vec{z}_k) \quad (253)$$

schreiben kann, und diese Beziehung war nachzuweisen. \square

Der Fall $\alpha = 1$: Hier sind die Koordinaten der Personen durch $L = Q\Lambda^{1/2}$ gegeben. Die Koordinaten der i -ten und der h -ten Person sind durch die i -te bzw. h -te Zeile von L gegeben, oder durch die i -te und die h -te Spalte von $L' = \Lambda^{1/2}Q'$. Weiter folgt aus $Z = LP'$ für L die Beziehung $LP'P = L = ZP$, denn P ist ja orthonormal. Also hat man ebenfalls $L' = P'Z'$. Mit \vec{F}_i wird der i -Spaltenvektor von L' bezeichnet, - dies ist natürlich die i -te Zeile von L .

Satz 3.5 Die Personenkoordinaten seien durch $L = Q\Lambda^{1/2}$ gegeben. Dann ist die Euklidische Distanz zwischen der i -ten und der h -ten Person

$$(\vec{F}_i - \vec{F}_h)'(\vec{F}_i - \vec{F}_h) = (\vec{z}_i - \vec{z}_h)'(\vec{z}_i - \vec{z}_h), \quad (254)$$

d.h. die euklidische Distanz zwischen den Personen entspricht der Euklidischen Distanz zwischen den Personen in den z -Koordinaten.

Beweis: Wegen $Z' = PL'$ hat man

$$(\vec{z}_i - \vec{z}_h)'(\vec{z}_i - \vec{z}_h) = (\vec{F}_i - \vec{F}_h)'P'P(\vec{F}_i - \vec{F}_h) = (\vec{F}_i - \vec{F}_h)'(\vec{F}_i - \vec{F}_h).$$

□

Die Beziehungen (252) und (254) machen deutlich, wie die "direkte" Betrachtung der Daten, d.h. die Berechnung der euklidischen Distanz zwischen Personen anhand der z_{ij} -Werte bzw. anhand der Koordinaten im Raum der latenten Dimensionen zu interpretieren ist. Betrachtet man die euklidischen Distanzen im Q -Raum, so entsprechen sie Mahalanobis-Distanzen im Datenraum, betrachtet man dagegen euklidische Distanzen im L -Raum, so entsprechen sie euklidischen Distanzen im Datenraum.

3.2.8 Die multivariate Normalverteilung

Die multivariate Normalverteilung ist, für die Variablen X_1, X_2, \dots, X_n , durch

$$f(x_1, \dots, x_n) = A \exp [-(\vec{x} - \vec{\mu})^t \Sigma^{-1} (\vec{x} - \vec{\mu})] \quad (255)$$

mit der Normierungskonstante

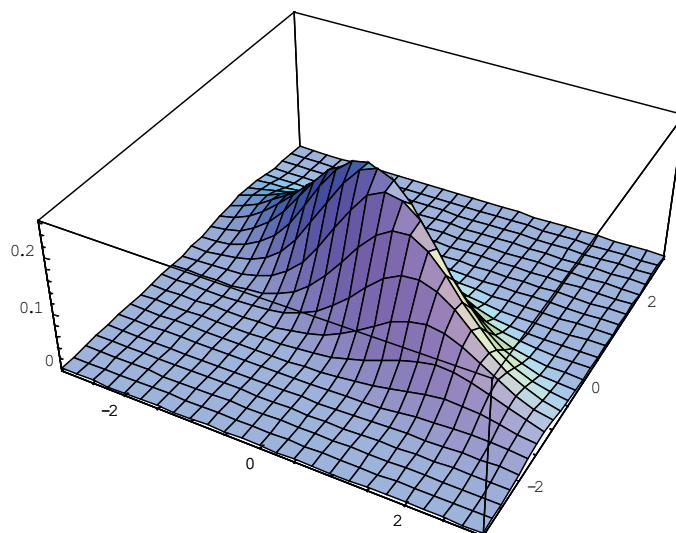
$$A = \frac{1}{(2\pi)^{1/2n} |\Sigma^{-1}|^{1/2}}. \quad (256)$$

definiert. Darin ist Σ die $n \times n$ -Varianz-Kovarianz-Matrix für die X_1, X_2, \dots, X_n . Der Erwartungswert der j -ten Variablen X_j sei μ_j ; dann ist $\vec{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^t$ der Vektor der Erwartungswerte. $|\Sigma^{-1}|$ ist die *Determinante* der Matrix Σ^{-1} ; die Determinante (vergl. Abschnitt 6.5 für den 2-dimensionalen Fall) ist stets eine einzelne Zahl (Skalar). $|\Sigma^{-1}|^{1/2}$ entspricht dem σ^{-1} im 1-dimensionalen Fall; man kann sagen, dass $|\Sigma^{-1}|^{1/2}$ die Gesamtvarianz des Vektors \vec{x} repräsentiert.

Die Definition (256) ist keineswegs willkürlich formuliert worden. Im Anhang 6.5 wird die 2-dimensionale Normalverteilung für korrelierte Variable explizit hergeleitet, und der n -dimensionale Fall folgt durch Verallgemeinerung.

Die Inspektion der Definition (255) zeigt, dass ein Ort gleicher Wahrscheinlichkeit für die X_1, \dots, X_n durch ein Ellipsoid gegeben ist. Die Orientierung

Abbildung 10: 2-dimensionale Normalverteilung; $\sigma_x = 1.25$, $\sigma_y = .75$, $r = -.75$



des Ellipsoids ist durch die Eigenvektoren von Σ^{-1} gegeben, wobei diese Eigenvektoren identisch mit denen der Matrix Σ sind. Die Eigenwerte sind die reziproken Eigenwerte von Σ . Für $k = 1$ ist die Länge einer Halbachse gerade gleich dem entsprechenden Eigenwert (vergl. Gleichung (??), Seite ??).

Denn der Exponent der Exponentialfunktion in (244) ist durch $(\vec{x} - \vec{\mu})' S^{-1} (\vec{x} - \vec{\mu})$ gegeben. Setzt man diese Größe gleich δ^2 ,

$$\delta^2 = (\vec{x} - \vec{\mu})' S^{-1} (\vec{x} - \vec{\mu}). \quad (257)$$

δ^2 ist also das Quadrat einer Mahalanobis-Distanz, und für $\delta^2 =$ eine Konstante ist $f(\vec{x}) =$ eine Konstante. Nach der Gleichung (161), Seite 60, sind die Längen der Halbachsen durch

$$a_1 = \delta \sqrt{\nu_1}, \quad a_2 = \delta \sqrt{\nu_2}, \quad \nu_1 \geq \nu_2 \quad (258)$$

gegeben, wobei, nach Satz 2.4, ν_1 und ν_2 die Eigenwerte von S sind.

3.2.9 Beispiele

Beispiel 3.1 In der Tabelle 2 werden "künstliche" Daten gegeben, wie sie bei Einschätzung irgendwelcher Objekte oder Personen von Personen auf Ratingskalen gegeben werden. Das Prinzip der Hauptachsentransformation als Lösung des Problems der Faktorenanalyse soll anhand dieser Daten illustriert werden. Die Mittelwerte der Spalten sind $\bar{x}_1 = 4.143$, $\bar{x}_2 = 4.000$, $\bar{x}_3 = 3.857$, die Standardabweichungen sind $\hat{s}_1 = 2.193$, $\hat{s}_2 = 2.449$, $\hat{s}_3 = .900$. Dementsprechend

Tabelle 2: Daten (X)

	Skalen		
Vp	S ₁	S ₂	S ₃
1	1	2	3
2	2	1	3
3	5	7	4
4	7	5	5
5	3	2	4
6	6	7	5
7	5	4	3

ergibt sich die Matrix der z -Werte Z und die Matrix R der Korrelationen:

$$Z = \begin{pmatrix} -1.433 & -0.817 & -0.953 \\ -0.977 & -1.224 & -0.953 \\ 0.391 & 1.225 & 0.159 \\ 1.302 & 0.050 & 1.270 \\ -0.521 & -0.816 & 0.159 \\ 0.847 & 1.224 & 1.270 \\ 0.391 & 0.000 & -0.953 \end{pmatrix}, \quad R = \begin{pmatrix} 1.000 & 0.806 & 0.772 \\ 0.806 & 1.000 & 0.681 \\ 0.772 & 0.681 & 1.000 \end{pmatrix}.$$

Die Matrix Λ der Eigenwerte von R , und die Matrix der P Eigenvektoren von R sind wie folgt:

$$\Lambda = \begin{pmatrix} 2.508 & 0 & 0 \\ 0 & 0.322 & 0 \\ 0 & 0 & 0.171 \end{pmatrix}, \quad P = \begin{pmatrix} -0.595 & 0.101 & 0.792 \\ -0.573 & 0.642 & -0.509 \\ -0.564 & -0.760 & -0.325 \end{pmatrix} \quad (259)$$

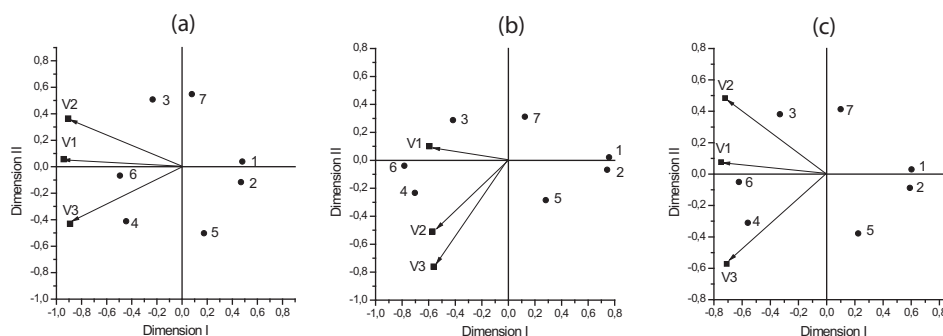
Die Zeilen in der Matrix P repräsentieren die Variablen.

Der Fall $\alpha = 0$: Für die Matrix Q der Eigenvektoren von ZZ' - also die Matrix der Faktorscores - und die Matrix $A = P\Lambda^{1/2}$ der Ladungen findet man

$$Q = \begin{pmatrix} -0.479 & 0.039 & -0.413 \\ 0.469 & -0.116 & 0.152 \\ -0.264 & 0.508 & -0.359 \\ -0.445 & -0.411 & -0.414 \\ 0.177 & -0.502 & 0.051 \\ -0.495 & -0.067 & 0.356 \\ 0.078 & 0.549 & -0.614 \end{pmatrix}, \quad A = \begin{pmatrix} -0.943 & -0.057 & 0.329 \\ -0.907 & 0.364 & -0.210 \\ -0.892 & -0.431 & -0.134 \end{pmatrix} \quad (260)$$

Die Zeilen in Q repräsentieren die Vpn, die in A repräsentieren sie die Variablen. In Abb. 11 findet man die entsprechenden graphischen Repräsentationen

Abbildung 11: Graphische Darstellung der Ergebnisse; (a) der Fall $\alpha = 0$, (b) der Fall $\alpha = 1$, (c) der Fall $\alpha = 1/2$.



für die Annahme $s = 2$. Die Personen liegen in guter Näherung auf einem Kreis. Dies entspricht der Tatsache, dass die Spaltenvektoren von Q die Länge 1 haben und die Daten gut durch 2 Dimensionen erklärt werden können. Denn die Komponenten des ersten Spaltenvektors \vec{q}_1 von Q sind ja die Koordinaten der Personen auf der ersten Achse, und wenn \vec{q}_1 die Länge 1 hat, dann sollte für die "Varianz" $\sum_i q_{i1}^2 \approx 1$ dieser Projektionen $\sum_i q_{i1}^2 \approx 1$ gelten, und der entspricht die Länge $\sqrt{\sum_i q_{i1}^2}$. Das gleiche gilt für die Projektionen auf die zweite Dimension, also für \vec{q}_2 .

Man rechnet leicht nach, daß die Längen der Vektoren für die Skalen durch .89 für Skala 1, .956 für Skala 2 und .982 für Skala 3 ist. Die Approximation durch 2 Dimensionen ist nicht schlecht. Die Berechnung der Korrelationen auf der Basis von 2 Dimensionen sei zur Übung empfohlen.

Der Fall $\alpha = 1$: Man fokussiert auf die Personen, betrachtet also die Matrix $L = Q\Lambda^{1/2}$ für die Personenkoordinaten und P für die Variablen:

$$L = \begin{pmatrix} .758 & .022 & -.171 \\ .743 & -.066 & -.063 \\ -.418 & .288 & -.148 \\ -.704 & -.233 & .171 \\ .281 & -.285 & -.021 \\ -.784 & -.038 & -.147 \\ .124 & .312 & .253 \end{pmatrix}, \quad P = \begin{pmatrix} -.595 & .101 & .792 \\ -.573 & .642 & -.509 \\ -.564 & -.760 & -.325 \end{pmatrix} \quad (261)$$

Die PC-Werte (PC für Principal Components = Hauptachsen) zeigen deutlich die elliptische Form der Punktekonfiguration der "Personen", im Gegensatz zur eher kreisförmigen Konfiguration des Falles $\alpha = 0$.

Der Fall $\alpha = 1/2$: Hier wird für die Personen die Matrix $\tilde{L} = Q\Lambda^{1/4}$ und für

die Variablen die Matrix $\tilde{A} = P\Lambda^{1/4}$ betrachtet:

$$\tilde{L} = \begin{pmatrix} .602 & .030 & -.265 \\ .590 & -.088 & .098 \\ -.332 & .382 & -.231 \\ -.560 & -.310 & .266 \\ .223 & -.378 & -.033 \\ -.623 & -.050 & -.229 \\ .099 & .414 & .394 \end{pmatrix}, \quad \tilde{A} = \begin{pmatrix} -.749 & .076 & .512 \\ -.721 & .484 & -.327 \\ -.709 & -.572 & -.209 \end{pmatrix}. \quad (262)$$

In Abbl. 11, (c) wird dieser Fall illustriert. □

Beispiel 3.2 Davenport und Studdert-Kennedy (1972)¹⁶ analysierten die ästhetischen Urteile des Kunstkritikers Roger de Pile über 56 Maler, von Albani, Dürer, Veronese, Holbein, Rembrandt, Rubens, Titian bis Van Dyck, Vanius und den Zuccaros, die de Pile Jahr 1743 notierte; diese Ratings liefern möglicherweise auch Informationen über die Kunstrezeption in der Mitte des 18-ten Jahrhunderts. Monsieur de Pile "ratete" die Maler in bezug auf vier Merkmale: "Komposition", "Zeichnung", "Farbe" und "Ausdruck", d.h. er schätzte die Maler bezüglich dieser Merkmale auf einer Skala von 0 bis 20 ein; die Ratingskala ist also keine Erfindung neuzeitlicher Psychologen. Die Skala wurde von ihm bei 20 "verankert": dieser Wert drückt "vollständige Perfektion, die noch kein Mensch je erreicht hat" aus. Die genannten vier Merkmale sind sicherlich komplexer Natur, aber M. de Pile war der Ansicht (zitiert nach Davenport et al (1972)), daß sie unabhängig voneinander die elementaren Qualitäten eines Gemäldes reflektieren. Die Maler übernehmen hier die Rolle der Vpn, die vier Merkmale die Rolle der Tests. Die Korrelationen zwischen den Merkmalen sind in der Tabelle 5 zusammengefaßt worden: Generell fällt auf, daß die drei Merkmale Komposition, Zeichnung und Ausdruck positiv miteinander korrelieren, daß aber alle drei Merkmale negativ mit dem Merkmal Farbe korrelieren. Die Eigenwerte dieser Matrix sind

$$\lambda_1 = 2.298, \quad \lambda_2 = 1.017, \quad \lambda_3 = .375, \quad \lambda_4 = .310$$

Man überzeuge sich, daß die Summe der Eigenwerte gleich 4 ist. Der Anteil der Varianz der ersten Komponente ist dann $\pi_1 = \lambda_1/4 = .57$, der der zweiten Komponente ist $\pi_2 = \lambda_2/4 = .25$, der der dritten Komponente $\pi_3 = \lambda_3 = .09$, und für die letzte Komponente erhält man schließlich $\pi_4 = \lambda_4/4 = .08$.

Davenport et al. betrachten nicht die Eigenvektoren der Korrelationsmatrix, sondern die der *Varianz-Kovarianzmatrix*. Die entsprechenden Faktorladungen werden in Tabelle 5 angegeben.

¹⁶Davenport, M., Studdert-Kennedy, H. (1970) Use of orthogonal factors for selection of variables in a regression equation. Appl. Statist. **21**, 324-333. Dem Titel entsprechend diskutieren die Autoren die Anwendung der Hauptachsentransformation (Principal Component Analysis - PCA) im Rahmen eines Regressionsproblems. Es sollen optimale Prädiktoren für die Ratings gefunden werden.

Tabelle 3: Roger de Piles Ratings und korrespondierende Faktorwerte

	Maler	de Piles Kategorien					Faktorwerte			
		Komp.	Zeich.	Farbe	Ausd.	Schule	ϕ_1	ϕ_2	ϕ_3	ϕ_4
1	Albani	14	14	10	6	e	-.038	-.012	.069	.173
2	Dürer	8	10	10	8	f	.066	-.051	.000	-.204
3	Del Sarto	12	16	9	8	a	-.073	-.071	-.077	.110
4	Barocci	14	15	6	10	c	-.123	-.076	.072	.011
5	Bassano	6	8	17	0	d	.250	.023	-.064	.057
6	Del Prombe	8	13	16	7	a	.074	.042	-.242	.017
7	Bellini	4	6	14	0	d	.281	-.055	.018	-.103
8	Bourdon	10	8	8	4	h	.101	-.076	.228	-.100
9	Le Brun	16	16	8	16	h	-.209	.042	-.015	-.106
10	Veronese	15	10	16	3	d	.083	.154	.134	.232
11	The Caracci	15	17	13	13	e	-.147	.100	-.168	.079
12	Corregio	13	13	15	12	e	-.041	.145	-.116	-.059
13	Volterra	12	15	5	8	b	-.086	-.149	.054	.025
14	Diepenbeck	11	10	14	6	g	.081	.070	.023	-.010
15	Domenichino	15	17	9	17	e	-.216	.046	-.118	-.119
16	Giorgione	8	9	18	4	d	.178	.103	-.106	.004
17	Guercino	18	10	10	4	e	-.002	.079	.358	.199
18	Holbein	9	10	16	13	f	.043	.142	-.162	-.298
19	Da Udine	10	8	16	3	a	.167	.093	.052	.032
20	J.Jordaens	10	8	16	6	g	.136	.119	.020	-.082
21	L.Jordaens	13	12	9	6	c	-.004	-.029	.127	.061
22	Josepin	10	10	6	2	c	.079	-.158	.221	.028
23	Giulio Romano	15	16	4	14	a	-.226	-.063	.047	-.176
24	Lanfranco	14	13	10	5	e	-.013	-.010	.108	.173
25	Da Vinci	15	16	4	14	a	-.205	-.080	.069	-.102
26	Van Leyden	8	6	6	4	f	.139	-.132	.288	-.249
27	Michelangelo	8	17	4	8	a	-.079	-.258	-.124	-.012
28	Caravaggio	6	6	16	0	e	.272	.023	.035	-.030
29	Murillo	6	8	15	4	d	.194	.014	-.061	-.119
30	Otho Venius	13	14	10	10	g	-.069	.006	-.015	-.004
31	Palma Vecchio	5	6	16	0	d	.284	.006	.003	-.055
32	Palma Giovane	12	9	14	6	d	.085	.097	.092	-.018
33	Parmigiano	10	15	6	6	b	-.036	-.176	-.011	.063
34	Fr. Penni	0	15	8	0	a	.151	-.349	-.307	.063
35	Perino del Vaga	15	16	7	6	a	-.099	-.083	.085	.238
36	Cortona	16	14	12	6	c	-.046	.065	.075	.246
37	Perugino	4	12	10	4	a	.123	-.172	-.158	-.077
38	Pordenone	8	14	17	5	d	.087	.036	-.282	.142
39	Pourbus	4	15	6	6	f	.030	-.275	-.199	-.088
40	Poussin	16	17	6	15	h	-.216	-.038	-.027	-.079
41	Primaticcio	15	14	7	10	b	-.112	-.027	.117	.011
42	Raphael	17	18	12	18	a	-.243	.143	-.174	-.035
43	Rembrandt	15	6	17	12	g	.053	.296	.165	-.248
44	Rubens	18	13	17	17	g	-.135	.315	-.059	-.101

Tabelle 4: Roger de Piles Ratings und korrespondierende Faktorwerte; Fortsetzung

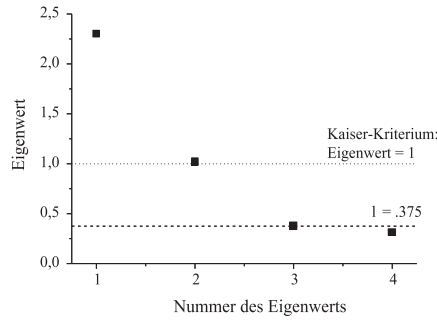
		de Piles Kategorien					Faktorwerte			
	Maler	Komp.	Zeich.	Farbe	Ausd.	Schule	ϕ_1	ϕ_2	ϕ_3	ϕ_4
45	Fr. Salviata	13	15	8	8	b	-.077	-.066	.015	.086
46	Le Sueur	15	15	4	15	h	-.200	-.061	.096	-.177
47	Teniers	15	12	13	6	g	.001	.092	.096	.158
48	Testa	11	15	0	6	c	-.089	-.293	.160	.018
49	Tintoretto	15	14	16	4	d	.014	.120	-.028	.244
50	Titian	12	15	18	6	d	.024	.122	-.228	.254
51	Van Dyck	15	10	17	13	g	-.016	.263	.003	-.136
52	Vanius	15	15	12	13	c	-.124	.099	-.069	-.007
53	T. Zuccaro	13	14	10	9	b	-.059	-.002	-.005	-.034
54	F. Zuccaro	10	13	8	8	b	-.014	-.094	-.004	-.065
	\bar{x}	11.56	12.46	10.94	7.70					
	Stand'abw. s	4.09	3.46	4.65	4.85					
		a: Renaissance, b: Manieristen, c: Seicento, d: Venezianisch, e: Lombardisch, f: 16-tes Jahrh., g: 17-tes Jahrh., g: französisch								

Tabelle 5: Merkmalskorrelationen und Faktorladungen

Korrelationen zwischen den Merkmalen				
	Kompos.	Zeichn.	Farbe	Ausdruck
Kompos.	1.00	.415	-.097	.656
Zeichn.	.415	1.00	-.517	.575
Farbe	-.097	-.517	1.00	-.209
Ausdruck	.656	.575	-.208	1.00
de Piles Ästhetik: Faktorladungen bezüglich der Hauptachsen				
Merkmal	ϕ_1	ϕ_2	ϕ_3	ϕ_4
Kompos.	.48	-.37	.78	.10
Zeichn.	.42	.19	-.28	.84
Farbe	-.38	-.85	-.21	.31
Ausdruck	.66	-.33	-.31	-.43
kum. Varianz	55.95	84.48	93.59	100.00

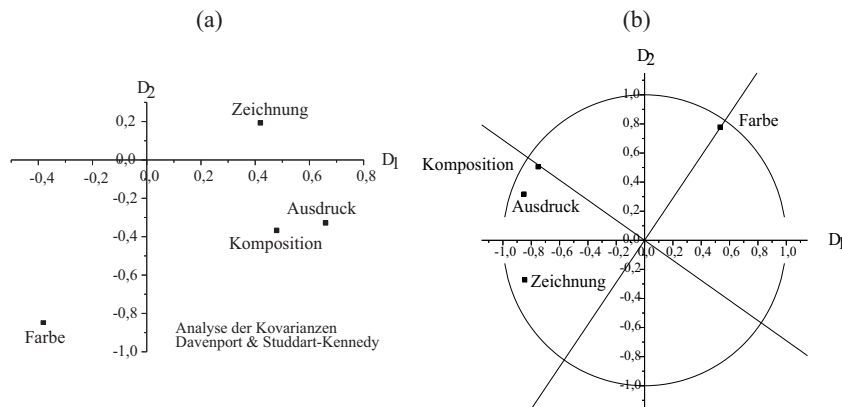
Bevor man die Dimensionen interpretiert, sollte man sich klarmachen, was die Faktorwerte und Faktorladungen in diesem Beispiel bedeuten. Die Rolle der Personen wird hier durch die Maler eingenommen, die der Tests durch die vier Merkmale. Die Faktorwerte sind die Ausprägungen der Personen, hier also der Maler, auf den latenten Beurteilungsdimensionen ϕ_s . Die Faktorladungen sind die Gewichte, mit denen die vier Merkmale die latenten Dimensionen erfassen bzw. reflektieren. Da die Merkmale Ausdruck, Zeichnung und Kompo-

Abbildung 12: Eigenwert-Diagramm



sition positiv miteinander korrelieren, kann man vermuten, daß de Pile zwar glaubte, die vier Merkmale unabhängig voneinander beurteilen zu können, sich aber gleichwohl auf latente, d.h. ihm nicht klar bewußte Kriterien bezog. Andererseits können aber die Merkmale für ihn tatsächlich völlig getrennte, unabhängige Dimensionen darstellen; dann treten sie aber bei den Malern in korrelierter Weise auf. Maler, die z.B. gut bezüglich des Merkmals Komposition sind, sind häufiger auch gut hinsichtlich des Merkmals Zeichnung.

Abbildung 13: Faktorladungen für de Piles Merkmale von Gemälden: (a) von Kovarianzen, (b) von standardisierten Werten



Die Abbildung 13 (a) zeigt die Merkmale in dem durch die ersten beiden Dimensionen definierten Koordinatensystem. Die erste Komponente D_1 "erklärt" sicherlich den größten Teil der Varianz, und alle vier Merkmale Komposition, Zeichnung, Farbe und Ausdruck "laden" auf dieser Achse, wobei "Ausdruck" den größten Wert auf der ersten Achse hat, "Farbe" den größten Wert (absolut gesehen) auf der zweiten Dimension. Die beiden ersten Dimensionen erklären

bereits 82% der Varianz in den Daten, während die beiden letzten Dimensionen nur jeweils $\approx 9\%$ erklären. Gängigen Abbruchkriterien zufolge könnte

Tabelle 6: Faktorenladungen bei standardisierten Ratings

Merkmal	Dimension			
	ϕ_1	ϕ_2	ϕ_3	ϕ_4
Komposition	-.749	.505	.367	.219
Zeichnung	-.841	-.275	-.373	.277
Farbe	.535	.777	-.310	.116
Ausdruck	-.850	.317	-.149	-.394

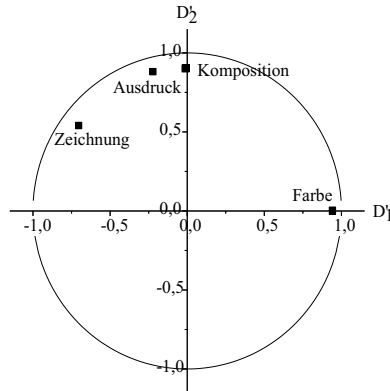
man die beiden letzten Dimensionen also als "Fehler"-Dimensionen vernachlässigen. Die Inspektion der Komponenten auf diesen Achsen zeigt aber, daß die Merkmale "Komposition" auf der dritten und "Zeichnung" auf der vierten noch verhältnismäßig hohe Koordinatenwerte haben. Diesem Befund zufolge kann man vermuten, daß in der Tat vier Dimensionen in die Urteile eingehen; der Scree-Test oder Kaiser und Dickmans Kriterium können also ein falsches Bild liefern (nach Guttman (1954) ist die Anzahl der Eigenwerte größer als 1 ja auch nur eine *untere* Grenze). Diese Dimensionen reflektieren aber nicht eindeutig die von de Pile gewählten Charakteristika. Alle Dimensionen reflektieren alle Charakteristika, wenn auch die letzten drei Dimensionen deutlich zu bestimmten, von de Pile benutzten Merkmalen korrespondieren.

In Gleichung (??), Seite ??, wird eine Beziehung zwischen den Faktorladungen A und den Korrelationen R hergestellt. Dieser Beziehung zufolge können die Korrelationen zwischen den Merkmalen aus den Ladungen zurückgerechnet werden. A ist insgesamt eine (n, n) -Matrix; berücksichtigt man nur $r < n$ Faktoren, so erhält man die Approximation $R \approx A_r A_r'$. Je geringer die angenommene Anzahl r von latenten Dimensionen ist, desto ungenauer ist diese "Vorhersage" der Korrelationen. Nimmt man $r = 2$ an, so sollten dementsprechend die Vektoren, die durch die Projektion der Vektoren $p_j = (p_{j1}, p_{j2}, \dots, p_{ps})'$ (also in die 2-dimensionale Ebene entstehen, bereits in guter Näherung die Länge 1 haben. In den Abbildungen 13 (b) und 14 ist der Einheitskreis eingezeichnet worden. Gilt die Annahme $r = 2$ exakt, sollten die Punkte, die die Merkmale repräsentieren, alle *auf* diesem Kreis liegen. Die Punkte liegen sehr nahe an diesem Kreis, was die Annahme $r = 2$ stützt.

Davenport et al. interpretieren die erste Achse (die erste *Komponente*) als einen "ersten Index" (rough index) von de Piles Gesamtreaktion auf einen Maler. Das negative Vorzeichen für die Farbe sei nicht Ausdruck eines negativen Stellenwerts der Farbe bei der Beurteilung, sondern Ausdruck von, nach de Pile, geringer Meisterschaft eines Malers.

In der Tabelle 6 werden die Faktorenladungen wiedergegeben, die sich bei der Analyse der standardisierten Ratings ergeben; Abbildung 13 (b) liefert

Abbildung 14: R. De Piles Kriterien: rotierte Achsen



ein Bild der Beziehungen zwischen den Merkmalen für diese Analyse. Da die Kovarianzen hier noch durch das Produkt der Standardabweichungen geteilt werden, ergibt sich eine andere Metrik. Gleichwohl ist die Punktekonfiguration ähnlich wie die im Falle der Analyse der Kovarianzen, sie ist überdies an den Achsen gespiegelt. Interessant ist die Rotation etwa von λ_1 durch das Merkmal "Farbe". Die dazu orthogonale Achse geht genau durch das Merkmal "Komposition"; in Abbildung 14 noch einmal dargestellt worden, die Werte der Ladungen auf den neuen Achsen findet man in Tabelle 7 (der Rotationswinkel beträgt -0.968 Radians bzw. -55.45°).

Die beiden Merkmale "Farbe" und "Komposition" sind in der Tat für die Stichprobe der betrachteten Maler unkorreliert: in der Matrix der Korrelationen findet man, daß die beiden Merkmale mit -0.097 korrelieren. Es hätte durchaus sein können, daß sich die beiden Merkmale nicht als unkorreliert erwiesen hätten, wäre die Stichprobe von Malern anders zusammengesetzt gewesen. Bei kleineren Stichproben kann es durchaus vorkommen, daß gerade Objekte - hier: Maler - in die Stichprobe aufgenommen werden, in der die Merkmale von einem Objekt zum nächsten kovariieren.

Das Interessante an der Rotation ist aber weniger die Tatsache, daß die empirische Korrelation gerade repliziert wird, sondern die Möglichkeit, (i) diese beiden Merkmale als "latente" Dimensionen zu betrachten und (ii) durch die Projektion der beiden anderen Merkmale "Zeichnung" und "Ausdruck" auf die neuen Achsen den Anteil zu bestimmen, den diese Merkmale an den Dimensionen "Komposition" und "Farbe" haben. Interessant ist auch, daß die Merkmale "Komposition" und "Ausdruck" nahezu gleiche Ladungen auf der rotierten Dimension λ'_2 haben: eine gute Komposition impliziert dann auch einen gut bewerteten Ausdruck, während eine mangelnde Komposition einen Zerfall des Ausdrucks nach sich ziehen wird (zumindest für R. de Piles).

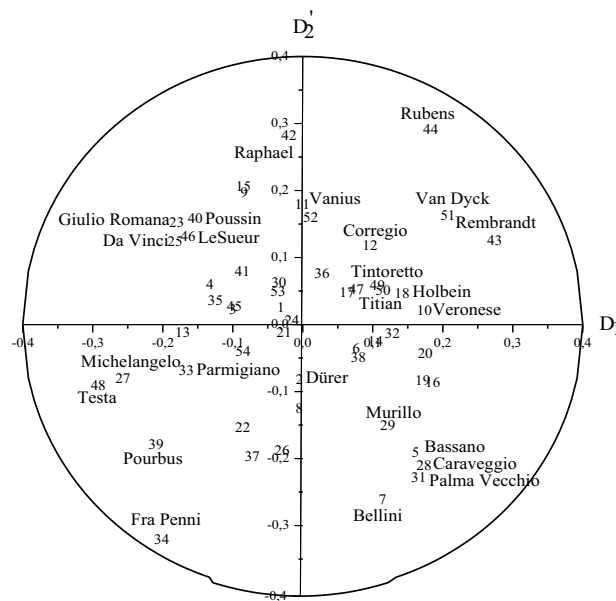
Es sei noch darauf hingewiesen, daß die Rotation zumindest partiell die

Kriterien der Einfachstruktur erfüllt. Es ist noch interessant, sich die Faktor-

Tabelle 7: Faktorenladungen für ϕ_1 und ϕ_2 , rotiert

Merkmal	Dimension	
	ϕ_1	ϕ_2
Komposition	-.008	.900
Zeichnung	-.703	.537
Farbe	.943	.000
Ausdruck	-.222	.880

Abbildung 15: R. De Piles Kriterien: die Maler (rotiert). D'_1 = "Farbe", D'_2 = "Komposition"



werte der einzelnen Maler anzusehen; Abbildung 15 zeigt die Maler bezüglich der Achsen *Farbe* und *Komposition*.

Auffallend ist, daß ein Meister wie Dürer nahe am Mittelpunkt der Konfiguration liegt. Dürer erscheint Herrn de Pile als mittelmäßig, ebenso wie Tintoretto und Tizian. Inspektion der Tabelle der Ratings zeigt, daß Caravaggio und Palma Vecchio gut bezüglich ihrer Farbgebung, aber schlecht hinsichtlich "Zeichnung" und "Komposition" beurteilt werden. Rembrandt wird bezüglich "Farbe", und "Komposition" und "Ausdruck" überdurchschnittlich beurteilt, fällt aber bei "Zeichnung" stark ab; Rubens erhält in allen Kategorien bis auf "Farbe" bessere Werte als Rembrandt (man muß nicht mit de Pile überein-

stimmen!). Da Vinci liegt zwar bei "Zeichnung" deutlich über dem Mittelwert, fällt aber in de Piles Meinung hinsichtlich "Farbe" stark ab und gerät so in eine Gegenposition zu Caravaggio und Palma Vecchio. überraschend ist die Beurteilung von Michelangelo, der nur bezüglich des Merkmals "Zeichnung" einen überdurchschnittlichen Wert erhält.

Es sei noch angemerkt, daß die Maler *nicht* wie die Merkmale auf dem oder nahe dem Einheitskreis liegen müssen; die Scores (Koordinaten) sind gleich den Komponenten der Eigenvektoren Q von ZZ' . Es gilt zwar für die s -te Dimension $Q'_s Q_s = 1$, d.h. die Einheitsvektoren haben zwar die Länge 1, aber dies heißt nicht, daß die Vektoren, die die Maler repräsentieren, die Länge 1 haben. Viele der Maler haben noch auf der dritten und vierten Dimension relativ hohe Scores.

Die Komponentenanalyse zeigt Strukturen in de Piles Urteilen auf, die sich einem nicht erschließen, wenn man sich seine Beurteilungen direkt, d.h. ohne weitere Analyse, anschaut. Gleichwohl zeigt ein Vergleich der Faktorwerte mit den Ratings durch de Pile, wie die Analyse diese Urteile aufschlüsselt: die gegensätzlichen Positionen etwa von Caravaggio und Poussin oder Rembrandt und Pourbus finden sich zwar in den komplementären Urteilen de Piles bei diesen Malern wieder, aber es gelingt nicht, aus den Urteilen direkt – eben ohne die formale Analyse – die relative Position der Maler zueinander herzuleiten.

Es ist möglich, daß de Pile seine vier Kategorien tatsächlich unabhängig voneinander benutzte, sie aber in einer korrelierten Weise (verg. Tabelle 5) bei den Malern bzw. ihren Gemälden zutreffen. Denkbar ist aber auch, daß die Merkmale der Gemälde in einer komplizierten, nichtlinearen Weise in die Urteile eingingen - die von de Pile benutzten Merkmale sind ja selbst hochkomplex! -, und die Hauptachsenanalyse vier linear wirkende Dimensionen nur vorspiegelt; anhand des Datenmaterials kann eine solche Hypothese weder verworfen noch bestätigt werden. Aus anderen Untersuchungen weiß man aber, daß Gestalteffekte durch nichtlineare Interaktion von Elementen der Wahrnehmung entstehen und entsprechend in die Urteile eingehen. Es wäre verwunderlich, wenn dies bei de Pile nicht der Fall wäre.

□

3.3 Die Hauptfaktorenanalyse

Nach dem Fundamentaltheorem (223) gilt, wenn Σ durch R ersetzt wird,

$$R = AA' + \Psi. \quad (263)$$

Man kann nun

$$R_H = R - \Psi = AA' \quad (264)$$

setzen und die Matrix R_H als Ausgangsgleichung wählen. Ψ ist die Diagonalmatrix der empirischen Restvarianzen $\hat{\sigma}_j^2$ (vergl. (224)). Mit

$$\hat{h}_i^2 = 1 - \hat{\sigma}_j^2 \quad (265)$$

wird die i -te Kommunalität bezeichnet. Man geht nun in zwei Schritten vor:

1. Anders als bei der Hauptkomponentenmethode werden zuerst die Restvarianzen $\hat{\sigma}_j^2$ bestimmt, indem man zunächst Schätzungen \hat{h}_j^2 für die Kommunalitäten bestimmt; die $\hat{\sigma}_j^2$ ergeben sich dann gemäß (265).
2. Für die reduzierte Matrix $R_H = R - \hat{\Psi}$ wird eine Hauptachsentransformation vorgenommen. Man berechnet also die Eigenwerte $\lambda_1, \dots, \lambda_n$ der Matrix R_H und die zugehörigen Eigenvektoren P und definiert wieder $A = P\Lambda^{1/2}$, so dass man wieder

$$R_H = AA'$$

gilt. Die Spalten $\vec{A}_1, \dots, \vec{A}_s$ heißen die ersten s *Hauptfaktoren*. Sie erklären

$$\sum_{u=1}^s \lambda_u = \sum_{u=1}^s h_u^2 \quad (266)$$

an Varianz.

R_H ist nicht die ursprüngliche Korrelationsmatrix. Sie entsteht durch Subtraktion von $\hat{\Psi}$ von R , und $\hat{\Psi}$ ist nur eine Schätzung. Dies bedeutet, dass bei nicht optimaler Wahl von $\hat{\Psi}$ sich negative Eigenwerte für R_H ergeben können, die natürlich keine inhaltliche Bedeutung haben. $\hat{\Psi}$ muß dann neu gewählt werden. Man geht deshalb im Prinzip *iterativ* vor: man beginnt z.B. mit der Wahl $\hat{\Psi}_0 = 0$ und erhält eine Lösung wie bei der Hauptkomponentenanalyse. Oder man beginnt mit $\hat{\Psi}_0 = I - K_0$, wobei I die Einheitsmatrix ist und K_0 ist eine Matrix, die erste Schätzungen der Kommunalitäten enthält. Man erhält dann

$$R - \hat{\Psi}_0 = A_{s1}A'_{s1} + E_1,$$

wobei E_1 einen nicht erklärten Rest bedeutet, d.h. man hat für R die Darstellung

$$R = A_{s1}A'_{s1} + E_1 + \hat{\Psi}_0.$$

$A_{s1}A'_{s1}$ definiert einen "erklärten" Anteil von R auf der Basis von s angenommenen latenten Dimensionen. Die Diagonalelemente von $A_{s1}A'_{s1}$ sind dann neue Schätzungen für die Kommunalitäten, so dass man $K_1 = \text{diag}(A_{s1}A'_{s1})$ definieren kann. Dann erhält man $\hat{\Psi}_1 = I - K_1$, und nun wendet man die Hauptachsentransformation auf die Matrix $R - \hat{\Psi}_1$ an. So fährt man weiter fort und erhält sukzessive bessere Schätzungen für die Matrix A der Ladungen. Am Ende hat man Schätzungen für die Ladungen und die dazu korrespondierenden Kommunalitäten.

3.4 Die Schätzung der Kommunalitäten

Eine schwierige Frage ist, wie die Kommunalitäten h_j^2 geschätzt werden können. Berechnet man die Matrix $Z'Z$, so stehen in der Diagonale ja nur Einsen, weil einfach eine Messwertreihe mit sich selbst korreliert wird. Andererseits ist die Kommunalität aber der Anteil der Korrelation, der nur durch die gemeinsamen Faktoren zustande kommt. Man hat verschiedene Schätzungen vorgeschlagen:

1. **Quadrate der multiplen Korrelation:** Dem Fundamentaltheorem der FA zufolge gilt ja

$$\Sigma = AA' + \Psi.$$

Die Diagonalzellen von AA' enthalten dann den Anteil von r_{jj} , der auf die gemeinsamen Faktoren zurückgeht. Man kann nun argumentieren (Guttman, 1956), dass die Vorhersage z.B. der Variable X_j durch die restlichen Variablen X_k , $k \neq j$ ebenfalls nur durch die Existenz gemeinsamer Faktoren möglich ist. Setzt man also eine multiple Regression der Form

$$X_j = b_{0j} + b_{1j}X_1 + \dots + b_{nj}X_n + e_j \quad (267)$$

an, wobei auf der rechten Seite alle Variablen ausser X_j stehen, so kann man hierfür einen multiplen Korrelationskoeffizienten R_j bestimmen, der die Güte der Vorhersage von x_j durch die übrigen Variablen angibt, und die wiederum geht auf die gemeinsamen Faktoren zurück. Also kann man diesen multiplen Korrelationskoeffizienten bzw. dessen Quadrat R_j^2 als Schätzung der Kommunalität einsetzen.

2. **Iterative Schätzung:** Man kann sich den Kommunalitäten auch iterativ nähern. Dazu geht man von der Matrix $Z'Z$ aus, in deren Diagonale Einsen stehen. Man führt eine Hauptachsenstransformation durch und bestimmt anhand des Scree-Tests eine plausible Schätzung der Anzahl der Faktoren. Aus diesen werden dann die ersten Schätzungen der Kommunalitäten bestimmt. Diese werden dann an die Stelle der Einsen in der Diagonale von $R = Z'Z$ eingesetzt, wodurch die Matrix R_1 entsteht. Man faktorisiert dann R_1 . Man nimmt wieder eine Schätzung der Faktoren vor, errechnet daraus eine neue Schätzung der Kommunalität, bildet die Matrix R_2 , etc. Man kann dann hoffen, dass die Faktorschätzungen und damit die geschätzten Kommunalitäten gegen bestimmte Werte konvergieren, und diese Werte sind dann die endgültigen Schätzungen. Häufig erfolgt die Konvergenz sehr schnell, d.h. nach nur wenigen Iterationen. Man kann dieses Verfahren mit dem der multiplen Korrelation kombinieren, indem man als Startwerte die multiplen Korrelationen einsetzt.
3. **Die größte Korrelation:** Eine andere Schätzung besteht darin, dass man einfach die größte Korrelation, die eine Variable mit den anderen aufweist, als erste Schätzung wählt.

Zur Problematik der Kommunalitätsschätzungen: Die Schätzung von Kommunalitäten ist mit gewissen Willkürlichkeiten verbunden; es gibt kein Verfahren, von dem man (etwa aufgrund eines mathematischen Beweises) sagen könnte, es sei das beste. Demnach ergibt sich die Frage, welchen Gewinn an Einsicht in die Struktur der Daten man tatsächlich aufgrund der Kommunalitätsschätzung hat. So hat Kalveram (1970) darauf hingewiesen, dass man zwar die empirischen Korrelationen in der Matrix $Z'Z$ durch die Ladungsmatrix A gewissermaßen vorhersagen kann (als Test der Hypothese, dass A mögliche Ladungen enthält), dass aber diese Vorhersage durch verschiedene Matrizen A erfolgen kann, d.h. A ist nicht eindeutig bestimmt. Verschiedene Matrizen A führen aber zu verschiedenen Schätzungen der Kommunalitäten. Hat man überdies relativ viele Variablen X_1, \dots, X_n , so geht der multiple Regressionskoeffizient bei der Vorhersage von X_j aufgrund von (267) schnell gegen 1, also gegen den Wert, den man sowieso in $Z'Z$ in den Diagonalzellen stehen hat. Die Schätzungen der Faktorladungen werden dann kaum noch von den Schätzungen der Kommunalitäten beeinflusst. Diese Situation hat man insbesondere bei der Analyse von Fragebögen, bei denen man schnell mehr als 20 Items hat.

Ein weiteres Problem bei der Schätzung der Kommunalitäten ergibt sich dadurch, dass die veränderten Matrizen R_1, R_2 etc nicht mehr positiv-definit sind, d.h. man errechnet für sie Eigenwerte, die kleiner als Null sind. Da die Eigenwerte aber Varianzanteile repräsentieren, sagt man auf diese Weise negative Varianzen voraus, die es aber, da Varianzen ja Summen von Quadraten sind, gar nicht geben darf.

Ein möglicher Ausweg ist die im folgenden besprochene Image-analyse.

3.5 Image-Analyse

Guttman (1953) schlug vor, die Daten, d.h. die Matrix Z zu zerlegen:

$$Z = G + E, \quad (268)$$

wobei G das *Image*, und E das *Antiimage* von Z ist. Intuitiv gesprochen ist das Image derjenige Teil der Daten, die durch die gemeinsamen Faktoren erklärt werden. Dementsprechend ist $G'G$ der durch die gemeinsamen Faktoren erklärte Teil der Korrelationen, so dass

$$G'G \cong R_H = R - \Psi. \quad (269)$$

Die Faktoren werden dann durch die Eigenvektoren und Eigenwerte von $G'G$ bestimmt, und die Anzahl der zu berücksichtigenden Faktoren wird wieder über die Eigenwerte bestimmt. Die Image-Analyse stellt eine Approximation der Hauptfaktorenanalyse dar. Es gilt dann

$$G'G \approx AA'. \quad (270)$$

Die aus dem Image berechneten Hauptkomponenten heißen *Imagefaktoren*.

Definition der Images: Da die Variablen miteinander korrelieren, kann man jede Variable durch die anderen über die multiple Regression "vorhersagen",

$$\vec{Z}_j = \sum_{k=1, k \neq j}^n \beta_{jk} \vec{Z}_k + \vec{E}_j, \quad \vec{E}_j' \vec{E}_j = \min. \quad (271)$$

Hier ist \vec{E}_j der übliche Fehlervektor, und die Forderung $\vec{E}_j' \vec{E}_j = \min$ besagt, dass die β_k so gewählt werden sollen, dass die Varianz der Fehler minimal sein soll; dies ist das übliche Kleinste-Quadrate-Kriterium. Das *Image* der Variablen V_j ist dann

$$G_j = \sum_{k=1, k \neq j}^n \beta_{jk} \vec{Z}_k. \quad (272)$$

G_j ist also der aufgrund der restlichen Variablen vorausgesagte Teil von \vec{Z}_j . Die Details der Herleitung der β -Parameter wird hier übergangen, man findet jedenfalls

$$G = Z(I - R^{-1}D^{-1}), \quad D^{-1} = \text{diag}(R^{-1}) \quad (273)$$

$$E = Z - G \quad (274)$$

Für die Korrelationsmatrix R ergibt sich die Zerlegung

$$R = G'G - E'E + 2D^{-1}. \quad (275)$$

Guttman hat gezeigt, dass sich die Images den gemeinsamen Faktoren annähern. Man kommt gewissermaßen automatisch zu "richtigen" Kommunalitäten.

3.6 Faktorentransformationen

Die Ladungsmatrix ist nicht eindeutig; ist T eine geeignet dimensionierte, orthonomale Matrix, d.h. gilt $T'T = TT' = I$, so kann man

$$\vec{x} - \vec{\mu} = A\vec{F} + \vec{e} = \underbrace{ATT'}_I \vec{F} + \vec{e} = \underbrace{AT}_{A^*} \underbrace{T'L}_{L^*} + \vec{e} \quad (276)$$

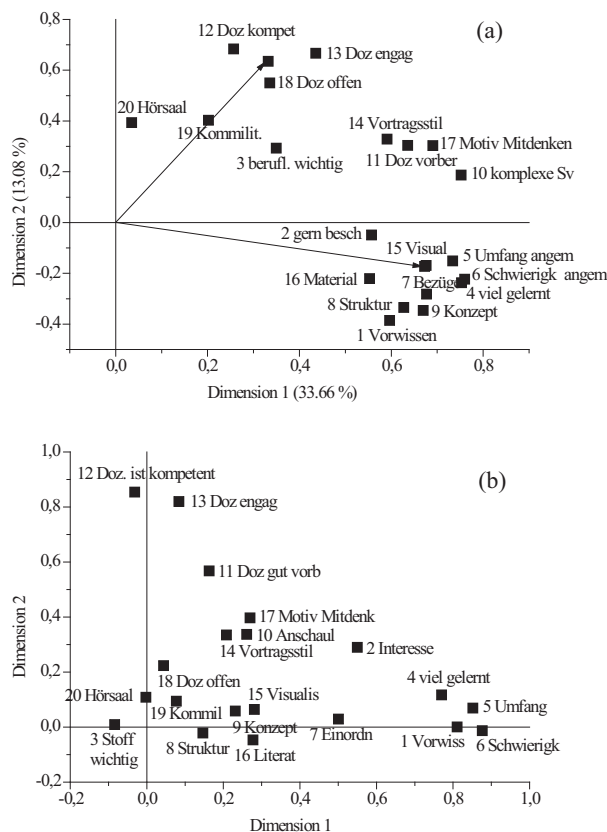
schreiben, also

$$A^* = AT, \quad L^* = T'\vec{F} \quad (277)$$

ebenfalls Lösungen darstellen. Man transformiert also nicht nur die Faktorenladungen, sondern korrespondierend dazu auch die Faktorscores. Die Transformation ist i.a. eine Rotation der Koordinatenachsen, bzw. die entsprechende Rotation der Punktekonfiguration.

Die Frage ist, nach welchen Kriterien eine Rotation ausgeführt werden soll. Die erste Möglichkeit ist, die Rotation einfach graphisch auszuführen. Ein Beispiel wird in Abbildung 16 gegeben. Die Punkte repräsentieren Fragen zur

Abbildung 16: Hauptachsentransformation der Fragen, (a) unrotiert, (b) Varimaxtransformation



Evaluation einer Lehrveranstaltung (Statistik III). Für die Punkte 12, 13 und 18 wurde der Schwerpunkt SP berechnet, und der Vektor vom Ursprung des Koordinatensystems zum Punkt SP könnte die Richtung der neuen "y"-Achse sein. Die zweite neue Achse könnte nach Maßgabe des Vektors, der durch den Schwerpunkt der Punkte im unteren, rechten "Kasten" geht, gewählt werden. Diese Punkte repräsentieren Fragen, die sich auf die Befindlichkeit der Studierenden beziehen, während die ersteren sich auf Eigenschaften des Dozenten beziehen. Die neuen Achsen sind dann nicht orthogonal, d.h. sie repräsentieren keine unabhängigen Dimensionen. Man kann die folgenden Betrachtungen anstellen:

1. Man behält die ursprünglichen, durch die Anwendung der SVD gewonnenen Dimensionen bei und interpretiert die Vektoren zu den beiden Schwerpunkten als Repräsentation der jeweiligen Item-"Cluster": die Beurteilung des Dozenten erweist sich als nicht unabhängig von der durchschnittlichen Befindlichkeit der Studierenden, die durch den Vektor zum Schwerpunkt des Befindlichkeits-"Clusters" repräsentiert wird. Studie-

rende, die die Schwierigkeit und den Umfang des Stoffes für angemessen halten, die etwas gelernt zu haben meinen und Bezüge zu anderen Lerninhalten sehen, tendieren dazu, den Dozenten für offen, kompetent und engagiert zu halten; Studierende, die die Schwierigkeit für zu groß und den Stoffumfang als nicht angemessen halten, sehen den Dozenten eher in einem anderen Licht, sie sehen ihn als weniger kompetent, weniger offen und weniger engagiert: der Dozent ist verantwortlich für die Schwierigkeiten und all das Ungemach, das die Vorlesung bereitet. Die latenten Achsen selbst sind aber nicht ganz durch die Befindlichkeit der Studierenden und der Offenheit etc des Dozenten bestimmt; die Frage ist, welche Eigenschaften sie repräsentieren.

2. Das Interpretationsproblem läßt sich vielleicht vereinfachen, wenn man die Koordinatenachsen rotiert. So könnte man die erste Achse so rotieren, dass sie durch den Schwerpunkt der Studierendenbefindlichkeit geht, und könnte dann dieses Cluster zur Interpretation der ersten Achse heranziehen. Die zweite Achse würde entsprechend mit rotiert und würde mehr durch die Dozenteneigenschaften definiert, - wenn auch jetzt die zweite Dimension noch nicht vollständig die Dozenteneigenschaften "offen", "kompetent" und "engagiert" repräsentieren würde.
3. Man führt eine Varimax-Rotation durch; das Resultat findet man in Abb. 16, (b). Das Verfahren wird weiter unten besprochen; klar ist allerdings bereits, dass nun die Dozenteneigenschaften und die Studierendenbefindlichkeiten in der Tat orthogonale, also unkorrelierte Dimensionen repräsentieren.

Allgemein kann man zwischen orthogonalen und obliquen Transformationen unterscheiden; die in bzug auf Abb. 16, (a) vorgeschlagene Rotation, also die Wahl der Vektoren, die durch die Schwerpunkte gehen, wäre ein Beispiel für eine oblique Transformation. In jedem Fall rotiert man die Dimensionen paarweise, etwa die Dimensionen D_1 und D_2 , eine eventuell existierende dritte Dimension bleibt davon unberührt, da sie zunächst senkrecht auf der $D_1 \times D_2$ -Ebene steht.

Orthogonale Rotationen: Für die Rotation um einen Winkel α ergibt sich die Transformationsmatrix

$$T = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix}. \quad (278)$$

Fasst man die beiden ersten Ladungsvektoren aus A , \vec{A}_1 und \vec{A}_2 , zu einer Matrix $\vec{A} = [\vec{A}_1, \vec{A}_2]$ zusammen, so ergibt sich

$$\vec{A}^* = \vec{A}T. \quad (279)$$

Die Komponenten der Spalten von \vec{A}^* enthalten die Koordinaten aller Punkte auf den um den Winkel α rotierten ursprünglichen Achsen.

Oblique Rotationen: Möchte man die ursprüngliche D_1 -Achse um einen Winkel α , und die D_2 -Achse um einen Winkel β rotieren, so muß man die Transformationsmatrix

$$T = \begin{pmatrix} \cos \alpha & -\sin \beta \\ \sin \alpha & \cos \beta \end{pmatrix}^{-1} \quad (280)$$

wählen. Die neue Ladungsmatrix errechnet sich dann wie in (279). Die Winkel α und β kann man frei wählen; zum Beispiel kann man diese Winkel für die beiden Vektoren in Abb. 16 bestimmen.

Eine Alternative zur Wahl bestimmter α und, im Falle obliquer Rotationen, β kann man ein globales Kriterium spezifizieren, demzufolge die Rotationswinkel rechnerisch bestimmt werden. Die im folgenden Abschnitt beschriebene Methode bezieht sich auf orthogonale Rotationen.

Varimax-Rotationen: Rotationen dieser Art (Kaiser, 1958) versuchen, orthogonale Rotationen der Achsen derart zu bestimmen, dass einige Variablen auf einer Dimension hoch und auf allen übrigen Dimensionen niedrig laden; dies ist das Kriterium der Einfachstruktur (simple structure, Thurstone (1945)). Das Verfahren bezieht sich auf die Spalten von A . Formal läßt sich das Kriterium so fassen: Die Varianz der Ladungsquadrate soll maximiert werden. Diese Formulierung erklärt den Namen Varimax-Rotation. Allgemein sei $A^* = AT$ die Matrix der Ladungen auf den rotierten Achsen. Es sei

$$\bar{a}_k = \frac{1}{n} \sum_{j=1}^n \tilde{a}_{jk} \quad (281)$$

die mittlere Ladung auf der k -ten transformierten Achse. Dann sei

$$V_1 = \sum_{k=1}^s \sum_{j=1}^n (\tilde{a}_{jk}^2 - \bar{a}_k)^2. \quad (282)$$

Die \tilde{a}_{jk} sollen so bestimmt werden, dass $V_1 = \max$. Die Maximierung erfolgt über alle orthogonalen Transformationsmatrizen T . Das Verfahren zur tatsächlichen Bestimmung der \tilde{a}_{jk} ist numerisch, d.h. iterativ, und ist in den gängigen Statistikpaketen implementiert.

Die Koordinatenachsen in Abb. 16, (b) sind durch Varimax-Rotation bestimmt worden. Ganz offenbar wird die Dimension 1 nun durch die Fragen 1 (Vorwissen), 4 (Ich habe viel gelernt) und 5 (angemessener Umfang des Stoffes) und 6 (angemessene Schwierigkeit des Stoffes) definiert. Die zweite Dimension wird in erster Linie durch die Frage 12 (Der Dozent erscheint auf seinem Themengebiet kompetent) charakterisiert, und durch 13 (Das Engagement des Dozenten ist stets deutlich). Die Kompetenz des Dozenten ist demnach ein Faktor, der unabhängig vom Vorwissen und der erfahrenen Schwierigkeit des Stoffes wirkt. Die hohe Korrelation zwischen der Frage 4 (viel gelernt) und

den Fragen 1, 5, und 6 zeit, dass diejenigen meinen, viel gelernt zu haben, wenn sie bereits über ein hinreichendes Vorwissen verfügen und dementsprechend die Schwierigkeit und den Umfang des Stoffes für angemessen halten. Für Personen ohne dieses Vorwissen ist die Schwierigkeit und der Umfang des Stoffes zu groß und es stellt sich der Eindruck, nicht viel gelernt zu haben, ein (nicht viel gelernt zu haben kann auch bedeuten, das Gefühl zu haben, der Klausur nicht gewachsen zu sein). Die Merkmale Vortragsstil, Fähigkeit zur anschaulichen Darstellung komplexer Zusammenhänge (10) und Motivation zum Mitdenken (17) erscheinen als Mischungen aus der Kompetenz des Dozenten und dem Faktor 1 (Vorwissen etc), ebenso das Merkmal 2 (Interesse: ich beschäftige mich gern mit dem Stoff), d.h. der Dozent kann dazu beitragen, dass sich Studierende sich (relativ) gern mit dem Stoff beschäftigen. Andere Merkmale (16 (Material wie Literatur und Skripten), 18 (Dozent ist offen für Fragen), 19 (Verhalten der Kommilitonen ist diszipliniert etc), 8 (Veranstalt. ist strukturiert) und 9 (Veranstaltung hat klares Konzept)) werden kaum durch diese Faktoren erklärt; die Analyse zeigt, dass diese Merkmale am besten durch jeweils spezifische Faktoren erklärt werden können.

Quartimax-Rotationen: Dieses Verfahren ist formal wie das Varimax-Verfahren definiert, bezieht sich aber auf die Zeilen von A . Es sei sic!

$$\bar{a}_i = \frac{1}{m} \sum_{u=1}^m \tilde{a}_{ik}. \quad (283)$$

Dann wird

$$V_2 = \sum_i \sum_r (\tilde{a}_{ir}^2 - \bar{a}_i)^2 \quad (284)$$

maximiert, $V_2 = \max$.

Promax-Rotationen: Es sei T eine orthogonale Varimax-Rotation. Das Ergebnis kann gelegentlich im Sinne der Einfachstruktur verbessert werden, in dem eine nachfolgende oblique Rotation durchgeführt wird (Hendrickson, 1964). Man geht also von $A^* = AT$ aus und sucht jetzt eine Transformation \tilde{T} , so dass

$$\tilde{A} = A^* \tilde{T} \quad (285)$$

dem Kriterium der Einfachstruktur noch besser entspricht, d.h. dass es noch mehr Ladungen gibt, deren Quadrate entweder nahe bei 1 oder bei 0 liegen. Auf die Details soll hier nicht eingegangen werden, vgl. etwa Brachinger und Ost (1996), p. 686.

4 Typen von Analysen

Catell (1952) hat eine Reihe von Typen von Analysen vorgestellt, auf die kurz eingegangen werden soll.

1. *R-Analyse*: Um die Idee der Faktorenanalyse bzw. der Hauptachsenanalyse zu fixieren, wurde für die Ausgangs- oder Datenmatrix Z angenommen, daß die Spalten oder Spaltenvektoren Z_j "Tests", die U_j also Testscores, repräsentieren, die Komponenten dieser Vektoren also die Meßwerte oder Scores der Vpn in einem gegebenen Test sind. Die Zeilen von Z enthalten die Scores oder Meßwerte der i -ten Versuchsperson in allen Tests. Die Matrix $R = Z'Z$ enthält dann die Korrelationen zwischen den Tests, oder zwischen den U_j . Die Analyse dieser Matrix zielt zunächst auf eine Repräsentation der Tests oder U_j in einem durch latente Dimensionen definierten Raum. Cattell (1952) nannte diese Art der Analyse *R-Analyse*.
2. *Q-Analyse*: Standardisiert man die Meßwerte nicht durch Mittelung über die Vpn, sondern über die Tests, so kann man die Korrelationen zwischen den Personen bestimmen; dem entspricht die Matrix ZZ' , wobei Z die in der eben genannten Weise erzeugte Matrix von Standardwerten ist. Die Analyse zielt nun primär auf *Typen* von Personen. Cattell (1952) spricht von *Q-Analyse*.
3. *O-Analyse*: Bei der R- und der Q-Analyse tritt die Zeit nicht explizit auf. Die Meßerte x_{ij} und ihre Standardwerte z_{ij} repräsentieren Größen, die von der Zeit zwar abhängen können, aber bei denen der zeitliche Verlauf nicht im Vordergrund des Interesses steht. Die Entwicklung von Merkmalen in der Zeit ist aber generell durchaus von Interesse. Die Faktorenbzw. Hauptachsenanalyse kann hier nützlich sein.

So kann man bei *einer* Person mehrere Variable U_j zu verschiedenen Zeitpunkten messen und dann Korrelationen zwischen den Zeitpunkten berechnen; dabei wird über die Variablen gemittelt. Cattell (1952) spricht dann von einer *O-Analyse*. Die Zeitpunkte spielen hier die Rolle der "Tests" bei der R-Analyse. Ob es sinnvoll ist, eine solche Analyse zu rechnen, wird von der Auswahl der Variablen abhängen; vermutlich sollten sie einen ähnlichen zeitlichen Verlauf haben. Die latenten Variablen repräsentieren dann jedenfalls Zeitpunkte, die unabhängig voneinander eine Variation in den Merkmalen erzeugen. So analysierten Evardsson und Vegelius (1981) Ratings der Stimmung in verschiedenen Situationen und dementsprechend zu verschiedenen Zeitpunkten; die Schlußfolgerung der Autoren ist, daß Personen ihre Zeit in bezug auf Stimmungen strukturieren.
4. *P-Analyse*: Andererseits kann man die Kovariation von Merkmalen über die Zeit untersuchen, entweder bei einer Person, oder man betrachtet die über Personen gemittelten Verläufe in den Variablen (diese Mittelung setzt natürlich voraus, daß die verschiedenen Personen im wesentlichen gleiche Verläufe haben). Diese Art der Analyse heißt nach Cattell (1952) *P-Analyse*. Die P-Analyse entspricht der multivariaten Zeitreihenanalyse, vorausgesetzt, daß die Messungen zu gleichabständigen Zeitpunkten

vorgenommen werden. Außerhalb der Psychologie ist diese Art der Analyse auch als Karhunen-Loève-Analyse bekannt (Papoulis (1968)). Die latenten Variablen sind hier in der Zeit unabhängig voneinander verlaufende Variablen, aus denen sich die beobachteten, über die Zeit kovariierenden Variablen durch Linearkombination "erklären" lassen. So könnte es unabhängig voneinander verlaufende Biorythmen geben, aus denen sich beobachtete Stimmungsverläufe linear vorhersagen lassen.

Die P-Analyse ist keineswegs problemlos anzuwenden. Holtzman (1962) wies wohl als erster auf die Problematik der *wiederholten Messungen* (mehrfache Messungen bei den gleichen Personen) bei dieser Art der Analyse hin. Die Kritik ist noch ausführlicher von Anderson (1963) formuliert worden. Molenaar (1985) hat den Ansatz in neuer Form wieder aufgenommen. Die latenten Faktoren sind hier zufällige Funktionen der Zeit, aus denen sich die beobachteten Zeitreihen zur Zeit t als gewogene Summe der Werte dieser Funktionen zu den Zeiten $t - 1, t - 2, \dots$ ergeben. Eine detaillierte Beschreibung der Molenaarschen Arbeit überschreitet den hier gegebenen Rahmen.

5. *S-Analyse*: Schließlich läßt sich noch die Kovariation von Personen über die Zeit analysieren. Cattell (1952) spricht dann von *S-Analyse*. So könnte es Typen von Personen geben, die durch spezielle Verläufe der eben genannten Biorythmen charakterisiert werden können. Man kann Fragestellungen angehen wie etwa die, ob sich vielleicht die psychisch angeleglich pastösen Athletiker von den psychisch mobileren Zyklothymen unterscheiden, wodurch auch immer.

5 Dichotome Variable und nichtlineare Faktorenanalyse

5.1 Dichtome Variable I

Eine unmittelbare Anwendung der Faktorenanalyse auf dichotome Daten ergibt sich, wenn man von den Korrelationen zwischen den Items ausgeht. Es gibt zwei Möglichkeiten: man geht vom ϕ -Koeffizienten oder vom tetrachorischen Koeffizienten aus. Die damit verbundenen Probleme werden im Folgenden vorgestellt.

Der ϕ -Koeffizient: Es seien X und Y dichotome Variable, d.h. es gelte $X = \{0, 1\}$ und $Y = \{0, 1\}$. Es werden N Messungen gemacht, d.h. es wird bei N Probanden der Wert sowohl von X als auch von Y bestimmt. Gesucht ist die Korrelation zwischen X und Y . Das Ergebnis der Messungen kann in einer

Vierfeldertafel zusammengefasst werden:

		Y		Σ	
		1	0		
X	1	a	b	a + b	(286)
	0	c	d	c + d	
Σ		a + c	b + d	N	

Wendet man hierauf die gewöhnliche Formel für den Produkt-Moment-Korrelationskoeffizienten an, so ergibt sich nach einigen Vereinfachungen der Ausdruck

$$r_{xy} = \phi_{xy} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}. \quad (287)$$

Das Vierfelder- χ^2 ist

$$\phi_{xy} = \sqrt{\frac{\chi^2}{N}}, \quad (288)$$

so dass der Determinationskoeffizient durch

$$\phi_{xy}^2 = \frac{\chi^2}{N}, \quad (289)$$

gegeben ist.

Für den Produkt-Moment-Korrelationskoeffizienten gilt generell $-1 \leq r_{xy} \leq 1$; diese Eigenschaft überträgt sich natürlich auf den ϕ -Koeffizienten. Andererseits läßt sich zeigen, dass der tatsächliche Wertebereich des ϕ -Koeffizienten auf ein Teilintervall von $[-1, 1]$ beschränkt sein kann. Dieser Sachverhalt soll kurz elaboriert werden, da er u.a. bei Anwendungen in der Faktorenanalyse von Bedeutung sein kann.

Es werde zunächst der Fall betrachtet, dass der Zusammenhang zwischen X und Y perfekt ist, so dass $\phi_{xy} = 1$; der Fall $\phi_{xy} = -1$ wird auf analoge Weise behandelt. Der Zusammenhang ist perfekt, wenn nur Paare $(X = 1, Y = 1)$ und $(X = 0, Y = 0)$ vorkommen; dann gilt $b = c = 0$ und (287) impliziert, dass $\phi_{xy} = 1$. Kommen nur die Kombinationen $(X = 1, Y = 0)$ und $(X = 0, Y = 1)$ vor, so ist $a = d = 0$ und $\phi_{xy} = -1$.

Aus $b = c = 0$ folgt für die Randsummen $a+b = a+c = a$ und $c+d = b+d = d$, d.h. die Randverteilungen für X und Y sind identisch. Dann aber folgt, dass $a+b \neq a+d$ und $c+d \neq b+d$ implizieren, dass $b \neq c$ und damit $\phi_{xy} < 1$. Die Frage ist nun, wie groß ϕ_{xy} überhaupt werden kann, wenn $b \neq c$ gilt. Um diese Frage zu diskutieren, ist es sinnvoll, in (287) die a, b, c und d durch die dazu korrespondierenden Wahrscheinlichkeiten (relative Häufigkeiten) $p_{11} = a/N$,

$p_{10} = b/N$ etc zu ersetzen, mit $N = a + b + c + d$. Man erhält

		Y		Σ
		1	0	
X	1	p_{11}	p_{10}	p_x
	0	p_{01}	p_{00}	$1 - p_x$
Σ		p_y	$1 - p_y$	1

(290)

Dann ist $a = Np_{11}$, $b = Np_{10}$, etc. In dieser Notation nimmt die Formel für ϕ_{xy} die Form

$$\phi_{xy} = \frac{N^2(p_{11}p_{00} - p_{10}p_{01})}{\sqrt{N^4 p_x(1-p_x)p_y(1-p_y)}} = \frac{p_{11}p_{00} - p_{10}p_{01}}{\sqrt{p_x(1-p_x)p_y(1-p_y)}} \quad (291)$$

an. Da X und Y dichotom sind, ist die Anzahl der Antworten $X = 1$ bzw. $Y = 1$ jeweils binomialverteilt, mit den Erwartungswerten¹⁷ Np_x und Np_y und den Varianzen $\mathbb{V}(X) = Np_x(1-p_x)$ und $\mathbb{V}(Y) = Np_y(1-p_y)$. Dementsprechend steht im Nenner des Ausdrucks für ϕ_{xy} auch das Produkt $p_x(1-p_x)p_y(1-p_y)$ (N kürzt sich ja heraus), denn in der allgemeinen Korrelationsformel $\rho_{xy} = \mathcal{K}(\mathcal{X}, \mathcal{Y})/\sigma_{\mathcal{X}}\sigma_{\mathcal{Y}}$ steht eben auch die Wurzel aus dem Produkt der Varianzen von X und Y . Der oben betrachtete Fall $\phi_{xy} = 1$ implizierte gleiche Randverteilungen und damit gleiche Varianzen für X und Y . Sind diese Varianzen ungleich, kann ϕ_{xy} nicht mehr den Wert $+1$ oder -1 annehmen. Hier unterscheidet sich der ϕ -Koeffizient vom allgemeinen Produkt-Moment-Korrelationskoeffizienten, obwohl er nur ein Spezialfall des Letzteren ist. Denn es sei $Y = \alpha X + \beta$, d.h. Y sei eine lineare Transformation von X . Da diese Transformation nicht den üblichen Fehlerterm ε enthält, muß die Korrelation zwischen X und Y gleich 1 sein, obwohl die Varianzen verschieden sind, denn es ist ja $\mathbb{V}(Y) = \alpha^2\mathbb{V}(X)$, d.h. nur für den Spezialfall $\alpha = 1$ sind auch die Varianzen identisch. Jedenfalls findet man

$$\rho_{xy} = \frac{\mathcal{K}(\mathcal{X}, \mathcal{Y})}{\sigma_x\sigma_y} = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{\sigma_x\sigma_y} = \frac{\mathbb{E}(\alpha X^2 + \beta X) - \mathbb{E}(X)(\alpha\mathbb{E}(X) + \beta)}{\sigma_x\sigma_y},$$

d.h.

$$\rho_{xy} = \frac{\alpha\mathbb{E}(X^2) + \beta\mathbb{E}(X) - \alpha\mathbb{E}^2(X) - \beta\mathbb{E}(X)}{\alpha\sigma_x^2} = \frac{\alpha\sigma_x^2}{\alpha\sigma_x^2} = 1.$$

Die X - und Y -Werte haben eine unterschiedliche Varianz, allerdings kürzt sich der unterscheidende Skalenfaktor heraus. Beim ϕ -Koeffizienten werden aber Häufigkeiten betrachtet, die auf einer Absolutskala definiert sind und also nicht mehr transformiert werden dürfen, d.h. es darf nur der Spezialfall $\alpha = 1$ und $\beta = 0$ betrachtet werden.

Der Fall $\mathbb{V}(X) = \mathbb{V}(Y)$ ist sicherlich ein Spezialfall, der höchst selten eintritt. Da aber $\mathbb{V}(X) \neq \mathbb{V}(Y)$ bereits $|\phi_{xy}| < 1$ impliziert, ohne dass etwas über

¹⁷Der Einfachheit halber ist hier die Unterscheidung von Erwartungswert und Stichprobenmittelwert unterlassen worden.

den tatsächlichen Zusammenhang ausgesagt wird, ist es von Interesse, den maximal möglichen Wert von $|\phi_{xy}|$ bei ungleichen Varianzen zu bestimmen. Dazu wird der folgende Satz bewiesen:

Satz 5.1 Die p_{ij} , p_x und p_y seien wie in (290) definiert, und es gelte $\mathbb{V}(X) \neq \mathbb{V}(Y)$. Der maximal mögliche ϕ_{xy} -Wert ist durch

$$\phi_{xy}^{\max} = \sqrt{\frac{p_x(1-p_y)}{p_y(1-p_x)}} < 1 \quad (292)$$

gegeben.

Folgerung: Für den Spezialfall $p_x = p_y$ und damit $\mathbb{V}(X) = \mathbb{V}(Y)$ folgt aus (292) sofort $\phi_{xy}^{\max} = \pm 1$.

Beweis: Aus (291) folgt

$$\phi_{xy}^2 = \frac{(p_{11}p_{00} - p_{10}p_{01})^2}{p_x(1-p_x)p_y(1-p_y)}$$

Nach Voraussetzung ist $p_x(1-p_x) \neq p_y(1-p_y)$. ϕ_{xy}^2 wird maximal, wenn die Differenz $p_{11}p_{00} - p_{10}p_{01}$ maximal wird. Für $p_{11} \neq 0$ und $p_{00} \neq 0$ wird die Differenz maximal, wenn $p_{10} = 0$ oder $p_{01} = 0$ ist (der Fall $p_{10} = p_{01} = 0$ entspricht dem oben bereits behandelten Fall $b = c = 0$, der gleiche Varianzen impliziert und deswegen der Voraussetzung ungleicher Varianzen nicht entspricht). Jedenfalls gilt dann

$$\phi_{xy}^2 = \frac{(p_{11}p_{00})^2}{p_x(1-p_x)p_y(1-p_y)}.$$

Es gelte insbesondere $p_{10} = 0$. Wegen $p_{11} + p_{10} = p_x$ folgt nun $p_{11} = p_x$, und ebenso folgt $p_{00} = 1 - p_y$, so dass $p_{11}p_{00} = p_x(1 - p_y)$. Dementsprechend hat man

$$\phi_{xy}^2 = \frac{(p_x(1-p_y))^2}{p_x(1-p_x)p_y(1-p_y)} = \frac{p_x(1-p_y)}{p_y(1-p_x)},$$

woraus (292) sofort folgt. Der Fall $p_{01} = 0$ führt zum gleichen Resultat. Die Betrachtung für ϕ_{xy}^{\min} ist analog. \square

Je kleiner die Korrelationskoeffizienten, desto geringer der statistische Zusammenhang zwischen den Items, desto mehr latente Dimensionen zeigt eine Faktorenanalyse an. So könnte man plausiblerweise argumentieren. In der Tat wird oft der Begriff der *Schwierigkeitsfaktoren* in diesem Zusammenhang eingeführt: dies sind Faktoren, die nur die unterschiedliche Schwierigkeit dichotomer Items reflektieren, nicht aber qualitativ verschiedene latente Dimensionen. Solche Faktoren wären also zu erwarten, wenn die Randhäufigkeiten der 4-Felder-Tafeln, die zur Berechnung des ϕ -Koeffizienten aufgestellt werden, für die jeweiligen Items nicht identisch sind. Man macht sich leicht klar,

dass die Identität der Randverteilungen für irgendzwei Items zusammen mit der Forderung, dass die Identität für alle Paare gelten soll, bedeutet, dass alle Items dann die gleiche Schwierigkeit haben müssen, – eine Forderung, die in aller Strenge schwer zu erfüllen sein dürfte. McDonald & Ahlawat (1974) haben allerdings nachgewiesen, dass das Problem der Schwierigkeitsfaktoren nicht überbewertet werden sollte: sind die Unterschiede zwischen den Schwierigkeiten (und damit den Randverteilungen) nicht allzu drastisch und sind die Regressionen zwischen den Items linear, so sind keine Schwierigkeitsfaktoren zu erwarten!

Der tetrachorische Koeffizient: Gegeben seien die zufälligen Veränderlichen X_1 und X_2 , die bivariat normalverteilt seien, d.h. die gemeinsame Dichtefunktion sei durch

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} \left(\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} \right) \right]. \quad (293)$$

Dabei sind μ_1 und μ_2 die Erwartungswerte von X_1 bzw. X_2 , und σ_1^2 und σ_2^2 sind die Varianzen von X_1 und X_2 . ρ ist die Korrelation zwischen den beiden Variablen.

Nun werde angenommen, dass X_1 und X_2 nicht direkt beobachtet werden können, sondern nur Dichotomisierungen: es seien

$$Y_1 = \begin{cases} 0, & X_1 \leq \gamma_1 \\ 1, & X_1 > \gamma_1 \end{cases}, \quad Y_2 = \begin{cases} 0, & X_2 \leq \gamma_2 \\ 1, & X_2 > \gamma_2 \end{cases} \quad (294)$$

Weiter sei

$$\pi_1 = P(Y_1 = 1) = P(X_1 > \gamma_1) = \int_{\gamma_1}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 dx_2 \quad (295)$$

$$\pi_2 = P(Y_2 = 1) = P(X_2 > \gamma_2) = \int_{-\infty}^{\infty} \int_{\gamma_2}^{\infty} f(x_1, x_2) dx_1 dx_2. \quad (296)$$

Weiter kann man die Wahrscheinlichkeit bestimmen, mit der sowohl $X_1 > \gamma_1$ als auch $X_2 > \gamma_2$ gilt; diese ist durch

$$\pi_{12} = P(X_1 > \gamma_1 \cap X_2 > \gamma_2) = \int_{\gamma_1}^{\infty} \int_{\gamma_2}^{\infty} f(x_1, x_2) dx_1 dx_2 \quad (297)$$

gegeben.

$$\bar{\pi}_{12} = P(X_1 \leq \gamma_1 \cap X_2 \leq \gamma_2) = \int_{-\infty}^{\gamma_1} \int_{-\infty}^{\gamma_2} f(x_1, x_2) dx_1 dx_2 \quad (298)$$

Die Vierfeldertafel illustriert diese Wahrscheinlichkeiten:

	$X_2 > \gamma_2$	$X_2 \leq \gamma_2$	Σ	
$X_1 > \gamma_1$	π_{12}	$\pi_1 - \pi_{12}$	π_1	(299)
$X_1 \leq \gamma_1$	$1 - \pi_1 - \bar{\pi}_{12}$	$\bar{\pi}_{12}$	$1 - \pi_1$	
Σ	π_2	$1 - \pi_2$	1	

Die Wahrscheinlichkeiten π_1 , π_2 , π_{12} und $\bar{\pi}_{12}$ können als Anteile aus den Daten geschätzt werden. Daraus ergibt sich die Schätzung r_{12} für ρ aus den obigen Gleichungen. Natürlich können die Gleichungen für die π -Werte nicht nach ρ und den anderen Parametern aufgelöst werden; die Schätzungen müssen numerisch bestimmt werden. Die Schätzung r_{12} ist der *tetrachorische Korrelationskoeffizient*.

Die Problematik des tetrachorischen Korrelationskoeffizienten besteht in der Annahme der bivariaten Normalverteilung. Gilt diese Annahme nicht, erhält man verzerrte Korrelationskoeffizienten. Gilt diese Annahme, so kann es natürlich sein, dass verschiedene Gruppen von Personen sich durch die Erwartungswerte μ_1 und μ_2 unterscheiden. Der Korrelationskoeffizient ρ ist aber nicht von diesen beiden Parametern abhängig, so dass die Schätzungen r_{xy} für ρ *invariant* gegenüber Veränderungen der Erwartungswerte sein sollten. Carroll (1961) hat gezeigt, dass die Normalverteilung die einzige Verteilung ist, die diese Invarianz zulässt.

5.2 Dichotome Variable II

Im Falle dichotomer Items nehmen die Komponenten von \mathbf{x}_j , $j = 1, \dots, p$ nur die Werte 1 oder 0 an; beantwortet eine Person das j -te Item "positiv" (löst sie die Aufgabe oder gibt sie das Vorhandensein eines Merkmals an, oder findet man das Merkmal bei der Person), so wird $x_{ij} = 1$ gesetzt, sonst $x_{ij} = 0$. Der folgende Ansatz wurde zuerst von Christofferson (1975) vorgeschlagen.

Man kann die Messungen x_{ij} mit den Werten einer entsprechenden latenten Variablen ξ in Verbindung bringen: Ist, für die i -te Person, $\xi_i \geq \tau_j$, so löst die Person die Aufgabe bzw. beantwortet sie positiv, und für $\xi_i < \tau_j$ löst sie die j -te Aufgabe nicht bzw. beantwortet sie negativ. τ_j ist ein Schwellenparameter, der für das j -te Item charakteristisch ist.

Im Fall von Messungen auf einem Kontinuum werden die (Mess-)Werte in X über die Beziehung $X = FA'$ zu latenten Variablen in F mit den Ladungen A in Beziehung gesetzt. Insbesondere hat man für den j -ten Spaltenvektor $\mathbf{x}_j = \vec{X}_j$ (dessen Komponenten die Messungen der j -ten Variablen sind) die Gleichung

$$\mathbf{x}_j = F\mathbf{a}_j,$$

wobei \mathbf{a}_j die j -te Spalte von A' , d.h. die j -te Zeile von A ist. Ist

$$\vec{\xi}_j = \begin{pmatrix} \xi_{1j} \\ \xi_{2j} \\ \vdots \\ \xi_{mj} \end{pmatrix}$$

der j -te Vektor für die latenten Variablen mit ξ_{ij} der Wert der i -ten Person für diese Variable, so kann man

$$\vec{\xi}_j = F\mathbf{a}_j + \varepsilon_j, \quad \xi = FA' + \varepsilon \quad (300)$$

schreiben, wobei jetzt ξ die aus den Spaltenvektoren $\vec{\xi}_j$ zusammengesetzte Matrix ist, und die Zeilen von A durch die \mathbf{a}_j gegeben sind. Fasst man (ξ_1, \dots, ξ_p) als zufälligen Vektor auf (die ξ_{ij} sind dann Realisationen der Komponente ξ_j) und nimmt man an, dass er multivariat normalverteilt sei, so dass

$$f(\xi) = \frac{1}{|\Sigma|^{1/2}(2\pi)^{p/2}} \exp(e\varepsilon'\Sigma^{-1}\varepsilon) \quad (301)$$

so ist Σ durch

$$\Sigma = \xi'\xi = A\Lambda A' + \Psi \quad (302)$$

gegeben. $F'F = F'F$ ist die Matrix der Korrelationen zwischen den Faktoren, falls der oblique Fall zugelassen wird. Im Falle orthogonaler Faktoren ist Λ eine Diagonalmatrix. Da die beobachteten Messungen nur die Werte 0 oder 1 annehmen kann, folgt, dass die Diagonalwerte von Σ nicht identifizierbar sind. Man setzt dann die Diagonalelemente gleich 1 und erhält dann

$$\Psi = I - \text{diag}(A\Lambda A'). \quad (303)$$

Um das Modell an die Daten anzupassen, müssen (i), der Vektor $\tau = (\tau_1, \dots, \tau_p)'$ der Schwellen sowie (ii) die Matrizen A und F aus den Daten geschätzt werden. Um die Maximum-Likelihood-Methode anwenden zu können, muß zunächst ein Ausdruck für die Wahrscheinlichkeit der Daten gefunden werden; für den Zufallsvektor $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_p)'$ findet man

$$g(\mathbf{x}_1, \dots, \mathbf{x}_p) = \int_{\tau_1}^{\infty} \int_{-\infty}^{\tau_2} \dots \int_{\tau_p}^{-\infty} f(\mathbf{x}) d\mathbf{x} \quad (304)$$

Ist hierin τ_j die untere Grenze eines Integrals, so bedeutet dies, dass die latente Variable einen Wert größer als τ_j hat, dementsprechend $x_{ij} = 1$ ist. Ist τ_j die obere Grenze des Integrals. Dies ist der Ausdruck für eine Person; die entsprechenden Ausdrücke für alle Personen müssen dann ebenfalls miteinander multipliziert werden. Es ist klar, dass die Minimalisierung der Gesamt-Likelihood eine formidable Aufgabe ist, zumal die Mehrfachintegrale in (304) implizieren einen hohen Rechenaufwand. Christoffersons Ansatz wurde von Muthén (1978) aufgenommen, der eine verbesserte Schätzung vorschlug; auf die Details kann hier nicht eingegangen werden.

5.3 Nichtlineare Modelle

Zunächst sei daran erinnert, dass der Ansatz (??) im Kern ein multipler Regressionsansatz ist, allerdings ein spezieller. Nimmt man an, dass x_j durch $r \leq p$ latente Variablen bestimmt wird, so kann allgemein $X_j = f_j(F_1, \dots, F_r)$ gelten, wobei f_j irgendeine Funktion ist. f_j ist nicht bekannt. Aber unter sehr allgemeinen Bedingungen kann eine unbekannte Funktion durch ein geeignet gewähltes Polynom approximiert werden¹⁸. Die einfachste Approximation ist dann (??). Darüber hinaus kann man Terme hinzufügen, die durch Potenzen der F_k sowie durch Produkte $F_k F_{k'}$ etc definiert sind. Die Kunst ist dann, die zugehörigen Koeffizienten zu schätzen.

Wie Busemeyer und Jones (1983) ausführen, sind viele psychologische Gesetzmäßigkeiten durch Produkte von Variablen definiert. So ist etwa die Arbeitsmotivation dem Modell von Vroom (1964) durch das Produkt von Erwartung und Valenz (expectency \times valence) definiert, und die Performanz einer Person durch das Produkt von Fähigkeit und Motivation (ability \times motivation). Sofern das Modell korrekt ist, sollten Tests etwa der Motivation oder der Performanz die korrespondierenden Produkte von entsprechenden latenten Variablen erfassen. Die Autoren zeigen dann allerdings, dass es außerordentlich schwierig ist, Modelle dieser Art über hierarchische Regressionsmodelle zu testen: sind die Prädiktorvariablen nicht messfehlerfrei, so werden interaktive Trends unterschätzt, und die Reliabilität von Produkttermen ist eine Funktion des Produktes der Einzelreliabilitäten, – da diese Zahlen kleiner als 1 sind, ist die Gesamtreliabilität kleiner als die Einzelreliabilitäten. Hinzu kommen Probleme, die sich aus den Skalenniveaus ergeben: Produktterme können u. U. durch geeignete Transformationen in additive Terme transformiert werden.

Kenny und Judd (1984) haben aber gezeigt, dass es u. U. möglich ist, tatsächlich ist, Modelle mit nichtlinear wirkenden latenten Variablen zu schätzen. Sie illustrieren ihren Ansatz zweier Gleichungen:

$$y = a_1x + a_2x^2 + u \quad (305)$$

$$y = b_1x + b_2z + b_3xz + v \quad (306)$$

Hierin x und z Variable, die auf y wirken, und a_1 , a_2 sowie b_1 , b_2 und b_3 sind Regressionskoeffizienten, u und v sind Residuen ("Fehler"). xz repräsentiert eine interaktive Wirkung von x und z auf y . Die Frage ist nun, wie diese Koeffizienten geschätzt werden können, wenn x und z nicht direkt gemessen werden können und als latente Variable in die Größe y eingehen. Es wird angenommen, dass alle Variablen Abweichungen vom jeweiligen Mittelwert repräsentieren.

Zuerst wird die Schätzung der Parameter für den Fall der Gleichung (??) illustriert. Dazu definieren Kenny & Judd zwei "Indikatoren" x_1 und x_2 für die

¹⁸Weierstraßscher Approximationssatz

latente Variable x :

$$x_1 = x + u_1 \quad (307)$$

$$x_2 = cx + u_2 \quad (308)$$

wobei c wieder ein Parameter ist. Dann wird angenommen, dass die Residuen u_1 , u_2 und x und die latente Variable x alle unabhängig voneinander sind. Um den Effekt von x^2 schätzen zu können, müssen entsprechende Indikatoren bestimmt werden. Dazu bieten sich die Größen x_1^2 , x_2^2 und x_1x_2 an. Aus (307) und (308) erhält man

$$x_1^2 = x^2 + u_1^2 + 2xu_1 \quad (309)$$

$$x_2^2 = c^2x^2 + u_2^2 + 2cxu_2 \quad (310)$$

$$x_1x_2 = cx^2 + cxu_1 + xu_2 + u_1u_2 \quad (311)$$

Man kann nun die zu schätzenden Parameter (= Ladungen) in einer Tabelle zusammenfassen. Offenbar muß nur ein einzelner Parameter, c , geschätzt werden.

Tabelle 8: Ladungen für das nichtlineare Faktorenmodell

Variable	x	x^2	u_1	u_2	u_1^2	u_2^2	xu_1	xu_2	u_1u_2
x_1	1	0	1	0	0	0	0	0	0
x_2	c	0	0	1	0	0	0	0	0
x_1^2	0	1	0	0	1	0	2	0	0
x_2^2	0	c^2	0	0	0	1	0	$2c$	0
x_1x_2	0	c	0	0	0	0	c	1	1

Die Gesamtheit der latenten Variablen ist $x, x^2, u_1, u_2, u_1^2, u_2^2, xu_1, xu_2, u_1u_2$. Die Struktur der Kovarianzmatrix für diese Variablen hängt allerdings von Annahmen über deren Verteilung ab. Unter der Annahme, dass x , u_1 und u_2 normalverteilt sind, ergeben sich die folgenden Beziehungen

$$\begin{aligned} \sigma_{x^2}^2 &= 2\sigma_x^4, & \sigma_{u_1^2}^2 &= 2\sigma_{u_1}^4 \\ \sigma_{u_2^2}^2 &= 2\sigma_{u_2}^4, & \sigma_{xu_1}^2 &= \sigma_x^2\sigma_{u_1}^2 \\ \sigma_{xu_2}^2 &= \sigma_x^2\sigma_{u_2}^2, & \sigma_{u_1u_2}^2 &= \sigma_{u_1}\sigma_{u_2} \end{aligned}$$

Die Kovarianzen erweisen alle als gleich Null, und die Varianzen sind Funktionen von σ_x^2 , $\sigma_{u_1}^2$ und $\sigma_{u_2}^2$. Aus der Annahme der Normalverteilung für x , u_1 und u_2 folgt allerdings *nicht*, dass auch x^2 normalverteilt ist, so dass ein Maximum-Likelihood-Ansatz auf der Basis einer multivariaten Normalverteilung nicht zur Schätzung der Parameter herangezogen werden kann. Einem Vorschlag von McDonald (1978) folgend verwenden die Autoren eine Verallgemeinerte Kleinste-Quadrate-Schätzung mit der Inversen der Stichprobenkovarianzmatrix als Gewichtsmatrix (die Verallgemeinerte KQ-Schätzung wird im Anhang vorgestellt).

In analoger Weise können Interaktionen zwischen latenten Variablen diskutiert werden. Dazu werden wieder Indikatoren für die Interaktionen definiert und dann wird die dazu korrespondierende Ladungsmatrix hergeleitet. Die Matrix für die Kovarianzen zwischen den latenten Variablen wird unter der Annahme der multivariaten Normalverteilung hergeleitet. Für den Ansatz (306) ergeben sich

$$x_1 = x + u_1 \quad (312)$$

$$x_2 = d_1x + u_2 \quad (313)$$

$$x_3 = z + u_3 \quad (314)$$

$$x_4 = d_2z + u_4 \quad (315)$$

Daraus ergeben sich die Indikatoren für das Produkt xz :

$$x_1z_1 = xz + xu_3 + zu_1 + u_1u_3 \quad (316)$$

$$x_1z_2 = d_2xz + xu_4 + d_2zu_1 + u_1u_4 \quad (317)$$

$$x_2z_1 = d_1xz + d_1xu_3 + zu_2 + u_2u_3 \quad (318)$$

$$x_2z_2 = d_1d_2xz + d_1xu_4 + d_2zu_2 + u_2u_4 \quad (319)$$

Damit hat man insgesamt 15 latente Variable: x , z , xz , xu_3 , xu_4 , zu_1 , zu_2 , u_1 , u_2 , u_3 , u_4 , u_1u_3 , u_1u_4 , u_2u_3 , u_2u_4 . Wird angenommen, dass x , z , u_1 , u_2 ,

Tabelle 9: Ladungen für das nichtlineare Faktorenmodell

Variable	x	z	xz	u_1	u_2	u_3	u_4	u_1u_3	u_1u_4	u_2u_3	u_2u_4	xu_3	xu_4	zu_1	zu_2
x_1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
x_2	d_1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
z_1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
z_2	0	d_2	0	0	0	0	1	0	0	0	0	0	0	0	0
x_1z_1	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0
x_1z_2	0	0	d_2	0	0	0	0	0	1	0	0	0	1	d_2	0
x_2z_1	0	0	d_1	0	0	0	0	0	0	1	0	d_1	0	0	1
x_2z_2	0	0	d_1d_2	0	0	0	0	0	0	0	1	0	d_1	0	d_2

u_3 , u_4 und v Abweichungen vom jeweiligen Mittelwert repräsentieren, multivariat normalverteilt sind und paarweise unkorreliert sind mit der Ausnahme von x und z , so sind die Diagonalelemente der Kovarianzmatrix der latenten Variablen durch

$$\begin{aligned} \sigma_{xz}^2 &= \sigma_x^2\sigma_z^2 + \sigma_{x,z}^2 & \sigma_{u_1,u_2}^2 &= \sigma_{u_1}^2\sigma_{u_2}^2, & \sigma_{u_1,u_4}^2 &= \sigma_{u_1}^2\sigma_{u_4}^2 \\ \sigma_{u_2u_3}^2 &= \sigma_{u_2}^2\sigma_{u_3}^2, & \sigma_{u_2u_4}^2 &= \sigma_{u_2}^2\sigma_{u_4}^2, & \sigma_{xu_3}^2 &= \sigma_x^2\sigma_{u_3}^2 \\ \sigma_{zu_1}^2 &= \sigma_z^2\sigma_{u_1}^2, & \sigma_{zu_2}^2 &= \sigma_z^2\sigma_{u_2}^2, & & \end{aligned} \quad (320)$$

Hier ist $\sigma_{x,z}^2$ die Kovarianz für x und z . Unter diesen Randbedingungen können die Parameter für das nichtlineare Modell über die Verallgemeinerte KQ-Methode gefunden werden. Der Fit des Modells ist bemerkenswert gut.

Muthén (2002) liefert eine Übersicht über nichtlineare Faktormodelle; in Muthén und Muthén (1998 – 2001) findet man einen *user's guide* für das Programm MPlus, mit dem verschiedene Modelle auf Daten angewendet werden können.

5.4 Latent-Class-Modelle

Eine häufige Fragestellung in der Diagnostik ist die Zuordnung von Probanden oder Patienten zu bestimmten Klassen. Die Klassen können Krankheiten oder <untergruppen von Krankheiten sein, oder Berufsgruppen, für die ein Proband besonders geeignet ist, oder Kulturen bzw. Zeitabschnitten, denen ein Archäologe seine Funde zuordnen möchte. Gegeben sind allgemein Symptome, und das Auftreten bestimmter Gruppen von Symptomen legt eine bestimmte Zuordnung nahe. Das diagnostische Ziel ist demnach eine Klassifikation, – wobei aber zumindest am Anfang noch nicht klar ist, ob solche Klassen überhaupt existieren. Bei einer Diskriminanzanalyse sind im Allgemeinen bereits bestimmte Klassen vorgegeben und man sucht nach einer Gewichtung der Symptome, die eine optimale Zuordnung zu den betrachteten Klassen erlauben. Dabei werden aber kontinuierliche Ausprägungen der Symptome vorausgesetzt. Will man anhand von dichotomen oder ordinalen Indikatoren klassifizieren und soll u. U. erst herausgefunden werden, ob Klassen existieren, muß nach anderen Verfahren gesucht werden.

Das LC-Modell für dichotome Variable wurde zuerst von Lazarsfeld und Henry (1968) eingeführt und dann von Goodman (1974) für den Fall nominaler Variablen verallgemeinert; weitere Literaturangaben findet man in Magidson und Vermunt (2003), die insbesondere den explikatorischen Fall diskutieren. Die folgenden Betrachtungen sind an dieser Arbeit orientiert.

In einer explorativen LC-Analyse wird zunächst ein Einklassenmodell, dann ein Zweiklassenmodell etc an die Daten angepasst. Solche Modelle heißen auch *LC Cluster Modelle*. Van der Ark und Van der Heijden (1998) und Van der Heijden, Gilula und Van der Ark (1999) haben gezeigt, dass eine LC Analyse dazu benutzt werden kann, die Anzahl der latenten Variablen zu bestimmen, die einer Menge von nominalen Variablen unterliegen. Darüber hinaus gibt es Beziehungen zur Korrespondenzanalyse und zur Faktorenanalyse.

Man kann zwischen LC Cluster Modellen und LC Faktormodellen unterscheiden. Beide Modelle können im Rahmen der log-linearen Modelle beschrieben werden.

LC Cluster Modelle Zur Illustration werden vier nominale Variablen A , B , C und D betrachtet, dazu eine latente Variable X mit T Kategorien. Das log-lineare Modell ist dann

$$\log F_{ijklt} = \lambda + \lambda_t^x + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{it}^{Ax} + \lambda_{jt}^{Bx} + \lambda_{kt}^{Cx} + \lambda_{lt}^{Dx} \quad (321)$$

i, j, k, l bezeichnen Stufen von A, B, C und D , und t die Stufen der latenten Variablen, $t = 1, \dots, T$. (321) hat die Form eines log-linearen Modells für eine 5-dimensionale Häufigkeitstabelle mit den Häufigkeiten F_{ijklt} . Es gibt Terme, die "Haupteffekte" (i) mit der latenten Variablen x repräsentieren und mit den vier "Symptomen" A bis D . Darüber hinaus gibt es "Interaktionseffekte" zwischen x und den Symptomen. Die Terme $\lambda_t^x, \lambda_i^A, \dots, \lambda_l^D$ werden mit in die Betrachtung einbezogen, um keine Vorannahmen über die Randverteilungen machen zu müssen. Es wird angenommen, dass die Reaktionen auf A, \dots, D gegeben die t -te Stufe von x entspricht der Annahme der lokalen Unabhängigkeit, die es ermöglicht, Interaktionsterme zwischen den A, \dots, D zu vernachlässigen. Für den Fall nur einer Klasse ($T = 1$) reduziert sich das Modell auf

$$\log F_{ijklt} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D. \quad (322)$$

Es gibt hier keine Wechselwirkungen, – dieser Fall repräsentiert die Nullhypothese H_0 . H_0 ist das Modell, gegen das andere Modelle, in denen Interaktionen angenommen werden, getestet werden. Es hat

$$N_{param}(H_0) = (I - 1) + (J - 1) + (K - 1) + (L - 1) \quad (323)$$

freie, also zu schätzende Parameter. Für das Modell (321) hat man

$$N_{param}(T) = (T - 1) + N_{param}(H_0) \times (1 + (T - 1)) = (T - 1) + N_{param}(H_0) \times T, \quad (324)$$

und die Anzahl der Freiheitsgrade für den Test zur Modellanpassung ist

$$DF_T = IJKL - N_{param}(T) - 1 = IJKL - (1 + N_{param}(H_0)) \times T \quad (325)$$

Man beginnt mit dem Grundmodell ($T = 1$), und jedesmal, wenn die Anzahl der latenten Variablen um 1 erhöht wird, erhöht sich die Anzahl verschiedener Parameter um $1 + N_{param}(H_0)$ und die Anzahl der Freiheitsgrade reduziert sich um diese Zahl.

Das Latent Class Faktorenmodell Es sei X eine latente Variable mit vier Kategorien, $X = \{1, 2, 3, 4\}$. Sie könnte erklärt werden durch Bezug auf auf zwei dichotome latente Variablen $V = \{1, 2\}$, $W = \{1, 2\}$, indem man

	$W = 1$	$W = 2$	(326)
$V = 1$	$X = 1$	$X = 2$	
$V = 2$	$X = 3$	$X = 4$	

Das LC Cluster-Modell der Gleichung (321) mit $T = 4$ Klassen kann nun als unbeschränktes LC Faktorenmodell mit zwei dichotomen Variablen V und W angeschrieben werden:

$$\begin{aligned} \log F_{ijklrs} &= \lambda + \lambda_r^V + \lambda_s^W + \lambda_{rs}^{VW} + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{ir}^{AV} + \lambda_{jr}^{BV} \\ &= +\lambda_{kr}^{CV} + \lambda_{ir}^{DV} + \lambda_{is}^{AW} + \lambda_{irs}^{AVW} + \lambda_{jrs}^{BVW} \\ &\quad +\lambda_{krs}^{CVW} + \lambda_{irs}^{DVW} \end{aligned} \quad (327)$$

Der Term λ_x^t in (321) taucht hier nicht mehr auf, – statt dessen kann

$$\lambda_{2(r-1)+s} = \lambda_r^V + \lambda_s^W + \lambda_{rs}^{VW}$$

geschrieben werden. Die 2-Variablensterme, die X enthalten, können in der Form

$$\lambda_{i,2(r-1)+s} = \lambda_{ir}^{AV} + \lambda_{is}^{AW} + \lambda_{irs}^{AVW}$$

und

$$\lambda_{j,2(r-1)+s} = \lambda_{js}^{BW} + \lambda_{js}^{BW} + \lambda_{jrs}^{BVW}$$

ausgedrückt werden, und so fort. Man hat hier eine Reparametrisierung des ursprünglichen Modells, bei dem aber die Anzahl der freien Parameter nicht reduziert wird.

Man kann nun ein *Basis-R-Faktorenmodell* definieren: darin werden R paarweise unabhängige, dichotome latente Variablen angenommen, bei dem die Faktorladungen entsprechenden Parameter die Assoziation der latenten Variablen mit den gemessenen "Indikator"variablen¹⁹. Beim Basis- R -Faktorenmodell werden bestimmte Restriktionen für die latenten Variablen spezifiziert. Die erste Restriktion besteht darin, dass alle Terme die Interaktionen zwischen mehr als 2 Variablen repräsentieren gleich Null gesetzt werden, so dass

$$\lambda_{irs}^{AVW} = \lambda_{irs}^{BVW} = \lambda_{irs}^{CVW} = \lambda_{irs}^{DVW} = 0$$

resultiert. Die 2-Variablensterme nehmen dann die Form

$$\lambda_{i,2(r-1)+s}^{AX} = \lambda_{ir}^{AV} + \lambda_{is}^{AW}, \quad \lambda_{j,2(r-1)+s}^{BX} = \lambda_{jr}^{BV} + \lambda_{js}^{BW}, \quad etc$$

an. Die Nullsetzung der 3-Variableninteraktionen entspricht insofern der Standardfaktorenanalyse, als dann jede latente Variable einen Einfluß auf jede beobachtete Variable hat und es keine Wechselwirkungen höherer Ordnung gibt (in Gleichung (??) treten keine Terme auf, die durch Produkte der F_k definiert sind).

Das LC-Faktorenmodell ist ein Spezialfall eines LC-Cluster-Modells; die Tabelle 10 zeigt die möglichen Äquivalenzen.

Eine Anwendung dieser Verfahren findet man in Rist, Glöckner-Rist und Demmel (2009).

6 Anhang

6.1 Eine alternative Herleitung

Die Forderung, die Datenvektoren als Linearkombination orthogonaler Vektoren darzustellen, führte auf die Hauptachsentransformation der Daten. Man

¹⁹Der Begriff der Indikatorvariablen ist üblicherweise etwas anders definiert, etwa $\chi = 1$, wenn ein Ereignis eingetreten ist, $\chi = 0$ sonst; hier bedeutet der Ausdruck soviel wie: ein Merkmal ist vorhanden – oder nicht.

Tabelle 10: Äquivalenzbeziehungen zwischen LC-Cluster-Modellen und LC-Faktormodellen

LC Cluster-Modelle			LC-Faktormodelle		
Anzahl der Lat. Klassen	Anzahl der Parameter	df	Anzahl der Faktoren	Anzahl der Parameter	df
1	5	26	0	5	26
2	11	20	1	11	20
3	17	14	2	17	14
4	23	8	3	23	8
5	29	2	4	29	2

kann alternativ fordern, dass die erste latente Variable eine maximale Varianz haben soll, die zweite dann die zweitgrößte Varianz, etc. Gemeint ist damit, dass die Koordinaten der Personen auf der ersten Achse eine maximale Varianz haben sollen, etc. Es wird vorausgesetzt, dass der Mittelwert der Koordinaten der Personen auf den Achsen jeweils gleich Null sein soll. Ist $\vec{F}_1 = (F_{11}, F_{21}, \dots, F_{m1})'$ der Vektor der Koordinaten der m Personen auf der ersten Achse, so soll also $Var(\vec{F}_1) = \vec{F}_1' \vec{F}_1$ maximiert werden, also das Quadrat $\|\vec{F}_1\|^2$ der Länge von \vec{F}_1 . Dazu muß eine Nebenbedingung eingeführt werden, da sich sonst das triviale Resultat $\|\vec{F}_1\|^2 = \infty$ ergäbe. Natürlich soll wieder $Z = LP'$ gelten, $L = [\vec{F}_1, \vec{F}_2, \dots, \vec{F}_r]$. Dementsprechend wird als Nebenbedingung

$$\sum_{j=1}^n p_{j1}^2 = 1 \quad (328)$$

eingeführt, d.h. $\phi(p_{11}, \dots, p_{n1}) = \sum_j p_{j1}^2 - 1 = 0$ (vergl. Anhang, Gleichung (370), Seite 131). Man definiert also die Hilfsfunktion

$$F(p_{11}, \dots, p_{n1}) = \sum_{i=1}^m F_{i1}^2 + \mu \left(\sum_{j=1}^n p_{j1}^2 - 1 \right). \quad (329)$$

Die F_{i1} sind aber Funktionen der p_{j1} ; es gilt ja $Z\vec{Y}_1 = \vec{F}_1$, so dass $\|\vec{F}_1\|^2 = \vec{p}_1' Z' Z \vec{Y}_1$, und mithin

$$F(p_{11}, \dots, p_{n1}) = \vec{Y}_1' Z' Z \vec{Y}_1 + \mu (\vec{Y}_1' \vec{Y}_1 - 1). \quad (330)$$

Differenziert man diese Gleichungen nach den Komponenten von \vec{p}_1 und setzt die entstehenden Ableitungen gleich Null, so findet die Lösung für \vec{Y}_1 und μ . Da $\vec{Y}_1 = (p_{11}, \dots, p_{n1})'$, ist

$$\frac{\partial \vec{Y}_1}{\partial p_{j1}} = (0, \dots, 0, 1, 0, \dots, 0)' = \vec{e}_j,$$

denn $dp_{k1}/dp_{j1} = 0$ für $j \neq k$, und $dp_{k1}/dp_{j1} = 1$ für $j = k$; \vec{e}_j ist der j -te Einheitsvektor. Da bei der rechten Seite von (330) die Produktregel der Differentiation anzuwenden ist, findet man

$$\frac{\partial F}{\partial p_{j1}} = \vec{e}'_j Z' Z \vec{Y}_1 + \vec{Y}'_1 Z' Z \vec{e}_j + \mu(\vec{e}'_j \vec{Y}_j + \vec{Y}'_j \vec{e}_j), \quad j = 1, \dots, n \quad (331)$$

Aber $\vec{e}'_j Z' Z \vec{Y}_1$ etc. sind Skalare, und mithin gilt

$$\frac{\partial F}{\partial p_{j1}} = 2\vec{e}'_j (Z' Z \vec{Y}_1 + \mu \vec{p}_j). \quad (332)$$

Für $\partial F / \partial p_{j1} = 0$, $j = 1, \dots, n$ ergebe sich die Lösung $\vec{\pi}_1$, d.h. für $\vec{Y}_1 = \vec{\pi}_1$ nimmt F ein Extremum an. Da $2\vec{e}'_j \neq 0$, folgt $Z' Z \vec{\pi}_1 + \mu \vec{p}_j = 0$, also

$$Z' Z \vec{Y}_1 = -\mu \vec{\pi}_j, \quad j = 1, \dots, n \quad (333)$$

und dies bedeutet, das $\vec{\pi}_j$ ein Eigenvektor von $Z' Z$ sein muß, und $-\mu = \lambda$ der zugehörige Eigenwert. Es werde nun der Einfachheit halber wieder \vec{Y}_j statt $\vec{\pi}_j$ geschrieben und mit P die Matrix, deren Spaltenvektoren eben diese \vec{Y}_j sind, bezeichnet. Da $Z' Z$ symmetrisch ist, muß $\lambda > 0$ gelten. Da $ZP = L$, ist $L'L = P'Z'ZP = \Lambda$, mithin

$$\vec{F}'_1 \vec{F}_1 = |\vec{F}_1|^2 = \lambda_1. \quad (334)$$

Zusammenfassend hat man also

Satz 6.1 Die Varianz $|\vec{F}_1|^2$ der Komponenten von \vec{F}_1 ist maximal, wenn P die Matrix der Eigenvektoren von $Z' Z$ ist, und $|\vec{F}_1|^2 = \lambda_1$, λ_1 der erste (d.h. der maximale) Eigenwert von $Z' Z$.

6.2 Lineare und statistische Unabhängigkeit

Es ist gezeigt worden, dass die Orthogonalität von Vektoren deren lineare Unabhängigkeit impliziert, aber umgekehrt linear unabhängige Vektoren nicht orthogonal sein müssen. Die Orthogonalität zweier latenter Variablen bzw. Vektoren \vec{F}_j und \vec{F}_k bedeutet, dass sie auch unkorreliert sind (da es sich um zufällige Vektoren handelt, macht es Sinn, von der Unkorreliertheit zu reden). Andererseits wird von latenten Variablen nicht mehr verlangt, als dass sie die beobachteten Variablen $\vec{X}_1, \dots, \vec{X}_n$ vorherzusagen gestatten und dass sie kein redundantes System sein sollen, d.h. keine der latenten Variablen soll sich durch die anderen erklären lassen. Man muß also eigentlich nur lineare Unabhängigkeit, nicht aber notwendig auch Orthogonalität fordern. Für l.u., aber nicht orthogonale Vektoren \vec{F}_j , $1 \leq j \leq s$ verschwinden die Skalarprodukte aber nicht. Da die Skalarprodukte in diesem Fall tatsächlich Korrelationskoeffizienten sind, sind nicht orthogonale, wenn auch linear unabhängige Vektoren korreliert.

Dieser Sachverhalt zeigt, dass der Begriff der Korrelation nicht mit den Begriffen Abhängigkeit bzw. Unabhängigkeit gleichzusetzen ist. Der Begriff der linearen Unabhängigkeit aus der linearen Algebra bedeutet für die Anwendung in der Faktorenanalyse, dass "linear unabhängige" Vektoren verschiedene, nicht eindeutig auseinander zu erklärende Qualitäten repräsentieren, die aber in der Stichprobe bis zu einem gewissen Grade - aber eben nicht vollständig - miteinander gekoppelt auftreten. Diese Kopplungen implizieren die von Null verschiedene Korrelation.

Umgekehrt bedeutet aber eine Korrelation gleich Null noch nicht, dass die nicht miteinander korrelierenden Variablen auch unabhängig voneinander sind. Es lassen sich Beispiele finden, in denen zwei Variablen perfekt - deterministisch - voneinander abhängen, die Korrelation aber gleich Null ist:

Beispiel: (Feller, 1968, p. 236) Die Variable X nehme die Werte -2, -1, 1, 2 an, und es sei $Y = X^2$; dann ist Y durch X eindeutig bestimmt. Die möglichen Messwertpaare (-2, 4), (-1, 1), (1, 1) und (2, 4) seien gleichhäufig aufgetreten. Dann ist $\bar{x} = 0$, $\bar{y} = 2.5$, $\text{Kov}(x, y) = (\sum_{i=1}^n x_i y_i - \bar{x}\bar{y})/n$, d.h. $\text{Kov}(x, y) = -8 - 1 + 1 + 4 - 0 \cdot 2.5 = 0$, also ist auch $r = 0$, obwohl die Variablen X und Y deterministisch voneinander abhängen.

Das Beispiel mag ein wenig gekünstelt erscheinen, aber es kommt auf die Grundsätzlichkeit des Arguments an. Der Schluß von einer Korrelation gleich Null auf die statistische Unabhängigkeit der Variablen ist i.a. nur bei der Normalverteilung gestattet. Die multivariate Normalverteilung ist in (156) definiert worden:

$$f(\vec{x}) = A \exp(-(\vec{x} - \vec{\mu})' S^{-1} (\vec{x} - \vec{\mu})),$$

wobei S die Varianz-Kovarianzmatrix ist. Nun sind allgemein die zufälligen Veränderlichen x_1, \dots, x_n , also die Komponenten von \vec{x} , paarweise unabhängig, wenn die gemeinsame Dichte $f(x_1, \dots, x_n) = f(\vec{x})$ sich als Produkt der Dichten für die einzelnen Komponenten darstellen läßt, wenn also

$$f(\vec{x}) = f_1(x_1) f_2(x_2) \cdots f_n(x_n) = \prod_{i=1}^n f_i(x_i) \quad (335)$$

gilt. Nun sei S eine Diagonalmatrix; dann ist auch S^{-1} eine Diagonalmatrix:

$$S^{-1} = \begin{pmatrix} 1/s_1^2 & 0 & \cdots & 0 \\ 0 & 1/s_2^2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & 1/s_n^2 \end{pmatrix}. \quad (336)$$

Für die Dichte $f(\vec{x})$ erhält man dann den Ausdruck

$$f(\vec{x}) = A \prod_{i=1}^n \exp\left(-\frac{(x_i - \mu_i)^2}{s_i^2}\right); \quad (337)$$

die x_1, \dots, x_n sind dann also auch stochastisch unabhängig.

Die Möglichkeit, von der paarweisen Unkorreliertheit auf die stochastische Unabhängigkeit schließen zu können, ist allerdings - gegeben einige sehr allgemeine Voraussetzungen - charakteristisch für die Normalverteilung (vergl. Feller (1966), p. 84). Für nicht normalverteilte Messungen muß der Schluß von der Unkorreliertheit auf die stochastische Unabhängigkeit nicht gelten (vergl. das obige Beispiel). Zusammenfassend kann man sagen:

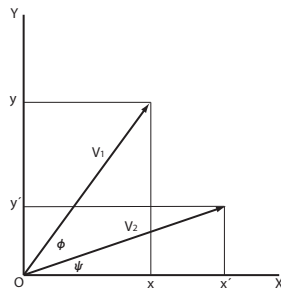
1. Lineare Unabhängigkeit schließt Unkorreliertheit nicht aus, und
2. Unkorreliertheit bedeutet nicht notwendig stochastische Unabhängigkeit.

6.3 Koordinatenrotation

6.3.1 Ansatz I

Gegeben seien zwei Vektoren \vec{v}_1 und \vec{v}_2 mit gleicher Länge $L = \|\vec{v}_1\| = \|\vec{v}_2\|$. Der Winkel zwischen den beiden Vektoren sei ϕ , und der Winkel zwischen \vec{v}_2 und der x -Achse sei ψ . Man kann \vec{v}_2 als eine Transformation von \vec{v}_1 auffassen; da die Länge der Vektoren identisch ist, ist die Transformation eine Rotation. Umgekehrt kann man von \vec{v}_1 annehmen, dass er durch Rotation aus \vec{v}_2 hervorgegangen ist. Also existiert eine Matrix T derart, dass $\vec{v}_2 = T\vec{v}_1$, und da die Rotation von \vec{v}_2 in den Vektor \vec{v}_1 gerade die Umkehrung der Rotation von \vec{v}_1 in \vec{v}_2 ist, muß $T^{-1}\vec{v}_2 = \vec{v}_1$ gelten, und T^{-1} ist die zu T inverse Transformation. Nun muß T bestimmt werden. Durch T werden die Koordinaten (x, y) für \vec{v}_1 in die Koordinaten (x', y') von \vec{v}_2 überführt. Man muß also die Gleichungen für diese Überführung herleiten, um T zu bestimmen. Offenbar

Abbildung 17: Rotation I



ist $\sin(\phi + \psi) = y/L$, und $\cos(\phi + \psi) = x/L$, etc. Insgesamt hat man die Beziehungen

$$x = L \cos(\phi + \psi) \quad (338)$$

$$y = L \sin(\phi + \psi) \quad (339)$$

$$x' = L \cos \psi \quad (340)$$

$$y' = L \sin \psi \quad (341)$$

Generell gelten die folgenden beiden Gleichungen:

$$\sin(\phi \pm \psi) = \sin \phi \cos \psi \pm \cos \phi \sin \psi \quad (342)$$

$$\cos(\phi \pm \psi) = \cos \phi \cos \psi \mp \sin \phi \sin \psi \quad (343)$$

Angewandt auf (338) und (339) erhält man

$$\cos(\phi + \psi) = \cos \phi \cos \psi - \sin \phi \sin \psi \quad (344)$$

$$\sin(\phi + \psi) = \sin \phi \cos \psi + \cos \phi \sin \psi. \quad (345)$$

Für $\sin \psi$ und $\cos \psi$ kann man nun die Ausdrücke in (340) und (341) einsetzen, so dass man eine Beziehung zwischen x und x' und y und y' erhält:

$$Lx = Lx' \cos \phi - Ly' \sin \phi \quad (346)$$

$$Ly = Lx' \sin \phi + Ly' \cos \phi \quad (347)$$

bzw.

$$x = x' \cos \phi - y' \sin \phi \quad (348)$$

$$y = x' \sin \phi + y' \cos \phi. \quad (349)$$

In Matrixform erhält man also

$$\begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}. \quad (350)$$

Hier wird \vec{v}_2 , also der Vektor mit den Koordinaten (des Endpunkts) (x', y') , in den Vektor \vec{v}_1 transformiert, vergl. Abb. 17, es ist also

$$\begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} = T^{-1}; \quad (351)$$

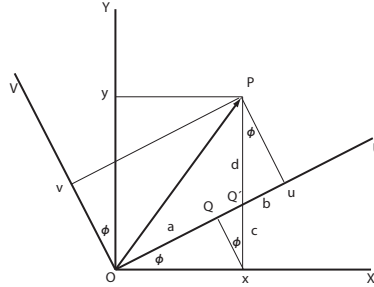
Die Drehung von \vec{v}_1 in den Vektor \vec{v}_2 muß diese Rotation invertieren, d.h. es muß um den Winkel $-\phi$ rotiert werden. Substituiert man $-\phi$ in (351), so erhält man

$$\begin{pmatrix} \cos(-\phi) & -\sin(-\phi) \\ \sin(-\phi) & \cos(-\phi) \end{pmatrix} = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} = T. \quad (352)$$

6.3.2 Ansatz II

Gegeben sei der Vektor \vec{x} , der durch die Verbindung \overline{OP} definiert wird. In Bezug auf die (X, Y) -Koordinatenachsen hat der Punkt P - der Endpunkt des Vektors \vec{x} - die Koordinaten (x, y) . Das X, Y -System werde um den Winkel

Abbildung 18: Rotation II



ϕ rotiert und geht in das (U, V) -System über. Gesucht sind die Koordinaten (u, v) des Vektorendpunkts P im neuen System.

Es werden zunächst einige Beziehungen zwischen Verbindungslinien und dem Winkel ϕ hergestellt. Es ist

$$a = \overline{OQ}, \quad b = \overline{QQ'} + \overline{Q'u} \quad (353)$$

$$c = \overline{xQ'}, \quad d = \overline{Q'P} \quad (354)$$

Zur Erinnerung: es gilt allgemein: $\cos \phi = \text{Ankathete}/\text{Hypotenuse}$, $\sin \phi = \text{Gegenkathete}/\text{Hypotenuse}$. Mithin lassen sich aus Abb. 18 die folgenden Beziehungen ableiten:

$$\sin \phi = \overline{QQ'}/c = \overline{Q'u}/d \quad (355)$$

$$\cos \phi = a/x = \overline{Pu}/d \quad (356)$$

$$\tan \phi = \sin \phi / \cos \phi = c/x \quad (357)$$

Weiter gelten sicherlich die Beziehungen

$$u = a + b, \quad y = c + d. \quad (358)$$

Dann folgt

$$b = \overline{QQ'} + \overline{Q'u} = c \sin \phi + d \sin \phi = (c + d) \sin \phi = y \sin \phi. \quad (359)$$

so dass

$$a = x \cos \phi \quad (360)$$

$$b = y \sin \phi \quad (361)$$

$$u = a + b = x \cos \phi + y \sin \phi \quad (362)$$

Weiter findet man

$$\begin{aligned} v &= d \cos \phi - (y - c) \cos \phi \\ &= y \cos \phi - c \cos \phi \end{aligned} \quad (363)$$

Aber wegen (357) hat man $x \cos \phi = c$, so dass man zusammen mit (362) die Gleichungen

$$u = x \cos \phi + y \sin \phi \quad (364)$$

$$v = -x \sin \phi + y \cos \phi \quad (365)$$

hat. In Matrixform erscheinen diese Gleichungen in der Form

$$\begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix} \quad (366)$$

Man invertiert diese Rotation, indem man die (U, V) -Achsen um den Winkel $-\phi$ zurückdreht. Also muß gelten

$$\begin{pmatrix} \cos(-\phi) & \sin(-\phi) \\ -\sin(-\phi) & \cos(-\phi) \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}. \quad (367)$$

Aber $\cos(-\phi) = \cos \phi$, und $\sin(-\phi) = -\sin \phi$, so dass

$$\begin{pmatrix} \cos(-\phi) & \sin(-\phi) \\ -\sin(-\phi) & \cos(-\phi) \end{pmatrix} = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}. \quad (368)$$

Bezeichnet man die Transformationsmatrix in (366) mit T , so hat man

$$T = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}, \quad T^{-1} = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}. \quad (369)$$

Man rechnet leicht nach, dass $TT^{-1} = T^{-1}T = I$, d.h. T ist orthonormal.

6.4 Extrema von Funktionen mit Nebenbedingungen

Es sei $f(x_1, x_2)$ eine Funktion der Variablen x_1 und x_2 ; so ist z.B. die Länge eines Vektors mit den Komponenten x und y durch die Funktion $f(x_1, x_2) = (x_1^2 + x_2^2)^{1/2}$ gegeben. Gelegentlich möchte man wissen, welche Extremwerte f annehmen kann. Ist z.B. $\vec{x} = (x_1, x_2)$ der Vektor vom Mittelpunkt einer Ellipse bis zur Ellipse selbst, so kann man fragen, welche maximale Länge \vec{x} annehmen kann. Ohne jede Nebenbedingung ist aber die Frage nach dem Extremwert einer Funktion oft gar nicht sinnvoll zu stellen. So ist die maximale Länge eines Vektors gleich unendlich. Vektoren, die einen gemeinsamen Ursprung haben und deren Endpunkte auf einer Ellipse liegen, sind aber nie unendlich lang, - aber es gibt Vektoren mit maximaler Länge. Die Tatsache, dass der Anfangspunkt eines Vektors im Mittelpunkt einer durch bestimmte Parameter definierten Ellipse liegt und der Endpunkt des Vektors auf der Ellipse, definiert eine Nebenbedingung, und bei der Frage nach der maximalen Länge muß diese Nebenbedingung mit berücksichtigt werden.

Die Nebenbedingung läßt sich durch eine Gleichung spezifizieren, denen die x_1 und x_2 gehorchen müssen. Diese Gleichung läßt sich allgemein durch

den Ausdruck $\phi(x_1, x_2) = 0$ angeben. Definieren die x_1, x_2 eine Ellipse, so gilt insbesondere

$$ax_1^2 + bx_2^2 + 2cx_1x_2 = k_0,$$

wobei k_0 eine Konstante ist. Dementsprechend hat man die Nebenbedingung

$$\phi(x_1, x_2) = ax_1^2 + bx_2^2 + 2cx_1x_2 - k_0 = 0. \quad (370)$$

$\phi(x_1, x_2)$ beschreibt eine Kurve in der Ebene - eben eine Ellipse. Das Quadrat der Länge des Vektors \vec{x} ist stets durch $f(x_1, x_2) = x_1^2 + x_2^2$ gegeben. Die Nebenbedingung ϕ bedeutet, dass z.B. x_2 eine Funktion von x_1 sein muß; etwa $x_2 = g(x_1)$. Diese Funktion kann man für x_2 in die zu maximierende Funktion f einsetzen: $f(x_1, x_2) = f(x_1, g(x_1))$, und damit hat man das Problem, ein Extremum für f unter der Nebenbedingung $\phi(x_1, g(x_1)) = 0$ zu finden, auf das Problem zurückgeführt, das Extremum für $f(x_1, g(x_1))$ zu finden. Dazu muß die Ableitung bezüglich x_1 gefunden und gleich Null gesetzt werden; Anwendung der Kettenregel liefert

$$\frac{df}{dx_1} = \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2} g'(x_1) = 0, \quad g'(x_1) = \frac{dg}{dx_1}. \quad (371)$$

Analog dazu muß

$$\frac{d\phi}{dx_1} = \frac{\partial \phi}{\partial x_1} + \frac{\partial \phi}{\partial x_2} g'(x_1) = 0 \quad (372)$$

gelten. Löst man die beiden Gleichungen nach g' auf, so erhält man

$$g'(x_1) = -\frac{\partial f/\partial x_1}{\partial f/\partial x_2} = -\frac{\partial \phi/\partial x_1}{\partial \phi/\partial x_2}. \quad (373)$$

Dies bedeutet, dass $\partial f/\partial x_1 \propto \partial \phi/\partial x_1$ und $\partial f/\partial x_2 \propto \partial \phi/\partial x_2$ sein muß, und der Proportionalitätsfaktor identisch sein muß, - sonst würde er sich in (373) nicht herausgekürzt haben. Setzt man diesen Proportionalitätsfaktor gleich $-\mu$ (das Minuszeichen ist keine Einschränkung der Allgemeinheit, sondern soll nur Also folgen die Gleichungen

$$\partial f/\partial x_1 = \mu \partial \phi/\partial x_1 \quad (374)$$

$$\partial f/\partial x_2 = \mu \partial \phi/\partial x_2, \quad (375)$$

die auch in der Form

$$\partial f/\partial x_1 + \mu \partial \phi/\partial x_1 = 0 \quad (376)$$

$$\partial f/\partial x_2 + \mu \partial \phi/\partial x_2 = 0 \quad (377)$$

geschrieben werden können (hierbei ist das Vorzeichen von - in + geändert worden, was keine Einschränkung der Allgemeinheit darstellt, da μ ja noch völlig unbekannt ist). Diese Gleichungen sind aber nichts weiter als die partiellen Ableitungen der Funktion

$$F(x_1, x_2) := f(x_1, x_2) + \mu \phi(x_1, x_2), \quad \phi(x_1, x_2) = 0. \quad (378)$$

Damit hat man zur Bestimmung eines Extremums unter Nebenbedingungen die

Lagrangesche Multiplikatorenregel: Bestimmt werden soll der Extremwert einer Funktion $f(x_1, x_2)$ unter der Nebenbedingung

$$\phi(x_1, x_2) = 0.$$

Dazu bildet man die neue Funktion $F(x_1, x_2) = f(x_1, x_2) + \mu\phi(x_1, x_2)$, wobei μ der *Lagrange Faktor* (oder *Lagrange Multiplikator*) ist. Die Werte (x_{01}, x_{02}) , für die f das gesuchte Maximum/Minimum annimmt, ergeben sich als Lösungen der Gleichungen (376) und (377) sowie der Nebenbedingung $\phi(x_{01}, x_{02}) = 0$. Der unbekannte Lagrange Faktor ergibt sich dabei ebenfalls als Teil der Lösung dieser Gleichungen.

Alternative Herleitung: Man kann die Gleichungen (371) und (372) auch zusammenschreiben:

$$\frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2} g'(x_1) = 0 \quad (379)$$

$$\frac{\partial \phi}{\partial x_1} + \frac{\partial \phi}{\partial x_2} g'(x_1) = 0 \quad (380)$$

Um den Term g' "loszuwerden" (abgesehen davon, dass man g nicht immer in expliziter Weise bestimmen kann, möchte man es auch gar nicht, wenn es sich vermeidenläßt!), kann man (380) mit einem geeignet gewählten Faktor λ multiplizieren und dann (380) und (379) addieren:

$$\frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2} g'(x_1) = 0 \quad (381)$$

$$\lambda \frac{\partial \phi}{\partial x_1} + \lambda \frac{\partial \phi}{\partial x_2} g'(x_1) = 0, \quad (382)$$

und λ soll so gewählt werden, dass

$$\frac{\partial f}{\partial x_2} g'(x_1) + \lambda \frac{\partial \phi}{\partial x_2} g'(x_1) = 0$$

woraus

$$\lambda = -\frac{\partial f / \partial x_2}{\partial \phi / \partial x_2} \quad (383)$$

folgt. Addition von (381) und (382) liefert dann

$$\frac{\partial f}{\partial x_1} + \lambda \frac{\partial \phi}{\partial x_1} = 0 \quad (384)$$

$$\frac{\partial f}{\partial x_2} + \lambda \frac{\partial \phi}{\partial x_2} = 0, \quad (385)$$

wobei sich die zweite Gleichung unmittelbar aus (383) ergibt. Die beiden Gleichungen sind offenbar wieder die partiellen Ableitungen der Funktion $F(x_1, x_2) = f(x_1, x_2) + \lambda\phi(x_1, x_2)$, womit die Lagrangesche Multiplikatorenregel wieder hergeleitet ist.

6.5 Herleitung der 2-dimensionalen Normalverteilung

X und Y seien normalverteilt. Gesucht ist die gemeinsame Verteilung $f(x, y)$, wenn für die Korrelation $r_{xy} \neq 0$ gilt; der Fall unabhängiger, normalverteilter Variablen ergibt sich dann als Spezialfall für $r_{xy} = 0$.

Sind A und B zwei zufällige Ereignisse, so ist die bedingte Wahrscheinlichkeit $p(A|B)$ durch

$$p(A|B) = \frac{p(A \wedge B)}{p(B)} \quad (386)$$

gegeben. Es sei $A = \{X = x\}$, $B = \{Y = y\}$, und $f(x, y)$ sei die gemeinsame Dichte für X und Y . $g(y|x)$ sei die bedingte Dichte für Y , gegeben $X = x$, und $f_X(x) = \int f(x, y)dy$ sei die Randverteilung für X . Dann hat man, analog zu (386),

$$g(y|x) = \frac{f(x, y)}{f_X(x)}. \quad (387)$$

Die gemeinsame Dichte ergibt sich daraus als

$$f(x, y) = g(y|x)f_X(x). \quad (388)$$

Es ist

$$f_X(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left(-\frac{(x - \mu_x)^2}{2\sigma_x^2}\right). \quad (389)$$

Um g zu bestimmen, werde die Regressionsgleichung

$$y = ax + b + \varepsilon$$

betrachtet. Für festes x ist g die Dichte für Y , gegeben x (kurz: $Y|x$), und $\sigma_{y|x}^2 = \sigma_\varepsilon^2$. Die Voraussetzung der Normalverteilung für Y bedeutet dann $Y \sim N(ax + b, \sigma_\varepsilon^2)$. Allgemein gilt für die unconditionierte Varianz von Y

$$\sigma_y^2 = a^2 \sigma_x^2 + \sigma_\varepsilon^2.$$

Es ist $r = r_{xy} = a\sigma_x/\sigma_y$, so dass $a = r\sigma_y/\sigma_x$. Dann ist

$$\sigma_y^2 = \frac{r^2 \sigma_x^2 \sigma_y^2}{\sigma_x^2} + \sigma_\varepsilon^2 = r^2 \sigma_y^2 + \sigma_\varepsilon^2,$$

woraus

$$\sigma_\varepsilon^2 = (1 - r^2)\sigma_y^2 \quad (390)$$

folgt, und für g findet man

$$g(y|x) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \exp\left(-\frac{(y - (ax + b))^2}{2\sigma_\varepsilon^2}\right) = \frac{1}{\sigma_y \sqrt{(1 - r^2)} \sqrt{2\pi}} \exp\left(-\frac{(y - (ax + b))^2}{2\sigma_\varepsilon^2}\right)$$

Es ist $\mu_y = a\mu_x + b$ und also

$$y - (ax + b) = y - \mu_y - (ax + b) + \mu_y = y - \mu_y - (ax + b - a\mu_x - b) = y - \mu_y - a(x - \mu_x)$$

und

$$(y - (ax + b))^2 = (y - \mu_y)^2 + a^2(x - \mu_x)^2 - 2a(x - \mu_x)(y - \mu_y).$$

Substituiert man für a wieder $r\sigma_y/\sigma_x$, so erhält man die gemeinsame Dichte

$$f(x, y) = A \exp \left[-\frac{1}{2(1-r^2)} \left(\frac{(y - \mu_y)^2}{\sigma_y^2} + \frac{(x - \mu_x)^2}{\sigma_x^2} - \frac{2r(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y} \right) \right], \quad (391)$$

mit

$$A = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}. \quad (392)$$

Es werde die Menge der Punkte $\{(x, y) | f(x, y) = k_0\}$, k_0 eine Konstante, betrachtet. Diese Menge ist durch die Bedingung

$$\frac{(x - \mu_x)^2}{\sigma_x^2} + \frac{(y - \mu_y)^2}{\sigma_y^2} - \frac{2r(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y} = k, \quad (393)$$

k eine Konstante, gegeben. Dies ist die Gleichung für eine Ellipse. Ellipsen sind die geometrischen Orte für Punkte mit gleicher Wahrscheinlichkeit, wenn (X, Y) gemeinsam normalverteilt sind. Dazu werde die Ellipsengleichung in Matrixschreibweise formuliert. Es sei $\xi_1 = x - \mu_x$, $\xi_2 = y - \mu_y$, und $\xi = (\xi_1, \xi_2)^t$. Dem Übergang von Gleichung (??) zu Gleichung (??) (Seite ??) entsprechend sei

$$M = \frac{1}{(1-r^2)} \begin{pmatrix} 1/\sigma_x^2 & -r/\sigma_x\sigma_y \\ -r/\sigma_x\sigma_y & 1/\sigma_y^2 \end{pmatrix}. \quad (394)$$

Dann kann die Ellipsengleichung in der Form

$$(x - \mu_x, y - \mu_y)M \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix} = k \quad (395)$$

geschrieben werden. Üblich ist allerdings eine andere Schreibweise, die die Struktur der Matrix M deutlicher macht und die sichtbar wird, wenn man die zu M inverse Matrix M^{-1} betrachtet: es ist

$$M^{-1} = \begin{pmatrix} \sigma_x^2 & r\sigma_x\sigma_y \\ r\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}. \quad (396)$$

Wegen $r = r_{xy} = \text{Kov}(x, y)/(\sigma_x\sigma_y) = \sigma_{xy}/(\sigma_x\sigma_y)$ ist $r\sigma_x\sigma_y = \text{Kov}(X, Y) = \sigma_{xy}$, so dass M^{-1} offenbar gerade gleich der Varianz-Kovarianz-Matrix für den Vektor $(x, y)^t$ ist. Diese Matrix wird üblicherweise mit Σ bezeichnet, so dass $M^{-1} = \Sigma$, d.h. aber $M = \Sigma^{-1}$. Ist $\vec{x} = (x, y)^t$, so hat man statt (395) nun die Darstellung

$$\vec{x}^t \Sigma^{-1} \vec{x} = k. \quad (397)$$

Die Orientierung der Ellipsen wird durch die Eigenvektoren von Σ^{-1} angegeben. Wegen

$$\Sigma = P\Lambda P', \quad \Sigma^{-1} = (P\Lambda P')^{-1} = P\Lambda^{-1}P' \quad (398)$$

sind dies die gleichen Orientierungen, die durch die Varianz-Kovarianz-Matrix festgelegt werden; nur die Eigenwerte von Σ und σ^{-1} stehen in einem reziproken Verhältnis zueinander.

Nun werde noch der Normalisierungsfaktor A für $f(x, y)$ (Gleichung (392)) betrachtet:

$$A = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}.$$

Die Summe aller Wahrscheinlichkeiten muß bekanntlich stets gleich 1 sein. Bei stetigen Variablen entspricht die Summe einem Integral. Die einfach Standard-normalverteilung ist bekanntlich durch die Dichte

$$f(z) = ae^{-z^2/2}, \quad -\infty < z < \infty$$

definiert, wobei a eine noch zu bestimmende Konstante ist. Es muß

$$\int_{-\infty}^{\infty} f(z)dz = \int_{-\infty}^{\infty} ae^{-z^2/2}dz = a \int_{-\infty}^{\infty} e^{-z^2/2}dz = 1$$

gelten. Aus der Analysis ist aber bekannt, dass

$$\int_{-\infty}^{\infty} e^{-z^2/2}dz = \sqrt{2\pi}$$

gilt, so dass $a\sqrt{2\pi} = 1$ folgt. Damit ergibt sich für die *Normierungskonstante* a

$$a = \frac{1}{\sqrt{2\pi}}.$$

Für die 2-dimensionale Dichte verfährt man analog und gelangt so auf den Ausdruck für die Normierungskonstante A .

Natürlich will man auch 3-, 4- und allgemein n -dimensionale Dichten anwenden können. Während man bei einer 1-dimensionalen Verteilung mit dem Integral die Fläche unter der Dichtefunktion berechnet, um die Normierungskonstante zu bestimmen, muß man bei 2- und mehrdimensionalen Dichten ein Volumen unter einer Fläche berechnen. Es wäre mühsam und unökonomisch, die Konstante für jedes n separat zu berechnen. Man macht also von einem allgemeinen Resultat der Analysis Gebrauch, demzufolge das Volumen unter einer Fläche, die durch die n -dimensionale Gaußsche Dichte definiert ist, durch die *Determinante* der Matrix gegeben ist, die der inversen Varianz-Kovarianz-Matrix Σ^{-1} entspricht. Die Determinante wird auf eine relativ komplizierte Weise aus den Elementen von Σ^{-1} berechnet, auf die hier nicht weiter eingegangen werden kann, zumal ein Modul zur Berechnung von Determinanten in jedem Statistikprogramm implementiert ist. Für den 2-dimensionalen Fall kann die Determinante allerdings leicht angegeben werden. Hat man etwa die Matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

so ist die Determinante durch

$$|M| = a \cdot d - b \cdot c$$

gegeben. Für Σ^{-1} erhält man demnach

$$|\Sigma^{-1}| = \frac{1}{(1-r^2)^2} \left(\frac{1}{\sigma_x^2 \sigma_y^2} - \frac{r^2}{\sigma_x^2 \sigma_y^2} \right) = \frac{1-r^2}{(1-r^2)^2} \left(\frac{1}{\sigma_x^2 \sigma_y^2} \right) = \frac{1}{(1-r^2) \sigma_x^2 \sigma_y^2}. \quad (399)$$

Der Vergleich mit der Definition von A zeigt, dass

$$A = \frac{1}{2\pi \sqrt{|\Sigma^{-1}|}} = \frac{1}{2\pi |\Sigma^{-1}|^{1/2}} \quad (400)$$

Für den allgemeinen, n -dimensionalen Fall erhält man

$$A = \frac{1}{(2\pi)^{(1/2)n} |\Sigma^{-1}|^{1/2}}, \quad (401)$$

wobei Σ^{-1} natürlich eine $(n \times n)$ -Matrix ist.

Literatur

- [1] Brachinger, HW, Ost F.: Modelle mit latenten Variablen: Faktorenanalyse, Latent-Structure-Analyse und LISREL-Analyse. In: Fahrmeier, L., Hamerle, A., Tutz, G. (Hrsg): Multivariate statistische Verfahren. Walter de Gruyter, Berlin, New York, 1996
- [2] Busemeyer, J.R., Jones, L. E. (1983) Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin*, 93 (3), 549 - 562
- [3] Cliff, N. (1988) The Eigenvalues-Greater-Than-One-Rule and the reliability of components. *Psychological Bulletin*, 103, 276-279
- [4] Christofferson, A. (1975) Factor analysis of dichotomized variables. *Psychometrika*, 40, 5-32
- [5] Davenport, M., Studdert-Kennedy, H. (1970) Use of orthogonal factors for selection of variables in a regression equation. *Applied Statistics*, 21, 324-333
- [6] Eckart, C., Young, G. (1936) The approximation of one matrix by another of lower rank. *Psychometrika*, 1, 211-218
- [7] Feller, W.: An introduction to probability theory and its applications, Vol. II, New York 1966
- [8] Gabriel, K.R. (1971) The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58, 453-467
- [9] Gabriel, K.R. (1978) Least squares approximation of matrices by additive and multiplicative models. *Journal of the Royal Society B*, 40, 186-196
- [10] Gould, J.: Der falsch vermessene Mensch. Frankfurt, 1988
- [11] Harman, H.H.: Modern Factor Analysis. Chicago, 1967
- [12] Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.*, 24 (6), 417 - 441
- [13] Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *Journal Educational Psychology*, 24 (7), 498-520
- [14] Gower, C., Hand, D.J.: Biplots. Chapman & Hill, London, 1996
- [15] Guttman, L: (1956) Image theory for the structure of quantitative variables, *Psychometrika*, 18, 277-296
- [16] Kelly, T.L. (1940) Comment on Wilson and Worcester's "Note on factor analysis". *Psychology*, 5, 117 - 120

- [17] Kenny, D. A., Judd, C. M. (1984) Estimating the nonlinear interactive effects of latent variables. *Psychological Bulletin*, 96 (1), 201 – 210
- [18] Magidson, J., Vermunt, J. K. (2002) Latent class models for clustering: a comparison with K-means. *Canadian Journal of Marketing Research*, 20, 37 – 44
- [19] Muthé, B. O. (2002) Beyond SEM: General latent variable modelling. *Behaviormetrika*, 29 (1), 81 – 117
- [20] Pearson, K. (1901) On lines and planes of closest fit to systems of points in space. *Phil. Mag.*, 6, 557–572
- [21] Rist, F., Glöckner-Rist, A., Demmel, R. (2009) The Alcohol Use Disorders Identification Test revisited: Establishing its structure using nonlinear factor analysis and identifying subgroups of respondents using latent factor analysis. *Drug and Alcohol Dependence*, 100, 71 –82
- [22] Spearman, C. (1904) General intelligence, objectively determined and measured. *American Journal of Psychology*, 15, 201 – 293
- [23] Thurstone, L.L. (1931) Multiple Factor Analysis, *Psychological Review*, 38, 406 – 427

Index

- Antiimage, 104
- Assoziativgesetz der Matrixmultiplikation, 40
- Basis, 30
- Basis eines Vektorraums, 32
- Basis- R -Faktorenmodell, 123
- Basisvektoren, 28
- Biplot, 81
- Cauchy-Schwarzsche Ungleichung, 24
- Datenreduktion, 10
- dual scaling, 71
- Eigenvektor, 49
- Eigenwert, 49
- eilbasis, 30
- Ellipse, 51
- Ellipsoid, 51
- Euklidische Distanz, 82
- Faktorenmuster, 78
- Faktorenstruktur, 78
- Faktorladungen, 79
- Faktorwerte, 10, 79
- Gleichungssystem
 - homogen, 44
 - inhomogen, 44
- Hauptachsentransformation, 56
- Hauptkomponentenanalyse, 78
- Hauptkomponentenmethode, 80
- idempotent, 42
- Image, 104
- Inverse, 59
- Karhunen-Loéve-Analyse, 111
- Kommunalität, 14, 77
- kommutativ, 40
- Korrelation
 - tetrachorische, 116
 - Vierfelder (ϕ -), 111
- Kosinussatz, 22
- Ladungen, 10
- LC Cluster Modelle, 121
- linear abhängig, 30
- linear unabhängig, 30
- Linearkombination, 19
- Mahalanobis-Distanz, 82
- Matrix
 - Diagonal, 44
 - symmetrische, 44
- multivariate Normalverteilung, 59
- Normalenvektor, 34
- normiert, 22
- O-Analyse, 110
- orthogonal, 23
- P-Analyse, 110
 - positiv definit, 50
 - positiv semi-definit, 50
- Pythagoras, 22
- Q-Analyse, 110
- quadratische Form, 50, 56
- R-Analyse, 110
- Rang, 43
 - Spalten-, 43
 - Zeilen-, 43
- Reifikation, 13
- Rotationsmatrix, 48
- S-Analyse, 111
- Schwierigkeitsfaktoren, 114
- scree-Test, 67
- Singularwertzerlegung, 63
- Skalarprodukt, 14, 20, 21
- Spaltenstandardisierung, 45
- Transformationsmatrix, 48
- Typen, 110

Varimax-Rotation, 108
Vektor, 17
Vektorraum (Definition), 29
Vektorraum, 28
Vektorraum, n -dimensionaler, 29
Zentrierungsmatrix, 42