

# DIE FAKTORENANALYSE — EIN MODERNES STATISTISCHES HILFSMITTEL DES GEOGRAPHEN FÜR DIE OBJEKTIVE RAUMGLIEDERUNG UND TYPENBILDUNG

DIETER STEINER

## *Einleitung*

Während eines längeren Aufenthaltes am Geographischen Institut der Universität Chicago hatte der Schreiber Gelegenheit, mit den dort gelehrten und geübten quantitativen Methoden bekannt zu werden. Die von Herrn Prof. Dr. Brian J. L. Berry erteilten Kurse in «Quantitative Methods» und «Regional Analysis» sowie die von ihm betreuten praktischen Übungen mit Verwendung der elektronischen Rechenanlage der Universität waren äußerst anregend. Sie machten es klar, daß sich für die geographische Forschung mit der Einführung statistischer Verfahren und dem Gebrauch von Computern ganz neue Horizonte eröffnen. Der Wunsch wurde wach, die Kenntnis dieser Methoden auch unter den Schweizer Geographen zu fördern, die, abgesehen von der rein beschreibenden Statistik, dieselben bis anhin recht selten benutzt haben<sup>1</sup>. Der vorliegende Artikel, der sich mit der erst vor wenigen Jahren in die Geographie eingeführten Faktorenanalyse befaßt, möchte dazu einen Beitrag leisten.

Ich möchte es nicht versäumen, an dieser Stelle Herrn Prof. Berry für das in den Kursen und in Diskussionen Gebotene, sowie Herrn Robert A. Murdie für seine Einführung in die Benützung des Rechenzentrums und in ein einfaches Programmieren herzlich zu danken.

## *Das Prinzip der Faktorenanalyse*

Die Faktorenanalyse ist für den Geographen von besonderem Interesse, ermöglicht doch ihre sinnvolle Anwendung die Lösung des Grundproblems der Geographie, nämlich die räumliche Gliederung, auf objektiver Grundlage. Es sei im folgenden versucht, das Wesen dieser Analyse kurz zu erläutern. Eine detaillierte Darstellung der mathematischen Grundlagen ist im Rahmen dieses Beitrages natürlich nicht möglich. Der Leser, der sich durch Selbststudium in die Faktorenanalyse einarbeiten möchte, sei auf die zuständige Fachliteratur verwiesen (Cooley und Lohnes (5, p. 151 und ff.), Fruchter (7), Harman (8), Kendall (9, p. 10 und ff.), und Thurstone (14).

Stellen wir uns vor, daß wir für eine Anzahl von Lokalitäten (Punkte auf, oder Flächenausschnitte aus der Erdoberfläche) numerische Beobachtungen für eine Serie von Veränderlichen zur Verfügung haben. Für  $m$  Orte und  $n$  Variablen lassen sich die Werte in einer  $m \times n$ -Matrix zusammenstellen:

$$\begin{array}{ccccccc}
 x_{11} & x_{12} & x_{13} & \cdots & \cdots & \cdots & x_{1n} \\
 x_{21} & x_{22} & x_{23} & \cdots & \cdots & \cdots & x_{2n} \\
 x_{31} & x_{32} & x_{33} & \cdots & \cdots & \cdots & x_{3n} \\
 \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 x_{m1} & x_{m2} & x_{m3} & \cdots & \cdots & \cdots & x_{mn}
 \end{array}$$

<sup>1</sup> Siehe dazu Jörg Roth: Mathematische Statistik und Geographie in der Schweiz. *Geographica Helvetica* 17 (3): 229—236, 1962.

Ein paarweiser Vergleich der an einer solchen Matrix beteiligten Variablen wird normalerweise ergeben, daß statistisch bedeutsame Korrelationen vorkommen. Eine hohe Korrelation zwischen zwei der Veränderlichen zeigt an, daß diese in starkem Maße in derselben Weise variieren und somit weitgehend das gleiche aussagen.

Es ist nun der Zweck der Faktorenanalyse, die Struktur dieser gegenseitigen Abhängigkeiten aufzudecken, Überflüssigkeiten zu eliminieren und, im Endeffekt, die Zahl der ursprünglichen Veränderlichen zu einer geringeren Zahl von neuen orthogonalen, d. h. nicht-korrelierten, Dimensionen zu reduzieren. Diese neuen Variablen werden Faktoren genannt. Sie können als die grundlegenden, voneinander unabhängigen Richtungen betrachtet werden, in denen die untersuchten Lokalitäten variieren können.

Die ursprünglichen Beobachtungen werden zunächst in standardisierte Werte<sup>2</sup> verwandelt. Danach erfolgt die Berechnung aller zwischen den n-Variablen möglichen Einfach-Korrelationen, was eine n x n-Korrelationsmatrix ergibt:

$$\begin{array}{cccccc}
 1,0 & r_{12} & r_{13} & \dots & r_{1n} \\
 r_{21} & 1,0 & r_{23} & \dots & r_{2n} \\
 r_{31} & r_{32} & 1,0 & \dots & r_{3n} \\
 \cdot & \cdot & \cdot & \dots & \cdot \\
 r_{n1} & r_{n2} & r_{n3} & \dots & 1,0
 \end{array}$$

Die Korrelation der Variablen mit sich selbst ist natürlich hundertprozentig, und somit sind die  $r_{jj}$ -Werte ( $j = 1, 2, \dots, n$ ) in der Hauptdiagonalen der Matrix alle 1,0. Da wir mit Standardwerten arbeiten, sind dies gleichzeitig auch die Streuungen der einzelnen Veränderlichen. Die Summierung über die Hauptdiagonale ergibt für die totale Streuung den Wert n.

Der Zusammenhang zwischen den alten und den neuen Variablen ist gegeben durch eine Serie linearer Transformationen:

$$\begin{array}{l}
 z_{1i} = a_{11} f_{1i} + a_{12} f_{2i} + \dots + a_{1p} f_{pi} \\
 z_{2i} = a_{21} f_{1i} + a_{22} f_{2i} + \dots + a_{2p} f_{pi} \\
 z_{3i} = a_{31} f_{1i} + a_{32} f_{2i} + \dots + a_{3p} f_{pi} \quad (i = 1, 2, \dots, m) \\
 \cdot \\
 z_{ni} = a_{n1} f_{1i} + a_{n2} f_{2i} + \dots + a_{np} f_{pi}
 \end{array}$$

Dabei sind die  $z_{ji}$  ( $j = 1, 2, \dots, n$ ) die standardisierten Beobachtungswerte für die ursprünglichen Variablen, die  $f_{hi}$  ( $h = 1, 2, \dots, p$ ) die Werte, die den einzelnen Lokalitäten in den neuen Dimensionen zukommen, und die  $a_{jh}$  ( $j = 1, 2, \dots, n$  und  $h = 1, 2, \dots, p$ ) die Koeffizienten, die das Gewicht anzeigen, mit dem die der gesamten Streuung zugrunde liegenden Faktoren zur Bildung der z-Werte beitragen. Diese Koeffizienten werden auch Faktorladungen genannt. p entspricht der Zahl der Faktoren und soll, wenn möglich, kleiner als n sein. Das obige Gleichungssystem kann in abgekürzter Matrix-Schreibweise wie folgt dargestellt werden:

$${}_n Z_m = {}_n A_p {}_p F_m$$

<sup>2</sup> Standardisierte Werte erhält man, indem man für jede Variable von den individuellen Beobachtungen den Mittelwert subtrahiert und die Ergebnisse durch die Standardabweichung dividiert. In standardisierter Form haben alle Veränderlichen den Mittelwert 0,0 und die Streuung 1,0.

Die Faktorladungen sind Elemente von Eigenvektoren der Korrelationsmatrix. Es muß eine Matrix  $A$  mit Elementen  $a_{jh}$  gefunden werden, so daß die Gleichung

$${}_n A_p \cdot {}_p A'_n = {}_n R_n$$

gilt. Für die praktische rechnerische Lösung müssen die Eigenvektoren einer nach dem andern bestimmt werden. Man berechnet in einem ersten Schritt die  $a$ -Koeffizienten für Faktor 1 so, daß die Summe ihrer Quadrate ein Maximum ist, und bildet danach das Produkt  $A(1) A(1)'$ . Mit  $A(1)$  bezeichnen wir den die Elemente  $a_{j1}$  ( $j = 1, 2, \dots, n$ ) enthaltenden Eigenvektor. Die resultierende Matrix wird von der Korrelationsmatrix  $R$  subtrahiert, womit eine erste Residualmatrix  $R(1)$  entsteht. Es ist also

$${}_n R(1)_n = {}_n R_n - {}_n A(1)_{1 \ 1} A(1)'_n .$$

Darauf verfährt man mit der Residualmatrix genau gleich wie mit der ursprünglichen Korrelationsmatrix und extrahiert Faktor 2, so daß die Summe der quadrierten Faktorladungen wiederum ein Maximum ist. Das geht so weiter, bis alle Faktoren gebildet sind<sup>3</sup>.

Diese Behandlung der Korrelationsmatrix führt zur Auscheidung von Faktoren mit absteigendem Grad der Generalität. Der erste Faktor erklärt den größtmöglichen Prozentsatz der gesamten Streuung, der zweite einen maximalen Anteil an der restlichen Streuung usw. Für jeden Faktor ergibt die Summe der quadrierten Faktorladungen die von ihm erklärte Streuung oder seinen Eigenwert  $\lambda_h$ . Die Matrixmultiplikation  $A' A$  liefert die Matrix  $\Lambda$ , deren Hauptdiagonale sämtliche Eigenwerte  $\lambda_h$  ( $h = 1, 2, \dots, p$ ) enthält, also

$${}_p A'_n \cdot {}_n A_p = {}_p \Lambda_p .$$

Die Summe aller Eigenwerte ist gleich der totalen Streuung  $n$ .

Es stellt sich die Frage, wann mit der Extraktion von Faktoren aufgehört werden soll. Faktoren, deren Anteil an der Streuung so klein ist, daß er keine statistische Bedeutung mehr besitzt, sollten offensichtlich unberücksichtigt bleiben. Das Problem geeigneter Prüfverfahren zur Beantwortung dieser Frage ist noch nicht völlig geklärt, und man folgt gewöhnlich dem einfachen Vorschlag Kaisers, mit der Faktorierung aufzuhören, wenn  $\lambda_h$  unter 1,0 sinkt<sup>4</sup>.

<sup>3</sup> Es ist an dieser Stelle darauf hinzuweisen, daß die Extraktion der Faktoren grundsätzlich von zwei verschiedenen Korrelationsmatrizes ausgehen kann. Der erste einfachere Fall entspricht der oben erläuterten Situation, in der alle Werte der Hauptdiagonalen der Korrelationsmatrix 1,0 sind. Die auf eine solche Matrix angewendete Faktorenanalyse führt zur Bildung der grundlegenden orthogonalen Dimensionen oder sogenannten Hauptkomponenten, die die gesamte vorkommende Streuung erklären («principal component analysis»). Im zweiten Fall setzt man für die Hauptdiagonalelemente der Korrelationsmatrix die sogenannten Kommunalitäten ein, d.h. die Streuungsanteile der einzelnen alten Veränderlichen, die durch die Abhängigkeit von den neuen Faktoren erklärt werden. Die Kommunalität von z. B.  $z_1$  entspricht den quadrierten Faktorladungen  $a_{1h}$  summiert über die  $p$  Faktoren. Da die Kommunalitäten primär unbekannt sind, müssen dafür in der Korrelationsmatrix Schätzwerte verwendet werden. Bei der am meisten verwendeten Lösung nimmt man als Näherungswerte für die Kommunalitäten die quadrierten multiplen Korrelationskoeffizienten. Man stützt sich also auf die Korrelationen, die zwischen der Variablen  $z_j$  und allen übrigen Veränderlichen  $z_1, z_2, \dots, z_{j-1}, z_{j+1}, \dots, z_n$  herrschen. Diese zweite Methode der Faktorenanalyse baut sich somit nur auf die gemeine, nicht auf die totale Streuung auf. Es ist dies, im Gegensatz zur Hauptkomponentenanalyse, die Faktorenanalyse i. e. S. («factor analysis»). Es hängt vom Verwendungszweck ab, welches der beiden Verfahren mit Vorteil gebraucht wird. Für viele Probleme ergeben beide fast identische Lösungen.

<sup>4</sup> Siehe Harman (8), p. 363.

Die nun vorliegende Lösung ist meist nicht optimal und kann durch eine Rotation der Achsen verbessert werden. Stellen wir uns die  $m$  Beobachtungen als Punkte im  $n$ -dimensionalen Raum vor. Wenn wir nun Orte gleicher Punktdichte miteinander verbinden, erhalten wir Hyperellipsoid-Schalen. Nach dem oben beschriebenen Verfahren erklärt der erste der ausgeschiedenen Faktoren ein Maximum der vorhandenen Streuung, und er entspricht geometrisch somit der längsten Achse, die durch diese Hyperellipsoide gelegt werden kann. Der zweite Faktor stellt die nächstlängste Achse dar, und so weiter. Die nach diesem Prinzip der größtmöglichen Allgemeinheit bestimmten neuen Veränderlichen werden nun hohe und tiefe, aber oft auch mittlere Faktorladungen aufweisen, was die Interpretation der Bedeutung der Faktoren erschwert. Die Faktorladungen sind nämlich identisch mit den Korrelationen zwischen alten und neuen Variablen und zeigen daher, wie die ursprünglichen Veränderlichen zur Bildung der Faktoren beitragen. Um die Struktur der Faktoren und damit ihre Interpretierbarkeit zu vereinfachen, wird eine abgeleitete Lösung angestrebt, bei der die alten Variablen wenn möglich nur auf einem Faktor eine hohe Ladung haben. Dies kann mit der Achsenrotation nach dem sogenannten Varimax-Kriterium erreicht werden<sup>5</sup>.

Es gilt nun noch, die  $f_{hi}$  ( $h = 1, 2, \dots, p$  und  $i = 1, 2, \dots, m$ ), d.h. die Werte zu bestimmen, die den  $m$  Lokalitäten in bezug auf die neuen Variablen zukommen. Es werden dafür verschiedene Verfahren angewendet. Das einfachste besteht darin, das Transformationsgleichungssystem wie folgt umzuordnen:

$$\begin{aligned} f_{1i} &= a_{11} z_{1i} + a_{21} z_{2i} + \dots + a_{n1} z_{ni} \\ f_{2i} &= a_{12} z_{1i} + a_{22} z_{2i} + \dots + a_{n2} z_{ni} \\ f_{3i} &= a_{13} z_{1i} + a_{23} z_{2i} + \dots + a_{n3} z_{ni} \\ &\dots \\ f_{pi} &= a_{1p} z_{1i} + a_{2p} z_{2i} + \dots + a_{np} z_{ni} \end{aligned} \quad (i=1, 2, \dots, m)$$

Dies entspricht der Matrixgleichung

$${}_p F_m = {}_p A' {}_n Z_m.$$

Es ist daran zu denken, daß die  $f$ -Werte keine tatsächlich beobachteten Meßwerte sind, sondern abstrakte Konstruktionen. Wenn ihre Bedeutung aus der Faktorenstruktur nicht ohne weiteres ersichtlich ist, gibt oft eine kartographische Darstellung ihrer Verteilung den nötigen Aufschluß.

#### *Ein einfaches hypothetisches Beispiel und eine geometrische Erklärung*

Das Gesagte sei mit einem einfachen Beispiel illustriert. Wir nehmen der Einfachheit halber an, daß wir es mit nur drei Beobachtungen für drei Ausgangsvariablen zu tun haben, die die in Tabelle 1 aufgeführten Werte ergeben haben. Wir können uns z.B. vorstellen, daß die drei Lokalitäten klimatische Stationen sind und daß die drei Veränderlichen irgendwelche Temperatur-, Niederschlags- und Sonnenscheindauereinheiten darstellen.

Tab. 1 Theoretische Werte für drei Orte und drei Variablen

Klimatische Station Nr.	$x_1$ Temperatur	$x_2$ Niederschlag	$x_3$ Sonnenschein
1	1,7	0,8	2,1
2	2,3	1,9	1,0
3	3,5	2,7	1,7

<sup>5</sup> Zur Definition des Varimax-Kriteriums siehe Harman (8), p. 301 u. ff.

Zunächst verwandeln wir die obige Matrix  $X$  mit Elementen  $x_{ij}$  in die Matrix  $Z$  mit standardisierten Werten  $z_{ij}$ :

$$Z: \begin{matrix} & -0,873 & -1,048 & 0,898 \\ & -0,218 & 0,105 & -1,077 \\ & 1,092 & 0,943 & 0,180 \end{matrix}$$

Danach berechnen wir die Korrelationen  $r_{jk}$  und erhalten die Matrix  $R$ :

$$R: \begin{matrix} & 1,000 & 0,961 & -0,176 \\ & 0,961 & 1,000 & -0,442 \\ & -0,176 & -0,442 & 1,000 \end{matrix}$$

Wie sich aus den Korrelationskoeffizienten ersehen läßt, sind  $x_1$  und  $x_2$  sehr stark miteinander assoziiert ( $r_{12} = 0,961$ ), während die Korrelationen zwischen  $x_1$  und  $x_3$  und  $x_2$  und  $x_3$  bedeutend geringer sind. Dieser Befund erlaubt schon jetzt die Feststellung, daß sich die Variablen 1 und 2 ohne wesentlichen Informationsverlust zu einer einzigen neuen Dimension zusammenschließen lassen.

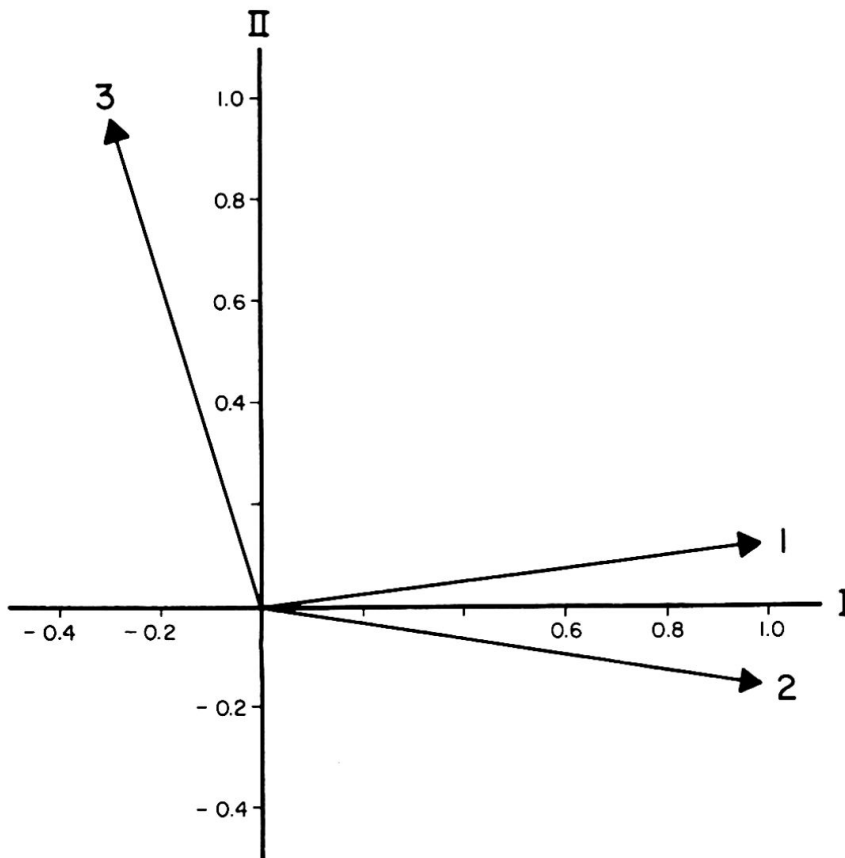
Nach dem bei Harman (8), p. 160 und ff. oder bei Kendall (9) p. 19 und ff. beschriebenen schrittweisen Verfahren werden nun die orthogonalen Faktoren extrahiert, was schließlich zur Matrix  $A$  der Faktorladungen führt, die in Tab. 2 dargestellt sind.

Tab. 2 Faktorladungen und Faktoreigenwerte

Alte Variablen	Faktoren			Streuung der alten Variablen
	I	II	III	
1	0,983	0,120	0,141	1,000
2	0,979	-0,159	0,130	1,000
3	-0,296	0,954	0,045	1,000
Eigenwert	2,012	0,949	0,039	3,000 = Total der Streuung der alten Variablen
% Streuung erklärt	67,1	31,6	1,3	= Summe der Faktoreigenwerte

Das Resultat zeigt eine sehr einfache Faktorenstruktur, die durch Rotation nicht mehr verbessert werden könnte. Jede der ursprünglichen Variablen hat eine hohe Ladung auf nur einem der Faktoren. Faktor I mit einem Eigenwert von rund 2 erklärt zwei Drittel der gesamten Streuung. Er ist sehr hoch mit den Ausgangsvariablen 1 (Temperatur) und 2 (Niederschlag) korreliert und faßt somit diese beiden Veränderlichen zu einer Dimension zusammen, die von Kalt-trocken zu Heiß-feucht geht. Der Eigenwert von Faktor II ist zwar kleiner als 1,0, jedoch nur unbedeutend. Wie aus der hohen Korrelation ersichtlich ist, stellt diese zweite neue Dimension fast unverändert die alte Variable 3 (Sonnenscheindauer) dar, und sie kann deshalb ohne wesentlichen Verlust an Information nicht vernachlässigt werden. Dagegen ist Faktor III keine grundlegend neue Dimension mehr, sondern baut sich im wesentlichen aus Fehlerschwankungen auf, die einen sehr kleinen Eigenwert ergeben. Zusammenfassend läßt sich somit sagen, daß die Variation von Ort zu Ort, die ursprünglich durch Messungen auf drei verschiedenen Skalen beschrieben war, sich zu praktisch hundert Prozent in nur zwei nicht-korrelierten Richtungen bewegt.

Fig. 1 zeigt eine geometrische Darstellung des Zusammenhanges zwischen alten und neuen Veränderlichen. Die Ausgangsvariablen erscheinen dabei als Vektoren im Achsen-system der Faktoren. Der Cosinus des Winkels zwischen zwei Vektoren ist identisch mit der Korrelation zwischen den zwei durch die Vektoren wiedergegebenen Variablen. Die Cosini der Winkel zwischen den Vektoren 1, 2 und 3 entsprechen also den Korrelationen zwischen den alten Variablen, diejenigen der Winkel zwischen denselben Vek-



Figur 1  
 Der geometrische Zusammenhang zwischen den alten Variablen (1, 2 und 3) und den neuen orthogonalen Dimensionen (I und II) bei dem im Text erläuterten hypothetischen Beispiel einer einfachen Faktorenanalyse

toren und den Achsen I und II stellen die Korrelationen zwischen den ursprünglichen Variablen und den Faktoren dar. Die beiden Faktoren selbst stehen senkrecht aufeinander. Der Winkel von  $90^\circ$  ergibt einen Cosinus (= Korrelation) von 0,0.

*Klassifikation der Beobachtungen mit Hilfe von Distanzgruppierung und Trennverfahren*

Die Werte  $f_{hi}$  ( $h = 1, 2, \dots, p$  und  $i = 1, 2, \dots, m$ ) zeigen die Lage der untersuchten Orte im Raum der  $p$  Faktoren als Punkte an. Die Distanz zwischen den Punkten ist ein Maß für die Ähnlichkeit der zugehörigen Beobachtungen. Eine Gruppierung der  $m$  Orte auf Grund dieser Distanzen im Faktorenraum führt somit zu einer objektiven Klassifikation. Es werden zunächst alle möglichen paarweisen Distanzen  $d_{ik}$  mit Hilfe der Beziehung

$$d_{ik} = \sqrt{(f_{1i} - f_{1k})^2 + (f_{2i} - f_{2k})^2 + \dots + (f_{pi} - f_{pk})^2}$$

berechnet. Dies ergibt eine  $m \times m$ -Distanzenmatrix:

0	$d_{12}$	$d_{13}$	...	$d_{1m}$
$d_{21}$	0	$d_{23}$	...	$d_{2m}$
$d_{31}$	$d_{32}$	0	...	$d_{3m}$
...	...	...	...	...
$d_{m1}$	$d_{m2}$	$d_{m3}$	...	0

Nun sucht man die kleinste vorkommende Distanz und vereinigt die beiden zugehörigen Orte zu einer Gruppe. Die Position dieser Gruppe im Koordinatensystem ist dabei

gegeben durch den Mittelpunkt zwischen den beiden vereinigten Punkten. Damit hat man die Zahl der Punkte um einen reduziert. Soweit sich die Distanzen verändert haben, werden sie neu berechnet, was zu einer  $(m-1) \times (m-1)$ -Matrix führt. Dann verfährt man gleich mit der nächstkleinsten Distanz und geht so weiter, bis alle  $m$  Lokalitäten in einer einzigen Gruppe vereinigt sind. Von diesem Endpunkt des Prozesses kann man sodann wieder rückwärts gehen. Ein erster Rückwärtsschritt ergibt die optimale Aufspaltung der Beobachtungen in zwei Gruppen, ein zweiter diejenige in drei Gruppen und so fort. Optimal heißt dabei, daß die derart ausgeschiedenen Gruppen eine maximale Einheitlichkeit aufweisen. Es ist klar, daß im Falle von geographischen Untersuchungen, bei denen Beobachtungen als Punkte (z. B. klimatische Stationen) auf oder als Flächenausschnitte (z. B. administrative Bezirke) aus der Erdoberfläche einen gegebenen Raum lückenlos erfüllen, eine solche Gruppierung zu einer geographischen Gliederung dieses Raumes führt. Falls sich benachbarte Lokalitäten bei diesem Prozeß zusammenschließen, erhält man eine Unterteilung in Regionen. Wenn dies dagegen für nicht-benachbarte Orte geschieht, ist das Resultat eine Typenausscheidung. Das letztere trifft beispielsweise zu für die im nächsten Abschnitt erwähnte Studie über die englischen Städte.

Allerdings ist die nun vorliegende Klassifikation noch nicht unbedingt eine endgültige. Wir haben bei der Distanzgruppierung nur auf die Entfernung der noch nicht klassierten Punkte von den Gruppenzentren abgestellt, jedoch die Streuung der schon klassierten Punkte um diese Zentren nicht berücksichtigt. Dies kann geschehen durch die Anwendung statistischer Trennverfahren («discriminant analysis» oder «discriminatory analysis»). Die nach dem obigen Verfahren erhaltene Gruppierung wird dabei geprüft und allenfalls korrigiert. Zudem vermittelt das Trennverfahren auch die Möglichkeit, neue Beobachtungen den bestehenden Gruppen zuzuordnen. Wir können an dieser Stelle nicht weiter auf diese Methoden eintreten und müssen den Leser auf die Darstellungen in Fisher (6), Kendall (9) und Linder (10) verweisen.

#### *Bisherige Anwendungen der Faktorenanalyse in geographischen Studien*

Die Faktorenanalyse ist bisher in erster Linie zur Klassifikation von geographischen Räumen in wirtschaftlicher Hinsicht angewendet worden. Ein kleineres Beispiel für Regionen der Vereinigten Staaten, dem vor allem methodische Bedeutung zukommt, ist von Berry in (3) dargelegt worden. Bei den größeren Arbeiten stand nicht nur die Gruppierung von Beobachtungseinheiten zu Regionen im Vordergrund, sondern ebenso sehr die strukturelle Analyse der Verflechtung der untersuchten Variablen. Man hoffte dabei, mit der Kristallisation von orthogonalen Dimensionen zu einer besseren Definition und einem besseren Verständnis von Begriffen wie «wirtschaftliche Gesundheit», «wirtschaftliche Stagnation», «Unterentwicklung» usw. zu gelangen. Einer zweiten Kategorie von Untersuchungen gehört sodann Mosers und Scotts Klassifikationsversuch der englischen Städte an (11). Im folgenden werden Problemstellungen und Resultate dieser Arbeiten in knapper tabellenartiger Form zusammengefaßt. Der an Details interessierte Leser kann die zugehörigen Publikationen im Original konsultieren.

##### 1 *Autoren: Thompson et al. (13):*

*Beobachtungseinheiten:* Die 58 Counties (inkl. New York City) des amerikanischen Staates New York.

*Anzahl und Typ der berücksichtigten Variablen:* 9 Veränderliche, die das Niveau des wirtschaftlichen Zustandes wie auch der wirtschaftlichen Entwicklung charakterisieren, z. B. Zunahme der Bevölkerung in den letzten Jahren, durchschnittliche Arbeitslosigkeit in den letzten Jahren, Zunahme der Beschäftigung in den letzten Jahren, Einkommen pro Kopf der Bevölkerung 1958.

*Problemstellung:* Objektive Erfassung von wirtschaftlichen Gesundheits- oder Krankheits-symptomen und regionale Gliederung der Counties;

*Typ der angewendeten Analyse:* Hauptkomponentenanalyse.

*Anzahl und Generalitätsgrad der resultierenden Faktoren:* 3 Faktoren, die zusammen 75,5% der gesamten Streuung erklären (Faktor I: 50,1%; Faktor II: 15,7%; Faktor III: 9,7%);

*Interpretation der Faktoren:* Faktor I erhält hohe Ladungen von fast allen alten Veränderlichen und ist somit ein allgemeiner wirtschaftlicher Zustands- und Entwicklungsmaßstab; er stellt

ferner ein Kontinuum dar. Faktor II hat dichotomen Charakter und repräsentiert den Gegensatz zwischen städtischen und ländlichen Gebieten. Counties mit großen städtischen Agglomerationen haben eine Tendenz zu höheren Werten auf diesem Faktor, die ungesunde wirtschaftliche Aspekte in ihrem Bereich anzeigen. Faktor III ist eine Dimension des ökonomisch-demographischen Wachstums, auf der Zuwachs, z. B. der Bevölkerung und des Detailhandels, mit Arbeitslosigkeit kontrastieren. Counties rund um New York und im Hudson/Mohawk-Korridor rangieren hoch auf dieser Skala und weisen auf den Bevölkerungszustrom hin, der noch stets in diese Gebiete hinein stattfindet, wobei aber die Beschäftigungsmöglichkeiten nachhinken.

## 2 Autor: Berry (1 und 2):

*Beobachtungseinheiten:* 95 Länder;

*Anzahl und Typ der berücksichtigten Variablen:* 43 sozio-ökonomische Indizes wie Bevölkerungsdichte, Bevölkerungszuwachs, Prozentsatz der kultivierten Landfläche, nationales Produkt, Energieproduktion und -konsumtion, internationaler Handel usw.

*Problemstellung:* Frage der «Unterentwicklung». Lassen sich die untersuchten Nationen eindeutig in eine Gruppe der weniger und eine Gruppe der höher entwickelten einordnen?

*Typ der angewendeten Analyse:* Hauptkomponentenanalyse.

*Anzahl und Generalitätsgrad der resultierenden Faktoren:* 3 Faktoren, die zusammen 90,9% der gesamten Streuung erklären (Faktor I: 84,2%; Faktor II: 4,2%; Faktor III: 2,5%);

*Interpretation der Faktoren:* Faktor I ist eine sehr starke allgemeine Dimension, die sich in erster Linie aus den technologischen und organisatorischen Indizes wie Transport, Handel, Energieproduktion und -konsumtion, nationales Produkt, Nachrichtenverkehr, Verstädterung usw. aufbaut. Er stellt eine kontinuierliche Skala dar, und eine Aufspaltung in deutlich differenzierte Entwicklungsgrade kann nicht festgestellt werden. Faktor II entspricht einer demographischen Skala, die sich aus Komponenten wie Bevölkerungsdichte, Bevölkerungszuwachs, Geburts- und Todesraten, Kindersterblichkeit, kultivierte Landfläche, Reisproduktion usw. zusammensetzt. Es läßt sich eine gewisse Tendenz erkennen, daß Länder mit tiefen Werten auf Faktor I, also mit geringem Industrialisierungsgrad, hohe Positionen auf der zweiten Dimension einnehmen, d. h. hohe Bevölkerungsdichte usw. aufweisen. Faktor III isoliert eine Besonderheit, die von diesen allgemeinen Tendenzen abweicht, und kombiniert hohen Bevölkerungszuwachs, hohe Geburtsraten, geringe Energiekonsumtion, kleines nationales Produkt und geringen Inland-Warentransport mit hohen Werten für internationalen Handel und Postsendungen. Es sind vor allem die Länder Zentral- und des nördlichen Südamerikas, die in dieser Richtung hervortreten, und man könnte sie zu einer Gruppe der «armen Handelsnationen» zusammenfassen;

*Regionale Gliederung der Beobachtungen:* Die untersuchten Länder wurden auf Grund ihrer Positionen auf den Faktoren I und II versuchsweise in regionale Gruppen eingeordnet und die Berechtigung dieser Zuordnung mit Hilfe statistischer Trennverfahren geprüft. Das Resultat ergab folgende Gruppierung: 1) Westeuropa, Nordamerika, Australien/Neuseeland und Südafrika; 2) Sowjetblock und Japan; 3) Südamerika; 4) Zentralamerika; 5) Nordafrika; 6) Asien; 7) Afrika südlich der Sahara.

## 3 Autoren: Ray und Berry (12):

*Beobachtungseinheiten:* 120 Counties der kanadischen Provinzen Ontario und Quebec.

*Anzahl und Typ der berücksichtigten Variablen:* Total 88 Veränderliche wirtschaftlicher und sozialer Art. Es wurden separate Faktorenanalysen für den ganzen Satz von Variablen und für vier Untergruppen durchgeführt. Diese umfaßten a) Kulturelle Daten (10 Veränderliche); b) Demographische Daten (13); c) Daten über Wohnverhältnisse (19), und d) Landwirtschaftliche Daten (29).

*Problemstellung:* Sozio-ökonomische Gliederung Zentralkanadas;

*Typ der angewendeten Analyse:* Faktorenanalyse i. e. S.;

*Anzahl und Generalitätsgrad der resultierenden bedeutsamen Faktoren:* a) Kulturelle Struktur: 2 Faktoren, die zusammen über 90% der gemeinen Streuung (Kommunalität) erklären. Faktor I gibt den Wechsel von Französisch/Katholisch zu Deutsch (und Englisch)/Protestantisch an. Faktor II stellt ein ländlich-städtisches Kontinuum dar.

b) Demographische Struktur: 4 Faktoren mit einem Anteil von über 92% an der gemeinen Streuung. Faktor I ist ein Maßstab für Unterschiede in Zuwachsraten; es handelt sich dabei wiederum um den Gegensatz zwischen ländlichen und städtischen Gebieten. Faktor II zeigt die Intensität der Einwanderung an und fällt räumlich mit der obigen Französisch/Englisch-Dimension zusammen. Counties mit französisch sprechender Bevölkerung haben, im Gegensatz zu den Counties in Ontario, eine sehr tiefe Einwanderungsrate. Faktor III ist ein Pionierfaktor und vereinigt geringe Bevölkerungsdichte mit Männerüberschuß und umgekehrt. Faktor IV zeigt die Verschiebung der Einwanderungsfrontier an und kontrastiert Gebiete mit hoher Einwanderung



vor 1931 mit solchen, die erst nach dem Zweiten Weltkrieg eine starke Zuwanderung erlebt haben.

c) Struktur der Wohnverhältnisse: 3 Faktoren erklären 84% der Kommunalität. Faktor I ist wiederum eine Stadt/Land-Dimension und drückt Unterschiede in Wert, Eigentumsverhältnissen und Zuwachs von Wohngebäuden aus. Faktor II gibt die Unterschiede in den Wohnverhältnissen zwischen den französisch und englisch sprechenden Gebieten an. Faktor III entspricht der Pionierdimension und stellt ältere und neuere Gebäude einander gegenüber.

d) Landwirtschaftliche Struktur: 5 Faktoren, die 84% der gemeinen Streuung erklären. Auf Faktor I erscheint wieder der Französisch/Englisch-Gegensatz, diesmal in Form eines Wechsels im landwirtschaftlichen Nutzungssystem. Faktor II separiert Counties auf Grund von Unterschieden in der Nutzungsintensität, die, zum Teil wenigstens, mit dem Stadt/Land-Gegensatz zusammenfallen. Die übrigen Faktoren identifizieren regionale Spezialisierungen, nämlich Milchwirtschaft (Faktor III), Früchte und Gemüse (Faktor IV) und Weizen und Geflügel (Faktor V).

e) Allgemeine Struktur: In den obigen Teilanalysen haben sich einige der Faktoren stets wiederholt. So erschienen mehrmals die Unterschiede zwischen den französisch und den englisch sprechenden Counties, zwischen den städtischen und den ländlichen Gebieten und zwischen den Pionier- und Altsiedelgebieten. Erwartungsgemäß sollten diese Faktoren deshalb auch in der allgemeinen Analyse erscheinen. Tatsächlich ergibt diese 7 Faktoren, die zusammen 74% der gemeinen Streuung erklären, wobei Faktor I (33,3%) dem Französisch/Englisch-Gegensatz entspricht, Faktor III (12,6%) den Wechsel von Altsiedel- zu Pioniergebieten anzeigt und Faktor VI (3,1%) die Stadt/Land-Dimension wiederholt. Die landwirtschaftlichen Spezialisierungen erscheinen auf Faktor II (13,6%), Faktor IV (4,5%) und Faktor V (4,3%). Faktor VII (2,6%) schließlich ist ein Industrialisierungsmaßstab.

#### 4 Autoren: Moser und Scott (11):

*Beobachtungseinheiten:* Die 157 Städte in England und Wales mit 50 000 und mehr Einwohnern (1951).

*Anzahl und Typ der berücksichtigten Variablen:* 60 Indizes für Bevölkerungszahl, -struktur und -veränderung, Wohnverhältnisse, Wirtschaft, soziale Zustände, Wahlen, Gesundheit und Erziehung.

*Problemstellung:* Objektive Klassifikation der Städte.

*Typ der angewendeten Analysen:* Hauptkomponentenanalyse.

*Anzahl und Generalitätsgrad der resultierenden bedeutsamen Faktoren:* 4 Faktoren, die zusammen für 60,4% der totalen Streuung verantwortlich sind. (Faktor I: 30,1%; Faktor II: 13,2%; Faktor III: 9,8%; Faktor IV: 7,3%).

*Interpretation der Faktoren:* Faktor I ist eine Dimension der sozialen Verhältnisse, auf der sich Unterschiede in der beruflichen Ausbildung, im Bildungsgrad, der politischen Orientierung und den Geburtsraten widerspiegeln. Faktor II erfaßt die Bevölkerungsveränderungen von 1931 bis 1951 und stellt die Vororte den alten Siedlungszonen gegenüber. Die ersteren haben bessere Wohnverhältnisse, eine höhere Zahl von Pendlern, eine stärkere Vertretung in den jüngeren Altersklassen, doch einen kleineren Umfang des Detailhandels, da der Ausbau der Einkaufsdienste nachhinkt. Faktor III entspricht der Bevölkerungsveränderung von 1951 bis 1958, und Faktor IV ist eine Dimension der Überbevölkerung, auf der die Zahl der Haushaltungen mit mehr als anderthalb Personen pro Zimmer und der Prozentsatz der 1- bis 3-Zimmer-Wohnungen hohe Ladungen haben.

*Klassifikation der Beobachtungen:* Auf Grund der Punktlagen im vierdimensionalen Faktorraum erfolgte eine zunächst vorläufige visuelle Zuordnung der Städte zu Gruppen, die dann statistisch geprüft und allenfalls korrigiert wurde, um die Distanzsummen innerhalb der Gruppen zu minimalisieren. Beispiele für die ausgeschiedenen Stadttypen sind: Küstenkurorte, administrative Zentren, Handelsstädte, Eisenbahnzentren, Häfen, Textilindustriestädte, Bergbaustädte, Metallindustriestädte, Wohnvororte.

Während in diesen Beispielen die Faktorenanalyse zur Ausscheidung von Einheitsregionen oder Typen in formaler Hinsicht benützt wurde, zeichnet sich die Verwendbarkeit derselben statistischen Methode zur Erfassung von funktionalen Räumen als weitere äußerst interessante Möglichkeit ab. Berry hat in (4) über einen Versuch berichtet, die wirtschaftsorganisatorische Struktur Indiens mit Hilfe der Faktorenanalyse herauszuarbeiten. Er verwendete Daten für den Inlandtransport von 63 verschiedenen Waren von und zu 36 Regionen, die die Staaten Indiens und größere Hafenstädte umfassen. Für das Zahlenmaterial des Jahres 1959 wurden verschiedene Faktorenanalysen (im ganzen 270!) durchgeführt, um folgendes zu definieren: a) Für jeden Warentyp die Gruppen von Sendern, von denen nach ähnlichen Zielen transportiert wird, und die Gruppen von Empfängern, die von ähnlichen Ursprungsorten Güter erhalten; b) Für jede Ursprungsregion die Gruppen von Waren, die nach gleichen Zielen gehen, und die Gruppen von Bestimmungsorten, in denen ähnliche Güterkombinationen ankommen; c) Für jede Empfängerregion die Gruppen von Gütern, die einen ähnlichen Ursprung haben, und die Gruppen von

Sendern, die übereinstimmende Kombinationen von Waren verschicken. Der Vergleich mit Stichproben aus andern Jahren ließ eine hohe Stabilität der funktionalen Struktur Indiens erkennen. Einzelheiten der Untersuchung, die noch weitergeführt wird, sind noch nicht veröffentlicht worden.

Es sei hier schließlich noch darauf hingewiesen, daß der Schreiber einige Versuche unternommen hat, die Faktorenanalyse für objektive klimatische Gliederungen zu benützen. Darüber wird jedoch später an anderer Stelle zu berichten sein.

### *Eine Gruppierung der Schweizer Kantone auf Grund von 7 Variablen*

Zum Schluß möchten wir zur weiteren Veranschaulichung des Wesens der Faktorenanalyse die statistische Klassifikation der Schweizer Kantone auf Grund von 7 Variablen demonstrieren. Die in Tab. 3 aufgeführten Werte für diese Veränderlichen wurden einer Faktorenanalyse unterworfen. Den Resultaten darf keine größere Bedeutung beigemessen werden, da erstens die Auswahl der berücksichtigten Daten eine recht zufällige war<sup>6</sup> und zweitens die Kantone als ausgedehnte Einheiten variabler Größe ohnehin keine vernünftige Basis für eine regionale Gliederung der Schweiz abgeben können. Der Wert dieses Beispiels ist in erster Linie illustrativ. Es wäre reizvoll, einmal eine detailliertere Analyse von gemeindeweißen Daten vorzunehmen, und wir glauben, daß sich z.B. das Problem der Abgrenzung der Berggebiete auf diese Weise objektiv lösen ließe.

Tab. 3 Datenmatrix: 25 Beobachtungen (Kantone) für 7 Veränderlichen

Kanton	Variablen						
	1	2	3	4	5	6	7
1 Zürich	551	22,5	3,2	1,36	96	14,2	296
2 Bern	129	10,9	3,5	0,69	99	11,9	264
3 Luzern	170	13,5	4,0	1,40	285	9,7	169
4 Uri	30	12,1	4,2	0,12	204	10,1	110
5 Schwyz	86	9,8	4,1	0,40	225	10,8	104
6 Obwalden	47	4,6	4,3	0,24	176	5,5	65
7 Nidwalden	81	14,4	4,3	0,40	319	8,3	104
8 Glarus	59	6,6	3,5	0,20	124	22,2	276
9 Zug	220	24,3	4,2	1,15	342	14,7	148
10 Fribourg	95	0,3	4,1	1,18	117	7,2	145
11 Solothurn	254	17,8	3,6	0,87	74	23,2	215
12 Baselstadt	6081	14,8	2,8	0,46	63	13,7	854
13 Baselland	346	37,9	3,5	1,04	71	17,9	434
14 Schaffhausen	221	14,7	3,3	0,96	63	23,0	255
15 Appenzell-AR	202	2,0	3,2	1,19	267	11,9	100
16 Appenzell-IR	75	3,6	3,9	0,64	307	4,3	119
17 St. Gallen	168	9,8	3,6	0,71	141	15,4	175
18 Graubünden	21	7,6	3,9	0,11	27	5,2	235
19 Aargau	257	20,0	3,7	1,21	66	19,5	192
20 Thurgau	165	11,1	3,6	2,29	117	17,8	168
21 Tessin	70	11,7	3,2	0,11	10	11,7	304
22 Waadt	134	13,8	3,1	0,80	98	8,4	289
23 Wallis	34	11,7	3,9	0,17	11	5,6	211
24 Neuenburg	185	15,2	3,0	0,66	155	18,6	304
25 Genf	919	27,8	2,8	1,55	135	10,2	594

Erklärung zu den Variablen: 1: Bevölkerungsdichte 1960 (Personen/km<sup>2</sup>); 2: Bevölkerungszuwachs in % 1950—1960; 3: Mittlere Anzahl von Personen pro Haushaltung 1960; 4: Intensität der agrarischen Landnutzung (Acker- und Wiesland/Weiden und Wald) 1955 resp. 1952; 5: Mittlere Größe einer landwirtschaftlichen Parzelle in a 1955; 6: Beschäftigte in der Industrie 1961 in % der Wohnbevölkerung 1960; 7: Steuereinnahmen pro Kopf der Bevölkerung in Fr. 1960. (Quelle: Statistisches Jahrbuch der Schweiz 1962.)

<sup>6</sup> Wir waren auf Daten angewiesen, die uns in Chicago aus dem Statistischen Jahrbuch der Schweiz zur Verfügung standen.

Die Daten wurden in folgender Weise verarbeitet: Zunächst zeichneten wir für jede Variable ein Häufigkeitsverteilungsdiagramm. Diese graphischen Darstellungen zeigten, daß einige der Verteilungen eine positive Asymmetrie aufwiesen. Diese konnte mit Hilfe einer Transformation zu logarithmischen Werten eliminiert werden. Auf die derart transformierten Werte erfolgte danach die Anwendung einer Hauptkomponentenanalyse auf der IBM 7090-Rechenanlage der Universität Chicago unter Benützung des Programmes MESA 20 (UCSM 303) von C. Bradford. Die folgende Korrelationsmatrix wurde produziert:

1,000	0,503	-0,653	0,594	0,098	0,472	0,649
0,503	1,000	-0,289	0,340	-0,082	0,443	0,542
-0,653	-0,289	1,000	-0,261	0,297	-0,429	-0,794
0,594	0,340	-0,261	1,000	0,434	0,386	0,119
0,098	-0,082	0,297	0,434	1,000	0,011	-0,466
0,472	0,443	-0,429	0,386	0,011	1,000	0,369
0,649	0,542	-0,794	0,119	-0,466	0,369	1,000

Die Endresultate der Faktorenanalyse, d. h. die Faktorladungen und Faktoreigenwerte sowie die Faktorscores sind in Tab. 4 und 5 dargestellt. Eine Achsenrotation wurde auch versucht. Diese lieferte jedoch praktisch identische Ergebnisse.

Tab. 4 Matrix der Faktorladungen, mit Faktor-Eigenwerten, Prozentsatz der erklärten Streuung und Kommunalität über 2 Faktoren.

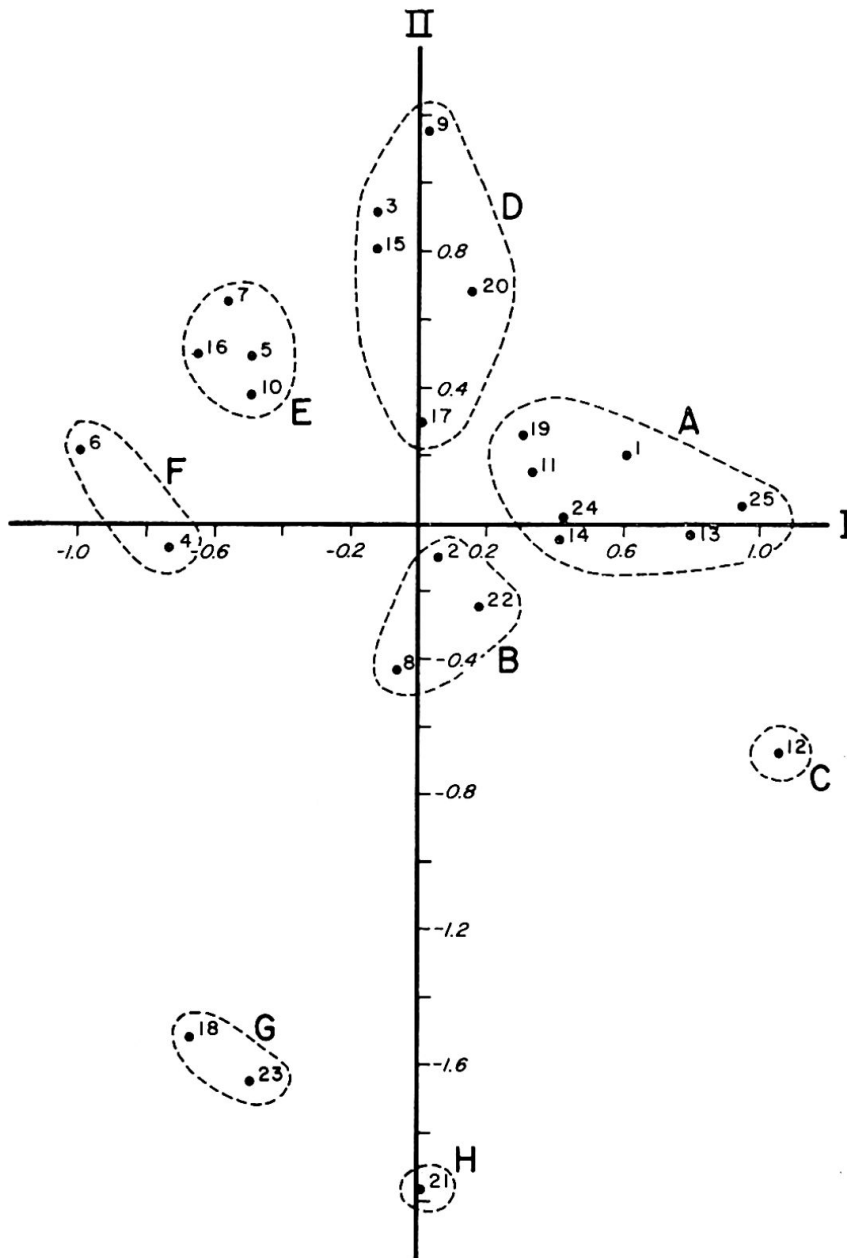
Alte Variablen	Faktoren							Kommunalität über zwei Faktoren
	I	II	III	IV	V	VI	VII	
1	0,868	0,225	-0,226	-0,155	-0,105	0,317	-0,092	0,804
2	0,689	0,058	0,619	-0,344	-0,045	-0,108	-0,083	0,478
3	-0,812	0,280	0,406	-0,122	0,055	0,249	0,133	0,738
4	0,534	0,712	-0,128	-0,077	0,421	-0,065	0,061	0,792
5	-0,155	0,901	-0,112	-0,049	-0,365	-0,116	0,055	0,836
6	0,672	0,190	0,315	0,640	-0,041	0,046	0,026	0,488
7	0,837	-0,450	-0,073	-0,148	-0,105	-0,016	0,243	0,903
Eigenwert	3,348	1,689	0,733	0,596	0,339	0,195	0,100	
% Streuung erklärt	47,8	24,1	10,5	8,5	4,8	2,8	1,4	

Tab. 5 Wertung der 25 Kantone durch die zwei ersten Faktoren

Kanton	Faktor		Kanton	Faktor	
	I	II		I	II
1 Zürich	0,614	0,203	14 Schaffhausen	0,432	-0,051
2 Bern	0,065	-0,105	15 Appenzell-AR	-0,117	0,795
3 Luzern	-0,123	0,906	16 Appenzell-IR	-0,669	0,497
4 Uri	-0,727	-0,081	17 St. Gallen	0,007	0,296
5 Schwyz	-0,488	0,488	18 Graubünden	-0,666	-1,516
6 Obwalden	-0,991	0,222	19 Aargau	0,314	0,265
7 Nidwalden	-0,560	0,660	20 Thurgau	0,163	0,683
8 Glarus	-0,062	-0,445	21 Tessin	0,006	-1,970
9 Zug	0,033	1,148	22 Waadt	0,180	-0,248
10 Fribourg	-0,493	0,384	23 Wallis	-0,494	-1,647
11 Solothurn	0,342	0,151	24 Neuenburg	0,426	0,015
12 Baselstadt	1,061	-0,675	25 Genf	0,953	0,051
13 Baselland	0,803	-0,027			

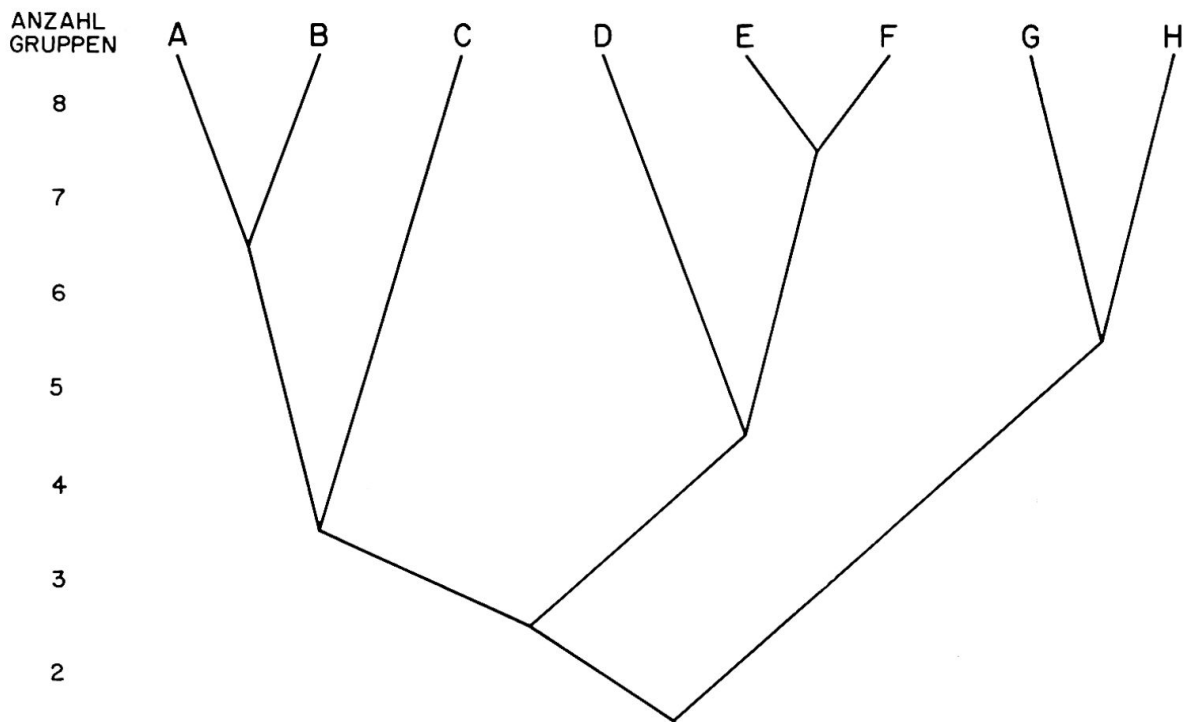
Zwei Faktoren haben Eigenwerte größer als 1,0. Der erste ist der allgemeinere: Er erklärt rund 50% der gesamten Streuung. Die beiden Faktoren zusammen sind für rund 72% der totalen Streuung verantwortlich. Hohe positive Ladungen auf Faktor I

Figur 2  
Die Lage der Kantone  
und ihre Gruppierung  
im zweidimensionalen  
Faktorenraum



haben Bevölkerungsdichte und Steuerzahlungen, eine hohe negative Ladung hat die Zahl der Personen pro Haushalt. Ferner sind auch Bevölkerungszuwachs und Prozentsatz der Beschäftigten in der Industrie in positivem Sinne wesentlich an der Bildung von Faktor I beteiligt. Auch die Landnutzungsintensität weist noch eine Ladung von 0,53 auf, wobei allerdings der entsprechende Wert für Faktor II größer ist. Faktor I ist somit eine Dimension der allgemeinen wirtschaftlichen Entwicklung, der Verstädterung und Industrialisierung. Faktor II zeigt die Veränderungen in landwirtschaftlicher Hinsicht an. Er setzt sich aus größeren positiven Komponenten der Landnutzungsintensität und des Parzellierungsgrades der Landwirtschaftsbetriebe zusammen.

In Fig. 2 ist die Position der Kantone im zweidimensionalen Faktorenraum graphisch dargestellt. Schon rein visuell lassen sich deutliche Gruppierungstendenzen erkennen. Um willkürliche Entscheide in Zweifelsfällen auszuschalten, muß jedoch eine rechnerische Gruppenbildung auf der Grundlage der zwischen den einzelnen Punkten gemessenen Distanzen vorgenommen werden. Fig. 3 zeigt die Hierarchie der Aufspaltung in



Figur 3 Aufspaltungsstammbaum für die in Figur 2 dargestellten Gruppen

Gruppen von maximaler Einheitlichkeit bis zum Detailniveau 8, wenn vom Endprodukt der Distanzgruppierung, einer einzigen Gruppe, wieder rückwärts gegangen wird. Aus dieser Darstellung ist ersichtlich, daß der größte Unterschied, der für die erste Aufspaltung in zwei Gruppen verantwortlich ist, zwischen den drei Bergkantonen Wallis, Graubünden und Tessin und allen übrigen Kantonen herrscht. Die nächste Aufteilung erfolgt im wesentlichen zwischen den Mittelland/Jura- und den randalpiner Kantonen. Bereits im nächsten Schritt wird Baselstadt von der Gruppe der Mittelland/Jura-Kantone einzeln abgetrennt, womit die Sonderstellung dieses Halbkantons als fast reines städtisches Gebiet zum Ausdruck kommt.

Die optimale Aufteilung in 8 Gruppen zeigt schließlich folgendes Bild: Gruppe A umfaßt die reinen Mittelland- oder Jurakantone mit hohen Werten für Faktor I (hoher Industrialisierungsgrad, starke Verstädterung usw.) und mittleren Werten für Faktor II (mäßige Nutzungsintensität und Güterzersplitterung). In Gruppe B befinden sich Bern, Waadt und Glarus mit einer Position nahe dem Nullpunkt des Koordinatensystems, also mit mittleren Skores in beiden Richtungen. Für Bern und Waadt erklärt sich dies leicht aus der flächenmäßigen Ausdehnung dieser Kantone über die drei Hauptnaturregionen der Schweiz. Die bestehenden Kontraste in den Bevölkerungs- und Wirtschaftsverhältnissen haben die Tendenz, sich gegenseitig aufzuheben. Für Glarus ist die Situation etwas anders. Obwohl in vieler Hinsicht ein eigentlicher Bergkanton, weist Glarus einen überdurchschnittlichen Industrialisierungsgrad und hohe Steuereinnahmen auf. Baselstadt erscheint allein in Gruppe C mit den höchsten Werten von allen Kantonen auf der demographisch-industriellen Skala, womit die bereits erwähnte Sonderstellung dieses Stadtkantons zum Ausdruck kommt. In Gruppe D finden wir eine Reihe von Kantonen im Übergangsgebiet vom Mittelland zu den Alpen. Sie haben eine mittlere Stellung auf der ersten Dimension und die höchste Position auf der zweiten und heben sich deshalb durch besondere Intensität der landwirtschaftlichen Nutzung und einen geringen Parzellierungsgrad heraus. Das letztere ist aus der Dominanz von Streusiedlung und Blockflur leicht erklärbar. Die angezeigte hohe Nutzungsintensität ist dagegen etwas irreführend, da der Intensitätsindex definiert war als Quotient von Acker- und Wiesland über Weideland und Wald. Die Intensitätsunterschiede

zwischen Ackerbau und Graswirtschaft kommen also nicht zur Geltung. Die Gruppen E und F umfassen randalpine Kantone. F skort am tiefsten auf Achse I und hat eine mittlere Stellung in bezug auf Achse II. E zeigt für beide Dimensionen etwas höhere Werte, womit der hier gegenüber F doch noch etwas stärkere wirtschaftliche Entwicklungsgrad angetönt ist. Wallis und Graubünden sind in Gruppe G vereinigt, die auf beiden Skalen sehr tief rangiert. Damit kommen die besonderen Probleme dieser Bergkantone gut zum Ausdruck, nämlich die kleine Bevölkerungsdichte, die geringe Industrialisierung, die extensive Landwirtschaft, die starke Güterzersplitterung, die eher kinderreichen Familien usw. Der Tessin schließlich erscheint wieder als ein Einzelgänger in Gruppe H mit stärkerer Industrialisierung und höheren Steuereinnahmen als Gruppe G. Dagegen hat der Tessin die tiefsten Werte von allen Kantonen auf der landwirtschaftlichen Skala, was auf die im Durchschnitt äußerst extensive Landwirtschaft und auf die extreme Güterzersplitterung zurückgeführt werden kann.

Es sei weiter noch darauf hingewiesen, daß sich in A die Sonderstellung der Kantone Baselland und Genf abzeichnet, obschon diese auf dem Detailniveau 8 noch nicht als eine separate Gruppe erscheinen. Nach Baselstadt haben diese beiden Kantone die höchsten Ränge für Faktor I, und dies insbesondere infolge eines sehr starken Bevölkerungszuwachses, womit der Suburbanisierungsprozeß in diesen Gebieten zum Ausdruck kommt.

Die oben beschriebene Gruppierung wurde in diesem Fall nicht mit statistischen Trennverfahren geprüft. Sonst wäre wahrscheinlich eine Korrektur erfolgt und der Kanton St. Gallen der Gruppe A, nicht der Gruppe D zugeschlagen worden. Wie aus Fig. 2 ersichtlich ist, ist der St. Gallen ähnlichste Kanton der Aargau, der sich in Gruppe A befindet. Trotzdem wird St. Gallen mit den Kantonen in D vereinigt, da im Laufe des Distanzgruppierungsprozesses die Entfernung zwischen den zu St. Gallen und dem Aargau gehörigen Punkten nie eine minimale ist. Es ist dies ein Beispiel dafür, daß die Distanzgruppierung oft nur eine Annäherung an eine optimale Lösung liefert.

Es ist zu beachten, daß die Faktorenanalyse nicht ein Allerweltsmittel zur Lösung geographischer Klassifikationsprobleme darstellt. Sie ist mit einer ganzen Reihe von Problemen behaftet, denen in Zukunft vermehrte Beachtung geschenkt werden sollte. Es sind dies in erster Linie:

1) Die Ergebnisse der Analyse hängen zweifellos von der Art der gewählten Beobachtungseinheiten ab. Bei Flächen spielen die absolute Größe der Einheiten an sich und die Variabilität der Größe in einem Satz von Einheiten eine Rolle. Bei Punkten sind ihre Dichte und ihre Verteilung ausschlaggebend.

2) Die Resultate sind natürlich beeinflusst von der Art der der Analyse unterworfenen ursprünglichen Veränderlichen. Die Auswahl der Variablen ist deshalb ein kritischer Schritt.

3) Die Frage ist bisher ungeklärt, ob man allen resultierenden Faktoren gleiche Wichtigkeit beimessen oder ob ihr Einfluß auf die Klassifikation der Beobachtungseinheiten nach der Größe ihrer Eigenwerte abgestuft werden soll.

#### LITERATUR

- (1) Berry, Brian J. L.: An inductive approach to the regionalization of economic development. Research Paper No. 62 («Essays on geography and economic development», herausg. von Norton S. Ginsburg): 78-107, Dept. of Geography, Univ. of Chicago, Chicago 1960. — (2) ders.: Basic patterns of economic development. Research Paper No. 68 («Atlas of economic development», herausg. von Norton S. Ginsburg): 110-119, Dept. of Geography, Univ. of Chicago, Chicago 1961. — (3) ders.: A method for deriving multi-factor uniform regions. *Przegląd Geograficzny* 23: 263-282, Warschau 1963. — (4) ders.: A multi-commodity framework for the economic regionalization of India. Mitt. an den 20. Int. Geographenkongreß, London 1964. Zusammenfassung in: 20th Int. Geogr. Congress, Abstracts of Papers: 178, Nelson, London 1964. — (5) Cooley, William W. und Paul R. Lohnes: *Multivariate procedures for the behavioral sciences*. J. Wiley and Sons, Inc., New York/London 1962. — (6) Fisher, R. A.: *Statistical methods for research*

workers. 11. Aufl., Oliver and Boyd, Edinburgh/London 1950. — (7) Fruchter, B.: An introduction to factor analysis. D. van Nostrand, Inc., New York 1954. — (8) Harman, Harry H.: Modern factor analysis. 2. Aufl., The Univ. of Chicago Press, Chicago 1962. — (9) Kendall, M. G.: A course in multivariate analysis. Griffin's Statistical Monographs and Courses No. 2, Hafner Publishing Co., New York 1961. — (10) Linder, Arthur: Statistische Methoden für Naturwissenschaftler, Mediziner und Ingenieure. 3. Aufl., Birkhäuser Verlag, Basel/Stuttgart 1960. — (11) Moser, C.A. und W. Scott: British towns. A statistical study of their social and economic differences. Oliver and Boyd, Edinburgh 1961. — (12) Ray, D. Michael und Brian J. L. Berry: Multi-variate socio-economic regionalization: A pilot study in central Canada. Mitt. an die Canadian Polit. Science Assoc. Conf. on Statistics, Charlottetown 1964. — (13) Thompson, John H., Sidney C. Sufrin, Peter R. Gould und Marion A. Buck: Toward a geography of economic health: The case of New York State. Annals of the Assoc. of Amer. Geogr. 52 (1): 1-20, 1962. — (14) Thurstone, L. L.: Multiple factor analysis. 6. Aufl., The Univ. of Chicago Press, Chicago 1961.

#### FACTOR ANALYSIS — A MODERN STATISTICAL TOOL OF THE GEOGRAPHER FOR OBJECTIVE REGIONALIZATION AND TYPE SEPARATION

The use of statistical methods has not been very popular among Swiss geographers. To further the knowledge of such methods their attention is drawn to the possibilities offered by the so-called factor analysis, a statistical technique introduced into geographical research only recently. Factor analysis, in conjunction with distance grouping and discriminant analysis, may provide an objective solution to the basic problem of geography: regionalization. This paper reviews the principle of factor analysis. This is followed by a summary of its practical applications to geographical research in recent years. The use of factor analysis has concentrated on problems of regionalization on the basis of measures of economic health, stagnation, and underdevelopment. By way of illustration, the author presents in the last section the results of an analysis of the 25 Swiss Cantons measured on 7 variables, such as population density, population increase, intensity of land use, percent employment in manufacturing, etc.

## GEOGRAPHISCHE ÜBUNGEN AN DER OBERSTUFE DER MITTELSCHULE

ERFAHRUNGEN AM KANTONALEN GYMNASIUM WINTERTHUR

FRITZ SCHIESSER

An vielen Maturitätsschulen der Schweiz werden die Lehrpläne überprüft, ob diese den heutigen Vorstellungen von der Bildungsaufgabe der höheren Schule genügen. Da die Kantonsschule Winterthur schon 1948 probeweise den Lehrplan des Gymnasiums einer Revision unterzog, die Zahl der Unterrichtsstunden reduzierte und dafür mehr praktische Übungen einführte, gelangen immer wieder Anfragen von Kollegen anderer Schulen zu uns, welche Erfahrungen bis jetzt gemacht worden seien. Wir kommen dem allgemeinen Interesse gerne entgegen, indem wir hier Bedeutung und Möglichkeiten des Faches Geographie im Rahmen der Schulreform umreißen.

#### A. DIE STELLUNG DER GEOGRAPHIE IM BILDUNGSPLAN DER MODERNEN MITTELSCHULE

Zu den wichtigsten Aufgaben einer Maturitätsschule gehören Vermittlung einer allgemeinen Bildung, Vorbereitung für die Hochschule und Heranbildung der Persönlichkeit. Neben der ethischen Erziehung sollen nicht nur Kenntnisse vermittelt, sondern ebenso stark muß die Fähigkeit gefördert werden, das erworbene Wissen zu verarbeiten. Praktische Übungen erhalten dabei die Funktion einer Ergänzung und Vertiefung des Fachunterrichts.