

JÜRGEN BÄHR

Regionalisierung mit Hilfe von Distanzmessungen

Will man ein größeres Gebiet nach bestimmten Gesichtspunkten in mehrere homogene Teilräume untergliedern, so läßt sich dafür nur in den seltensten Fällen ohne die Hilfe von mathematisch-statistischen Verfahren eine befriedigende Lösung angeben. Vielfach sind die dazu nötigen Berechnungen nur durch den Einsatz moderner Großrechenanlagen zu bewältigen.

Im folgenden sollen zwei verschiedene Verfahren zur Regionalisierung mit Hilfe von Distanzmessungen näher beschrieben und an einem Beispiel erläutert werden. Die dabei durchgeführten Rechnungen wurden auf einer IBM 7090 der Gesellschaft für Mathematik und Datenverarbei-

tung im Institut für Angewandte Mathematik der Universität Bonn durchgeführt.

Gewöhnlich setzt sich ein Untersuchungsraum aus einer größeren Anzahl von kleinräumlichen Einheiten zusammen (Landkreise, Gemeinden), auf die sich alle statistischen Angaben beziehen. Um zu größeren, möglichst homogenen Teilregionen des Untersuchungsraumes zu kommen, müssen jeweils mehrere der kleinsten Bezugseinheiten zusammengefaßt werden. Solange man bei einer derartigen Gruppenbildung nur eine einzige Variable (z. B. Einwohnerzahl, Beschäftigte in der Landwirtschaft) zugrunde legt, läßt sich das Problem noch relativ leicht lösen. Die einzige Schwie-

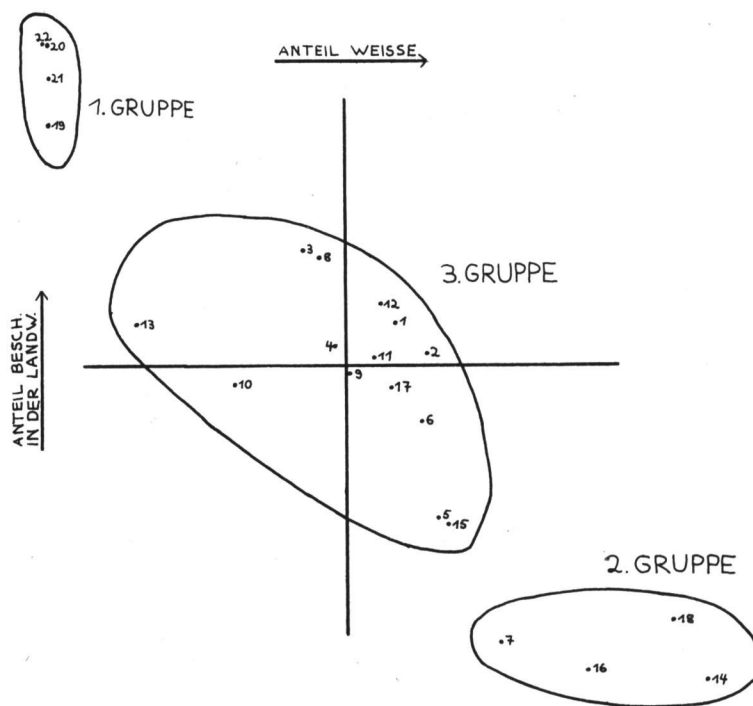


Abb. 1: Lage der Magistratsbezirke Südwesafrikas
im zweidimensionalen Koordinatensystem (standardisierte Werte)
und Zusammenfassung zu Gruppen

rigkeit liegt hier in der Wahl geeigneter Schwellenwerte. Sehr häufig ist man aber bestrebt, bei einer Regionalisierung mehrere Gesichtspunkte zu berücksichtigen. In derartigen Fällen haben sich Distanzmessungen in einem n -dimensionalen Raum zur Feststellung von Ähnlichkeiten gut bewährt.

Während diese Methode bisher in deutschsprachigen Arbeiten nur ganz vereinzelt angewandt wurde (Kilchenmann [8]*, Steiner [10]), lassen sich aus der neueren amerikanischen Literatur zahlreiche Beispiele anführen (insbes. Berry und seine Schule), die auf dem gleichen Grundgedanken – der Messung von Abständen im mehrdimensionalen Raum – beruhen, sich im einzelnen aber in der Wahl des Gruppierungskriteriums etwas unterscheiden.

Ein recht einfaches und überschaubares Verfahren gab Berry [2] in der kürzlich auch in deutscher Übersetzung erschienenen Untersuchung von neun amerikanischen Zensusbezirken an. In einer späteren Arbeit (Berry [4]) schlug er eine Verfeinerungsmöglichkeit vor, um zu möglichst gleich großen Teilregionen zu kommen. Eine genaue und vollständige Darstellung der einzelnen Gruppierungsschritte ist besonders für das zweite, etwas aufwendigere Verfahren nicht angegeben.

Im folgenden werden beide Methoden ausführlich und schrittweise behandelt, um ein späteres Nachvollziehen mit anderem Zahlenmaterial zu erleichtern. Beide Algorithmen sollen schließlich am gleichen Beispiel erprobt werden, da-

mit die unterschiedlichen Ergebnisse besser verdeutlicht werden können.

Ein beliebiger Untersuchungsraum möge sich in n Beobachtungseinheiten möglichst gleicher Größe untergliedern, und zur Auffindung homogener Teilgebiete sollen r Merkmale (Variable) herangezogen werden. Folgende Schritte müssen dann der Reihe nach ausgeführt werden:

1. Prüfung der Korrelation zwischen den Variablen

Bei beiden Regionalisierungsverfahren wird vorausgesetzt, daß die Variablen voneinander unabhängig sind und somit als die orthogonalen Koordinatenachsen in einem r -dimensionalen Raum aufgefaßt werden können. Anders gesagt, bedeutet diese Forderung, daß zwischen den für die Untergliederung ausgewählten Merkmalen keine oder doch nur eine geringfügige Korrelation bestehen darf. Im 2-dimensionalen Fall ließe sich eine Korrelation zwischen 2 Variablen x und y noch relativ einfach durch eine Achsendrehung ausgleichen (vgl. [1]). Der Winkel φ zwischen den beiden Achsen x und y muß dabei so gewählt werden, daß der Korrelationskoeffizient r mit dem $\cos \varphi$ übereinstimmt.

Sollen jedoch mehr als 2 korrelierende Variable die Grundlage für die Regionalisierung bilden, so müßte zunächst eine Faktorenanalyse durchgeführt werden, um die Ausgangsvariablen auf voneinander unabhängige und somit orthogonale Faktoren zurückzuführen. Zur Faktorenanalyse sei auf die einschlägige Literatur [11] verwiesen. Im weiteren wird davon ausgegangen, daß die zur Gliederung herangezogenen Größen voneinander unabhängig sind.

* Die Zahlen in [] beziehen sich auf die Literaturangaben am Schluß des Beitrages.

2. Umwandlung in standardisierte Variable

Um den Einfluß von verschiedenen Maßeinheiten auszuschalten, ist eine Überführung der Ausgangsdaten in ihre standardisierte Form nötig. Das erfüllt die folgende einfache Transformation:

$$(1) \quad z_i = \frac{x_i - \bar{x}}{s_x}$$

wobei x_i Wert der Variablen x für den i -ten Beobachtungsräum; $\bar{x}$ Mittelwert der x_i ; s_x Standardabweichung der x_i .

Wird diese Transformation für sämtliche r Variable ausgeführt, so werden damit alle Mittelwerte 0 und alle Standardabweichungen 1.

3. Distanzmessungen

Die n Beobachtungseinheiten werden als Punkte im r -dimensionalen Raum dargestellt. Für den 2dimensionalen Fall ist ein Beispiel angegeben (Abb. 1). Je näher 2 Punkte zusammenliegen, desto größer ist die Ähnlichkeit zwischen den Beobachtungseinheiten, die sie darstellen. Umgekehrt ähneln sich 2 Raumeinheiten in bezug auf die herangezogenen Merkmale nur sehr wenig, wenn sie im Koordinatensystem weit voneinander entfernt liegen.

Für die Abstandsmessung zwischen zwei Punkten P und Q im r -dimensionalen Raum kann eine einfache Erweiterung des Satzes von *Pythagoras* herangezogen werden:

$$(2) \quad PQ^2 = \sum_{i=1}^r (x_{i1} - x_{ij})^2$$

wobei x_{i1} die Koordinaten von P und x_{ij} die Koordinaten von Q sind.

Die Distanzquadrate zwischen je zwei Punkten lassen sich am einfachsten in Matrixform wie folgt darstellen:

Beob. Einheiten

$$\begin{matrix} & \begin{matrix} 1 & 2 & \dots & i & \dots & j & \dots & n \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ i \\ \vdots \\ j \\ \vdots \\ n \end{matrix} & \begin{bmatrix} d_{11}^2 = 0 & & & & & & \\ d_{21}^2 & 0 & & & & & \\ d_{i1}^2 & d_{i2}^2 & 0 & & & & \\ d_{j1}^2 & d_{j2}^2 & d_{ji}^2 & 0 & & & \\ d_{n1}^2 & d_{n2}^2 & d_{ni}^2 & d_{nj}^2 & 0 & & \end{bmatrix} \end{matrix} \quad \text{1. Schritt} \rightarrow$$

$$\begin{matrix} & \begin{matrix} 1 & 2 & i/j & n \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ i/j \\ \vdots \\ n \end{matrix} & \begin{bmatrix} 0 & & & \\ d_{21}^2 & 0 & & \\ d_{i/j1}^2 & d_{i/j2}^2 & 0 & \\ d_{n1}^2 & d_{n2}^2 & d_{ni/j}^2 & 0 \end{bmatrix} \end{matrix}$$

Die Matrix der Distanzquadrate enthält in der Diagonalen jeweils Nullen, und sie ist symmetrisch, da bei n Punkten insgesamt nur $n \cdot (n-1) : 2$ verschiedene Abstände errechnet werden können ($d_{ij}^2 = d_{ji}^2$).

4a. 1. Gruppierungsalgorithmus: Minimierung der Distanzquadrate zwischen den Gruppen

Ausgangspunkt dieses Verfahrens ist die oben angegebene Matrix der Distanzquadrate (d_{ik}^2); es genügt dabei, nur die Hälfte ober- oder diejenige unterhalb der Diagonalen zu betrachten. Im 1. Schritt wird in dieser Matrix der kleinste Wert der d_{ik}^2 gesucht und die beiden Punkte, zwischen denen dieses kleinste Abstandsquadrat gemessen wurde, zu einer Gruppe zusammengefaßt. Die i -te und die j -te Zeile und Spalte werden anschließend durch eine einzige, neue Zeile bzw. Spalte ersetzt, in der die Distanzquadrate vom Mittelpunkt der gebildeten Gruppe zu allen anderen Punkten (in späteren Schritten auch Gruppen) enthalten sind. Dazu können folgende Formeln dienen:

Berechnung der Koordinaten des Mittelpunktes (Schwerpunktes) von 2 Gruppen (bzw. Punkten):

$$(3) \quad s_l = \frac{x_{i1} \cdot n_i + x_{j1} \cdot n_j}{n_i + n_j} \quad l = 1 \dots r$$

wobei x_{i1} und x_{j1} die Koordinaten der Mittelpunkte der Gruppen i und j ; s_l die Koordinaten des Schwerpunktes; n_i und n_j Anzahl der Punkte in den Gruppen i bzw. j sind.

Berechnung der Distanzquadrate vom Mittelpunkt der zusammengefaßten Gruppe zu allen anderen Punkten bzw. Gruppen:

$$(4) \quad d_{i/jk}^2 = \sum_{i=1}^r (x_{ki} - s_i)^2$$

wobei x_{ki} Koordinate des Mittelpunktes einer beliebigen Gruppe k ; $d_{i/jk}^2$ Distanzquadrat zwischen Mittelpunkt der aus i und j neu gebildeten Gruppe und dem Mittelpunkt der Gruppe k sind.

Während vor Beginn der Rechnung alle Punkte noch nicht gruppiert waren, entsteht im 1. Gruppierungsschritt aus der $n \cdot n$ -Matrix eine $(n-1) \cdot (n-1)$ -Matrix, danach eine $(n-2) \cdot (n-2)$ -Matrix; schließlich wird im $(n-2)$ -ten Schritt eine $2 \cdot 2$ -Matrix errechnet, die die Distanzquadrate zwischen zwei Gruppen, in die die ursprünglichen Beobachtungseinheiten zusammengefaßt wurden, enthält. Es besteht nach der Durchführung dieses Algorithmus die Möglichkeit, den Untersuchungsraum in eine beliebige Anzahl von Teilräumen aufzugliedern.

Mit einer Zusammenfassung der Originalbeobachtungen in nur wenige Gruppen geht jedoch jeweils eine immer stärker werdende Generalisation einher. Es ist natürlich anzustreben, eine Gruppierung so vorzunehmen, daß der Informationsverlust, der jeweils durch die Verallgemeinerung er-

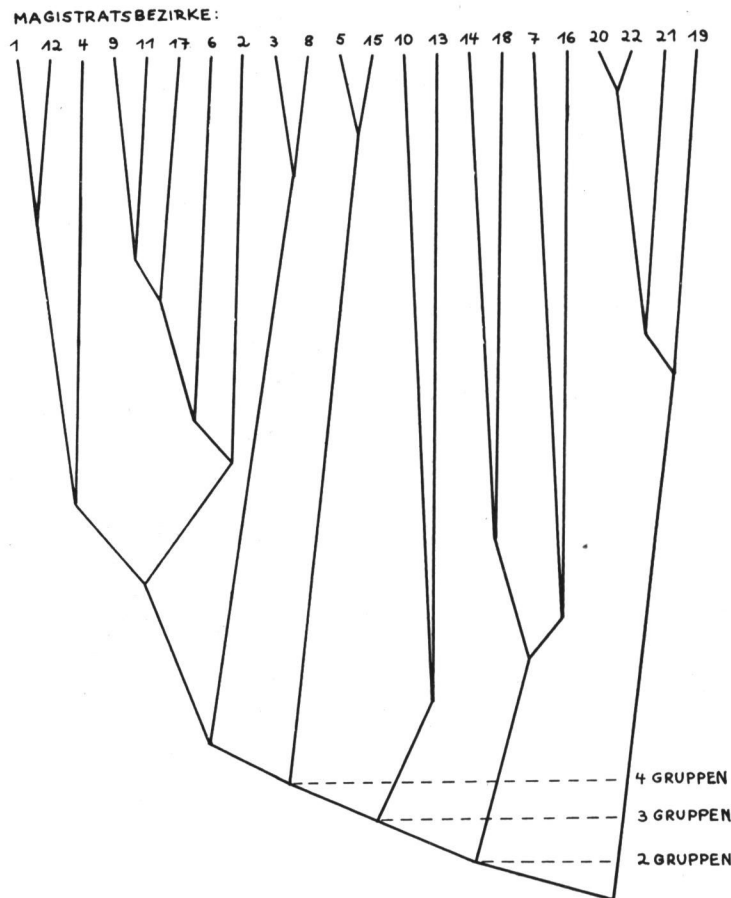


Abb. 2: Stammbaum beim 1. Gruppierungsverfahren

kauft werden muß, nicht zu groß wird. Deshalb ist es ratsam, diesen Verlust an Information für jeden Schritt zu berechnen und erst danach zu entscheiden, in wieviel Gruppen aufgeteilt werden sollte. Eine Maßzahl dafür gibt die Summe der Distanzquadrate innerhalb der Gruppen (SG) und außerhalb der Gruppen (SA) an. Vor der Gruppierung ist $SG = 0$, da ja noch keine Gruppe gebildet wurde, SA umfaßt demgegenüber die Summe aller zwischen den Einzelpunkten möglichen Distanzquadrate. Nach dem ersten Schritt fällt SG mit dem Quadrat des Abstandes zwischen den gruppierten Punkten zusammen, während SA um den gleichen Betrag vermindert werden muß. Sobald nicht mehr Einzelpunkte, sondern bereits bestehende Gruppen zu größeren Einheiten zusammengefaßt werden, erhöht sich SG nach jedem Schritt um mehrere Werte, da ja in diesem Fall die Anzahl der Distanzquadrate innerhalb der Gruppen um mehr als einen Wert zunimmt. Beispiel: 2 Gruppen A, B aus 2 bzw. 3 Punkten mögen zu einer neuen Gruppe vereinigt werden, die somit aus 5 Punkten besteht. In ihr sind insgesamt 10 Distanzmessungen möglich. Zu SG sind jedoch nur 6 Distanzquadrate hinzuzufügen, da sämtliche innerhalb der Gruppen A und B vorkommenden Abstände (insgesamt 4) bereits in früheren Schritten berücksichtigt wurden.

Gibt man nun SG und SA in Prozent von $SG+SA$ an, so kann man aus dem mehr oder weniger starken prozentualen Anwachsen von SG ablesen, ob bei der gerade

vorgenommenen Gruppeneinteilung viel oder wenig an Detailinformation verlorengegangen ist. In Abb. 2 und 3 sind für das gewählte Anwendungsbeispiel SG % und SA % angegeben; ein ins Auge fallendes Abknicken der Kurve ist ein Hinweis darauf, daß der Informationsverlust zwischen 2 Schritten besonders hoch ist und eine günstige Gruppeneinteilung vor diesem Einschnitt vorgenommen werden sollte.

4b. 2. Gruppierungsalgorithmus: Minimierung des Gruppendistanzzuwachses

Das unter Punkt 4a geschilderte Verfahren führt nicht in allen Fällen zu ungefähr gleich großen Gruppen (siehe Anwendungsbeispiel), da bei der Neuberechnung der Distanzquadrate nach jedem Schritt vom Gruppenmittelpunkt ausgegangen und die Anzahl der Punkte in der Gruppe nicht berücksichtigt wird. Eine gewisse Abwandlung der unter 4a angegebenen Methode begünstigt die Bildung von etwa gleich großen Gruppen.

Für jede Gruppe wird die sogenannte Einzelgruppendistanz (ED) definiert, und zwar als die Summe der Distanzquadrate zwischen den Punkten der Gruppe dividiert durch die Anzahl der Punkte in der Gruppe.

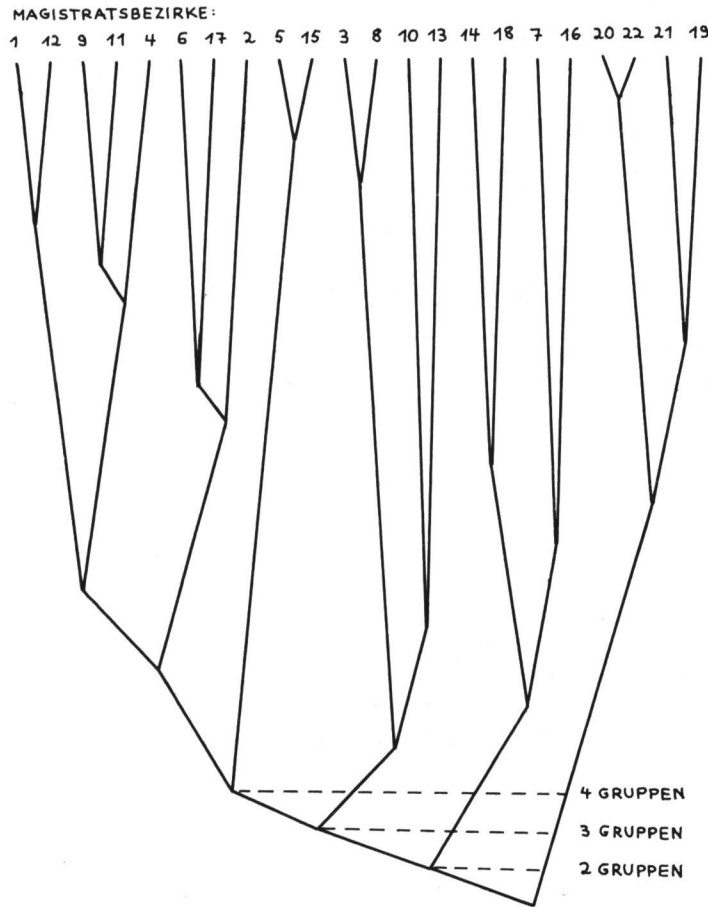


Abb. 3: Stammbaum beim 2. Gruppierungsverfahren

$$(5) \quad ED = \left(\sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij}^2 \right) : n \quad \times \left((n_i + n_k) \cdot d_{ik}^2 + (n_j + n_k) \cdot d_{jk}^2 - n_k \cdot d_{ij}^2 \right)$$

Die Summe aller einzelnen Gruppendifferenzen ergibt die (Gesamt-) Gruppendifferenz (GD). Bei der Gruppierung wird gefordert, daß der Zuwachs zur Gruppendifferenz ein Minimum wird. Dazu ist zunächst die Matrix der Distanzquadrate (d_{ij}^2) in die Matrix der Distanzzuwächse $(d_{ij}^{\prime 2})$ umzuwandeln, die in der i-ten Zeile und j-ten Spalte gerade den Zuwachs zur Gruppendifferenz enthält, der sich ergeben würde, wenn die Gruppen (bzw. Punkte) i und j zusammengefaßt würden. Da vor dem 1. Schritt die Gruppendifferenz 0 ist, ergibt sich die Ausgangsmatrix $(d_{ij}^{\prime 2})$, indem man jedes Element der Matrix (d_{ij}^2) durch 2 teilt (GD vor dem 1. Schritt = 0, GD nach dem ersten Schritt = $d_{ij}^2 : 2$). Im folgenden wird nun von der so erhaltenen Matrix $(d_{ij}^{\prime 2})$ ausgegangen und in jedem Schritt ihr kleinstes Element gesucht. Dadurch können die Punkte bzw. Gruppen zusammengefaßt werden, die den kleinsten Zuwachs der Gruppendifferenz (GD) verursachen. Wiederum werden anschließend 2 Zeilen und Spalten zu einer einzigen zusammengefaßt und deren Elemente neu berechnet. Dafür gibt Kilchenmann [8] die folgende Rechenregel an:

$$(6) \quad d_{ijk}^{\prime 2} = \frac{1}{n_i + n_j + n_k}$$

wobei $d_{ijk}^{\prime 2}$ Distanzzuwachs, der sich ergeben würde, wenn im nächsten Schritt die neu aus den Gruppen i und j gebildete Gruppe mit der Gruppe k vereinigt würde; n_i, n_j, n_k Anzahl der Punkte in der Gruppe i, j, k.

Ganz analog zu Fall 4a entsteht so schrittweise aus einer $n \cdot n$ -Matrix eine $2 \cdot 2$ -Matrix. Auch hier läßt sich die allmählich fortschreitende Verallgemeinerung berechnen. Die oben definierte Gruppendifferenz verändert sich mit jedem Schritt wie folgt:

$$(7) \quad GD = GD + d_{ij}^{\prime 2}$$

wobei $d_{ij}^{\prime 2}$ das kleinste Element der Matrix ist und die Punkte bzw. Gruppen i und j zusammengefaßt werden.

Entsprechend der Zunahme der Gruppendifferenz nimmt die analog definierte Zwischengruppendifferenz (ZD) ab. Gruppendifferenz und Zwischengruppendifferenz ergeben zusammen die totale Distanz, die als Bezugsgrundlage für eine prozentuale Angabe dienen kann (vgl. Abb. 3). Ein starker Abfall der Kurve weist wiederum auf einen hohen Informationsverlust hin, so daß eine günstige Gruppeneinteilung besser vor diesem Schritt vorgenommen werden sollte.

Im Unterschied zum 1. Gruppierungsverfahren geht jetzt

	1 12	2 (2 6 9 11 17)	3 8	4	5 15	6 9 11 17 (-)	7	10	13	14	16	18	19 20 21 22
1 12	0												
2 (2 6 9 11 17)	0,1 (0,2)	0											
3 8	0,2	0,6 (0,6)	0										
4	0,1	0,3 (0,1)	0,2	0									
5 15	1,5	0,8 (0,6)	2,5	1,3	0								
6 9 11 17 (-)	0,2 (-)	0,1 (-)	0,6 (-)	0,1 (-)	0,6 (-)	0							
7	3,6	2,5 (2,2)	5,0	3,2	0,5	2,2 (-)	0						
10	0,8	1,1 (0,7)	0,6	0,3	1,8	0,7 (-)	3,8	0					
13	1,6	2,4 (1,9)	1,0	1,1	3,9	1,9 (-)	6,6	0,4	0				
14	6,7	5,0 (5,0)	8,9	6,8	2,4	5,0 (-)	1,2	8,4	12,4	0			
16	4,7	3,3 (3,1)	6,4	4,5	1,0	3,1 (-)	0,2	5,5	8,7	0,4	0		
18	5,1	3,6 (3,7)	7,0	5,2	1,6	3,7 (-)	0,8	6,8	10,3	0,1	0,2	0	
19 20 21 22	4,2	5,9 (5,7)	2,8	4,1	9,8	5,7 (-)	14,3	3,5	1,8	21,5	17,1	18,6	0

Matrix 1

die Gruppengröße in die Formel zur Berechnung der Gruppendistanz und damit auch des Gruppendistanzzuwachses ein. Das hat beispielsweise zur Folge, daß ein Punkt bei nahezu gleichem Abstand zu 2 Gruppenmittelpunkten jeweils zu der Gruppe mit der geringen Punktezahl geschlagen wird (vgl. 7. Gruppierungsschritt im Anwendungsbeispiel). In der Regel werden dadurch zu unterschiedliche Gruppengrößen vermieden. Beide Algorithmen führen nicht unbedingt zu zusammenhängenden Räumen; es kann durchaus vorkommen, daß räumlich weit auseinanderliegende Beobachtungseinheiten zu einem Typ zusammengefaßt werden (vgl. Anwendungsbeispiel).

*

Um die theoretischen Ausführungen zu den verschiedenen Regionalisierungsverfahren etwas zu verdeutlichen, möchte ich beide Methoden an einem Beispiel aus meinem Arbeits-

gebiet Südwestafrika durchrechnen. Die Unterschiede und Gemeinsamkeiten werden so noch klarer hervortreten. Damit die Ausgangssituation ohne Schwierigkeiten auch zeichnerisch dargestellt werden kann, will ich mich darauf beschränken, bei der Gruppierung 2 Merkmale zugrunde zu legen. Für sämtliche Berechnungen auf der IBM 7090 wurden eigene Programme verwendet, die unter Benutzung der zitierten Literatur geschrieben wurden.

Südwestafrika ist verwaltungsmäßig in 22 Magistratebezirke eingeteilt. Im folgenden sollen diese Beobachtungseinheiten (n = 22) zu größeren, möglichst homogenen Regionen im Hinblick auf ihre Bevölkerungs- und Wirtschaftsstruktur zusammengefaßt werden. Bei der Gruppierung wurden folgende Gesichtspunkte berücksichtigt (r = 2):

- 1. Anteil der Weißen an der Bevölkerung jedes Magistratebezirkes
- 2. Anteil der Beschäftigten in der Landwirtschaft

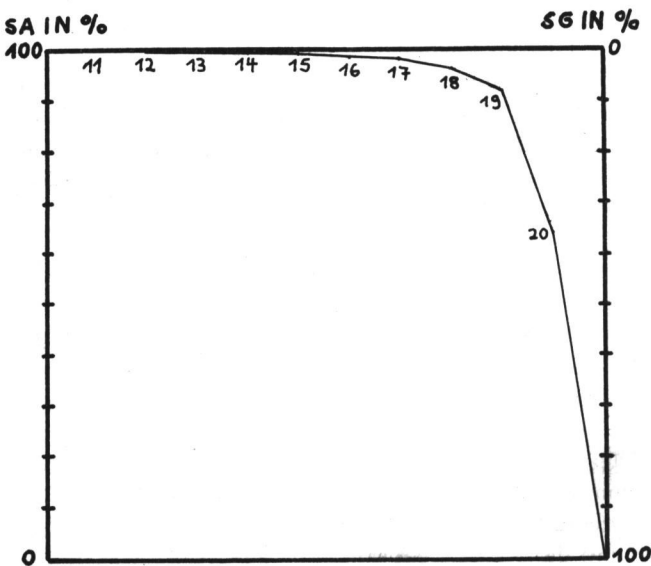


Abb. 4: Informationsverlust beim 1. Gruppierungsverfahren vom 11. zum 21. Schritt (zur Erklärung der Abkürzungen vgl. S. 14)

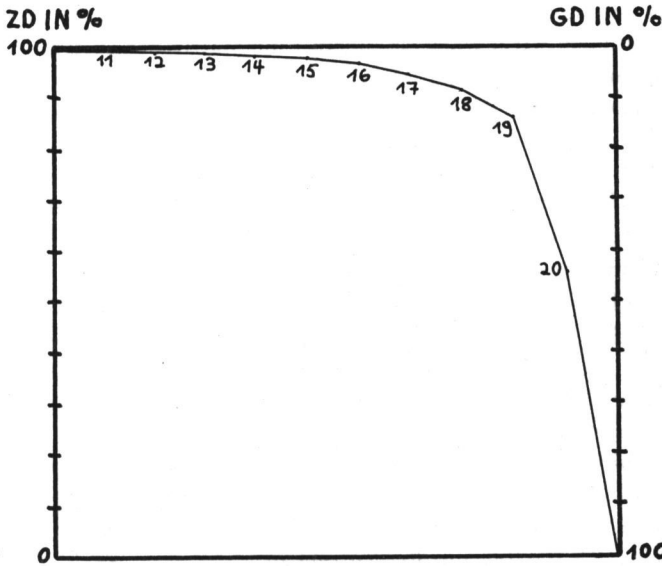


Abb. 5: Informationsverlust beim 2. Gruppierungsverfahren vom 11. zum 21. Schritt (zur Erklärung der Abkürzungen vgl. S. 15)

	1 12	2 6 17	3 8	4 9 11	5 15	7	10	13	14 (14 18)	16	18 (-)	19 21	20 22
1 12	0												
2 6 17	0,2	0											
3 8	0,3	0,9	0										
4 9 11	0,1	0,2	0,4	0									
5 15	1,3	0,6	2,5	1,2	0								
7	2,3	1,5	3,5	2,1	0,3	0							
10	0,5	0,7	0,4	0,3	1,2	1,9	0						
13	1,2	1,7	0,6	1,0	2,5	3,3	0,2	0					
14 (14 18)	4,3(5,6)	3,5(4,8)	6,1(8,2)	4,7(6,5)	1,7(2,1)	0,6(0,6)	4,2(5,0)	6,2(7,6)	0				
16	2,9	2,2	4,4	3,0	0,7	0,1	2,7	4,4	0,2(0,2)	0			
18 (-)	3,2 (-)	2,5 (-)	4,9 (-)	3,5 (-)	1,1 (-)	0,4 (-)	3,4 (-)	5,2 (-)	0,1 (-)	0,1 (-)	0		
19 21	4,4	7,0	2,5	5,2	9,1	9,2	2,1	1,0	13,9(19,4)	11,0	12,0 (-)	0	
20 22	5,2	8,3	3,1	6,3	10,6	10,4	2,8	1,5	15,3(21,3)	12,3	13,2 (-)	0,1	0

Matrix 2

Die Korrelation zwischen beiden Variablen ist äußerst gering ($r = 0,087$), so daß sie unberücksichtigt bleiben konnte. Die beiden standardisierten Variablen bilden die Koordinatenachsen eines 2dimensionalen Systems, in dem jede Beobachtungseinheit als Punkt wiedergegeben ist (Abb. 1). Sämtliche Distanzen zwischen den Punkten wurden berechnet (vgl. Formel (2)) und daraus die Matrix der Distanzquadrate und die Matrix der Distanzzuwachse als Ausgangszustände für das erste bzw. zweite Gruppierungsverfahren gebildet. Anschließend wurde in beiden Fällen das kleinste Element der Matrix gesucht, die beiden Punkte zu einer ersten Gruppe zusammengefaßt sowie aus 2 Zeilen und 2 Spalten eine einzige neue Zeile bzw. Spalte berechnet. Bei einer 22 · 22-Matrix kann der allmähliche Gruppierungsprozeß aus Platzgründen nicht vollständig und im einzelnen angegeben werden. Ein Beispiel möge dazu genügen (vgl. Matrix 1, Schritt 10, 1. Verfahren).

$$\text{Min } d_{ij}^2 = d_{26}^2 = 0,069,$$

d. h. eine aus den Punkten 6, 9, 11 und 17 bestehende Gruppe wird mit Punkt 2 zusammengefaßt. Damit ergibt sich nach dem 10. Schritt folgende Gruppeneinteilung:

- Gruppe 1: 2, 6, 9, 11, 17
Gruppe 2: 19, 20, 21, 22
Gruppe 3: 1, 12
Gruppe 4: 3, 8
Gruppe 5: 5, 15
Gruppe 6: 4
- Gruppe 7: 7
Gruppe 8: 10
Gruppe 9: 13
Gruppe 10: 14
Gruppe 11: 16
Gruppe 12: 18

Vor einem weiteren Schritt müssen die Distanzquadrate vom Mittelpunkt der neuen Gruppe zu allen anderen Punkten bzw. Gruppen nach Formel (4) neu berechnet werden (Werte in Klammern).

Vgl. nunmehr Matrix 2 (Schritt 10, 2. Verfahren). Hiernach ergibt sich $\text{min } d_{ij}^2 = d_{1418}^2 = 0,059$, d. h., aus den Punkten 14 und 18 wird eine neue Gruppe gebildet. Somit ergibt sich nach dem 10. Schritt folgende Gruppeneinteilung:

- Gruppe 1: 4, 9, 11
Gruppe 2: 2, 6, 17
Gruppe 3: 14, 18
Gruppe 4: 20, 22
Gruppe 5: 19, 21
Gruppe 6: 1, 12
- Gruppe 7: 3, 8
Gruppe 8: 5, 15
Gruppe 9: 7
Gruppe 10: 10
Gruppe 11: 13
Gruppe 12: 16

Nach Formel (6) erfolgt die Neuberechnung des Distanzzuwachses zwischen der neuen Gruppe 14 und 18 und allen übrigen Gruppen, d. h., Zeile und Spalte 14 der Matrix ändern sich, während Zeile und Spalte 18 ganz wegfallen (Werte in Klammern).

Ein Vergleich mit den Zwischenergebnissen nach dem 10. Schritt bei der 1. Methode läßt bereits deutlich erkennen, daß die Gruppengröße jetzt sehr viel einheitlicher ist. Noch deutlicher wird der unterschiedliche Gruppierungsalgorithmus am Beispiel des 7. Schrittes (vgl. Abb. 2 und 3). Die Ausgangsposition vor diesem Schritt ist bei beiden Verfahren gleich (vgl. Abb. 1). Sehr dicht zusammen liegen die Punkte 19 und 21 sowie die Gruppe 20, 22. Während Punkt 21 beim 1. Verfahren zu der Gruppe 20, 22 geschlagen wird, führt der 2. Algorithmus zum Zusammenlegen der Punkte 19 und 21 zu einer weiteren aus 2 Elementen bestehenden Gruppe.

Im einzelnen läßt sich der Gruppierungsprozeß für beide Verfahren an dem sogenannten „Stammbaum“ (Abb. 2 und 3) verfolgen. Der schrittweise zunehmende Informationsverlust, der bei jeder Zusammenfassung erfolgt, geht aus den Abb. 4 und 5 hervor. Es ist bemerkenswert, daß sich der Verlust an Detailinformation bis etwa zum 18. Schritt in ganz engen Grenzen hält (ca. 8%). Die beiden Abb. (2 und 3) sind allerdings nicht unmittelbar miteinander vergleichbar, da der Informationsverlust jeweils etwas anders definiert wurde. Aus beiden Abbildungen läßt sich aber klar ablesen, daß eine Gliederung Südwesafrikas in 3 Regionen am ehesten zu vertreten ist. (Hier erfolgt jeweils ein starker Knick der Kurve, die den Informationsverlust wiedergibt.)

Damit würden die beiden Verfahren zunächst die im Norden des Landes gelegenen Magistratsbezirke Kaokoveld, Ovamboland, Okavango und Caprivi zu einer Re-

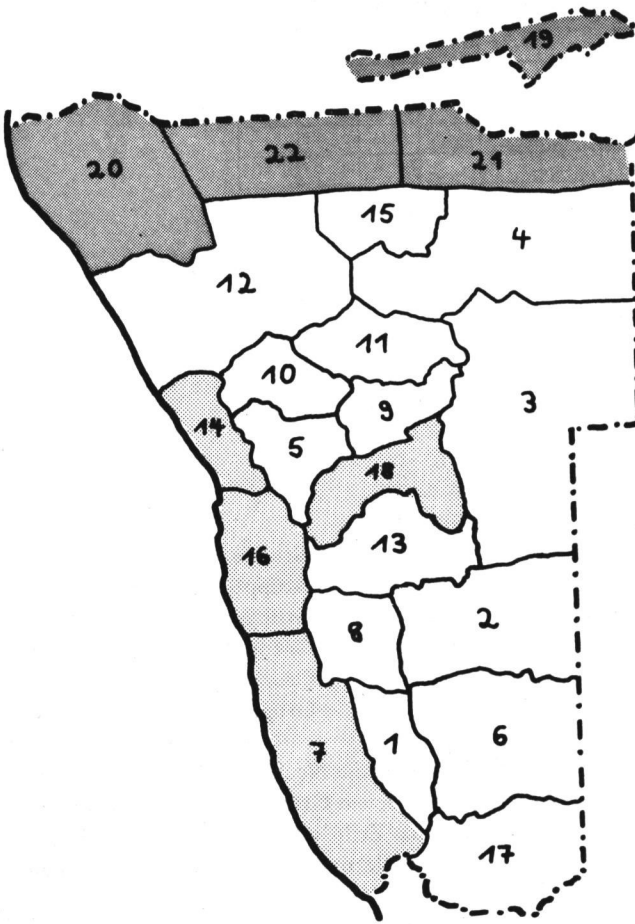


Abb. 6: Einteilung Südwestafrikas in drei verschiedene Raumtypen (Gruppe 1 dunkel, Gruppe 2 mittel, Gruppe 3 weiß). Die Nummern bezeichnen die verschiedenen Magstratsbezirke: 1. Bethanien; 2. Gibeon; 3. Gobabis; 4. Grootfontein; 5. Karibib; 6. Keetmanshoop; 7. Lüderitz; 8. Maltahöhe; 9. Okahandja; 10. Omaruru; 11. Otjiwarongo; 12. Outjo; 13. Rehoboth; 14. Swakopmund; 15. Tsumeb; 16. Walfischbucht; 17. Warmbad; 18. Windhoek; 19. Caprivi; 20. Kaokoveld; 21. Okavango; 22. Ovamboland

gion zusammenfassen, die sich grundlegend vom übrigen Südwestafrika unterscheidet. Dieses Gebiet ist ausschließlich von Eingeborenen bewohnt, die hauptsächlich Ackerbau und Viehzucht zur Selbstversorgung treiben (vgl. Abb. 6).

Die zweite entstehende Gruppe enthält die wichtigsten Industrie- und Versorgungszentren des Landes: Die Häfen Walfischbucht und Lüderitz mit der dort ansässigen fischverarbeitenden Industrie, das Gebiet des Diamantenabbaus bei Oranjemund sowie die Bezirke Windhoek und Swakopmund, in denen die Landwirtschaft gegenüber dem tertiären Sektor eine untergeordnete Rolle spielt.

In der letzten Gruppe werden schließlich diejenigen Distrikte zusammengestellt, die im wesentlichen durch die europäische Landwirtschaft geprägt sind.

Nimmt man demgegenüber eine Einteilung Südwestafrikas in 4 Raumtypen vor, so erhält man bei beiden Verfahren unterschiedliche Ergebnisse. Die 4. Gruppe entsteht in beiden Fällen durch Teilung der 3. Gruppe. Aus der Zone europäischer Landwirtschaft werden Bezirke ausgegliedert

(1. Methode: Omaruru, Rehoboth; 2. Methode: Omaruru, Rehoboth, Maltahöhe, Gobabis), die sich durch einen höheren Anteil der Eingeborenen und Farbigen an der Gesamtbevölkerung auszeichnen (große Eingeborenenreservate in den Distrikten Gobabis und Omaruru, bzw. das Siedlungsgebiet der Farbigen um Rehoboth).

Mit Hilfe einer Diskriminanzanalyse läßt sich überprüfen, ob eine gewählte Gruppeneinteilung bestehen bleiben kann oder ob einzelne Elemente einer anderen Gruppe zugeordnet werden müssen. Bei dieser Methode wird eine Abbildung der Datenpunkte in einen Unterraum gesucht, in dem die vorgegebenen Gruppen „optimal“ getrennt werden. In dem angewandten Verfahren (Programm des Deutschen Rechenzentrums Darmstadt von E. Faber und W. Nollau) wurde folgendes Kriterium für eine optimale Trennung zugrunde gelegt:

$$\frac{\text{Streuungsmaß zwischen den Gruppen}}{\text{Streuungsmaß innerhalb der Gruppen}} = \text{Maximum}$$

Eine Überprüfung der erhaltenen Einteilung Südwestafrikas in vier Regionen bestätigte in beiden Fällen die vorgenommene Gruppierung, d. h., beide vorgeschlagenen Regionalisierungsverfahren nehmen nach der oben angeführten Forderung eine „optimale“ Gruppierung vor.

Es war nicht der Sinn dieser Arbeit, eine eingehende Interpretation der am einfachen Anwendungsbeispiel erzielten Ergebnisse zu geben. Vielmehr sollten zwei verschiedene Möglichkeiten einer objektiven Raumgliederung ausführlich dargestellt werden, um weitere Anwendungen derartiger Methoden auch in komplizierteren und weniger übersichtlichen Fällen zu ermöglichen.

Literatur

- [1] Berry, Brian J. L.: The impact of expanding metropolitan communities upon the central place hierarchy. In: *Annals of the Association of American Geographers*. 50. Jg. (1960) Nr. 2, S. 112–116.
- [2] Berry, Brian J. L.: A method for deriving multi-factor uniform regions. In: *Przegląd Geograficzny*. Bd. 33 (1961) S. 263–279. (Dt. Übers.: Eine Methode zur Bildung homogener Regionen mehrdimensionaler Definition. In: *Wirtschafts- und Sozialgeographie*. S. 212–227. – Köln 1970. = Neue wissenschaftliche Bibliothek, 35, Wirtschaftswissenschaften.)
- [3] Berry, Brian J. L.: The mathematics of economic regionalization. S. 77–106 in: *Economic Regionalization*. Hrsg.: Miloslav Macka. – Prag 1967.
- [4] Berry, Brian J. L.: A synthesis of formal and functional regions using a general field theory of spatial behaviour. In: *Spatial analysis. A reader in statistical geography*. S. 419–428. – Englewood Cliffs (N. J.) 1968.
- [5] Computer applications in the earth sciences: Colloquium on classification procedures. In: *Computer Contribution*. (Lawrence/Kan.) (1966) H. 7.
- [6] Forde, E. R.: The population of Ghana. – Evanston (Ill.) 1968. = *Studies in Geography*, 15.

- [7] *Haggett, P.*: Locational analysis in human geography. – London 1965.
- [8] *Kilchenmann, A.*: Untersuchungen mit quantitativen Methoden über die fremdenverkehrs- und wirtschaftsgeographische Struktur im Kanton Graubünden (Schweiz). – Diss. Zürich 1968.
- [9] *King, L. J.*: Statistical analysis in geography. – Englewood Cliffs (N. J.) 1969.
- [10] *Steiner, Dieter*: Die Faktorenanalyse – ein modernes statistisches Hilfsmittel des Geographen für die objektive Raumgliederung und Typenbildung. In: *Geographica Helvetica*. 20. Jg. (1965) Nr. 1, S. 20–34.
- [11] *Überla, Karl*: Faktorenanalyse. Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler. – Berlin, New York 1968.