

## Welche Gemeinde repräsentiert die "Mitte" des Landes Nordrhein-Westfalen?

Gerß, Wolfgang

Veröffentlichungsversion / Published Version  
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:  
GESIS - Leibniz-Institut für Sozialwissenschaften

### Empfohlene Zitierung / Suggested Citation:

Gerß, W. (2000). Welche Gemeinde repräsentiert die "Mitte" des Landes Nordrhein-Westfalen? *ZUMA Nachrichten*, 24(46), 74-88. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-211162>

### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

## **WELCHE GEMEINDE REPRÄSENTIERT DIE „MITTE“ DES LANDES NORDRHEIN-WESTFALEN?**

*WOLFGANG GERß<sup>1</sup>*

Durch ein multivariates statistisches Modell wird der „Mittelpunkt“ eines Gebietes in Abhängigkeit von beliebig vorzugebenden Merkmalen bestimmt. Die Mitte ist dabei im übertragenen Sinn – wie typisch, normal oder durchschnittlich – gemeint, das Modell kann jedoch anschaulich aus der geographischen Mitte abgeleitet werden. Die Modellkonstruktion geht von den multivariaten Distanzen zwischen den Teilen des Gebietes aus. Als Distanzmaß wird die Mahalanobis-Distanz der Euklid-Distanz vorgezogen. Als Mitte des Gebietes wird derjenige Gebietsteil angesehen, für den die Summe der – nach bestimmten Merkmalen berechneten – Mahalanobis-Distanzen zu allen anderen Gebietsteilen am kleinsten ist. Dieses Modell wird auf Daten für die Gemeinden des Landes Nordrhein-Westfalen angewendet. Die Daten sind 40 Merkmale aus dem Bestand der amtlichen Landesdatenbank, die in vier Merkmalsgruppen zusammengefaßt werden. Nach jeder Gruppe ergibt sich eine Gemeinde, die den demographischen, sozialen, ökonomischen oder ökologischen Mittelpunkt des Landes darstellt. Die Programmierung ermöglicht auch die praktikable Einbeziehung einer noch größeren Anzahl von Merkmalen.

Using a multivariate statistical model, the “central point” of a region is determined dependent on variables specified in content and number as needed. In this connection, the centre is defined in a figurative sense – as typical, normal or average -, but the model can be derived, by way of illustration, from the geographical centre. The construction of the model proceeds from the multivariate distances between the parts of the region. The Mahalanobis-distance is preferred to the Euklid-distance as measure of distance. That part of the region taken as its centre is that for which the sum of the Mahalanobis-distances – computed with certain variables – is minimal from all other parts. This model is applied to data on the municipalities of the Land Nordrhein-Westfalen, consisting of 40 variables from the official Land data bank, catego-

---

1 Ich danke meinem Sohn Dipl.-Stat. Joachim Gerß für die Programmierung.

rized in four groups of variables. One municipality is determined for each group which represents the demographical, social, economical or ecological centre of the Land. The programming makes it possible to include an even larger number of variables.

## **1. Anlaß**

Die „Gesellschaft für Konsumforschung“ hat ein Verfahren entwickelt, durch das Marketingpläne in einem Testort erprobt werden (GfK 1998). In diesem Ort werden Daten über das reale Kaufverhalten ermittelt und anschließend auf den nationalen Gesamtmarkt hochgerechnet. Der ausgewählte Testort muß dazu als normale Durchschnittsgemeinde typisch für den Gesamtmarkt sein. Er muß den sozio-demographischen Mittelpunkt der Bundesrepublik Deutschland darstellen. Diese Mitte sehen die Marktforscher in der Gemeinde Haßloch in Rheinland-Pfalz. Mit etwa einem Drittel weist ein besonders hoher Anteil der Haßlocher Haushalte nach verschiedenen Merkmalen (Lebensalter, Einkommen, Anzahl der Familienmitglieder) Durchschnittswerte aller deutschen Haushalte auf. Die Kaufkraft der Haßlocher Haushalte entspricht ungefähr dem Bundesdurchschnitt. Die wesentlichen Gründe zur Auswahl von Haßloch sind aber der hohe Grad der Bedarfsdeckung der Haushalte vor Ort (circa 90 bis 95 Prozent), die Existenz aller relevanten Geschäftstypen des Lebensmittel Einzelhandels und der frühzeitige Anschluß an das Kabelfernsehen (briefliche Mitteilung der GfK). Der letztgenannte Grund war die Voraussetzung dafür, daß hier der erste elektronische Testmarkt in Europa installiert werden konnte.

## **2. Methode**

Die Methode der GfK geht von einer verhältnismäßig geringen Anzahl inhaltlich sehr unterschiedlicher Merkmale aus, deren als verschieden angesehene Wichtigkeit vorgegeben wird und aus denen ein einziger Durchschnittswert aller betrachteten Orte – als „mittlerer“ Ort – abgeleitet wird. Die folgenden Ausführungen wurden von den Arbeiten der GfK angeregt. Anders als bei der GfK wird hier eine Methode dargestellt, bei der die beliebig vielen einzubeziehenden Merkmale nach ihren Inhalten in Merkmalsgruppen klassifiziert werden, für die sich verschiedene gruppenspezifische „Mitten“ ergeben. Dabei sollen die einzelnen Merkmale nicht von vornherein als mehr oder weniger wichtig behandelt werden; das Gewicht, mit dem jedes Merkmal in den Durchschnitt eingeht, soll sich vielmehr aus der Methode selbst ergeben. Damit ist sichergestellt, daß das Rechenergebnis nicht mehr von subjektiven Vorgaben beeinflusst wird, nachdem die Merkmalsgruppen definiert und die zugehörigen Merkmale be-

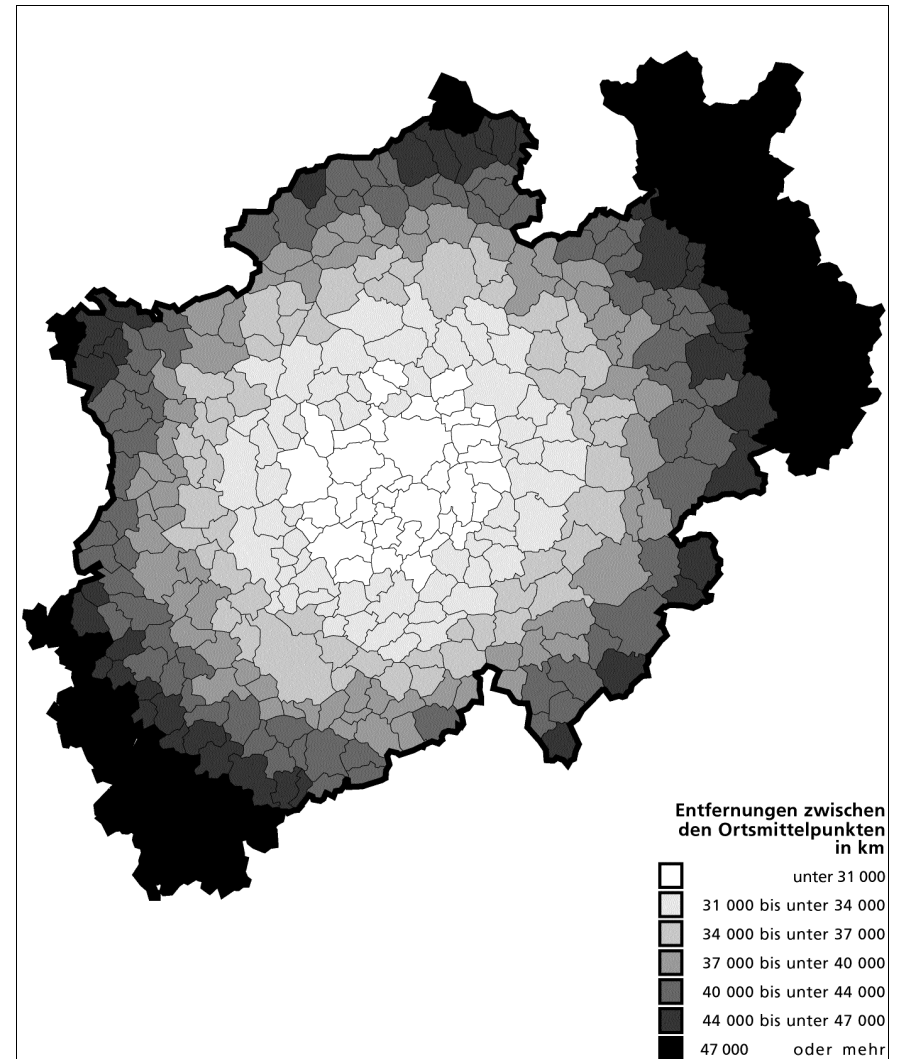
stimmt worden sind. Die gestellte Aufgabe besteht darin, ein multivariates statistisches Modell zu konstruieren und anzuwenden, das den „Mittelpunkt“ eines Gebietes in Abhängigkeit von beliebig vorzugebenden Merkmalen, im übrigen aber automatisch bestimmt. Das „Gebiet“ ist das Land Nordrhein-Westfalen, die „Punkte“ - von denen einer in der „Mitte“ liegt - sind die einzelnen Gemeinden. Wenn auch die Lage in der Mitte im übertragenen Sinn - wie typisch, normal oder durchschnittlich - gemeint ist, so liegt es doch nahe, zur Veranschaulichung zunächst von der geographischen Mitte auszugehen. Die geographische Mitte hat mit den im folgenden definierten Mitten inhaltlich nichts zu tun, sie ist jedoch zur Erläuterung des statistischen Modells wegen ihrer einfachen formalen Eigenschaften - die dann gegebenenfalls komplizierteren anderen Mitten angepaßt werden müssen - besonders gut geeignet. Der Mittelpunkt eines kreisförmigen Gebietes ist dadurch gekennzeichnet, daß die Entfernungssumme zu allen anderen Punkten am kleinsten ist. Dementsprechend wird die geographische Mitte eines sich wie Nordrhein-Westfalen nach allen Richtungen einigermaßen gleichmäßig ausdehnenden Landes von derjenigen Gemeinde eingenommen, für die die Summe der geradlinigen Entfernungen zu allen anderen Gemeinden am kleinsten ist (Abbildung 1).

Die Entfernungen zwischen den 396 Gemeinden des Landes sind in der Landesdatenbank Nordrhein-Westfalen gespeichert (Plewa 1985). Die Lage jeder Gemeinde wurde dabei nicht nach dem geometrischen Schwerpunkt der gesamten Gemeindefläche, sondern nach dem Ortsmittelpunkt des besiedelten Bereichs bestimmt. Die siedlungsgeographische Mitte des Landes Nordrhein-Westfalen ist danach die Stadt Herdecke. Die Entfernung zwischen zwei Gemeinden kann nach dem Satz des Pythagoras mit den Rechts- und Hochwerten der Gauß-Krüger-Koordinaten berechnet werden. Die Variable „geradlinige Entfernung“ wird dadurch auf zwei voneinander unabhängige Variablen - die horizontale und die vertikale Strecke - zurückgeführt. Auf diese Weise wird aus dem univariaten ein bivariates Modell der Entfernungsmessung. So haben zum Beispiel die Städte Bielefeld und Düsseldorf die folgenden Gauß-Krüger-Koordinaten (siehe Abbildung 2).

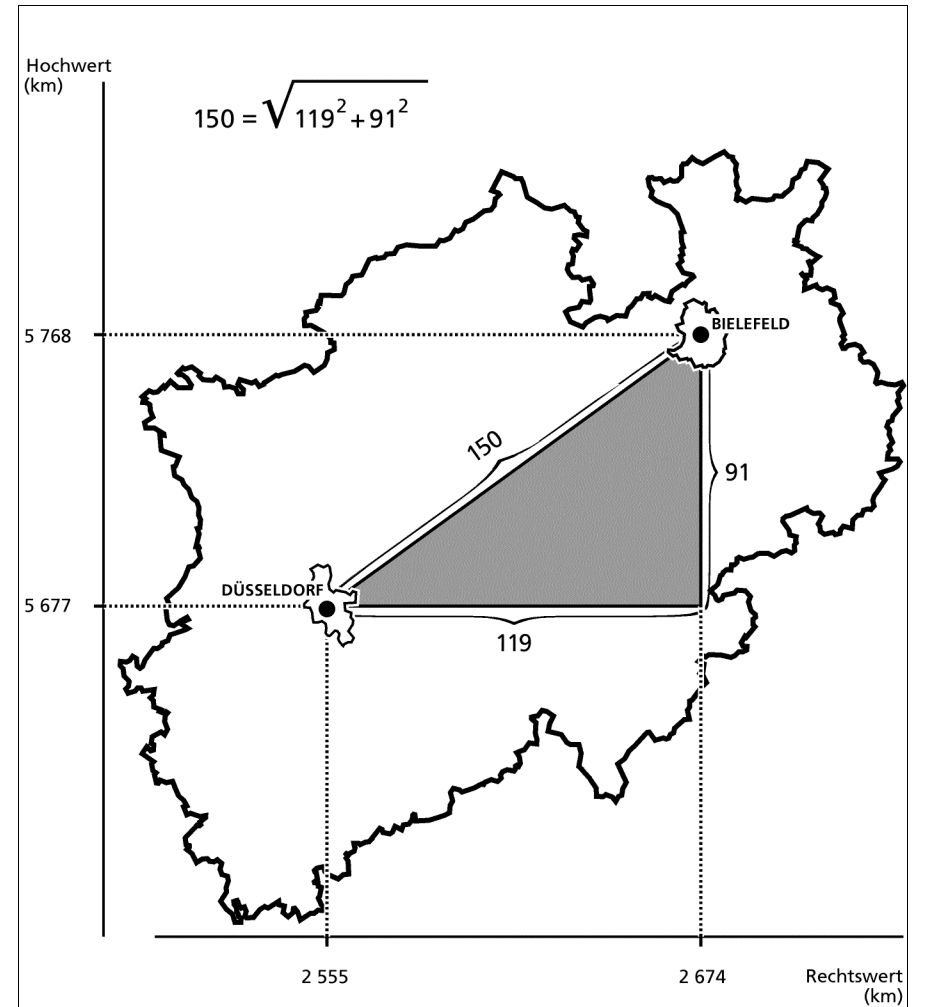
Bielefeld:      Hochwert  $B_H = 5\,768\,324$  m, Rechtswert  $B_R = 2\,673\,790$  m

Düsseldorf:    Hochwert  $D_H = 5\,677\,000$  m, Rechtswert  $D_R = 2\,555\,200$  m

**Abbildung 1: Summe der Entfernungen jeder einzelnen von den 395 anderen Gemeinden des Landes Nordrhein-Westfalen**



**Abbildung 2:** Entfernung zwischen Bielefeld und Düsseldorf sowie Hoch- und Rechtswerte der Gauß-Krüger-Koordinaten



Die Entfernung zwischen Bielefeld und Düsseldorf ( $\overline{BD}$ ) wird entweder univariat (als Länge der Hypotenuse) direkt gemessen oder bivariat (nach den Längen der beiden Katheten) aus den Koordinaten berechnet:

$$\overline{BD} = 149 \quad 679 \quad m = \sqrt{(B_H - D_H)^2 + (B_R - D_R)^2}$$

Die so berechnete Entfernung zwischen den beiden Gemeinden ist – in Anlehnung an geometrische Bezeichnungen – deren geographische „Euklid-Distanz“. Im übertragenen Sinn – wie in der Clusteranalyse (Bock 1974) – ist die Euklid-Distanz  $\Delta_E$  zwischen zwei Gemeinden A und B nach zwei beliebigen Variablen  $X_1$  und  $X_2$  wie folgt definiert:

$$\Delta_E = \sqrt{(x_{1A} - x_{1B})^2 + (x_{2A} - x_{2B})^2}$$

In der Clusteranalyse wird das ursprüngliche bivariate Modell der Euklidischen Distanzmessung wie folgt zu einem multivariaten Modell mit  $m$  Variablen  $X_j$  erweitert ( $j = 1, 2, \dots, m$ ):

$$\Delta_E^2 = \sum_{j=1}^m (x_{jA} - x_{jB})^2$$

Die quadrierte Euklid-Distanz kann zur Vereinfachung der Formelschreibweise verwendet werden, weil die kleinste Summe der quadrierten Distanzen zu derselben Gemeinde führt wie die hier interessierende kleinste Summe der nichtquadrirten Distanzen. In Matrix-Schreibweise ist die quadrierte Euklid-Distanz zwischen zwei Gemeinden A und B ( $d$  = Spaltenvektor,  $d^t$  = Zeilenvektor der Merkmalsdifferenzen aller  $m$  Variablen zwischen den beiden Gemeinden):

$$\Delta_E^2 = d^t d$$

In Abbildung 2 werden die Werte der Gemeinden in zwei Variablen als Punkte im ersten Quadranten eines Koordinatensystems dargestellt. Zur Anwendung des Satzes des Pythagoras müssen die beiden Achsen je Variablenpaar senkrecht aufeinander stehen. Wenn viele Gemeinden – zum Beispiel die 396 Gemeinden des Landes Nordrhein-Westfalen – gemeinsam betrachtet werden, sind die beiden Variablen dann voneinander unabhängig, wenn die beiden Achsen – gegebenenfalls nach ihrer Parallelverschiebung – außerdem die gesamte dargestellte Punktwolke eng umschließen. Die Erfüllung dieser beiden Forderungen zusammen ist dann möglich, wenn die Punkte – die den (hier sämtlich positiven, allgemein aber beliebigen) Merkmalswerten der Ge-

meinden entsprechen - über den ganzen ersten Quadranten (beschränkt auf die jeweiligen variablenspezifischen Wertebereiche) gleichmäßig weit verbreitet sind. In diesem Fall hat der Korrelationskoeffizient der beiden betrachteten Variablen etwa den Wert null. Nur dann ist die direkte Verwendung der Euklid-Distanz als Maß des bi- beziehungsweise multivariaten Unterschieds zwischen jeweils zwei Gemeinden unmittelbar aus der geometrischen Vorstellung ableitbar. Da aber in der Realität viele Variablen paarweise voneinander abhängig sind, die Korrelationskoeffizienten also Werte ungleich null haben, wird die Euklid-Distanz von diesen Korrelationen beeinflusst. Die Werte der Euklid-Distanz sollen hier von diesem Einfluß bereinigt werden. Unter Verwendung der Inversen der Kovarianzmatrix  $C$  – die nach der hier (siehe unten) notwendigen Standardisierung der Variablen mit der Korrelationsmatrix übereinstimmt – erhält man dann die – in der Terminologie der Clusteranalyse so bezeichnete – (quadrierte) „Mahalanobis-Distanz“:

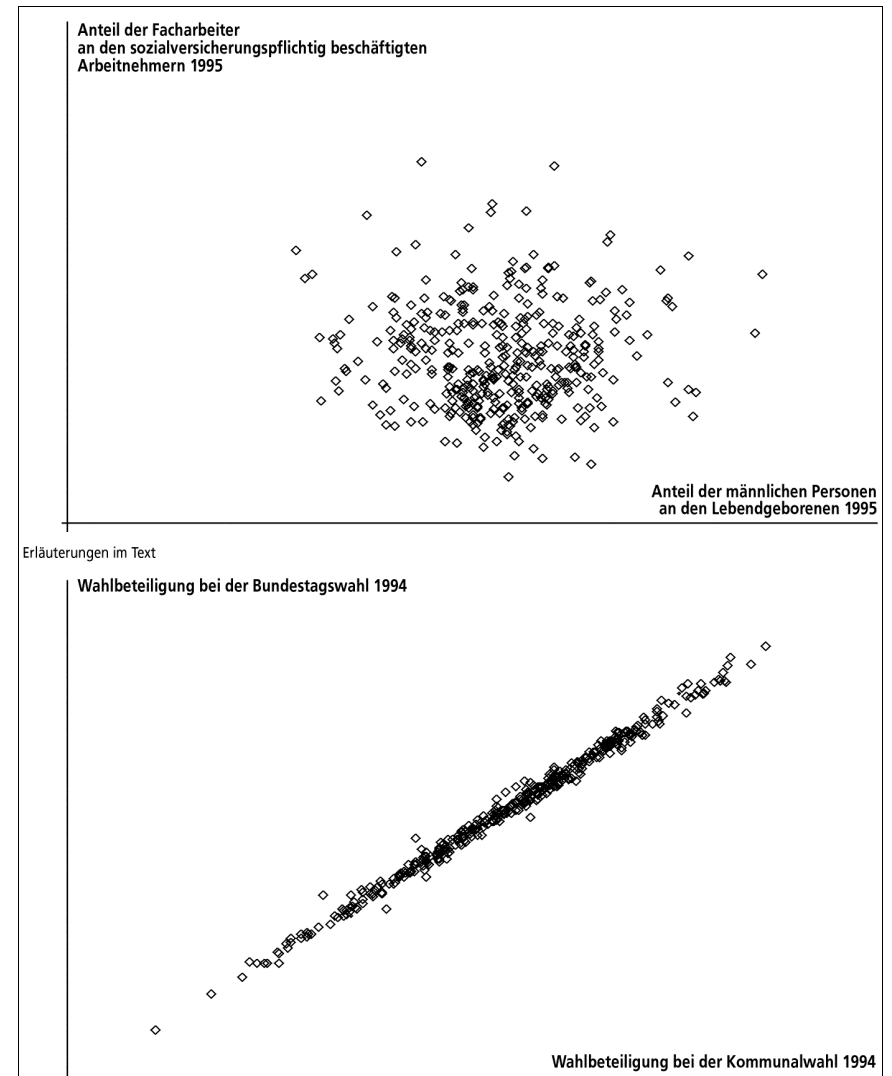
$$\Delta_M^2 = \mathbf{d}'\mathbf{C}^{-1}\mathbf{d}$$

Der Übergang von der Euklid- auf die Mahalanobis-Distanz sei an einem Beispiel demonstriert. In Abbildung 3 sind die 396 Gemeinden des Landes Nordrhein-Westfalen als Punkte im ersten Quadranten eines rechtwinkligen Koordinatensystems dargestellt. Die Koordinaten der Punkte sind (oben) die unkorrelierten ( $r = +0,00005$ ) Merkmale „Anteil der männlichen Personen an den Lebendgeborenen“ (Prozent; Jahr 1995) (horizontale Achse) und „Anteil der Facharbeiter an den sozialversicherungspflichtig beschäftigten Arbeitnehmern“ (Prozent; Jahr 1995) (vertikale Achse) und (unten) die hochkorrelierten ( $r = +0,99\ 593$ ) Merkmale „Wahlbeteiligung bei der Kommunalwahl 1994“ (Prozent) (horizontale Achse) und „Wahlbeteiligung bei der Bundestagswahl 1994“ (Prozent) (vertikale Achse). Die Skalierung ist in Abbildung 3 – zur Konzentration auf das Wesentliche – weggelassen worden, weil es nicht auf bestimmte Merkmalsausprägungen, sondern nur auf die Lage der Punkte zueinander ankommt.

Zur Berechnung der Euklid- und der Mahalanobis-Distanzen müssen die – wenn auch nicht in diesem Beispiel, so doch in der Praxis meistens – mit unterschiedlichen Maßeinheiten versehenen Merkmale durch Standardisierung auf arithmetisches Mittel null und Standardabweichung 1 dimensionslos gemacht werden. Aus den 396 Gemeinden können insgesamt 78 210 Paare gebildet beziehungsweise Distanzen berechnet werden.



**Abbildung 3: Werte der 396 Gemeinden des Landes Nordrhein-Westfalen in zwei unabhängigen Merkmalen (oben) und in zwei hochkorrelierten Merkmalen (unten)**



So ergibt sich zum Beispiel für das Gemeindepaar Düsseldorf/Mönchengladbach

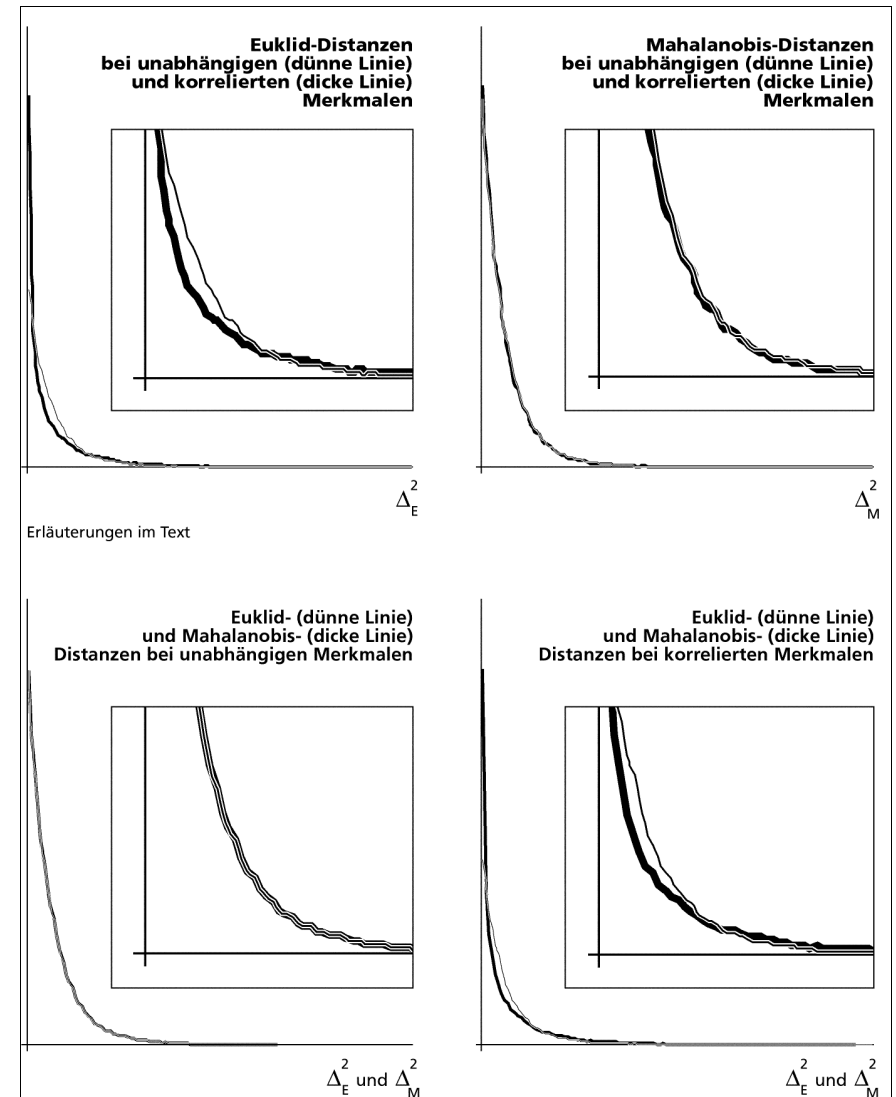
- aus den standardisierten unkorrelierten Merkmalen  $\Delta_E^2 = 0,8648 = \Delta_M^2$
- aus den standardisierten hochkorrelierten Merkmalen dagegen  $\Delta^2 = 4,6853$  und  $\Delta_M^2 = 2,3815$ .

Die Eliminierung der Variablenkorrelation führt – wo sie vorhanden ist – in diesem Fall in Abhängigkeit von der Korrelationsmatrix zu einer Verringerung des Distanzwertes. Die Häufigkeitsverteilung (Abbildung 4) der Distanzen sämtlicher Gemeindepaare hat dementsprechend bei den hochkorrelierten Merkmalen eine deutlich geringere Streuung, wenn sie nach den Mahalanobis-Distanzen (Standardabweichung  $s = 4,73$ ) statt nach den Euklid-Distanzen ( $s = 5,52$ ) gebildet wird. Bei den unkorrelierten Merkmalen ist die Streuung dagegen nach den Mahalanobis- und den Euklid-Distanzen gleich ( $s = 4,24$ ). Auch in Abbildung 4 ist die Skalierung zur besseren Übersichtlichkeit weggelassen worden, weil es hier nur auf die Form der Verteilung ankommt. Allgemein gewährleistet nur das Modell der Mahalanobis-Distanzen solche Meßwerte der multivariaten Unterschiede zwischen den Gemeinden, deren Streuung nicht infolge von Variablenkorrelationen verzerrt ist.

### 3. Ergebnisse

Diesen Überlegungen entsprechend wird im folgenden als „Mitte“ des Landes Nordrhein-Westfalen diejenige Gemeinde angesehen, für die die Summe der – nach bestimmten Merkmalen berechneten – Mahalanobis-Distanzen zu allen anderen Gemeinden am kleinsten ist. Dazu werden je zehn demographische, soziale, wirtschaftliche und umweltbezogene Merkmale verwendet, die bereits in früheren multivariaten Regionalanalysen die Datengrundlage bildeten (Gerß 1994). Dort wurde ausführlich begründet, warum die einzelnen Merkmale ausgewählt wurden. Nach jeder der vier Merkmalsgruppen ergibt sich eine Gemeinde, die den demographischen beziehungsweise sozialen beziehungsweise ökonomischen beziehungsweise ökologischen Mittelpunkt des Landes darstellt. Auf die Bestimmung der mittleren Gemeinde nach der Gesamtheit der 40 Merkmale wird verzichtet, da eine derartige Zusammenfassung zu keinen zusätzlichen Erkenntnissen führen würde und das Ergebnis schwierig interpretierbar wäre.

**Abbildung 4: Häufigkeitsverteilung der Euklid- und Mahalanobis-Distanzen zwischen den Gemeinden des Landes Nordrhein-Westfalen bei unabhängigen und hochkorrelierten Merkmalen**



In der folgenden Liste ist jedes verwendete Merkmal mit einem die Merkmalsgruppe bezeichnenden Buchstaben und einer laufenden Nummer versehen. In Klammern ist das Berichtsjahr der in die Untersuchung einbezogenen Daten angegeben. Alle Daten stammen als Ergebnisse der amtlichen Statistik vom Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen und sind überwiegend in der dort geführten Landesdatenbank gespeichert.

**Demographische Merkmale (D):**

- D1 Anteil der Ausländer an der Gesamtbevölkerung (Prozent) (1997)
- D2 Anteil der Ausländer an den Lebendgeborenen (Prozent) (1998)
- D3 Anteil der Einwohner im Alter von unter 18 Jahren (Prozent) (1997)
- D4 Anteil der Einwohner im Alter ab 65 Jahren (Prozent) (1997)
- D5 Anteil der Erwerbspersonen an den Zuzügen (Prozent) (1997)
- D6 Anteil der Erwerbspersonen an den Fortzügen (Prozent) (1997)
- D7 Anzahl der Lebendgeborenen je 1000 Einwohner (1998)
- D8 Anzahl der Gestorbenen je 1000 Einwohner (1998)
- D9 Anzahl der Zuzüge je 1000 Einwohner (1997)
- D10 Anzahl der Fortzüge je 1000 Einwohner (1997)

**Soziale Merkmale (S):**

- S1 Anteil der Ausländer an den Grundschulern (Prozent) (1997)
- S2 Anteil der Hauptschüler an den Schülern des siebten Schuljahrgangs (Prozent) (1997)
- S3 Anteil der Personen ohne Hauptschulabschluß an den Schulabgängern aller weiterführenden allgemeinbildenden Schulen (Prozent) (1997)
- S4 Anteil der Ein- und Zweifamilienhäuser an den fertiggestellten Wohngebäuden (Prozent) (1997)
- S5 Verfügbares Einkommen je Einwohner (DM) (1996)
- S6 Bruttoerwerbs- und -vermögenseinkommen je Einwohner (DM) (1996)
- S7 Anzahl der Wohnräume je vorhandene Wohnung (1997)
- S8 Anzahl der Wohnungen je vorhandenes Wohngebäude (1997)
- S9 Stimmenanteil der CDU bei der Bundestagswahl 1998 (Prozent)
- S10 Stimmenanteil der SPD bei der Bundestagswahl 1998 (Prozent)

**Wirtschaftliche Merkmale (W):**

- W1 Anteil der Landwirtschaft, Forstwirtschaft und Fischerei an der Gesamtzahl der Beschäftigten (Prozent) (1996)
- W2 Anteil von Energieversorgung, Wasserversorgung und Bergbau an der Gesamtzahl der Beschäftigten (Prozent) (1996)
- W3 Anteil von Verarbeitendem Gewerbe und Baugewerbe an der Gesamtzahl der Beschäftigten (Prozent) (1996)
- W4 Anteil von Handel, Nachrichtenübermittlung und Verkehr an der Gesamtzahl der Beschäftigten (Prozent) (1996)
- W5 Anteil von Kreditinstituten, Versicherungsgewerbe, Dienstleistungsunternehmen und freien Berufen an der Gesamtzahl der Beschäftigten (Prozent) (1996)
- W6 Anteil von Gebietskörperschaften, Sozialversicherung, Organisationen ohne Erwerbszweck und privaten Haushalten an der Gesamtzahl der Beschäftigten (Prozent) (1996)
- W7 Anteil der sozialversicherungspflichtig beschäftigten Arbeitnehmer an der Gesamtzahl der Beschäftigten (Prozent) (1996)
- W8 Anzahl der Beschäftigten je Betrieb in Bergbau und Gewinnung von Steinen und Erden sowie im Verarbeitenden Gewerbe (Jahresdurchschnitt 1998)
- W9 Bruttowertschöpfung zu Faktorkosten je Erwerbstätigen (DM) (1996)
- W10 Bruttoanlageinvestitionen je Erwerbstätigen (DM) (1996)

**Umweltbezogene Merkmale (U):**

- U1 Anteil der Gebäudefläche und der Gebäuden zugeordneten Freifläche an der Katasterfläche (Prozent) (1997)
- U2 Anteil der Verkehrsfläche an der Katasterfläche (Prozent) (1997)
- U3 Anteil der Erholungsfläche (überwiegend Grünanlagen) an der Katasterfläche (Prozent) (1997)
- U4 Anteil der Waldfläche an der Katasterfläche (Prozent) (1997)
- U5 Anteil des Dauergrünlands an der landwirtschaftlichen Nutzfläche (Prozent) (1997)
- U6 Anzahl (Bestand) der Kraftfahrzeuge je 1000 Einwohner (1998)
- U7 Anzahl der Straßenverkehrsunfälle je 1000 Einwohner (1997)
- U8 In Anlagen der Entsorgungswirtschaft und in betriebseigenen Anlagen (Deponien und thermische Behandlungsanlagen) beseitigte Abfälle (ohne besonders überwachungsbedürftige Abfälle) je Einwohner (kg) (1996)
- U9 Anzahl der Einwohner je km<sup>2</sup> (1997)
- U10 Anteil der Investitionen für den Umweltschutz an den Investitionen insgesamt der Betriebe im Verarbeitenden Gewerbe (Prozent) (1996)

Die Gemeinde mit der kleinsten Summe der demographischen Mahalanobis-Distanzen zu den anderen Gemeinden ist die Stadt Goch (Kreis Kleve). Diese Gemeinde repräsentiert somit die demographische Mitte des Landes Nordrhein-Westfalen. Es sei betont, daß die Mitte keinen Idealzustand darstellen muß, sondern die landesdurchschnittliche Situation im Hinblick auf die Gesamtheit der ausgewählten Merkmale beschreibt. Während aber die für das Land insgesamt ermittelten landesdurchschnittlichen Merkmalswerte in keiner existierenden Gemeinde wirklich vorkommen, sind die Werte der mittleren Gemeinde real. Die Werte der Stadt Goch in den berücksichtigten demographischen Merkmalen sind (in Klammern zum Vergleich landesdurchschnittliche Werte):

D 1	8,2 Prozent (11,4 Prozent)
D 2	9,4 Prozent (16,1 Prozent)
D 3	21,5 Prozent (19,5 Prozent)
D 4	14,7 Prozent (15,9 Prozent)
D 5	39,1 Prozent (30,2 Prozent)
D 6	31,9 Prozent (30,0 Prozent)
D 7	3,0 ( 2,7)
D 8	2,3 ( 2,5)
D 9	49,8 (48,5)
D10	40,6 (47,0)

Goch hat also in jeweils fünf demographischen Merkmalen über und unter dem Landesdurchschnitt liegende Werte. Zu den Merkmalen D5, D6, D9 und D10 ist zu bemerken, daß die landesdurchschnittlichen Vergleichswerte die Zu- und Fortzüge nicht nur über die Landesgrenze, sondern auch über alle Gemeindegrenzen innerhalb des Landes betreffen.

Die kleinste Summe der sozialen Mahalanobis-Distanzen zu den anderen Gemeinden ergibt sich für die Stadt Greven (Kreis Steinfurt), die somit die soziale Mitte des Landes repräsentiert. Die Werte von Greven in den berücksichtigten sozialen Merkmalen sind (in Klammern landesdurchschnittliche Werte):

S 1	8,3 Prozent (15,4 Prozent)
S 2	25,3 Prozent (25,4 Prozent)
S 3	5,1 Prozent ( 6,1 Prozent)
S 4	86,5 Prozent (82,4 Prozent)
S 5	29 981 DM (29 853 DM)

S 6	35 804 DM (34 904 DM)
S 7	4,9 ( 4,3)
S 8	1,8 ( 2,4)
S 9	37,7 Prozent (33,8 Prozent)
S10	42,3 Prozent (46,9 Prozent)

Greven hat in jeweils fünf sozialen Merkmalen über und unter dem Landesdurchschnitt liegende Werte. Die Merkmale S5 und S6 unterscheiden sich dadurch von den anderen Merkmalen, daß sie nicht direkt Ergebnisse statistischer Erhebungen sind, sondern auf Schätzmodellen beruhen (Gerß 1986b).

Als wirtschaftliche Mitte des Landes Nordrhein-Westfalen – mit der kleinsten Summe der wirtschaftlichen Mahalanobis-Distanzen zu den anderen Gemeinden – wird die Stadt Büren (Kreis Paderborn) ausgewiesen. Die Werte von Büren in den berücksichtigten wirtschaftlichen Merkmalen sind (in Klammern landesdurchschnittliche Werte):

W 1	4,6 Prozent (1,7 Prozent)
W 2	0,6 Prozent (2,1 Prozent)
W 3	41,8 Prozent (31,9 Prozent)
W 4	20,0 Prozent (21,1 Prozent)
W 5	18,8 Prozent (23,3 Prozent)
W 6	14,3 Prozent (19,9 Prozent)
W 7	73,0 Prozent (80,7 Prozent)
W 8	86,2 (145,9)
W 9	81 487 DM (100 239 DM)
W10	21 538 DM ( 17 369 DM)

Büren hat zwar nur in drei Merkmalen über und in sieben Merkmalen unter dem Landesdurchschnitt liegende Werte, jedoch gleichen sich insgesamt die über- und die unterdurchschnittlichen Werte relativ fast aus. Die Merkmale W9 und W10 beruhen auf speziellen Schätzmodellen im Rahmen der Sozialproduktberechnung (Gerß 1986a).

Die ökologische Mitte des Landes Nordrhein-Westfalen – mit der kleinsten Summe der umweltbezogenen Mahalanobis-Distanzen zu den anderen Gemeinden – wird durch die Stadt Werne (Kreis Unna) repräsentiert. Die Werte von Werne in den berücksichtigten umweltbezogenen Merkmalen betragen (in Klammern landesdurchschnittliche Werte):

U 1	12,3 Prozent (11,9 Prozent)
U 2	6,0 Prozent ( 6,6 Prozent)
U 3	0,3 Prozent ( 0,8 Prozent)
U 4	16,4 Prozent (24,7 Prozent)
U 5	13,0 Prozent (13,9 Prozent)
U 6	605 (581)
U 7	5,6 (5,7)
U 8	18,4 (1191,0)
U 9	405 (527)
U10	1,1 Prozent ( 4,3 Prozent)

Wenn auch Werne nur nach zwei Merkmalen über und nach acht Merkmalen zum Teil erheblich unter dem Landesdurchschnitt liegt, ist die Zuordnung dieser Gemeinde zur „Mitte“ des Landes nicht unplausibel. Die Merkmale, in denen Werne relativ am weitesten vom Landesdurchschnitt abweicht – insbesondere U8, U10 und U3 –, streuen extrem stark über die 396 Gemeinden. Die groß erscheinenden Abweichungen von Werne nach diesen Merkmalen sind somit tatsächlich vergleichsweise gering. Im übrigen ist für die umweltschutzstatistischen Merkmale im engeren Sinne – wie U8 – charakteristisch, daß sie oft sehr komplex sein müssen, um für Gemeinden aussagefähig sein zu können.

#### 4. Resümee

Mit der hier angewendeten Methode werden „mittlere“ Gemeinden ermittelt, die unter bestimmten variierenden Gesichtspunkten jeweils typisch sind. Dabei muß definiert werden, welche Art von „Mitte“ gemeint ist. Der Anspruch der GfK, einen einzigen „sozio-demographischen Mittelpunkt“ zu ermitteln, soll und kann wohl auch so allgemein nicht erfüllt werden. Die Merkmale, die die jeweils relevante Mitte kennzeichnen, werden hier nach Verfügbarkeit und Bewährung in anderen Untersuchungen festgelegt, aber nicht nach angenommener oder behaupteter Relevanz bevorzugt ausgewählt beziehungsweise zurückgestellt. Die Korrelation zwischen den Merkmalen spielt bei deren Auswahl keine Rolle. Die notwendige Praktikabilität der Datenbeschaffung und –verrechnung ist durch den direkten Zugriff auf die Landesdatenbank Nordrhein-Westfalen gegeben. Das geschilderte Verfahren ist mit Hilfe eines zu diesem Zweck erstellten EDV-Programms einfach anwendbar. Die Berechnung von Mahalanobis-Distanzen innerhalb einer hochdimensionalen Datenmatrix stellt besondere Anforderungen an die Rechnerkapazität. Die Bearbeitung einer derartigen multivaria-



ten Problemstellung wurde durch optimierte Ausnutzung der technischen Gegebenheiten bewerkstelligt. Dies betrifft insbesondere die Verwendung eines geeigneten Softwarepakets (S-Plus) sowie die Programmierung effizienter und kapazitätsschonender Algorithmen. Damit ist auch die Einbeziehung einer Anzahl von Merkmalen möglich, die größer ist als in den hier vorgestellten Beispielen.

### **Korrespondenzadresse**

Prof. Dr. Wolfgang Gerß  
Eifelstraße 14  
42579 Heiligenhaus

### **Literatur**

Bock, H.H., 1974: Automatische Klassifikation. Göttingen: Vandenhoeck und Ruprecht.

Gerß, W., 1986a: Ergebnisse der Volkswirtschaftlichen Gesamtrechnungen für Gemeinden. Statistische Rundschau für das Land Nordrhein-Westfalen 38(4): 211-218.

Gerß, W., 1986b: Schätzung aggregierter privater Einkommen in den Gemeinden. Statistische Rundschau für das Land Nordrhein-Westfalen 38 (12): 817-826.

Gerß, W., 1994: Kennzeichen der Planungsregionen des Landes Nordrhein-Westfalen. Statistische Rundschau Nordrhein-Westfalen 46 (11): 565-574.

GfK (Gesellschaft für Konsumforschung), 1998: Der erste experimentelle Mikro-Testmarkt Europas mit Targetable TV. Nürnberg: Selbstverlag.

Plewa, H.-E., 1985: Die Landesdatenbank Nordrhein-Westfalen – Ein modernes Auskunftsinstrument der amtlichen Statistik. Statistische Rundschau für das Land Nordrhein-Westfalen 37(6): 383-390.