

Statistische Analysen und Studien Nordrhein-Westfalen

Band 1

Impressum

Herausgeber:

Landesamt für Datenverarbeitung
und Statistik Nordrhein-Westfalen

Redaktion:

Bianca Klose, Hans Lohmann

Preis dieser Ausgabe: 3,20 EUR

Erscheinungsfolge: unregelmäßig

Bestellungen nehmen entgegen:

das Landesamt für Datenverarbeitung
und Statistik NRW,
Postfach 10 11 05,
40002 Düsseldorf,
Mauerstraße 51,
40476 Düsseldorf
Telefon: 0211 9449-2516/3516
Telefax: 0211 442006
Internet: <http://www.lids.nrw.de>
E-Mail: poststelle@lds.nrw.de

sowie der Buchhandel.

Pressestelle:

0211 9449-2521/2518

Zentraler Informationsdienst:

0211 9449-2495/2525

© Landesamt für Datenverarbeitung
und Statistik NRW, Düsseldorf, 2002

Für nicht gewerbliche Zwecke sind
Vervielfältigung und unentgeltliche
Verbreitung, auch auszugsweise, mit
Quellenangabe gestattet. Die Verbrei-
tung, auch auszugsweise, über elek-
tronische Systeme/Datenträger bedarf
der vorherigen Zustimmung. Alle üb-
rigen Rechte bleiben vorbehalten.

Bestell-Nr. Z 08 1 2002 51

ISSN 1619-506X

Inhalt

Haushaltstypisierung im Rahmen des registergestützten Zensus Dr. Marco Scharmer	3
Die Ermittlung des Wohnungsinhabers im Rahmen der Gebäude- und Wohnungszählung (GWZ) und deren Bedeutung für die Haushaltegenerierung Dr. Henrik Egbert und Dr. Marco Scharmer	10
Räumliche Mobilität und Arbeitsmarktprozesse Eine Analyse auf der Basis des Sozio-oekonomischen Panels Dr. Sylvia Zühlke und Uwe Hetke	15
Sicherung persönlicher Angaben in Tabellendaten Dipl.-Physiker Rüdiger Dietz Reipsilber	24
Das Personal des Landes am 30. Juni 2000 Regionalisierte Ergebnisse der Personalstandstatistik Dipl.-Volkswirt Heinz-Peter Emmerich	36

Zeichenerklärung

(nach DIN 55 301)

- 0 weniger als die Hälfte von 1 in der letzten besetzten Stelle,
jedoch mehr als nichts
- nichts vorhanden (genau null)
- . Zahlenwert unbekannt oder geheim zu halten
- ... Zahlenwert lag bei Redaktionsschluss noch nicht vor
- () Aussagewert eingeschränkt, da der Wert Fehler aufweisen kann
- / keine Angabe, da der Zahlenwert nicht sicher genug ist
- x Tabellenfach gesperrt, weil Aussage nicht sinnvoll
- p vorläufige Zahl
- r berichtigte Zahl

Abweichungen in den Summen erklären sich
aus dem Runden der Einzelwerte.

Haushaltstypisierung im Rahmen des registergestützten Zensustests

Dr. Marco Scharmer

Zensen sind der zentrale Bestandteil des Systems der amtlichen Statistik in der Bundesrepublik Deutschland. Sie liefern Basisdaten über die Bevölkerung, Erwerbstätigkeit und Wohnsituation, die als Grundlage für politische, wirtschaftliche und gesellschaftliche Planungsprozesse unentbehrlich sind. Daneben dienen Zensen im statistischen Gesamtsystem auch als Basis für nachfolgende Stichprobenerhebungen, wie beispielsweise den Mikrozensus. Die letzte Volkszählung wurde in Deutschland im früheren Bundesgebiet 1987, in der DDR im Jahre 1981 durchgeführt. Seitdem fanden beträchtliche Wanderungsbewegungen vor und nach der Wiedervereinigung sowie eine starke Zuwanderung von Aussiedlern und Ausländern zu Beginn der 90er-Jahre statt. Dieses führte dazu, dass die zur Verfügung stehenden Daten immer weniger die tatsächlichen Verhältnisse abbilden.

Die Bundesregierung hat sich jedoch aus Kosten- und Akzeptanzgründen bereits 1996 gegen eine herkömmliche Vollerhebung nach dem Vorbild der Volkszählung von 1987 entschieden. Vor dem Hintergrund von Plänen der Europäischen Union, im Jahre 2001 eine unionsweite Volks- und Wohnungszählung durchzuführen, wurde in Deutschland von den Statistischen Ämtern des Bundes und der Länder ein Modell entwickelt, das die Nutzung von Daten aus den Registern der Einwohnermeldeämter sowie von Daten der Bundesanstalt für Arbeit vorsieht. Den zentralen Bestandteil dieses Modells bildet die Zusammenführung von Melderegisterangaben mit postalisch erhobenen Wohnungs- und Gebäudeangaben.

Da die amtliche Statistik in der Bundesrepublik Deutschland mit diesem Methodenwechsel von einer primärstatistischen Vollerhebung zur Nutzung von bestehenden Registern Neuland betritt, sind umfangreiche Qualitäts- und Verfahrenstests hinsichtlich der Statistikauglichkeit dieser Register und des zu entwickelnden Verfahrens vorzunehmen.

Um Qualität und Validität der aus den Registern gewonnenen Daten sowie den Verfahrensablauf zu überprüfen, sind in einer Stichprobe auch Befragungen erforderlich, die bei einem künftigen Zensus weitgehend entfallen können. Die Voraussetzung für die Überprüfung des neuentwickelten Verfahrens bildet das am 3. August 2001 in Kraft getretene Gesetz zur Vorbereitung eines registergestützten Zensus.¹⁾

Mit dem vorliegenden Beitrag wird ein Überblick über die Haushaltstypisierung im Rahmen des Zensustests gegeben. Es werden die aufgetretenen methodischen Schwierigkeiten der Entwicklung eines konsistenten Typisierungskonzepts, deren Lösung und die bei einem zukünftigen registergestützten Zensus hieraus zu erwartenden Probleme mit der nachfolgenden Stichprobenerhebung Mikrozensus dargestellt.²⁾

1) Vgl. Gesetz zur Vorbereitung eines registergestützten Zensus (Zensusvorbereitungsgesetz), Bundesgesetzblatt, Jg. 2001, Teil 1 Nr. 40, S. 1882ff. – 2) Vgl. hierzu auch: Forster, Michael, Die Zukunft der Volkszählung in Deutschland, Traditionelle Zählung oder registergestützter Zensus?, in: Statistische Analysen und Studien Nordrhein-Westfalen, Ausgabe 4/01, S. 23ff und Fürnrohr, Michael/Rimmelpacher, Birgit, Testuntersuchungen zur Vorbereitung eines registergestützten Zensus, in: Bayern in Zahlen; Heft 1/2001

Die Idee der Haushaltegenerierung

Für Planungsaufgaben öffentlicher und privatwirtschaftlicher Institutionen sind aktuelle und tief gegliederte Daten über Größe und Struktur von Haushalten, welche die sozialen und ökonomi-

schen Verhältnisse der Bevölkerung widerspiegeln, von grundlegendem Interesse. Aus diesem Grund ist die Erhebung haushaltsstatistischer Daten zentraler Bestandteil der amtlichen Statistik im Rahmen jeder Volkszählung sowie des jährlichen Mikrozensus. Die regional tief gegliederten Haushaltsergeb-

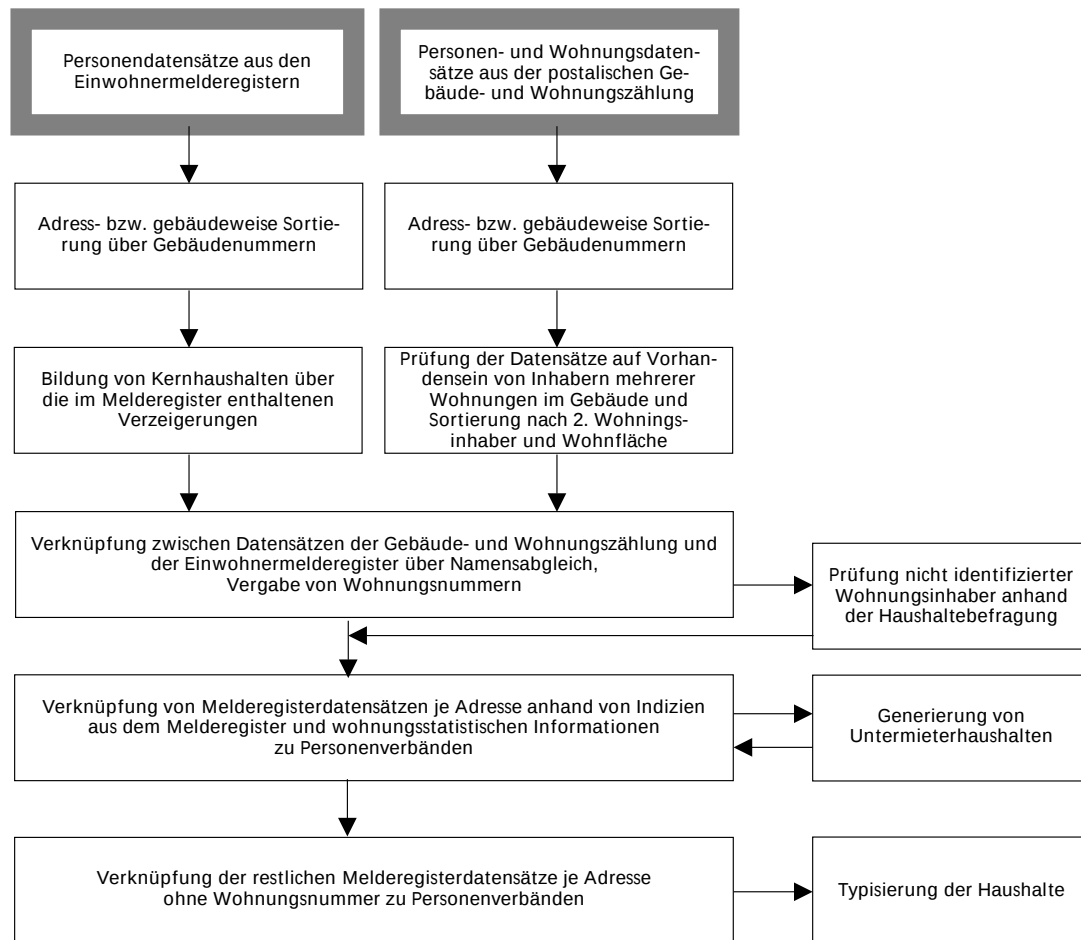
nisse einer Volkszählung standen jedoch nur in sehr großen Zeitabständen zur Verfügung und bei den aktuellen Mikrozensusergebnissen handelt es sich um eine 1 %-Stichprobe, welche Aussagen lediglich auf Ebene von Anpassungsschichten zulässt.³⁾

Mit dem Übergang von einer traditionellen Vollerhebung zu einem registergestützten Zensus stellt sich die Frage, wie künftig Haushaltszahlen und -strukturen erhoben werden können. Bei einem herkömmlichen Zensus werden alle Angaben im Haushaltszusammenhang direkt durch Befragungen erhoben. Diese Möglichkeit besteht in einem registergestützten Zensus nicht. Da insbesondere im kommunalen Bereich, aber auch in der Marktforschung eine große Nachfrage nach aktuellen und kleinräumig gegliederten Haushaltsdaten besteht, wurde die Idee, mit Hilfe von Informationen aus den Melderegistern haushaltsstatistische Ergebnisse zu generieren, bereits Ende der 70er-Jahre geboren.⁴⁾

Die grundsätzliche Idee des registergestützten Zensus ist ein von den statistischen Ämtern der Länder entwickeltes statistisches Konzept, welches die in den Personendatensätzen⁵⁾ der Einwohnermelderegister gespeicherten individuellen Informationen nutzt. Zurückgegriffen werden konnte auf die seit den 80er-Jahren entwickelten Programme zur Haushaltegenerierung, mit denen Haushaltszahlen und -strukturen ermit-

3) Unter Anpassungsschichten sind i. d. R. benachbarte Kreise zu verstehen. Da auf Kreisebene die Ergebnisse des Mikrozensus teilweise so dünn besetzt sind, dass der Stichprobenfehler in dieser tiefen regionalen Gliederung zu groß sein würde, sind alle Kreise bzw. kreisfreien Städte, die weniger als 300 000 Einwohner haben, zu Anpassungsschichten zusammengefasst worden. – 4) Vgl. von Klitzing, Friedrich/Osenberg, Hanno, Haushaltsdaten aus dem Melderegister, in: Beiträge zur Stadtforschung und Statistik, Köln 1995, Heft 1, S. 15ff. – 5) Als Personendatensatz wird der Einzeldatensatz bezeichnet, der sich aus den bei der Meldebehörde gespeicherten Daten (z.B. Vor-, Geburts-, Ehe- und Familienname Geburtsdatum, Anschrift, Einzugsdatum, usw.) zusammensetzt.

Ablaufschema des Haushaltegenerierungsverfahrens



Grafik: LDS NRW

telt werden können.⁶⁾ Die Ergebnisse der bislang angewandten Verfahren weisen gegenüber denen von primärstatistischen Erhebungen jedoch erhebliche Differenzen auf. Untersuchungen zeigten, dass bei der Haushaltegenerierung lediglich auf Basis von Angaben aus den Einwohnermelderegistern zu viele Haushalte gebildet werden, da Personen, die nicht in einen Haushaltszusammenhang zu bringen waren, als Einpersonenhaushalt ausgewiesen wurden.⁷⁾

6) So wird von zahlreichen Kommunen ein vom KOSIS-Verbund (Kommunales Statistisches Informations System) entwickeltes Verfahren angewendet, welches nur auf der Basis der Melderegisterdaten eine Haushaltegenerierung vornimmt. Vgl. von Klitzing, Friedrich, Haushaltegenerierungsverfahren HHGen98, Dokumentation im Auftrag des KOSIS-Verbunds, 01. 01. 1998. – 7) Zum Aufbau und zur Bewertung der bisherigen Haushaltegenerierungsverfahren vgl. Fest, Manfred/Herzog, Petra, Möglichkeiten einer Haushaltegenerierung unter Verwendung von Indizien aus Einwohnermelderegistern und wohnungsstatistischen Daten, in: Daten und Analysen, Heft 5/99, und Fürnrohr, Michael/König, Markus, Möglichkeiten einer Haushaltegenerierung im Rahmen der Zusammenführung von Einzeldaten aus Melderegistern mit primärstatistisch gewonnenen Wohnungsdaten, in: Bayern in Zahlen, Heft 4/99.

Die Erfahrungen mit den bisher entwickelten Haushaltegenerierungsverfahren haben gezeigt, dass die Qualität der Ergebnisse mit einer steigenden Anzahl von Informationen, vor allem der Einbeziehung von gebäude- und wohnungsstatistischen Daten, deutlich zunimmt. Damit das Haushaltegenerierungsverfahren als Teil des registergestützten Zensus nicht die in den Praxistests zutage getretenen Schwächen der bereits entwickelten Verfahren besitzt, wurde für den Zensus ein erheblich erweitertes Programm entwickelt. Die entscheidende Änderung gegenüber den bisherigen Verfahren ist jedoch, dass das Haushaltegenerierungsverfahren im Rahmen des Zensus Personendatensätze aus den Melderegistern systematisch ausgewertet und adressweise mit gebäude- und wohnungsstatistischen Informationen der zusätzlich bei den Gebäudeeigentümern durchzuführenden postalischen Gebäude- und

Wohnungszählung (GWZ) zusammenführt. Hierbei sollen die in der GWZ erhobenen Wohnungsinhaber und deren Wohnungsdatensätze mit den entsprechenden Datensätzen im Melderegister verknüpft werden. Anschließend werden über die Algorithmen des Haushaltegenerierungsverfahrens die übrigen Personensätze im Melderegister mit den Wohnungsinhabern zu Haushalten verknüpft. Dabei gibt die in der GWZ ebenfalls erfragte Zahl der bewohnten Wohnungen die Gesamtzahl der Haushalte vor (Wohnhaushalte).

Ziel der Haushaltegenerierung ist es, haushaltstypische Personenbeziehungen zu bilden und damit Erkenntnisse über Größe und Struktur von Haushalten sowie deren soziale und wirtschaftliche Verhältnisse zu erlangen. Soweit in den Personendatensätzen der Melderegister Verzeigerungen enthalten sind, ist eine Zusammenführung von Personen zu

Haushalten unkompliziert. Unter Verzeigerungen sind Merkmale in einem Personendatensatz zu verstehen, die auf einen Datensatz einer anderen Person hinweisen. Dieses ist der Fall bei Ehepartnern, minderjährigen Kindern und deren Eltern(-teilen) sowie minderjährigen Kindern und deren gesetzlichen Vertretern (einschließlich Stief- und Pflegeeltern).⁸⁾ Bei diesen Personenpaaren wird jeweils im Datensatz des einen auf den Datensatz des anderen verwiesen und vice versa. Die über Verzeigerungen zusammengeführten Personen bilden den so genannten Kernhaushalt, welcher ein vollständiger Haushalt, aber auch Teil eines Haushalts mit weiteren nicht verzeigten Personen sein kann.⁹⁾

Wesentlich schwieriger gestaltet sich hingegen eine Zusammenführung von Personen zu Haushalten, deren Melderegisterdatensätze keine Verzeigerungen aufweisen. Hier liegt dem statistischen Verfahren der Haushaltegenerierung die Überlegung zugrunde, dass jede Personenbeziehung, die einen Haushaltszusammenhang bedingen kann, eine bestimmte Übereinstimmung oder Kombination der im Personendatensatz der Melderegister enthaltenen Merkmale voraussetzt. Für jede potenzielle Personenbeziehung wird also eine Anzahl von Bedingungen in Form von Merkmalsübereinstimmungen bzw. Merkmalskombinationen zwischen den Personendatensätzen definiert, die erfüllt sein müssen, damit im Rahmen der Haushaltegenerierung die Zugehörigkeit der betreffenden Personen zum gleichen Haushalt mit hoher Wahrscheinlichkeit unterstellt werden kann.

Häufig bestehen mehrere Möglichkeiten, Personen zu verschiedenen Haushalten zusammenzuführen.¹⁰⁾

8) Teilweise enthalten die Melderegister in Deutschland auch Verzeigerungen zwischen Eltern und deren erwachsenen Kindern bis zum Alter von 27 Jahren. – 9) Ein Beispiel hierfür wäre, wenn im Haushalt einer Kernfamilie (Mutter, Vater, Kind) der Elternteil eines Ehepartners lebt. – 10) So können mit dem Verfahren der Haushaltegenerierung solche Haushalte wie nichteheliche Lebensgemeinschaften oder Wohngemeinschaften nur sehr schwer gebildet werden, da hier Merkmalsübereinstimmungen, die auf Namensgleichheiten oder gemeinsamer Mobilität aufbauen, nur selten gegeben sind.

Neben der Entwicklung geeigneter Merkmalsübereinstimmungen bzw. Merkmalskombinationen steht die Haushaltegenerierung somit vor dem Problem, die verschiedenen Bedingungen für die Merkmalsübereinstimmungen bzw. -kombinationen in eine Reihenfolge zu bringen, mit denen die in der Realität vorliegenden Haushaltszusammenhänge valide gebildet werden können.

Zur Notwendigkeit von Qualitäts- und Verfahrenstests

Bedeutende Voraussetzungen für verlässliche haushaltsstatistische Ergebnisse sind zum einen eine qualitativ hochwertige Datengrundlage und zum anderen das Vorhandensein eines ausgereiften Verfahrens zur Haushaltegenerierung.

Untersuchungen der Einwohnermelderegister haben gezeigt, dass Einschränkungen im Hinblick auf die Statistikauglichkeit bestehen.¹¹⁾ Dieses liegt u. a. daran, dass die Register zu Verwaltungszwecken geführt werden und bisher kein umfassender Einsatz in der amtlichen Statistik erfolgte. Als „lebende Register“ unterliegen sie einer laufenden, auch rückwirkenden Veränderung und Berichtigung, wodurch zu keinem Stichtag ein exakter Bestand aller Personen wiedergegeben wird.

Angesichts der Bedeutung der Einwohnermelderegister als zentrale Datengrundlage in einem registergestützten Zensus ist es notwendig, die Qualität der Melderegister und damit auch die erreichbare Qualität eines registergestützten Zensus zu überprüfen. Im Rahmen des Zensus-tests wurden Verfahren entwickelt, mit denen die Qualität der Melderegister hinsichtlich der Anzahl von so

genannten Mehrfachfällen¹²⁾, der Über-¹³⁾ und Untererfassung¹⁴⁾ sowie hinsichtlich der Vollständigkeit und des Grades des Zutreffens der zu einer Person gespeicherten Merkmale überprüft werden kann.

Zusätzlich wird im Rahmen des Zensus-tests in den ausgewählten Gebäuden eine Befragung aller Personen zu demografischen, erwerbs- und wohnungsstatistischen Merkmalen durchgeführt. Diese Befragung entspricht dem Vorgehen in einer herkömmlichen Volkszählung und soll als Maßstab für die Überprüfung der Qualität der Melderegister sowie der im Verfahren der Haushaltegenerierung gebildeten Haushaltszusammenhänge dienen.¹⁵⁾ Der Vergleich der Ergebnisse der Haushaltegenerierung mit denen der Befragung stellt damit eine Kontrolle der Zensus-testergebnisse dar. Der Zensus-test wird zeigen, ob die auf der Kombination von Informationen aus den Melderegistern mit wohnungsstatistischen Daten beruhende Haushaltegenerierung geeignet ist, Haushalts- und Wohnungsbelegungsdaten in zensustypischer Qualität zu ermitteln. In einem späteren registergestützten Zensus soll eine Befragung jedoch nur bei Haushaltsmitgliedern stattfinden, wenn die entsprechende Generierung dieser Haushalte zu unplausiblen Ergebnissen führt.

Haushaltstypisierung im Zensus-test

Die Ermittlung von Daten über die Struktur und Größe privater Haushalte war von jeher wichtiger Bestandteil von Volkszählungen, die in der Vergangenheit die einzig verfügbare Quelle von Haushaltsergeb-

12) Ein Mehrfachfall liegt dann vor, wenn eine Person in mehreren Gemeinden gleichzeitig mit alleinigem Wohnsitz, mit Hauptwohnsitz oder nur mit Nebenwohnsitz gemeldet ist. – 13) Unter Übererfassung (Karteileichen) versteht man Personen, die noch an Adressen gemeldet sind, an denen sie inzwischen aber weder Haupt- noch Nebenwohnsitz haben. Es handelt sich somit um eine Übererfassung im Melderegister. – 14) Als Untererfassung (Fehlbestände) werden Personen bezeichnet, die an dem Ort ihrer Wohnung nicht gemeldet sind. Es handelt sich damit um eine Untererfassung im Melderegister. – 15) Vgl. Forster, Michael, a. a. O. 4/01, S. 23ff., und Fürnrohr, Michael/Rimmelspacher, Birgit, Testuntersuchungen zur Vorbereitung eines registergestützten Zensus, in: Bayern in Zahlen, Heft 1/01; S. 13ff.

nissen in tiefer regionaler Gliederung waren. Für viele Fragestellungen sind Haushaltszusammenhänge von Bedeutung. So wird unter anderem die Nachfrage nach Wohnraum und nach einer Vielzahl von Konsumgütern von der Anzahl und Größe der Haushalte bestimmt. Auch eine Reihe staatlicher Maßnahmen und Leistungen stellen auf den Haushalt als Bezugsgröße ab. Haushaltstatistische Ergebnisse der Volkszählungen gaben darüber hinaus in Verbindung mit Wohnungsdaten auch Aufschluss über die Wohnsituation bestimmter Bevölkerungsgruppen.

Der Übergang zum Wohnhaushalt im registergestützten Zensus

Ziel der Haushaltegenerierung ist es, über die Nutzung der individuellen Informationen aus den Einwohnermelderegistern Erkenntnisse über haushaltstypische Personenbeziehungen sowie über die Zugehörigkeit von Personen zu bzw. deren Stellung innerhalb eines Haushaltes zu erlangen.¹⁶⁾ Hierbei wird von der in der amtlichen Statistik bisher üblichen Haushaltsdefinition¹⁷⁾ „ein Haushalt ist eine Personengruppe, die gemeinsam wohnt und wirtschaftet, d. h. insbesondere ihren Lebensunterhalt gemeinsam finanziert“ abgegangen. Im Rahmen einer Haushaltegenerierung wird nur noch auf das gemeinsame Wohnen der haushaltzugehörigen Personen abgestellt (Wohnhaushalte), da die Melderegister als zentrale Datengrundlage keine Informationen über die Finanzierung des Lebensunterhalts liefern können. Insofern führt der Übergang von einem herkömmlichen Zensus zu einem registergestützten Zensus zu einem Informationsverlust hinsichtlich der ökonomischen Zusammenhänge innerhalb der Wohnhaushalte.

16) Haushaltstypische Personenbeziehungen sind beispielsweise Ehepartner, nichteheliche Partner, Eltern(-teil) und erwachsene Nachkommen, Enkel und Großeltern(-teil) oder Mitglieder einer Wohngemeinschaft. – 17) so beispielsweise die Haushalte-Definition in der Volkszählung 1987 oder im Mikrozensus

Haushaltstypisierung im Rahmen der Generierung

Durch die Haushaltegenerierung werden Haushaltszusammenhänge von Personen, die zusammen wohnen, aufgrund bestimmter Übereinstimmungen oder Kombinationen der im Personendatensatz der Melderegister enthaltenen Merkmale gebildet. Die somit entstehenden Haushalte entsprechen der Definition von Wohnhaushalten.

Aufgrund der festgestellten Personen-, Haushalts- und Wohnungszusammenhänge werden Merkmale zur Stellung der Person im Personenverband ermittelt und gespeichert. Dabei handelt es sich um die folgenden Merkmale:

- Paareigenschaft (Partner in Ehepaar oder in nichtehelichem Paar, keine Partnereigenschaft)
- Elterneigenschaft (Elternteil, kein Elternteil)
- Vorfahreigenschaft (Vorfahre/Erziehender, kein Vorfahre)
- Nachkommeneigenschaft (Kind bei Elternteil oder ohne Elternteil/Vorfahre/Erzieher, erwachsener Nachkomme bei/mit Elternteil, Nachkomme bei Großelternteil, Kind bei sonstigem Erzieher, nicht als Kind/Nachkomme eingestuft)
- Geschwistereigenschaft (Geschwister, keine Geschwistereigenschaft)
- Generationsziffer der Person (erste Generation = älteste Generation im Haushalt usw.)

Zum Zwecke der Haushaltstypisierung wird in jedem Haushalt an eine Person die Ausprägung „Bezugsperson“ vergeben. Dieses ist in der Regel der im Melderegister identifizierte (erste) Wohnungsinhaber aus dem Datensatz der GWZ, also die Person, über die eine Verknüpfung zwischen den GWZ-Wohnungssätzen und den Melderegisterdatensätzen erfolgte. Für alle Personen, die nicht mit einem Wohnungssatz verknüpft wurden, unterbleibt die Vergabe der Ausprägung „Bezugsperson“. Diese Personen werden, unter Beachtung bestimmter Kriterien gegebenenfalls mit weiteren Personen, in eine unver-

knüpfte Wohnung¹⁸⁾ generiert und als „Sonstiger Haushalt“ klassifiziert. Außerdem wird jeder Person in einem Haushalt mit Bezugsperson eine Ausprägung zugeordnet, aus der sich deren Stellung zur festgelegten Bezugsperson ableiten lässt:

- Ehepartner der Bezugsperson
- Nichtehelicher Partner der Bezugsperson
- Nachkomme der Bezugsperson oder des Partners der Bezugsperson
- Enkel, Urenkel der Bezugsperson oder des Partners der Bezugsperson
- Sonstiges Kind (nicht mit der Bezugsperson verzeigert und im Alter von unter 18 Jahren)
- Elternteil der Bezugsperson oder des Partners der Bezugsperson
- Großelternteil der Bezugsperson oder des Partners der Bezugsperson
- Geschwisterteil der Bezugsperson oder des Partners der Bezugsperson
- Ehepartner einer Person im Haushalt (nicht der Bezugsperson)
- Sonstige Person

Unter der Voraussetzung valider Angaben in den Eingangsdatensätzen der Melderegister und der GWZ wird es im Rahmen der Haushaltegenerierung möglich, die Stellung jeder Person zu jeder weiteren Person innerhalb des Wohnhaushaltes darzustellen.

Probleme des Vergleichs von Haushaltstypen der Befragung und der Generierung

Die Typisierung im Rahmen der neu entwickelten Haushaltegenerierung führt im Zensustest zu erheblichen Problemen hinsichtlich einer Vergleichbarkeit mit entsprechenden Ergebnissen der Haushaltebefragung. So ist zu beachten, dass in der Haushaltebefragung, entsprechend dem Vorgehen in einer herkömmlichen Volkszählung, Haushalte in der Abgrenzung von Personen, die gemeinsam wohnen und wirtschaften, erfragt werden. Die Beibehaltung des

18) Unverknüpfte Wohnungen sind Wohnungen, bei denen kein Wohnungsinhaber angegeben ist oder der Wohnungsinhaber laut GWZ im Melderegister nicht zu identifizieren ist.

Haushaltstypenkonzepts in der Befragung wurde u. a. gefordert, damit sich die Ergebnisse des neuen Verfahrens eines registergestützten Zensus an den Ergebnissen der Befragung nach dem Vorbild einer herkömmlichen Volkszählung bewerten lassen. Hiernach bildet eine Person einen eigenen (Wirtschafts-)Haushalt, wenn sie alleine wirtschaftet, und zwar auch dann, wenn sie mit anderen Personen gemeinsam eine Wohnung bewohnt.

Das herkömmliche Konzept der Befragung hat zur Folge, dass bestimmte Haushaltsstrukturen nicht immer eindeutig abgebildet werden können. So kann eine Lebenspartnerschaft aus unverheirateten Personen nur festgestellt und typisiert werden, wenn einer der Lebenspartner die Bezugsperson (erste Person) des Haushaltes ist¹⁹⁾. Ist eine andere verwandte Person im Haushalt die Bezugsperson, z. B. ein Elternteil der Lebenspartner, so kann die nichteheliche Lebensgemeinschaft nicht zweifelsfrei erkannt werden. Besteht zur ersten Person kein Verwandtschaftsverhältnis, so wird die nichteheliche Lebensgemeinschaft überhaupt nicht erkannt. Der Haushalt kann nur noch der Kategorie „Sonstiger Haushalt“ zugeordnet werden.

Wie problematisch eine eindeutige Abgrenzung von Haushaltstypen ist, zeigt sich auch an dem folgenden Sachverhalt. So sind selbst Ehepaare nicht zweifelsfrei zu identifizieren, wenn die Bezugsperson nicht einer der Ehepartner ist, sondern ein Elternteil von einem der Ehepartner. In diesem Fall zeigen beide Ehepartner im Merkmal der Beziehung zur Bezugsperson die Ausprägung (Schwieger-)Tochter/Sohn. Die nahezu zweifelsfreie Unterscheidung zwischen einem Ehepaar und einem Geschwisterpaar unterschiedlichen Geschlechts ist dann nur möglich, wenn die Geburtsnamen der beiden Personen verglichen werden. Besteht zur ersten Person kein Verwandtschaftsverhältnis, so kann auch hier kein

Ehepaar erkannt werden. Der Haushalt wird dann zwangsläufig als „Sonstiger Haushalt“ typisiert.

Zusammenfassend lässt sich festhalten, dass im Rahmen der herkömmlichen Befragung eheliche wie nicht-eheliche Lebenspartnerschaften nur dann eindeutig erkannt werden, wenn einer der Partner die Bezugsperson ist. Dieses ist aber insbesondere in Mehrgenerationenhaushalten mit Großelternteilen, Elternteilen und Kindern nicht zwingend der Fall. Es ist durchaus wahrscheinlich, dass zwei Mehrgenerationenwohnhäuser in unterschiedliche Haushaltstypen kategorisiert werden, obwohl identische Personenbeziehungen innerhalb beider Haushalte bestehen. Gehört beispielsweise die Bezugsperson im ersten Haushalt der Großelterngeneration und im zweiten Haushalt der verheirateten Elterngeneration an, so werden die Personenbeziehungen nicht als identisch erkannt. Dieses führt dazu, im ersten Fall einen „Sonstigen Haushalt“ und im zweiten Fall den Haushaltstyp „Ehepaar mit Kind(ern) und weiteren verwandten Person(en)“ auszuweisen.

Maßgebend für die Typisierung in der Befragung ist damit, wer als Bezugsperson des befragten Haushalts in den Fragebogen eingesetzt wird. In der Haushaltebefragung im Rahmen des registergestützten Zensus werden, in Analogie zum Befragungsschema des Mikrozensus, Ausfüllhinweise zur ersten Person gegeben²⁰⁾. Sofern es sich um einen Zweigenerationenhaushalt mit minderjährigen Kindern handelt, ist eine Typisierung der Haushalte eindeutig. Leben in dem betrachteten Haushalt jedoch Großelternteile oder erwachsene Kinder mit eigenem Einkommen, dann ist die Entscheidung über die Bezugsperson i. d. R. dem Interviewer und bei Selbstausfüllern den Haushaltsmitgliedern überlassen. Damit ist eine eindeutige Haushaltstypisierung mit den Befragungsergebnissen nicht immer gewährleistet.

20) Erste Person oder Bezugsperson ist die Person, die maßgeblich zum Lebensunterhalt des Haushaltes beiträgt. Reihenfolge: Ehegatten/Lebenspartner, Kinder, Verwandte, sonstige Personen.

Da die Haushaltstypisierung aufgrund des abweichenden methodischen Ansatzes in der Generierung auch bei Mehrgenerationenhaushalten eindeutige Ergebnisse liefern wird – jede Beziehung einer Person zu allen anderen Personen im Haushalt wird durch das Verfahren erkannt – können Inkonsistenzen zwischen den Ergebnissen der Befragung und der Generierung auftreten. Ein Vergleich sowie eine Bewertung des Verfahrens registergestützter Zensus ist damit auf Haushaltsebene zunächst nicht möglich.

Haushaltstypen im Zensustest

Die Haushaltebefragung soll u. a. als Maßstab und Kontrollinstrument für die im Haushaltegenerierungsverfahren gebildeten Haushaltszusammenhänge dienen. Das Ergebnis dieses Vergleichs wird Aussagen darüber zulassen, ob die Kombination von Melderegisterdaten mit wohnungsstatistischen Angaben im Haushaltegenerierungsverfahren geeignet ist, Haushalts- und Wohnungsbelegungsdaten in zensustypischer Qualität zu ermitteln. Bei der Entwicklung eines Konzepts zur Haushaltstypisierung im Zensustest war es daher das Ziel, die durch die unterschiedlichen methodischen Ansätze von Befragung und Generierung auftretenden Inkonsistenzen bei der Bildung von Haushaltstypen auszuschalten. Eine Auswertung der Ergebnisse des Zensustests nach Haushaltstypen setzt damit eine einheitliche Haushaltstypisierung sowohl im Datensatz der Haushaltegenerierung (Personendaten aus dem Melderegister) als auch in den Personendaten der Haushaltebefragung voraus.

Eine deckungsgleiche Typisierung von Haushalten konnte aufgrund der unterschiedlichen methodischen Ansätze nur auf einer hochaggregierten Ebene erreicht werden. Für einen Vergleich der Haushalte zwischen Haushaltebefragung und Haushaltegenerierung werden die erfassten Personen in den Haushalten der Haushaltebefragung ent-

19) Im sehr ähnlichen Fragekatalog des Mikrozensus ist die Beantwortung der Frage zur Lebenspartnerschaft sogar freiwillig.

sprechend der laufenden Wohnungsnummer je Gebäude zu Wohnhaushalten zusammengefasst. Im Rahmen des Zensustests wird eine Auswertung von haushaltsstatistischen Ergebnissen lediglich über die folgenden Haushalte vorgenommen, wobei mit Ausnahme der Einpersonenhaushalte jeder Haushaltstyp nach der Anzahl von Kindern ausgewertet werden kann:

- Einpersonenhaushalte
- Ehepaar mit und ohne Kinder(n)
- Nichteheliche Lebensgemeinschaften mit und ohne Kinder(n)
- Elternteil mit Kind(ern)
- Sonstige Haushalte mit und ohne Kinder(n)

Fazit und Ausblick

Das Vorgehen, alle über den Kernhaushalt im Sinne der Generierung²¹⁾ hinausgehenden Mehrgenerationenhaushalte als „Sonstige Haushalte“ einzuordnen, führt im Ergebnis dazu, dass diese Haushaltskategorie ein Konglomerat unterschiedlichster Haushaltstypen darstellt. Die Reduktion auf nur wenige Haushaltstypen mit so genannten klassischen Kernfamilien war bis zur zweiten Hälfte des 20. Jahrhunderts vielleicht ausreichend, der anhaltende Wandel in den Familien- und Haushaltsstrukturen ist jedoch mit dieser Haushaltstypologie nicht darstellbar. Gerade Verschiebungen in der Größe und Zusammensetzung von Haushalten sowie daraus abgeleitet die tatsächlichen individuellen Lebens- bzw. Wohnverhältnisse sind aus politischer und wissenschaftlicher Sicht von hohem Interesse. Insofern ist die Eingrenzung auf lediglich fünf sehr eng definierte Haushaltstypen keine hinreichende Lösung.

Aufgrund des relativ geringen Stichprobenumfangs erlauben die Ergebnisse des Zensustests keine Aussagen

zu tief gegliederten Haushaltsstrukturen. Insofern ist eine Beschränkung auf einen engen Haushaltsbegriff im Zensustest für die Beurteilung der Ergebnisse zulässig. So wird aufgrund des relativ geringen Umfangs dieser Stichprobe der Zensustest keine statistischen Ergebnisse liefern. Vielmehr soll eine Überprüfung des Verfahrens „registergestützter Zensus“ sowie der Qualität der herangezogenen Register ermöglicht werden. Außerdem wird in einem künftigen registergestützten Zensus, anders als im Zensustest, auf eine Haushaltebefragung weitgehend verzichtet werden. Damit liefert ein registergestützter Zensus nur Haushaltstypen aus den Ergebnissen der Generierung, womit das oben dargestellte Problem der Inkonsistenz der Haushaltstypisierung entfällt.

Sollte in der Bundesrepublik Deutschland in Zukunft jedoch ein Verzicht auf eine herkömmliche primärstatistische Erhebung nach dem Vorbild der Volkszählung und ein Umstieg auf einen flächendeckenden registergestützten Zensus stattfinden, dann ergeben sich eine Reihe von Fragen und einschneidenden Problemen. Im Hinblick auf die Intention dieses Beitrags werden hier Aspekte hinsichtlich einer Haushaltstypisierung benannt.

Wie jeder Zensus dient auch ein registergestützter Zensus der Neujustierung des statistischen Gesamtsystems. Ergebnisse eines Zensus bilden die Quelle für kleinräumige Ergebnisse über Haushaltsgrößen und -strukturen sowie für wohnungsstatistische Daten. Die oben dargestellte Problematik, Wirtschaftshaushalte abzugrenzen, trifft auch auf die dem Zensus nachfolgende Stichprobenerhebung Mikrozensus zu.

So ist der Fragekatalog des Mikrozensus hinsichtlich des Haushaltszusammenhangs, also der Stellung aller im Haushalt lebenden Personen zueinander, ähnlich dem der Haushaltebefragung im Zensustest. Entspre-

chend dem Vorgehen in einer herkömmlichen Volkszählung werden im Mikrozensus Haushalte in der Abgrenzung von Personen, die gemeinsam wohnen und wirtschaften, definiert. Folglich ist eine Vergleichbarkeit von Zensus- und Mikrozensusergebnissen auf Haushaltsebene nicht gewährleistet oder es ist ein Informationsverlust aufgrund des hohen Aggregationsniveaus von Haushaltstypen hinzunehmen.

Ein Übergang auf die Definition „Wohnhaushalt“ dürfte im Mikrozensus möglich sein, ohne auf die Darstellung von Wirtschaftshaushalten auf einer tieferen Aggregationsebene, d. h. unterhalb der Wohnhaushalte, verzichten zu müssen. Vor dem Hintergrund eines flächendeckend durchgeführten registergestützten Zensus würde eine Beibehaltung des methodischen Ansatzes zur Bezugsperson im Mikrozensus zu Problemen hinsichtlich der Vergleichbarkeit von Haushaltsergebnissen und im Ergebnis zur unakzeptablen Reduktion auf nur wenige Haushaltstypen führen. Eine Anpassung des Befragungskonzepts im Mikrozensus erscheint somit unumgänglich. Denkbar wäre z. B. eine Lösung über die eindeutige Definition, welches Haushaltsmitglied Bezugsperson ist. Hier böte sich die Alternative an, immer den (ersten) Wohnungsinhaber, eventuell in Kombination mit dessen Alter und/oder Geschlecht, als Bezugsperson zu definieren. Auch könnte gänzlich auf die Kennzeichnung einer Bezugsperson verzichtet werden, wenn statt dessen ein völlig neues Erhebungsschema, z. B. entsprechend der in der §7-Erhebung „Zeitverwendung in Deutschland 2001/02“ verwendeten Beziehungsmatrix aller Haushaltsmitglieder zueinander, Eingang in den Mikrozensus fände. Unabhängig von den Ergebnissen des Zensustests und der Entscheidung über den möglichen Übergang zu einem registergestützten Zensus in der Bundesrepublik Deutschland sollte die Klärung dieser Fragen im Zentrum weiterer methodischer Untersuchungen stehen.

21) Unter Kernhaushalten werden in diesem Zusammenhang alle Haushalte verstanden, deren Haushaltszusammenhang über die in den Melderegistern enthaltenen Verzeigerungen gebildet wurden.

Befragungsschema in der §7-Erhebung „Zeitverwendung in Deutschland 2001/02“

Geschlecht		Staatsangehörigkeit		Familienstand		Beziehung aller Haushaltsmitglieder zueinander	
1 = männlich 2 = weiblich		1 = deutsch 2 = nicht deutsch		1 = verheiratet 2 = ledig 3 = verwitwet 4 = dauernd getrennt lebend		1 = Haupteinkommensbezieher/-in 2 = Ehepartner/-in 3 = Lebensgefährte/-in 4 = Tochter/Sohn 5 = Schwester/Bruder 6 = Mutter/Vater 7 = Enkelkind 8 = Großmutter/-vater 9 = Schwiegertochter/-sohn 10 = Schwiegermutter/-vater 11 = Sonstiges Verwandtschaftsverhältnis 12 = Sonstige Beziehung (nicht 1 - 11)	

Nummer der Haushaltsmitglieder	Vorname	Geburtsdatum	Nr. der Haushaltsmitglieder												
			1	2	3	4	5	6	7	8	9	10			
1	Jan	03.06.1958	1	1	1	1									
2	Petra	11.02.1960	2	1	1	2									
3	Julia	26.10.1985	2	1	2	4	4								
4	Felix	03.08.1992	1	1	2	4	4	5							
5															
6															
7															
8															
9															
10															

Julia ist die Tochter von Petra = 4

Felix ist der Bruder von Julia = 5

Petra ist die Ehefrau von Jan = 2

Grafik: LDS NRW, in Anlehnung an den Haushaltsfragebogen „Zeitverwendung in Deutschland 2001/02“

Kreisstandardzahlen – Die Gemeinden – Gemeindedaten

Kreisstandardzahlen 2001

Dieses Taschenbuch enthält die wichtigsten Zahlen aus vielen Bereichen der amtlichen Statistik, übersichtlich zusammengestellt für alle kreisfreien Städte und Kreise des Landes, außerdem Summenwerte für die Regierungsbezirke und das Land NRW (Bestell-Nr. Z 03 1 2001 00).



Die Gemeinden Nordrhein-Westfalens 2001 Informationen aus der amtlichen Statistik / Gemeindedaten NRW 2001

Als Taschenbuch sind die Gemeinden NRWs vergleichbar den Kreisstandardzahlen – jedoch mit einer Auswahl von Merkmalen für alle Gemeinden des Landes, einschl. Summenwerten für die Kreise, Regierungsbezirke und das Land NRW (Bestell-Nr. Z 04 1 2001 00; ISBN 3-935372-12-4).

Der Inhalt des Taschenbuchs ist auch auf CD-ROM („Gemeindedaten NRW“; Bestell-Nr. Z 04 8 2001 00; ISBN 3-935372-13-2) lieferbar – einschl. der Daten der Jahre 1990 bis 2000.

Die Gemeindeergebnisse erstrecken sich auf folgende Bereiche der amtlichen Statistik:

- Katasterfläche nach Nutzungsarten
- Verarbeitendes Gewerbe sowie Bergbau und Gewinnung von Steinen und Erden
- Bevölkerung, Altersstruktur der Bevölkerung, Bevölkerungsbewegung

- Bauhauptgewerbe
- Wahlen
- Kraftfahrzeugbestand, Straßenverkehrsunfälle
- Allgemein bildende Schulen
- Wohngebäude- und Wohnungsbestand, Baufertigstellungen
- Arbeitslose

- Beherbergungskapazität, Gästeankünfte und Gästeübernachtungen
- Pendler; Sozialversicherungspflichtig Beschäftigte
- Gemeindefinanzen
- Landwirtschaftliche Betriebe
- Vollzeit beschäftigtes Personal der öffentlichen Verwaltung



Die Ermittlung des Wohnungsinhabers im Rahmen der Gebäude- und Wohnungszählung (GWZ) und deren Bedeutung für die Haushaltegenerierung

Dr. Henrik Egbert und Dr. Marco Scharmer

Einleitung

Der Methodenwechsel von einer herkömmlichen Vollerhebung nach dem Vorbild früherer Volkszählungen zu einem registergestützten Zensus führte zur Entwicklung des Modells „registergestützter Zensus“ durch die Statistischen Ämter des Bundes und der Länder. Der Kombination von vorhandenen Daten aus Registern mit primärstatistisch erhobenen Angaben kommt in diesem Modell eine besondere Bedeutung zu. Die Zusammenführung von Personendatensätzen aus den Melderegistern mit Datensätzen einer Gebäude- und Wohnungszählung (GWZ) zum Zwecke einer Haushaltegenerierung ist eine unverzichtbare Voraussetzung, um Haushaltsstrukturen realistisch abbilden zu können.

Haushaltegenerierungsverfahren stellen eine Alternative zu Totalerhebungen dar und können aktuelle und kleinräumig gegliederte Daten über Größe und Struktur von Haushalten liefern. Zur Beurteilung der Eignung eines registergestützten Zensus als Alternative zu einer Totalerhebung sind jedoch Maßnahmen zur Eruierung, Prüfung und Bewertung der verfügbaren Quellen zur Feststellung der Auskunftspflichtigen und des Verfahrensablaufs notwendig.

Im Rahmen des Zensus-tests¹⁾ mit Stichtag 5. Dezember 2001 wurde in einer bundesweiten Stichprobe eine postalische GWZ als Testerhebung durchgeführt. Der Test dieser GWZ dient vor allem der Klärung folgender Fragen: Sind Verfahrensprobleme bei einer bundesweit einheitlichen Durchführung zu erwarten? Welche Unterschiede in Bezug auf Wohnungsmerkmale ergeben sich zwischen postalisch erhobenen Angaben von Gebäudeeigentümern einerseits und den durch Befragungen

erhobenen Angaben von Haushalten andererseits? Sind den Gebäudeeigentümern bzw. -verwaltern die Namen der aktuellen Wohnungsinhaber (d. h. der Bewohner), ihr Einzugsdatum und die Zahl der in der Wohnung lebenden Personen bekannt?²⁾

Insbesondere die Beantwortung der letzten Frage ist von außerordentlicher Bedeutung für das Verfahren der Haushaltegenerierung als zentralem Element des registergestützten Zensus.

Die GWZ im Zensus-test

Der Erhebungsablauf der GWZ stellt sich für Nordrhein-Westfalen wie folgt dar: Zunächst wurden 18 Gemeinden durch ein statistisches Zufallsverfahren bestimmt.³⁾ In diesen Gemeinden wurden insgesamt 1 650 Adressen ausgewählt. Je Gemeindegrößenklasse waren dies zwischen 40 und 150 Adressen. Nachdem diese Stichprobe gezogen worden war,

mussten für die 1 650 Adressen die Gebäudeeigentümer (bzw. Gebäudeverwalter, Erbbauberechtigten oder sonstigen Verfügungsberechtigten) vom Landesamt für Datenverarbeitung und Statistik NRW ermittelt werden. Für die Erhebungsorganisation war vorgesehen, jeweils einen Auskunftspflichtigen je Adresse anzuschreiben. Dieser Auskunftspflichtige sollte in der Lage sein, Angaben über die Gebäude und Wohnungen sowie die Namen der Wohnungsinhaber zu übermitteln.

Es erfolgte zunächst die Eruierung möglicher Quellen für die Anschriften von Gebäudeeigentümern bzw. Gebäudeverwaltern. Mit einer solchen Eruierung sollte geklärt werden, ob bei den im Zensus-testgesetz genannten Quellen⁴⁾ geeignete Anschriftendateien vorliegen, die für eine anschließende postalische Befragung genutzt werden können. In einem nächsten Schritt wurde die Qualität der vorhandenen Anschriften, d. h. die Frage nach deren Aktualität und Vollständigkeit, im Rahmen einer Vorbefragung der benannten Gebäudeeigentümer bzw. der Gebäudeverwalter kontrolliert.

Bei der Eruierung der Quellen wurde geprüft, ob neben den Grundsteuerstellen – von denen die vollständigen und aktuellsten Angaben erwartet wurden – auch andere Quellen

1) Zum Zensus-test vgl. grundlegend Eppmann, H., Köster, G., Rost, R., Zensus 2001 – Stand der Methodendiskussion; Statistische Rundschau Nordrhein-Westfalen, Ausgabe August 1998, S. 389 – 405. Grohmann, H.: Geschichte und Zukunft der Volkszählung in Deutschland; Berliner Statistik, Monatsschrift 7 – 12/00, S. 216 – 223. Fürnrohr, M.; Rimmelpacher, B.: Testuntersuchungen zur Vorbereitung eines registergestützten Zensus, Bayern in Zahlen, 1/2001, S. 13 – 18

2) Vgl. Forster, M.: Die Zukunft der Volkszählung in Deutschland. Traditionelle Zählung oder registergestützter Zensus? Statistische Analysen und Studien Nordrhein-Westfalen, Ausgabe 4/2000, S. 19. – 3) In NRW sind 36 Gemeinden für den Zensus-test ausgewählt. Von diesen Gemeinden ist genau die Hälfte in einer Unterstichprobe für den so genannten Verfahrenstest. Die GWZ ist zentraler Bestandteil des Verfahrenstests.

4) Vgl. BGBl. I S. 1882 – 1886, Zensus-testgesetz § 11 (2): „Für die Durchführung der Gebäude- und Wohnungstichprobe nach § 7 übermitteln die Gemeinden, die für die Führung der Grundbücher zuständigen Stellen, die für die Führung der Liegenschaftskataster zuständigen Stellen, die Finanzbehörden, die für die Gebäudebrandschutzversicherung zuständigen juristischen Personen des öffentlichen Rechts oder die Versorgungs- und Entsorgungsbetriebe den zuständigen statistischen Ämtern der Länder auf Anforderung Namen und Vornamen oder Bezeichnung und Anschrift der Eigentümer, Erbbauberechtigten, Verwalter oder sonstigen Verfügungsberechtigten der in die Erhebung einbezogenen Gebäude“.

über geeignete Anschriftendateien verfügen. Dazu wurden die Amtsgerichte als grundbuchführende Stellen, Katasterämter und ausgewählte Ver- und Entsorgungsunternehmen in den Zensus-testgemeinden NRWs befragt.

Die drei genannten Quellen erwiesen sich jedoch als nicht geeignet, flächendeckend die benötigten Anschriften liefern zu können. Von den bei den Amtsgerichten geführten Grundbüchern ist erst mittelfristig zu erwarten, dass die dort vorliegenden Informationen für einen registergestützten Zensus zu nutzen sind.⁵⁾ Die aktuellen und IT-gestützten Anschriftendateien von Ver- und Entsorgungsbetrieben können zwar für Einzelfallprüfungen herangezogen werden; aufgrund fehlender Differenzierung der Kundenanschriften nach Gebäudeeigentümern bzw. -verwaltern einerseits und Mietern andererseits, können diese Anschriftendateien aber nicht flächendeckend eingesetzt werden. Darüber hinaus wird eine weitere Marktliberalisierung im Ver- und Entsorgungsbereich⁶⁾ (und der damit einhergehende Verlust regionaler Monopolstellungen von Betrieben und evtl. auch eine wachsende Kundenmobilität) mittelfristig den Rückgriff auf diese Quellen zusätzlich erschweren. Die Katasterämter andererseits verfügen zwar über die Namen der Eigentümer, deren aktuelle Postanschriften sind aber i. d. R. nicht bekannt. Sollte das Automatisierte Liegenschaftsbuch (ALB) in Zukunft um die Eigentümeranschriften erweitert werden, so stände damit eine zusätzliche Quelle zur Verfügung.

Durchführung der GWZ im Rückgriff auf die Grundsteuerstellen

Aufgrund der Ergebnisse dieser Erhebung verschiedener Quellen in NRW erwies es sich als richtig, die

Grundsteuerstellen der Gemeinden als Hauptlieferanten für die Anschriften von Gebäudeeigentümern zu nutzen. In einem anschließenden Schritt wurde die Qualität der von den Grundsteuerstellen übermittelten Adressen geprüft. Dazu wurden die Gebäudeeigentümer in Form einer Vorbefragung angeschrieben, um Aktualisierung bzw. Vervollständigung ihrer Anschriften gebeten und über die bevorstehende Befragung informiert. Bei dieser Vorbefragung der Eigentümer stellte sich ein Korrekturbedarf von 10% der ausgewählten Adressen heraus.⁷⁾

Es zeichnet sich im Hinblick auf die spätere Haushaltegenerierung allerdings ein Problemfeld ab: Bei einer Reihe von Gebäuden können die jeweiligen Wohnungsinhaber gar nicht oder nur mit einem unverhältnismäßigen Aufwand ermittelt werden. Dieses Problem taucht insbesondere bei Gebäuden mit Eigentumswohnungen auf, da die Konzeption der GWZ im Zensus-test davon ausgeht, dass die jeweiligen Gebäudeverwalter über die angefragten Daten verfügen und die Fragebögen entsprechend ausfüllen können. Es stellte sich jedoch heraus, dass nicht allen Grundsteuerstellen die Verwalter von Gebäuden bekannt sind, sondern dass sie mitunter nur die Wohnungseigentümer benennen konnten. Dies bedeutete für die Erhebungsorganisation, dass – oftmals unter großem Zeitaufwand – die Gebäudeverwalter recherchiert werden mussten.

Ein gravierenderes Problem bestand aber darin, dass einige der Gebäudeverwalter keine vollständigen Angaben zu den Wohnungen machten. Für den Zensus-test ist hierbei von Bedeutung, dass den Verwaltern zwar die Wohnungseigentümer, nicht aber die tatsächlichen Wohnungsinhaber bekannt sind. Für eine erfolgreiche Haushaltegenerierung ist es aber unbedingt erforderlich, die je-

weiligen Wohnungsinhaber durch die postalische GWZ zu ermitteln. Dieses hatte zur Folge, dass einige Verwalter solcher Gebäude dem LDS NRW Listen mit den Wohnungseigentümern zur Verfügung stellten. Anschließend wurden alle Wohnungseigentümer dieser Gebäude gesondert angeschrieben und um Auskünfte zu ihrer Wohnung (u. a. zum Wohnungsinhaber) gebeten. Hieraus resultierte ein erheblicher Mehraufwand für die gesamte Erhebungsorganisation.

Diese Schwierigkeiten traten zwar nur bei 33 von 1650 ausgewählten Adressen auf und entsprechen somit einem Anteil von lediglich 2 %; allerdings mussten für diese 33 Gebäude ca. 500 Wohnungsinhaber zusätzlich befragt werden.

Eine andere Schwierigkeit trug ebenfalls dazu bei, dass bei weiteren Adressen keine Wohnungsinhaber ermittelt werden konnten: So waren den Grundsteuerstellen nicht alle aus den Melderegistern ausgewählte Adressen bekannt, und aus diesem Grund konnten auch nicht die Anschriften von Gebäudeeigentümern ermittelt werden. Dies traf für circa 3 % der ausgewählten Adressen zu. Zum Teil ergaben Rückfragen in den Gemeinden, dass bei einigen Eckhäusern die Adressen des Melderegisters von denen der Grundsteuerstellen abweichen, d. h. dasselbe Gebäude wird unter zwei unterschiedlichen Adressen in den Registern der Meldebehörden und der Grundsteuerstellen geführt.

Nach dem – mit erheblichem Zeitaufwand verbundenen – Heranziehen anderer Quellen blieben letztlich 1 % aller Adressen übrig, für die in der GWZ kein Gebäudeeigentümer bzw. Gebäudeverwalter ermittelt werden konnte. Durch die gezielte Nachfrage der Interviewer aus der Haushaltebefragung gelang es aber, weitere 0,8 % der Fälle zu klären, so dass im Test lediglich circa 0,2 % der Adressen aufgrund nicht ermittelbarer Gebäudeeigentümer bzw. -verwalter ausfielen. Die bisherigen Schwierigkeiten bei der Durchführung der GWZ haben direkte Folgen

5) Vgl. hierzu die Entwicklung bzgl. der Einführung des „Elektronischen Grundbuches“ in NRW, <http://www.justiz-online.nrw.de> (dort: „IT-Verfahren“ / „EGB“). – 6) Vgl. u. a. Hans-Jürgen Ewers, Bettina Mankel, Wolf Pluge, Petra Roth, Eric Heymann, Liberalisierung der Wasserversorgung – soll das letzte Versorgungsmonopol in Deutschland fallen? Zur Diskussion gestellt, ifo-Schnelldienst, 14, 2001, S. 3 – 14.

7) Der Rücklauf bei dieser freiwilligen Vorbefragung betrug 63,5 %. Bei 8,9 % bestand ein Korrekturbedarf des Namens oder der Anschrift, bei weiteren 1,4 % wurde festgestellt, dass der Eigentümer verzogen war (Prozentzahlen bezogen auf die Gesamtzahl von 1 650 Adressen).

für die Zusammenführung von Haushalten im Rahmen des Haushaltegenerierungsverfahrens.

Die Haushaltegenerierung im Zensustest

Ziel des Haushaltegenerierungsverfahrens ist es, über bestimmte Übereinstimmungen oder Kombinationen der im Personendatensatz der Melderegister enthaltenen Merkmale Personenbeziehungen und Wohnverhältnisse abzubilden.⁸⁾ Im Ergebnis werden zensustypische Datensätze generiert, welche Angaben zur Demographie, zur Haushaltsgröße und -struktur, zum Erwerbsleben⁹⁾ sowie zum Wohnverhältnis liefern.

Die im Rahmen einer postalischen GWZ bei den Gebäudeeigentümern erhobenen gebäude- und wohnungstatistischen Daten werden hierbei mit personenbezogenen Angaben aus den Einwohnermelderegistern zunächst adressweise zusammengeführt. Im Anschluss wird versucht, die Datensätze der unter einer Adresse gemeldeten Personen zu so genannten Kernhaushalten zusammenzuführen. Die Bildung von Kernhaushalten wird über die in den Melderegisterdatensätzen enthaltenen Verzeigerungen vorgenommen. Verzeigerungen sind Merkmale in einem Personendatensatz, die auf einen Datensatz einer anderen Person hinweisen.¹⁰⁾ Wenn die Datensätze der unter einer Adresse gemeldeten Personen Verzeigerungen zueinander enthalten, dann lassen sich maschinell sehr valide Kernhaushalte bilden.

8) Für jede potenzielle Personenbeziehung wird also eine Anzahl von Bedingungen in Form von Merkmalsübereinstimmungen bzw. Merkmalskombinationen zwischen den Personendatensätzen definiert, die erfüllt sein müssen, damit im Rahmen der Haushaltegenerierung die Zugehörigkeit der betreffenden Personen zum gleichen Haushalt mit hoher Wahrscheinlichkeit unterstellt werden kann. – 9) Im Rahmen des Zensustests sollen erwerbsstatistische Daten der Bundesanstalt für Arbeit mit den Personensätzen aus den Melderegistern verknüpft werden. Hierauf wird im Hinblick auf die Thematik der vorliegenden Ausführungen aber nicht eingegangen. – 10) Verzeigerungen sind i. d. R. Verwandtschaftsverhältnisse, wie zwischen Ehepartnern, Eltern(-teilen) und deren minderjährigen Kindern sowie zwischen minderjährigen Kindern und deren gesetzlichen Vertretern (einschließlich Stief- und Pflegeeltern).

In dem eigentlichen Haushaltegenerierungsverfahren wird in weiteren Schritten versucht, die im Rahmen der GWZ erhobenen Wohnungsinhaber in den Personensätzen des Melderegisters der jeweiligen Adresse zu identifizieren. In verschiedenen Algorithmen des Verfahrens werden dann mit den übrigen Personensätzen aus den Melderegistern Wohnhaushalte um die Wohnungsinhaber bzw. um die zuvor gebildeten Kernhaushalte generiert. Mit diesem Verfahren ist es somit möglich, über den Wohnungsinhaber verschiedene Personensätze mit einem Wohnungssatz zu verknüpfen, d. h. Haushalte zu bilden und diese einer Wohnung zuzuordnen. Die Namen der Wohnungsinhaber sind damit das zentrale Merkmal für die Haushaltegenerierung.

Wohnungsinhabernamen als zentrales Merkmal

In der Haushaltegenerierung werden somit GWZ- und Melderegisterdatensätze adressweise verknüpft. Das Verfahren setzt immer an den Wohnungsinhabern laut GWZ an und bildet um diese Personen und ihre Wohnung herum Haushalte.

Hierzu wurde ein mehrstufiger Namensabgleich entwickelt, den alle Wohnungsinhabernamen durchlaufen. Der Namensabgleich startet mit einer sehr restriktiven Stufe, die einen Wohnungsinhaber aus der GWZ nur dann mit einem Personensatz des Melderegisters verknüpft, wenn in beiden Datenbeständen die Vor- und Zunamen einschließlich etwaiger Namensbestandteile exakt übereinstimmen. In den weiteren Stufen werden die Zuordnungskriterien zunehmend „weicher“, d. h. eine Verknüpfung findet schließlich auch dann statt, wenn nur eine Namensähnlichkeit vorliegt (phonetischer Abgleich).

Falls der Namensabgleich keine (auch nur phonetische) Übereinstimmung zwischen dem Wohnungsinhabernamen aus der GWZ und dem Personensatz im Melderegister ergeben hat, wird ein Abgleich zwischen

dem Wohnungsinhabernamen laut GWZ und den Namensangaben in den Verzeigerungen des Melderegisters durchgeführt.¹¹⁾ Abschließend ist vorgesehen, die wenigen noch übrigen Wohnungsinhaber aus der GWZ in einer manuellen Dialogprüfung an allen unverknüpften Personensätzen des Melderegisters (der jeweiligen Adresse) vorbeizuführen und hierüber eine Identifikation des Wohnungsinhabers zu erreichen.

Ist es nach diesen Schritten nicht möglich, einen von der GWZ genannten Wohnungsinhaber im Melderegisterdatenbestand aufzufinden, so handelt es sich, unabhängig von der Ursache, um eine Unplausibilität bei der Zusammenführung, die in Fehlbeständen im Melderegister oder fehlerhaften Angaben der Gebäudeeigentümer begründet sein kann. An dieser Stelle des Verfahrens ist es möglich, eine Verknüpfung auch dann durchzuführen, wenn unter einer Adresse nur ein Wohnungsinhaber noch nicht im Melderegister identifiziert wurde und laut Melderegister auch nur ein Haushalt existiert. Hier wird eine Verknüpfung des Wohnungsinhabers und seiner Wohnung mit dem noch unverknüpften Haushalt ausgeführt.¹²⁾

Befragungsausfälle der GWZ und deren Bedeutung für die Haushaltegenerierung

Die ersten Ergebnisse der im Rahmen des Zensustests durchgeführten GWZ zeigten folgende Befragungsausfälle: Für einzelne Wohnungen oder Gebäude waren überhaupt keine Angaben erhältlich, da der Eigentümer nicht zu ermitteln war oder keine Auskunft erteilte. In diesem Fall handelt es sich um einen Totalausfall. Weiterhin kam es auch zu Ausfällen einzelner Merkmale, wobei fehlende Wohnungsinhabernamen für die Haushaltegenerierung die

11) Das Verfahren berücksichtigt auch diejenigen Wohnungsinhaber, die Inhaber mehrerer Wohnungen im selben Gebäude, aber i. d. R. nur einmal unter einer Adresse gemeldet sind. – 12) Weiterhin ist vorgesehen, unter bestimmten Bedingungen eine Verknüpfung ohne Namensabgleich hinsichtlich der Einzugsdaten vorzunehmen.

schwerwiegendsten Folgen haben. Schließlich wurde von einigen der angeschriebenen Gebäudeverwalter zwar die Auskunft zu den Gebäude- und Wohnungsangaben erteilt, gleichzeitig wiesen diese aber darauf hin, dass weder die Wohnungsinhabernamen noch die Zahl der Bewohner in allen Fällen aktuell seien.

Die Identifikation des Wohnungsinhabernamens im Melderegister, aber auch die Qualität des Befragungsergebnisses der GWZ über den Wohnungsinhaber sind jedoch im Verfahren der Haushaltegenerierung von zentraler Bedeutung. Ein Nichtauffinden des durch die GWZ genannten Wohnungsinhabers in den Datensätzen des Melderegisters sowie ein Befragungsausfall, insbesondere dieses Merkmals, macht eine maschinelle Haushaltegenerierung i. d. R. unmöglich. In diesen Fällen können die Angaben zur Demographie, zur Haushaltsgröße und -struktur sowie zum Wohnverhältnis nur über eine herkömmliche Erhebung mittels einer Befragung durch Interviewer festgestellt werden. Diese Angaben ersetzen für einzelne Wohnungen oder ganze Gebäude die Angaben aus dem Melderegister sowie aus der GWZ. Im Zensustest besteht für diese Fälle die Möglichkeit, die entsprechenden Informationen durch eine Übernahme der Ergebnisse der Haushaltebefragung zu ersetzen. Diese Option bestünde in einem flächendeckenden registergestützten Zensus nicht. Hier müssten die primärstatistischen Informationen sehr zeit- und kostenintensiv erhoben werden.

Folgerungen für einen zukünftigen registergestützten Zensus

Aufgrund der bisherigen Erfahrungen aus der postalischen GWZ im Zensustest kann unterstellt werden, dass es auch bei einer GWZ in einem zukünftigen Zensus problematisch werden wird, die Namen der tatsächlichen Wohnungsinhaber und aktuelle und vollständige Angaben zu den Wohnungen von allen Gebäudeverwaltungen zu erhalten. Bei solchen Adressen können die Gebäude-

verwalter u. U. nur eine Hilfsfunktion übernehmen und die Anschriften der Wohnungseigentümer übermitteln. Für eine hieraus resultierende Befragung der einzelnen Wohnungseigentümer ist allerdings mit einem erheblichen Mehraufwand bei einer nicht zu unterschätzenden Anzahl von Adressen zu rechnen.

Die relativ kleine Stichprobe von 1 650 Gebäuden für den Verfahrenstest entspricht circa 0,05 % der in NRW vorhandenen Wohngebäude.¹³⁾ Die Stichprobe ist insofern nicht repräsentativ, als einerseits Adressen mit relativ vielen gemeldeten Personen überdurchschnittlich oft ausgewählt wurden, andererseits aber Anstalten nicht berücksichtigt worden sind, da in diesen bei einem zukünftigen Zensus ohnehin eine primärstatistische Erhebung notwendig sein wird. Aufgrund der bisher ermittelten Zahlen lässt sich aber bereits jetzt für NRW feststellen, dass bei einem zukünftigen Zensus Begehungen vor Ort in einer beträchtlichen Anzahl von Gebäuden notwendig werden könnten.

Im Zensustest stellte sich bei mindestens 5 % aller Adressen heraus, dass die Gebäudeeigentümer- bzw. -verwalteradressen der Grundsteuerstellen für eine schriftliche Befragung nicht genutzt werden konnten. Alle Angaben beziehen sich auf einen Bearbeitungsstand der GWZ von knapp 90 %. Selbst wenn bei einem zukünftigen Zensus dieser Prozentsatz niedriger wäre und z. B. nur von einem Prozent aller Adressen ausgegangen werden müsste, so ergäbe sich eine Zahl von 34 300 Adressen, bei denen eine Vorortbegehung notwendig werden würde. Zu diesen kämen die Anstalten und ein bisher unbekannter Prozentsatz an Gebäuden, in denen die postalische GWZ keine oder nicht alle Wohnungsinhaber ermitteln können. Des Weiteren wird sich durch eine schriftliche Befragung sämtlicher Wohnungseigentümer in solchen Gebäuden, in denen der Verwalter nicht in der Lage ist, die aktuellen Wohnungsinhaber zu

¹³⁾ Der Bestand der Wohngebäude in NRW beträgt circa 3,43 Mill.; vgl. Statistisches Jahrbuch NRW 2001, S.396.

übermitteln, der Aufwand für die postalische GWZ erhöhen. Die im Zensustest festgestellten 33 Adressen (2 % der Stichprobe) wiesen durchschnittlich 10 Wohnungen auf. Da in der Stichprobenauswahl große Gebäudeüberproportional ausgewählt wurden, läge bei einem zukünftigen Zensus dieser Prozentsatz niedriger als 0,7 %. Bezogen auf die Gesamtzahl der Wohngebäude in NRW wäre aber ein Mehraufwand zu erwarten¹⁴⁾.

Die hier genannten Fallzahlen, in denen eine Begehung vor Ort in einem flächendeckenden registergestützten Zensus notwendig werden könnte, sind zwar nur grobe Schätzungen aufgrund erster Erfahrungen aus der GWZ. Doch stellen sie auch nur einen Teil der zusätzlichen primärstatistischen Erhebungen dar. Hinzu kommen noch Fälle, für die eine plausible Haushaltegenerierung aufgrund der noch zu prüfenden Melderegisterqualität (Fehlbestände und Überhänge) oder des Verfahrens selbst nicht möglich ist. Der tatsächliche Umfang der zusätzlich notwendig werdenden Begehungen vor Ort kann zum gegenwärtigen Zeitpunkt noch nicht quantifiziert werden. Aus dem hier aufgezeigten Zusammenhang lassen sich für die GWZ eines zukünftigen registergestützten Zensus jedoch Folgerungen ableiten.

Aufgrund des unverhältnismäßig hohen Rechercheaufwandes für die vorliegende zahlenmäßig kleine GWZ-Stichprobe im Zensustest, kann für einen zukünftigen Zensus abgeleitet werden, dass eine flächendeckende GWZ dezentral, d. h. in den Gemeinden, durchgeführt werden sollte. Gleiches gilt auch für die durch Unplausibilitäten notwendig werdenden Begehungen vor Ort durch Interviewer. Der Vorteil einer in den Gemeinden durchgeführten GWZ läge insbesondere darin begründet, dass aufgrund der räum-

¹⁴⁾ Im Test wurde zunächst von 1650 Ansprechpartnern für die ausgewählten Gebäude ausgegangen. Dadurch, dass für 33 Gebäude die Verwalter keine Angaben zu den Wohnungsinhabern machen konnten, erhöhte sich die Zahl der Ansprechpartner um 30 %. Hieraus ergibt sich ein entsprechender Mehraufwand für alle mit einer primärstatistischen Erhebung verknüpften Arbeitsschritte.

lichen Nähe zu den verschiedenen Quellen für Anschriftendateien die Eruiierung und Recherche der Gebäudeverwalter- und -eigentümerschriften vereinfacht und zeitlich verkürzt werden könnte. Hierdurch ließen sich der Aufwand und damit die Kosten für eine ohnehin zahlenmäßig erhebliche nachträgliche Vortortbefragung durch Erhebungsbeauftragte reduzieren.

Gleichwohl ist jedoch die Durchführung einer GWZ für die valide Bil-

dung von Haushalten im Rahmen der Haushaltgenerierung unerlässlich. Aufgrund der bisherigen Ergebnisse des Verfahrens sollte in einem flächendeckenden registergestützten Zensus den hier aufgezeigten Schwierigkeiten (Befragung der Wohnungseigentümer; Problem von Eckgebäuden) bereits im Vorfeld Rechnung getragen werden, beispielsweise durch die Einführung einer weiteren Befragungsstufe für die Wohnungseigentümer in Großgebäuden oder den Abgleich von Ad-

ressen aus dem Melderegister mit denen der Grundsteuerstellen. Langfristig ist vor dem Hintergrund der Abkehr von einer primärstatistischen Vollerhebung, dem Ersatz durch registergestützte Auswertung sowie der Kostenminimierung auch über den Aufbau eines Gebäude- und Wohnungsregisters als sinnvolle Option nachzudenken.

Statistisches Jahrbuch NRW 2001



Handlich, dick, blau, informativ, mit beige packter CD-ROM: Mit dem Jahreswechsel ist die aktuelle Ausgabe 2001 des Statistischen Jahrbuches für Nordrhein-Westfalen erschienen. Dieses „Buch des Wissens“ bietet auf fast 800 Seiten eine Fülle von Informationen aus den verschiedensten Bereichen der amtlichen Statistik. Das Zahlenwerk zeichnet nicht nur ein Bild der gesellschaftlichen und wirtschaftlichen Situation NRWs, sondern es zeigt zugleich auch viele Strukturen und Entwicklungstendenzen auf.

Mit der dem Buch beige geklebten CD-ROM-Fassung bietet das Landesamt PC-Nutzern die Möglichkeit, die Inhalte des Statistischen Jahrbuchs auch elektronisch weiterzuverarbeiten; Informationen zur Nutzung sowie notwendige „Browser“ werden auf der CD-ROM mitgeliefert.

Im Jahrbuch erfährt man unter anderem, dass

• Frankreich der bedeutendste Abnehmer für die NRW-Wirtschaft ist: Im Jahr 2000 beliefen sich die Exporte nach Frankreich auf 10,7 Milliarden Euro, das war mehr als auf dem gesamten asiatischen Markt abgesetzt wurde;

- im vergangenen Jahr in NRW – rein rechnerisch – alle zwölf Minuten eine Ehe geschieden wurde und bei den 45 200 Scheidungswaisen“ wurden;
- die durchschnittliche Klassengröße an den Grund- und Hauptschulen NRWs im Schnitt bei 23, an den Gymnasien bei 27 und an den Real- und Gesamtschulen bei 28 Schülern je Klasse liegt;
- die Zahl der NRW-Unternehmen, die im Jahr 2000 ein Insolvenzverfahren beantragten, mit 5 500 um zehn Prozent höher war als im Jahr zuvor;
- in NRW im Juli 2000 mehr als zehn Millionen Kraftfahrzeuge amtlich zugelassen waren – damit waren rein rechnerisch 60 Prozent der Bevölkerung im Besitz eines Kfz; in Gelsenkirchen lag dieser Anteil (49 Prozent) am niedrigsten, im Kreis Minden-Lübbecke (69 Prozent) am höchsten;
- eine nordrhein-westfälische „Durchschnittskuh“ im Jahr 2000 mit 6 856 Litern Milch pro Jahr 2 700 Liter mehr gibt als ihre „Ahnin“ aus dem Jahr 1970;
- von den über 2,5 Millionen ausländischen Gästen, die im Jahr 2000 in den über 5 000 nordrhein-westfälischen Beherbergungsbetrieben übernachteten, 211 000 US-Amerikaner und 81 000 Japaner waren;
- an den NRW-Hochschulen im letzten Wintersemester mehr Student(inn)en Informatik studierten als allgemeine Medizin und Zahnmedizin zusammen; die Zahl der Informatik-Studenten war mit 21 078 um nahezu ein Viertel (24,5 Prozent) höher als im vorangegangenen Wintersemester (1999/2000);
- fast jede zweite (48,2 Prozent) der 65 743 im Jahr 2000 in NRW eingebürgerten Personen ursprünglich die türkische Staatsangehörigkeit besaß;
- 2000 in Nordrhein-Westfalen mehr Menschen an AIDS (114 Todesfälle: 98 Männer, 16 Frauen) als an Tuberkulose (103 Todesfälle) starben;
- im Jahr 2000 nach Angaben der Kreditinstitute in NRW fast 214 000 Schecks im Gesamtwert von 565 Millionen Euro nicht eingelöst wurden;
- 650 000 Personen (3,4 Prozent der Bevölkerung) im Jahr 2000 laufende Hilfe zum Lebensunterhalt im Rahmen der Sozialhilfe empfangen haben;
- Auto- und Motorradfahrer(innen) an Rhein und Ruhr immer tiefer in ihre Geldbörse greifen müssen: Die Kosten für Anschaffung und Unterhaltung (sog. Kraftfahrerpreisindex) waren im September 2001 um über 17 Prozent höher als 1995;
- sich 1998 für jeden Erwerbstätigen in NRW ein Bruttoinlandsprodukt zu Marktpreisen von 54 551 Euro errechnet – die höchste Wirtschaftsleistung im Lande erzielte hier Düsseldorf mit 76 312 Euro, die niedrigste Bottrop mit 40 873 Euro;
- von den 396 Städten und Gemeinden NRWs nur 30 mehr als 100 000 Einwohner haben; Köln hat mit 963 000 die höchste, Dahlem im Kreis Euskirchen mit 4 300 die niedrigste Bevölkerungszahl des Landes.

Erhältlich ist das Statistische Jahrbuch NRW 2001 mit CD-ROM, das 29 EUR kostet und neben neuen statistischen Ergebnissen für das Land selbst auch einige Übersichten mit Resultaten für die kreisfreien Städte und Kreise NRWs sowie Eckdaten für die 16 deutschen Bundesländer präsentiert, entweder direkt beim LDS NRW oder über den Buchhandel (ISBN 3-935372-11-6).

Räumliche Mobilität und Arbeitsmarktprozesse

Eine Analyse auf der Basis des Sozio-oekonomischen Panels

Dr. Sylvia Zühlke und Uwe Hetke

Da in der wirtschafts- und sozialwissenschaftlichen Forschung die Analyse dynamischer Prozesse in den letzten Jahren zunehmend an Bedeutung gewonnen hat, wird gegenwärtig im Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen der Mikrozensus (MZ) als Längsschnittdatensatz aufbereitet. Bei den Längsschnittanalysen aus dem MZ besteht ein zentrales methodisches Problem darin, dass Haushalte und Personen, die zwischen zwei Erhebungsjahren fortziehen, nicht wieder befragt werden. Diese Ausfälle können dann zu Ergebnisverzerrungen führen, wenn die räumliche Mobilität mit denjenigen haushalts- und personenbezogenen Veränderungen zusammenhängt, die auf der Basis des Mikrozensus abgebildet werden sollen. Um die Richtung und das Ausmaß der Verzerrungen von Arbeitsmarktanalysen auf der Basis des Mikrozensus annähernd bestimmen zu können, werden im Folgenden die Zusammenhänge zwischen der räumlichen Mobilität und verschiedenen Arbeitsmarktprozessen auf der Basis des Sozio-oekonomischen Panels (SOEP) untersucht. Berücksichtigt werden dabei zunächst Entwicklungen, die sich zwischen zwei Erhebungsjahren vollziehen. Die Ergebnisse weisen darauf hin, dass die räumliche Mobilität nur mit sehr wenigen Arbeitsmarktprozessen zusammenhängt. Eine enge Beziehung zeigt sich insbesondere bei den Übergängen aus Arbeitslosigkeit und Nichterwerbstätigkeit in Erwerbstätigkeit. Gleichzeitig weisen die Analysen darauf hin, dass selbst bei diesen Übergängen im Rahmen der Längsschnittanalysen des Mikrozensus nur geringfügige Verzerrungen zu erwarten sind, da die Gruppe der räumlich mobilen Personen relativ klein ist und daher die Ergebnisse insgesamt von dieser Gruppe nicht sehr stark bestimmt werden.

Einleitung

Da die Analyse von personen- und haushaltsbezogenen Veränderungen in den letzten Jahren zunehmend an Bedeutung gewonnen hat, wird zurzeit im Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen für die Jahre 1996 bis 1999 der Mikrozensus als Längsschnittdatensatz aufbereitet. Für solche Analysen steht somit erstmals ein Datensatz zur Verfügung, der auf einer Teilnahmepflicht basiert und eine sehr große Anzahl an Fällen aufweist. Im Gegensatz zu einer normalen Paneluntersuchung, in der die Haushalte und Personen auch nach einem Umzug wieder befragt werden, können solche Erhebungseinheiten im Rahmen des Mikrozensus nicht in die Analysen einbezogen werden. Dieses Problem geht darauf zurück, dass der Mikrozensus als Rotationspanel konzipiert ist, das auf

einer Flächenstichprobe beruht. Danach werden nicht ausgewählte Haushalte, sondern ausgewählte Bezirke viermal in die Erhebung einbezogen. Dies hat zur Folge, dass anstelle der fortgezogenen Haushalte jeweils die zugezogenen Haushalte befragt werden.

Die systematischen Ausfälle infolge von Fort- und Zuzügen können im Rahmen von Längsschnittanalysen dann zu Ergebnisverzerrungen führen, wenn ein enger Zusammenhang zwischen räumlichen und anderen Mobilitätsprozessen besteht. Daher werden im Folgenden die Zusammenhänge zwischen räumlicher Mobilität und unterschiedlichen Arbeitsmarktprozessen auf der Basis des Sozio-oekonomischen Panels untersucht. Das Ziel der Analysen ist es, Aussagen darüber treffen zu können, ob räumlich mobile Personen beispielsweise häufi-

ger ihre berufliche Tätigkeit wechseln als räumlich immobile Personen. Die Analysen können somit Hinweise zu der Richtung und näherungsweise auch zu dem Ausmaß der Ergebnisverzerrungen geben, die bei den Längsschnittanalysen auf der Basis des Mikrozensus aus den fortzugsbedingten Ausfällen resultieren. Als Datengrundlage wurde das Sozio-oekonomische Panel ausgewählt, da es als einer der wenigen Datensätze der empirischen Sozialforschung die Analyse der räumlichen Mobilität erlaubt und zusätzlich den zu untersuchenden Zeitraum (1996 bis 1999) abdeckt. Zu berücksichtigen ist dabei, dass auch die Daten des Sozio-oekonomischen Panels eine spezifische Selektivität aufweisen, die zum einen aus der freiwilligen Teilnahme an der Erhebung und zum anderen aus der Panelmortalität resultiert.

Im Folgenden werden zunächst das Erhebungsdesign sowie der Stichprobenumfang des Sozio-oekonomischen Panels vorgestellt und das Problem der Panelmortalität beschrieben. Anschließend wird dargestellt, in welchem Umfang räumliche Mobilitätsprozesse beobachtet werden und wie diese mit anderen Arbeitsmarktprozessen zusammenhängen. Berücksichtigt werden dabei Übergänge zwischen unterschiedlichen Erwerbszuständen sowie Wechsel zwischen unterschiedlichen beruflichen Tätigkeiten, Wirtschaftszweigen und Einkommenspositionen, die sich zwischen zwei Erhebungsjahren vollziehen. Abschließend wird diskutiert, welche Schlussfolgerungen auf der Basis der vorliegenden Ergebnisse für die Längsschnittanalysen auf der Basis des Mikrozensus gezogen werden können.

Das Sozio-oekonomische Panel

Erhebungsmethode

Bei dem Sozio-oekonomischen Panel handelt es sich um eine repräsentative Längsschnittbefragung, die seit 1984 einmal im Jahr von der Projektgruppe „Sozio-oekonomisches Panel“ am Deutschen Institut für Wirtschaftsforschung durchgeführt wird¹⁾. Thematische Schwerpunkte der Befragung sind die Erwerbstätigkeit, die regionale und berufliche Mobilität, die Einkommenssituation, die Haushaltszusammensetzung und Wohnsituation, die gesundheitliche Situation, die Zufriedenheit mit einzelnen Lebensbereichen, die Zeitverwendung sowie die Bildung und Weiterbildung²⁾. Die Grundgesamtheit, die der Stichprobenziehung für die erste Welle zugrunde lag, umfasste die in Privathaushalten lebende Wohnbevölkerung der Bundesrepublik Deutschland einschließlich West-Berlins. Die Stichprobe wurde auf der Haushaltsebene gezogen, wobei zwei Teilstichproben gebildet wurden:

Stichprobe (A): Auf der Basis einer Random-Route-Stichprobe wurden Personen in Privathaushalten ausgewählt, deren Haushaltsvorstand nicht die Staatsangehörigkeit einer der Haupteinwanderergruppen (Personen mit türkischer, griechischer, jugoslawischer, spanischer oder italienischer Staatsangehörigkeit) besitzt. In der Regel handelt es sich dabei um Haushalte mit einem Haushaltsvorstand, der die deutsche Staatsangehörigkeit aufweist. 1984 umfasste diese Stichprobe 4 528 Haushalte.

Stichprobe (B): Auf der Basis des Ausländerzentralregisters wurden Personen in Privathaushalten mit Hilfe eines Zufallsverfahrens ausgewählt, deren Haushaltsvorstand die türkische, griechische, jugoslawische, spanische oder italienische Staatsangehörigkeit besitzt. Für die einzelnen Nationalitäten wurde jeweils eine repräsentative Stichprobe gezogen.

Um eigenständige Analysen der Haushalte mit einem ausländischen Familienvorstand zu ermöglichen, ist diese Gruppe in der Stichprobe überrepräsentiert. 1984 wurden 1 393 Haushalte mit einem ausländischen Haushaltsvorstand befragt.

Diese beiden Ursprungsstichproben wurden in den folgenden Jahren um einige Gruppen erweitert:

Stichprobe (C): Im Juni 1990 wurden erstmals Personen in 2 179 Privathaushalten befragt, deren Haushaltsvorstand zu diesem Zeitpunkt die Staatsbürgerschaft der DDR besaß. Diese Stichprobe wurde aus dem Zentralen Einwohnermelderegister der DDR gezogen.

Stichprobe (D): In den Jahren 1994 und 1995 wurden in zwei Schritten insgesamt 531 Haushalte in die Untersuchung einbezogen, in denen mindestens ein Haushaltsmitglied lebt, das nach 1984 nach Westdeutschland zugewandert ist. Diese Haushalte wurden in einem mehrstufigen Auswahlverfahren ermittelt (siehe hierzu im Einzelnen Rendtel, Pannenberg und Daschke 1997).

Stichprobe (E): Im Jahr 1998 wurden zusätzlich 1 067 Haushalte in die Untersuchung einbezogen, um die Panelmortalität, die im Zeitablauf zu einer Reduzierung des Stichprobenumfangs der Stichproben A bis D geführt hat, zu kompensieren. Die Haushalte wurden im Rahmen eines Random-Route-Verfahrens ausgewählt.

Befragt werden in jedem Haushalt alle Personen, die älter als 15 Jahre sind. Es liegen aber auch Informationen über die in den Haushalten lebenden Kinder vor, die im Rahmen des Haushaltsinterviews ermittelt werden. In jedem Jahr werden dieselben Haushalte und Personen befragt, auch wenn sie ihre Wohnung gewechselt haben. Eine Ausnahme bilden lediglich Umzüge in das Ausland. Neu aufgenommen in die Studie werden Personen dann, wenn sie in die Befragungshaushalte einzie-

hen und wenn Kinder im Befragungshaushalt das „Untersuchungsalter“ erreichen. Sofern einzelne Personen aus einem Haushalt ausziehen, bleiben sie Mitglieder der Stichprobe und werden weiter befragt³⁾. Ziehen diese Personen wiederum mit anderen zusammen, so gehören diese automatisch zur Untersuchungsgruppe. Trotz dieser durch das Stichprobendesign hervorgerufenen jährlichen Aufstockung nimmt die Anzahl der Befragungshaushalte und -personen infolge der so genannten Panelmortalität von Jahr zu Jahr ab. Die Hauptursache für die Abnahme der Stichprobe liegt neben der natürlichen Mortalität in der Verweigerung der weiteren Teilnahme sowie in der Nichterreichbarkeit der Haushalte. Um diese Verluste zu kompensieren, wurde die Stichprobe im Jahr 1998 aufgestockt (siehe oben „Stichprobe E“).

Stichprobenumfang der ausgewählten Datensätze

Um die ermittelten Ergebnisse in Beziehung zu den späteren Analysen auf der Basis des Mikrozensus setzen zu können, werden für die folgenden Analysen die Stichproben A bis E für die Jahre 1996 bis 1999 genutzt. Einbezogen werden zunächst alle Personen, die in einem privaten Haushalt leben, einschließlich der Kinder. Da die Haushalte mit einem ausländischen Haushaltsvorstand in dem gesamten Datenbestand überrepräsentiert sind, werden die hochgerechneten Ergebnisse dargestellt. Die Hochrechnung des Sozio-oekonomischen Panels orientiert sich am Mikrozensus, und zwar für die Haushalte an der wohnberechtigten Bevölkerung in Privathaushalten und für die Personen an der Bevölkerung in Privathaushalten am Hauptwohnsitz (siehe Pischner 2000). Der Berechnung des Hochrechnungsfaktors für das erste Erhebungsjahr kommt eine besondere Bedeutung zu, da sie die Ermittlung aller weiteren Faktoren bestimmt (siehe Pischner 1994; Rendtel 1995; Rendtel, Wagner und Frick 1995). Die Bildung der Hochrechnungsfaktoren beruht auf der Überlegung, dass die Gewichtungsfakto-

1) Detaillierte Informationen über die Inhalte, Anlage und methodische Umsetzung der Studie sind bei Hanefeld (1987) sowie in den jeweils aktuellen Benutzerhandbüchern der Projektgruppe zu finden (siehe Haisken-DeNew/Frick 2000). – 2) Zusätzlich wird in jedem Jahr ein Schwerpunktthema behandelt.

3) Von 1984 bis 1988 wurden nur die so genannten Stammpersonen „weiterverfolgt“. Hierbei handelt es sich um Personen, die zum Zeitpunkt der ersten Befragung (1984) Mitglied eines Haushaltes der Nettostichprobe waren, sowie um deren Kinder.

ren durch den Kehrwert der Auswahlwahrscheinlichkeit einer Stichprobeneinheit bestimmt werden. Die Auswahlwahrscheinlichkeiten für die unterschiedlichen sozio-demographischen Gruppen werden in einem mehrstufigen Verfahren ermittelt, in welches neben dem spezifischen Erhebungsdesign des Sozio-oekonomischen Panels ein Vergleich der Randverteilungen mit den Daten des Mikrozensus eingeht (siehe hierzu Haisken-DeNew und Frick 2000: 146).

Für die weiteren Wellen werden anschließend Bleibewahrscheinlichkeiten ermittelt, indem einerseits die Wahrscheinlichkeit einer erneuten Kontaktaufnahme und andererseits die Wahrscheinlichkeit einer erneuten Antwortgewährung geschätzt wird. Da aus den vorangegangenen Wellen Informationen über die ausgefallenen Erhebungseinheiten vorliegen, können die bekannten Merkmale zur Ermittlung der genannten Wahrscheinlichkeiten genutzt werden. Die Wahrscheinlichkeit, an allen Panelwellen teilzunehmen, wird anschließend aus dem Produkt der Einzelwahrscheinlichkeiten berechnet. Sollen die Angaben aus den folgenden Wellen nicht nur als Längsschnitt, sondern unabhängig davon auch als Querschnitt hochgerechnet werden, ist bei der Bildung der Gewichte zu beachten, dass neben den Personen, die an den vorangegangenen Wellen teilgenommen haben, auch die neu in das Panel eingetretenen Personen zu berücksichtigen sind⁴⁾.

Im Jahr 1996 wurden 6 817 private Haushalte im Rahmen des Sozio-oekonomischen Panels befragt. Diese Anzahl ist im Folgejahr zunächst gesunken, lag jedoch in den Jahren 1998 und 1999 bedingt durch die Aufstockung der Stichprobe deutlich darüber. Im Jahr 1996 lebten 17 401 Privatpersonen einschließlich Kindern in den befragten Haushalten. Auch hier ist 1997 zunächst eine Verringerung der Anzahl der befragten Personen zu verzeichnen, während in den beiden folgenden Jahren aufgrund der Aufstockung der

4) Für die Erhebungen ab 1990 wird bei den Querschnittsgewichten zusätzlich eine Randanpassung an den Mikrozensus vorgenommen (siehe hierzu Pischner 2000).

Anzahl der befragten und hochgerechneten Haushalte und Personen in den Querschnittdatensätzen des Sozio-oekonomischen Panels 1996 bis 1999				
Merkmal	1996	1997	1998	1999
Anzahl der befragten Haushalte				
Bundesgebiet	6 817	6 730	7 601	7 319
NRW	1 321	1 308	1 534	1 463
Prozentualer Anteil von NRW	19,4	19,4	20,2	20,0
Hochgerechnete Anzahl der Haushalte				
Bundesgebiet	37 235 792	37 435 253	37 544 941	37 787 279
NRW	8 295 860	8 520 874	8 469 558	8 465 041
Prozentualer Anteil von NRW	22,3	22,8	22,6	22,4
Anzahl der befragten Personen				
Bundesgebiet	17 401	17 015	18 660	17 876
NRW	3 406	3 333	3 753	3 577
Prozentualer Anteil von NRW	19,6	19,6	20,1	20,0
Hochgerechnete Anzahl der Personen				
Bundesgebiet	77 171 230	81 121 955	80 720 722	81 200 864
NRW	17 323 732	18 349 099	18 363 866	18 265 648
Prozentualer Anteil von NRW	22,4	22,6	22,7	22,5

Quelle: Das Sozio-oekonomische Panel 1996 bis 1999; eigene Berechnungen

Stichprobe eine deutlich höhere Anzahl an Personen erfolgreich befragt werden konnte. Auf Nordrhein-Westfalen entfallen in den einzelnen Jahren etwa ein Fünftel aller befragten Haushalte und Personen, so dass im Durchschnitt 1 400 Haushalte und 3 500 Personen pro Jahr in das Sozio-oekonomische Panel einbezogen wurden.

Bei der Anzahl der hochgerechneten Haushalte und Personen ist insgesamt eine relativ gute Anpassung an die im Mikrozensus ausgewiesenen Zahlen festzustellen (siehe Pischner 2001: 9f.; Tabelle 1). Eine Ausnahme

bildet lediglich das Jahr 1996, in dem die hochgerechnete Anzahl der Personen etwa 6 % unter der im Mikrozensus ausgewiesenen liegt⁵⁾.

Für die Analyse der räumlichen Mobilität wurden zunächst mehrere Längsschnittdatensätze gebildet, die die Zeiträume 1996/97, 1997/98 und

5) Die Ursache für diese Abweichung konnte bislang auch nicht mit Hilfe von Rückfragen an die Projektgruppe „Sozio-oekonomisches Panel“ geklärt werden. Da die Abweichungen keine Auswirkungen auf die Analysen des Zusammenhangs zwischen räumlicher Mobilität und Arbeitsmarktprozessen haben dürften, werden die weiteren Analysen auf der dargestellten Datenbasis vorgenommen.

Anzahl der befragten und der hochgerechneten Haushalte und Personen in den Längsschnittdatensätzen des Sozio-oekonomischen Panels für die Zeiträume 1996/1997, 1997/1998 und 1998/1999			
Merkmal	1996/97	1997/98	1998/99
Anzahl der befragten Haushalte			
Bundesgebiet	6 423	6 217	6 987
NRW	1 251	1 219	1 404
Prozentualer Anteil von NRW	19,5	19,6	20,1
Hochgerechnete Anzahl der Haushalte			
Bundesgebiet	36 493 663	37 051 685	36 985 089
NRW	8 253 306	8 666 433	8 425 323
Prozentualer Anteil von NRW	22,6	23,4	22,8
Anzahl der befragten Personen (mit Kinder)			
Bundesgebiet	16 014	15 301	16 814
NRW	3 159	3 004	3 390
Prozentualer Anteil von NRW	19,7	19,6	20,2
Hochgerechnete Anzahl der Personen (mit Kinder)			
Bundesgebiet	77 818 134	77 739 727	77 608 192
NRW	17 497 927	17 677 786	18 020 378
Prozentualer Anteil von NRW	22,5	22,7	23,2

Quelle: Das Sozio-oekonomische Panel 1996 bis 1999; eigene Berechnungen

1998/99 abdecken. Berücksichtigt wurden jeweils die Haushalte bzw. Personen, die in beiden Jahren an der Befragung teilgenommen haben. Der erste Längsschnittdatensatz (1996/97) umfasst 6 483 Haushalte und 16 351 Personen. In den beiden darauf folgenden Zeiträumen liegt die Fallzahl etwas darunter, während sich für den letzten Zeitraum – wiederum bedingt durch die Aufstockung – eine etwas höhere Fallzahl ergibt. Auf NRW entfällt auch hier im Durchschnitt ein Fünftel der befragten Haushalte und Personen.

Zur Aussagekraft der Ergebnisse

Zur Bewertung der Aussagekraft der Ergebnisse werden in der Regel zum einen die Standardfehler berechnet und zum anderen mögliche systematische Fehler analysiert. Für die Beurteilung der Standardfehler werden in der amtlichen Statistik üblicherweise Fehlerrechnungen durchgeführt, die auch für die Nutzer der Daten zugänglich sind. Vergleichbare Informationen liegen für das Sozio-oekonomische Panel jedoch nicht vor, so dass die Standardfehler für die folgenden Ergebnisse nicht ohne weiteres angegeben werden können.

Detaillierte Analysen wurden von der Projektgruppe „Sozio-oekonomisches Panel“ jedoch zu den systematischen Fehlern durchgeführt, die aus der Panelmortalität resultieren können (siehe Rendtel 1995; Pannenberg 2000). Wie bereits erwähnt, wird die Panelmortalität durch die natürliche Mortalität sowie durch die Nichterreichbarkeit der Haushalte und die Verweigerung der weiteren Teilnahme verursacht. Während die natürliche Mortalität die Ergebnisse nicht beeinflusst, können die Nichterreichbarkeit und die Verweigerung der weiteren Teilnahme die Aussagekraft der Ergebnisse beeinträchtigen. Dies ist dann der Fall, wenn sich die ausgefallenen Haushalte von den nicht ausgefallenen Haushalten hinsichtlich der darzustellenden Merkmale systematisch unterscheiden. Die Untersuchungen der Projektgruppe „Sozio-oekonomi-

sches Panel“ weisen darauf hin, dass in den Wellen der Jahre 1996 bis 1998 nur ein sehr geringer Anteil derjenigen Haushalte, die in der vorangegangenen Welle befragt wurden, in der Folgewelle nicht wieder aufgefunden wurden. Der Anteil liegt in den genannten Jahren und in den einzelnen Teilstichproben unter 1 % (Pannenberg 2000: 16f., Tabellen 2 bis 4). Deutliche Unterschiede ergeben sich allerdings hinsichtlich einzelner Haushaltstypen. So werden insbesondere Haushalte, die sich zwischen zwei Wellen aufgeteilt haben, deutlich seltener wieder aufgefunden als Haushalte, auf die dies nicht zutrifft.

Von den wieder aufgefundenen Haushalten haben in den Jahren 1996 bis 1998 zwischen 5 % und 10 % die weitere Teilnahme an der Panelerhebung verweigert (Pannenberg 2000: 16f., Tabellen 7 bis 9). Weitere Analysen der Projektgruppe „Das Sozio-oekonomische Panel“ zeigen, dass die Teilnahmebereitschaft neben den sozio-ökonomischen Merkmalen der Haushaltsvorstände durch Veränderungen in der Haushaltszusammensetzung sowie durch erhebungsbedingte Ereignisse beeinflusst werden (Pannenberg 2000: 16f., Tabellen 7 bis 9). Für die folgende Untersuchung ist insbesondere die Frage von Bedeutung, ob und in welchem Ausmaß Haushalte, die zwischen zwei Panelwellen umgezogen sind, häufiger die Teilnahme am Panel verweigern als immobile Erhebungseinheiten. In diesem Zusammenhang weisen die Analysen darauf hin, dass Haushalte, die umgezogen sind, seltener wieder befragt werden als andere Haushalte. So konnten beispielsweise im Jahr 1996 9,8 % aller umgezogenen Haushalte der Stichproben A und B nicht wieder befragt werden, während der entsprechende Wert bei den Haushalten ohne Umzug nur 5,3 % betrug (Pannenberg 2000: 16f., Tabellen 7 bis 9). Ähnliche Verhältnisse zeigen sich auch in den folgenden Jahren sowie für die anderen Teilstichproben. Dieses Ergebnis legt zunächst die Vermutung nahe, dass das *Ausmaß* an räumlicher Mobilität auf der Basis des Sozio-oekonomischen

Panels unterschätzt wird. Hier ist jedoch zu bedenken, dass die räumliche Mobilität bei der Bildung der Längsschnittgewichte berücksichtigt wird und dies wiederum Auswirkungen auf die Schätzung des Ausmaßes an räumlicher Mobilität hat.

Frick (1996) untersucht in diesem Zusammenhang für die Jahre 1984 bis 1994, ob die Längsschnittgewichtung des Sozio-oekonomischen Panels tatsächlich die geringere Teilnahmebereitschaft der räumlich mobilen Haushalte kompensiert. Er kommt zu dem Ergebnis, dass die Anwendung der Gewichtungsfaktoren für die gesamte Population zu einer nahezu perfekten Nachbildung der Teilnahmebereitschaft führt. Eine Einschränkung besteht lediglich darin, dass die Teilnahmequote der mobilen Gruppe in einigen Jahren etwas überschätzt wird (Frick 1996: 131)⁶.

Ein weiteres Problem bei der Ermittlung des Ausmaßes an räumlicher Mobilität ergibt sich im Sozio-oekonomischen Panel weiterhin daraus, dass durch die Längsschnittgewichtung die höheren Verweigerungsquoten der räumlich mobilen Haushalte in den Längsschnittdaten kompensiert werden, dies jedoch nicht auf die Untererfassung dieser Gruppe in der ersten Panelwelle zutrifft. Hier greifen die beschriebenen Gewichtungsverfahren nicht, da die erste Welle nur anhand externer Datenquellen kalibriert werden kann und hier in der Regel keine Informationen über Haushaltsveränderungen vorliegen. Im Ergebnis ist daher eher von einer leichten Unterschätzung als von einer Überschätzung der räumlich mobilen Haushalte im Sozio-oekonomischen Panel auszugehen. Sofern die Unterschätzung der räumlichen Mobilität Auswirkungen auf die Ergebnisinterpretationen hat, wird darauf entsprechend eingegangen.

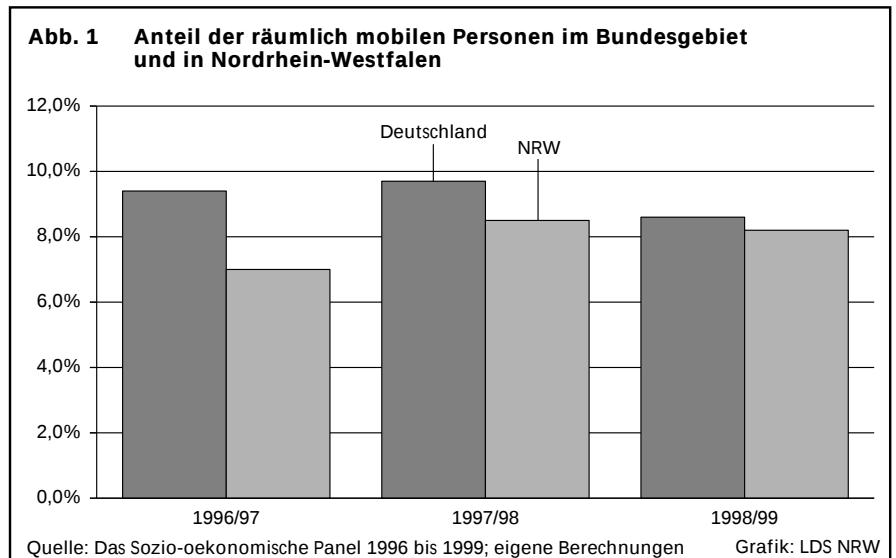
⁶ Frick (1996) prüft daher, ob die Abweichungen der geschätzten Teilnahmequote vom Erwartungswert signifikant sind. Er kommt zu dem Ergebnis, dass dies in einigen Jahren der Fall ist. Er korrigiert die Abweichungen durch eine weitere Kalibrierung der Bleibwahrscheinlichkeiten, so dass das selektive Teilnahme- bzw. Ausfallverhalten kontrolliert und ausgeglichen wird (Frick 1996: 132).

Bei der Bildung der Längsschnittgewichte wird unterstellt, dass die räumlich mobilen Haushalte, die nicht mehr befragt werden, keine systematischen Unterschiede im Vergleich zu den räumlich immobilen Haushalten, die weiterhin an der Erhebung teilnehmen, aufweisen. Lagen beispielsweise systematische Unterschiede hinsichtlich der Arbeitsmarktmobilität zwischen beiden Gruppen vor, so würden die Analysen des Zusammenhanges zwischen der räumlichen Mobilität und der Arbeitsmarktmobilität trotz Längsschnittgewichtung Verzerrungen aufweisen. Auf der Basis der Analysen der Projektgruppe „Das Sozio-oekonomische Panel“ erscheint jedoch die Annahme plausibler, dass die Teilnahmebereitschaft der räumlich mobilen Haushalte eher von feldbedingten Ereignissen beeinflusst wird. So weisen die Analysen insbesondere darauf hin, dass die Teilnahmebereitschaft der Haushalte sehr stark durch einen Interviewerwechsel beeinflusst wird – ein Ereignis, das eng mit den Umzügen der Haushalte verknüpft ist. Es ist daher davon auszugehen, dass auf der Basis der Daten des Sozio-oekonomischen Panels durchaus *systematische Unterschiede* zwischen räumlich mobilen und räumlich immobilen Personen hinsichtlich der Arbeitsmarktprozesse dargestellt werden können.

Zusammenhänge zwischen räumlicher Mobilität und Arbeitsmarktprozessen

Ausmaß der räumlichen Mobilität im Zeitablauf

Auf der Basis des Sozio-oekonomischen Panels soll im Folgenden der Zusammenhang zwischen räumlicher Mobilität und anderen personen- und haushaltsbezogenen Veränderungen analysiert werden. Da bei dieser Betrachtung der Zusammenhang zwischen räumlicher Mobilität und Arbeitsmarktprozessen im Vordergrund steht, gehen in die folgenden Analysen alle Personen ein, die zwischen 16 und 65 Jahre alt sind. Hierzu wird zunächst für die Bundes- und Landesebene berechnet, wie hoch der Anteil der Personen ist, die



zwischen 1996 und 1997, 1997 und 1998 oder 1998 und 1999 ihre Wohnung gewechselt haben. Dabei wird jeder Wechsel der Wohnung erfasst, selbst wenn dieser innerhalb eines Hauses stattfand. Außerdem werden solche Personen als räumlich mobil gekennzeichnet, die aus einem Haushalt fortgezogen sind, um einen neuen Haushalt zu gründen⁷⁾.

Im Zeitraum 1996/97 haben im Bundesgebiet 9,4 % aller Personen ihre Wohnung gewechselt (siehe Abbildung 1)⁸⁾. Für die folgenden Zeiträume variiert dieser Anteil leicht, so dass sich im Durchschnitt eine Mobilitätsquote in Höhe von 9,3 % ergibt. In NRW ist der Anteil der mobilen Personen im Zeitraum 1996/97 geringer, nähert sich jedoch in den folgenden Jahren den bundesweiten Zahlen an. Hier ergibt sich im Durchschnitt eine Mobilitätsquote in Höhe von 7,3 %. Wie bereits im vorangegangenen Abschnitt erwähnt, wird das Ausmaß an räumlicher Mobilität im Sozio-oekonomischen Panel voraussichtlich unterschätzt. Diese Vermutung kann auf der Basis einer ersten vorläufigen Auswertung des Mikrozensus-Panels bestätigt werden. Danach waren zwischen 1996 und 1997 etwa 10 % aller Personen im Alter von 15 bis 65 Jahren räumlich mobil.

7) Das Konzept ist somit mit der Erfassung der räumlichen Mobilität im Mikrozensus gut vergleichbar. – 8) Eine Analyse von Frick (1996) zeigt, dass es sich bei den im SOEP erfassten Umzügen vorrangig um kleinräumliche Mobilitätsprozesse handelt, die sich innerhalb einer kreisfreien Stadt oder eines Landkreises vollziehen.

Aus Abbildung 1 wird ersichtlich, dass die Gruppe der räumlich mobilen Personen relativ klein ist. Um Ergebnisse zum Zusammenhang zwischen räumlicher Mobilität und Arbeitsmarktprozessen darstellen zu können, werden die räumlichen Wechsel aus den einzelnen Zwei-Jahres-Längsschnitten für die folgenden Analysen kumuliert, und die Ergebnisse werden nur für das Bundesgebiet ausgewiesen. Bei den Ergebnissen handelt es sich somit um Durchschnittswerte für den Zeitraum 1996 bis 1999 auf der Bundesebene. Da jedoch zu vermuten ist, dass der Zusammenhang zwischen räumlicher Mobilität und Arbeitsmarktprozessen in Nordrhein-Westfalen ähnlich strukturiert ist wie im Bundesgebiet, sollten die Analysen auch Rückschlüsse auf Nordrhein-Westfalen zulassen.

Wechsel zwischen unterschiedlichen Erwerbszuständen

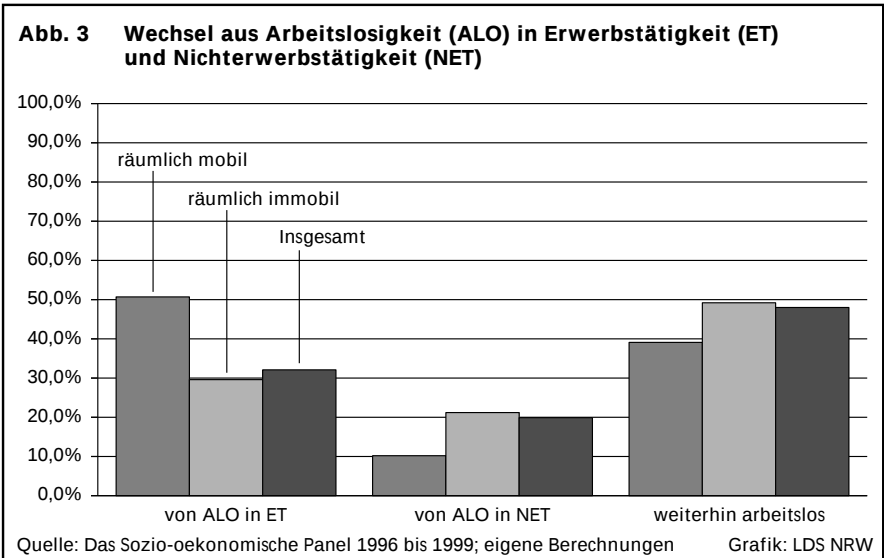
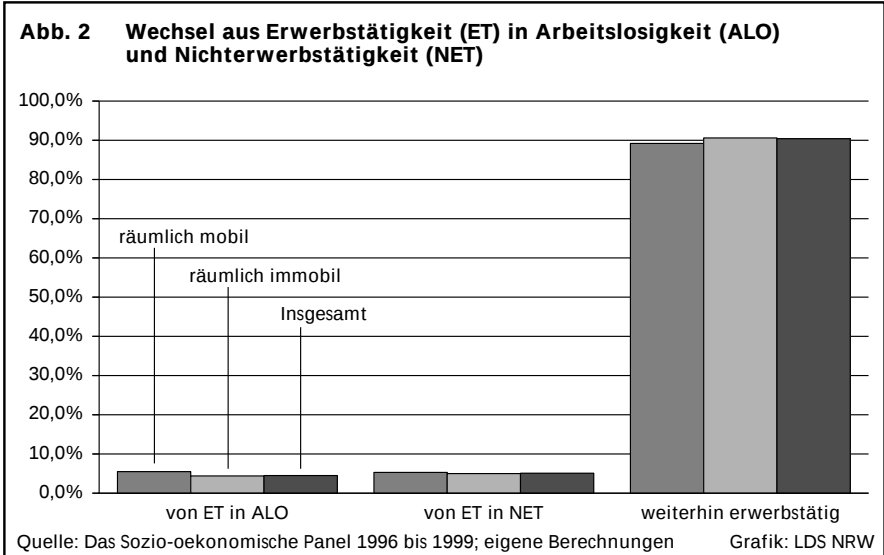
Im Folgenden wird zunächst der Zusammenhang zwischen der räumlichen Mobilität und den Wechseln zwischen mehreren Erwerbszuständen untersucht. Unterschieden werden dabei die Zustände (1) erwerbstätig, (2) arbeitslos und (3) nicht-erwerbstätig. Als erwerbstätig gelten Personen dann, wenn sie entweder einer Vollzeit- oder Teilzeitarbeit nachgehen, sie geringfügig beschäftigt sind oder sie sich in einer betrieblichen Ausbildung oder im Mut-

terschafts- oder Erziehungsurlaub befinden. Als arbeitslos werden solche Personen bezeichnet, die nicht erwerbstätig und gleichzeitig beim Arbeitsamt arbeitslos gemeldet sind. Nichterwerbstätig sind schließlich solche Personen, die weder beschäftigt noch beim Arbeitsamt arbeitslos gemeldet sind.

Auf der Basis der definierten Kategorien werden Wechsel zwischen den einzelnen Erwerbszuständen untersucht, die sich von einem Befragungszeitpunkt zum anderen vollzogen haben. Bei der Betrachtung solcher Wechsel ist zu berücksichtigen, dass die Mobilität insgesamt unterschätzt wird, da mehrfache Wechsel nicht erfasst werden. So kann beispielsweise eine Person zum ersten und zweiten Befragungszeitpunkt eine Erwerbstätigkeit ausgeübt haben, in der Zwischenzeit jedoch arbeitslos gewesen sein.

Abbildung 2 weist aus, wie hoch jeweils der Anteil der räumlich mobilen und der räumlich immobilen Personen ist, die aus einer Erwerbstätigkeit in Arbeitslosigkeit oder in Nichterwerbstätigkeit übergegangen sind⁹⁾. Es zeigt sich, dass von den räumlich Mobilen 5,5 % einen Wechsel aus einer Erwerbstätigkeit in Arbeitslosigkeit vorgenommen haben. Dieser Anteil lag bei den räumlich Immobilen mit 4,4 % nur geringfügig darunter. Bei dem Übergang von Erwerbstätigkeit in Nichterwerbstätigkeit sind die Unterschiede zwischen den räumlich mobilen und den räumlich immobilen Personen noch geringer. Ein solcher Wechsel ist von 5,3 % der räumlich Mobilen und von 5,0 % der räumlich Immobilen vollzogen worden.

Während sich bei den Übergängen aus einer Erwerbstätigkeit in andere Erwerbszustände kaum Unterschiede zwischen den beiden Gruppen zeigen, sind diese bei den Wechseln aus Arbeitslosigkeit in eine Erwerbstätigkeit und in Nichterwerbstätigkeit stark ausgeprägt. Deutlich werden



diese Unterschiede in Abbildung 3, die den Anteil derjenigen Personen ausweist, die von Arbeitslosigkeit in eine Erwerbstätigkeit und in Nichterwerbstätigkeit übergegangen sind. Während gut die Hälfte der räumlich mobilen Personen, die im Ausgangszustand arbeitslos waren, im Folgejahr eine Erwerbstätigkeit ausüben, liegt dieser Anteil bei den räumlich immobilen Personen bei nur knapp 30 %. Umgekehrt sind von den räumlich mobilen Personen, die sich in Arbeitslosigkeit befanden, innerhalb eines Jahres lediglich 10 % in Nichterwerbstätigkeit übergegangen, während dieser Anteil bei den räumlich immobilen Personen in etwa doppelt so hoch ist.

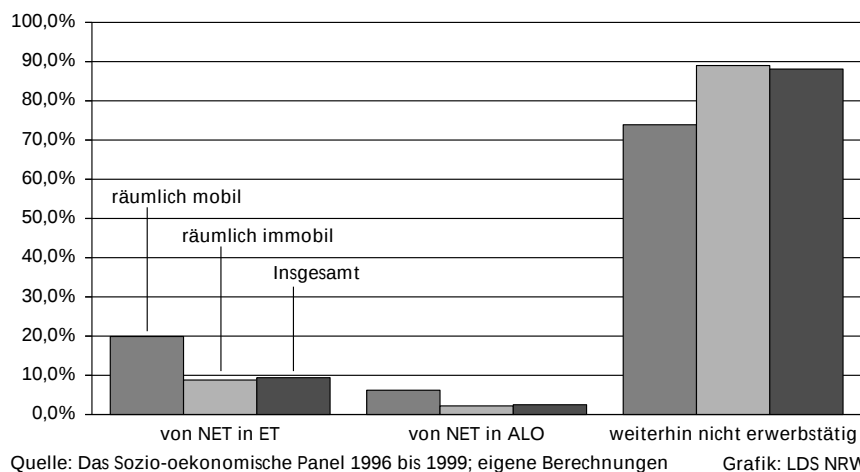
Deutliche Unterschiede zeigen sich auch in der folgenden Abbildung 4. Sie weist die Anteile derjenigen Per-

sonen aus, die jeweils innerhalb eines Jahres aus Nichterwerbstätigkeit in eine Erwerbstätigkeit oder in Arbeitslosigkeit gewechselt sind. Während von den räumlich Mobilen 19,9 % eine Erwerbstätigkeit aufgenommen haben, hat ein solcher Wechsel nur bei 8,8 % der räumlich Immobilen stattgefunden. Außerdem sind 6,2 % der räumlich Mobilen in Arbeitslosigkeit gewechselt, während dieser Anteil bei den räumlich Immobilen mit 2,2 % deutlich geringer ist.

Die bisher dargestellten Ergebnisse zeigen, dass die räumliche Mobilität eng mit der Aufnahme einer neuen Beschäftigung zusammenhängt. Da die räumlich mobilen Personen im Mikrozensus nicht enthalten sind, ist zu erwarten, dass die Ergebnisse, die auf der Basis der Längsschnittda-

9) Einbezogen werden in die Analysen nur diejenigen Personen, die bei den betrachteten Merkmalen im Vor- und Folgejahr gültige Werte aufweisen.

Abb. 4 Wechsel von Nichterwerbstätigkeit (NET) in Erwerbstätigkeit (ET) und in Arbeitslosigkeit (ALO)



ten des Mikrozensus gewonnen werden, für solche Übergänge Verzerrungen aufweisen. Um das Ausmaß dieser Verzerrungen annähernd bestimmen zu können, kann die Arbeitsmarktmobilität der räumlich immobilen Personen mit der Arbeitsmarktmobilität aller Personen verglichen werden. Führt man diesen Vergleich auf der Basis der Angaben in den Graphiken 3 und 4 durch, so ist zu erkennen, dass die beiden Werte sehr nah beieinander liegen. Dies bedeutet, dass der Beitrag der räumlich mobilen Personen zu der Arbeitsmarktmobilität insgesamt sehr gering ist. Zurückzuführen ist dies darauf, dass die Gruppe der räumlich mobilen Personen relativ klein ist und somit bei der Ermittlung der Arbeitsmarktmobilität insgesamt nicht sehr stark ins Gewicht fällt. Für die Längsschnittanalysen auf der Basis des Mikrozensus folgt daraus, dass selbst bei denjenigen Übergängen, bei denen sich die räumlich mobilen und die räumlich immobil Personen deutlich unterscheiden, eher geringe Ergebnisverzerrungen zu erwarten sind. Zu berücksichtigen ist hierbei jedoch, dass der Anteil der räumlich mobilen Personen im Mikrozensus höher ist als im Sozio-oekonomischen Panel und somit auch die Ergebnisverzerrungen, die aus den fortzugsbedingten Ausfällen resultieren, etwas größer ausfallen, als es die dargestellten Ergebnisse nahe legen. Da jedoch der Unterschied in den Anteilen der räumlich mobilen Personen im Mikrozensus und im Sozio-

oekonomischen Panel nicht sehr stark ausgeprägt ist, dürften sich die Ergebnisverzerrungen nur geringfügig erhöhen. Zu beachten ist weiterhin, dass bislang nur Entwicklungen betrachtet wurden, die sich zwischen zwei Erhebungszeitpunkten vollzogen haben. Daher soll in weiterführenden Analysen untersucht werden, wie sich die fortzugsbedingten Ausfälle auf die Ergebniserstellung auswirken, wenn längere Untersuchungszeiträume betrachtet werden.

Wechsel zwischen beruflichen Tätigkeiten und Wirtschaftszweigen

Die beruflichen Tätigkeiten werden im Sozio-oekonomischen Panel anhand der Internationalen Standardklassifikation der Berufe (ISCO) aus

dem Jahre 1968 klassifiziert, wobei drei Varianten zur Verfügung stehen, die sich nach dem Grad der Differenzierung unterscheiden. Im Folgenden werden die Ergebnisse für den „ISCO-Zweisteller“ ausgewiesen, der die Zusammenfassung der beruflichen Tätigkeiten zu 97 größeren Tätigkeitsfeldern vorsieht. Auf der Basis dieser Klassifikation wird der Anteil derjenigen Personen berechnet, die zwischen zwei Erhebungszeitpunkten ihr Tätigkeitsfeld gewechselt haben. Da die Bezugsgröße alle Personen sind, die zum Ausgangszeitpunkt erwerbstätig waren, werden nicht nur die Wechsler den Nichtwechslern gegenübergestellt, sondern es kann auch der Fall eintreten, dass eine Person in Nichterwerbstätigkeit oder Arbeitslosigkeit übergeht.

Abbildung 5 zeigt, dass bei 14,3 % derjenigen Personen, die räumlich mobil gewesen sind, gleichzeitig ein Wechsel des beruflichen Tätigkeitsfeldes zu beobachten ist. Damit weisen die räumlich mobilen Personen eine Wechselhäufigkeit auf, die nur etwa vier Prozentpunkte über derjenigen der räumlich immobil Personen liegt. Legt man den Analysen eine gröbere Klassifikation der beruflichen Tätigkeiten zugrunde, so nimmt die Wechselhäufigkeit insgesamt ab; die systematischen Unterschiede zwischen beiden Gruppen bleiben jedoch erhalten. Hingegen nimmt bei einer feineren Klassifikation der beruflichen Tätigkeiten die Wechselhäufigkeit zu; die systemati-

Abb. 5 Wechsel zwischen unterschiedlichen beruflichen Tätigkeitsfeldern und Wechsel in Nichterwerbstätigkeit bzw. Arbeitslosigkeit

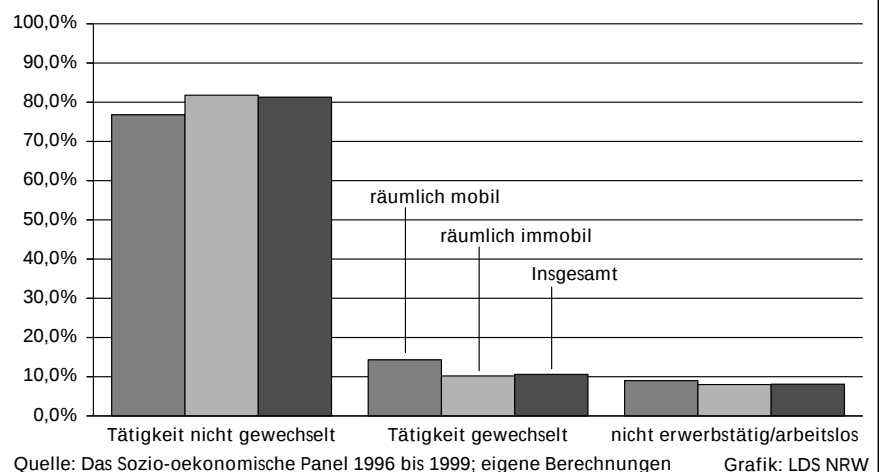
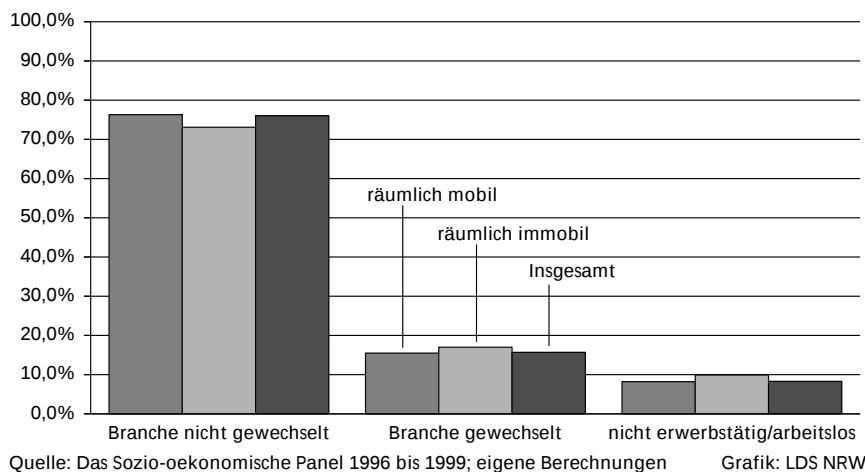


Abb. 6 Wechsel zwischen Wirtschaftszweigen und Wechsel in Nichterwerbstätigkeit bzw. Arbeitslosigkeit



schen Unterschiede zwischen beiden Gruppen sind jedoch auch hier zu beobachten.

Die Wirtschaftszweige werden im Sozio-oekonomischen Panel anhand der vom Zentrum für Umfragen, Methoden und Analysen (ZUMA) entwickelten Wirtschaftszweigssystematik klassifiziert, die 35 Branchen unterscheidet. So wie bei den beruflichen Tätigkeiten wird im Folgenden der Anteil derjenigen Personen betrachtet, die zwischen zwei Erhebungszeitpunkten den Wirtschaftszweig gewechselt haben. Aus der Abbildung 6 wird ersichtlich, dass sich auch bei diesen Wechseln keine großen Unterschiede zwischen den räumlich mobilen und den räumlich immobil Personen zeigen. So haben 15,5 % der räumlich mobilen und 17,0 % der räumlich immobil Personen den Wirtschaftszweig gewechselt.

Einkommensentwicklungen

Abschließend wird untersucht, wie sich das Einkommen der räumlich mobilen und der räumlich immobil Personen zwischen zwei Erhebungszeitpunkten verändert hat. Zugrunde gelegt werden dabei die Angaben zum Bruttogehalt, so dass nur Personen in die Analysen eingehen, die zum Ausgangszeitpunkt erwerbstätig waren. Die Einkommensentwicklung wurde berechnet, indem zunächst der mittlere Einkommenszuwachs ermittelt wurde. Dieser beträgt für den untersuchten

Zeitraum 9,1 %. Anschließend wurden die individuellen Einkommenszuwächse in Relation zu dem durchschnittlichen Einkommen kategorisiert, und zwar so, dass niedrige, mittlere und hohe Einkommenszuwächse unterschieden werden können¹⁰⁾. Personen, deren Einkommen sich negativ entwickelt hat, wurden zu der Gruppe mit einer negativen Einkommensentwicklung zusammengefasst. Da sich auch hier die Bezugsgruppe aus den Personen zusammensetzt, die im Ausgangsjahr erwerbstätig waren, werden auch hier die Wechsel in Nichterwerbstätigkeit und Arbeitslosigkeit ausgewiesen.

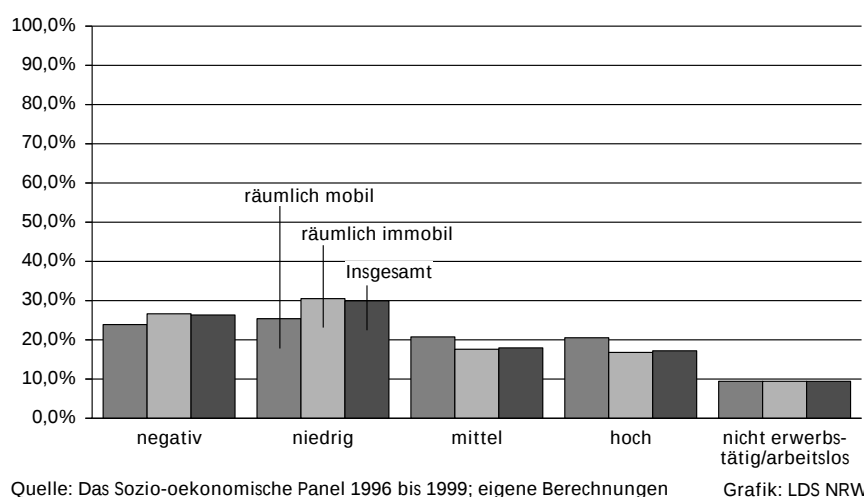
10) Die Gruppe mit einem niedrigen Einkommenszuwachs weist eine Zuwachsrate auf, die weniger als die Hälfte des Durchschnittswertes beträgt. Die Gruppe mit einem hohen Einkommenszuwachs hingegen weist eine Zuwachsrate auf, die 50 % über dem durchschnittlichen Wert liegt.

Aus Abbildung 7 wird ersichtlich, dass die räumlich immobil Personen etwas häufiger negative und niedrige Einkommenszuwächse aufweisen als die räumlich mobilen Personen. So nahm das Bruttogehalt bei 26,6 % der räumlich immobil Personen ab, während der entsprechende Anteil bei den räumlich mobilen Personen bei 23,9 % liegt. Gleichzeitig weisen nur 25,4 % der räumlich mobilen Personen, aber 30,5 % der räumlich immobil Personen niedrige Einkommenszuwächse auf. Im Gegensatz dazu konnten die räumlich mobilen Personen in rund 21 % aller Fälle mittlere und hohe Einkommenszuwächse erzielen, während die entsprechenden Anteile bei den räumlich Immobil nur 17,6 % bzw. 16,7 % betragen. Werden diese Analysen für das Nettoeinkommen durchgeführt, zeigen sich vergleichbare Ergebnisse.

Schlussfolgerungen für die Längsschnittanalysen auf der Basis des Mikrozensus

In den vorangegangenen Abschnitten wurde der Zusammenhang zwischen räumlicher Mobilität und verschiedenen Arbeitsmarktprozessen untersucht. Betrachtet wurden dabei Wechsel zwischen Erwerbszuständen, Wechsel zwischen beruflichen Tätigkeiten und Wirtschaftszweigen sowie Veränderungen der Einkommenssituation, die sich zwischen zwei Erhebungszeitpunkten vollzo-

Abb. 7 Entwicklung des Bruttogehaltes und Wechsel in Nichterwerbstätigkeit bzw. Arbeitslosigkeit



gen haben. Arbeitsmarktprozesse, die nur kleine Gruppen wie befristet oder geringfügig Beschäftigte betreffen, konnten auf der Basis des Sozio-oekonomischen Panels aufgrund zu geringer Fallzahlen in den ausgewählten Datensätzen nicht untersucht werden. Ebenso konnten aufgrund der geringen Fallzahlen bei den räumlich mobilen Personen keine Zusammenhänge zwischen den sozio-ökonomischen Merkmalen und den Arbeitsmarktprozessen untersucht werden.

Die durchgeführten Analysen weisen darauf hin, dass zwischen der räumlichen Mobilität und den untersuchten Arbeitsmarktprozessen kein genereller Zusammenhang besteht. Nur bei einer sehr geringen Anzahl an Wechselarten sind deutliche Unterschiede zwischen räumlich mobilen und räumlich immobilen Personen zu beobachten. Dies betrifft insbesondere Wechsel von Arbeitslosigkeit in Erwerbstätigkeit sowie Wechsel von Nichterwerbstätigkeit in Erwerbstätigkeit. Um annähernd abschätzen zu können, welche Auswirkungen das Fehlen der räumlich mobilen Personen auf die Ermittlung von Mobilitätsquoten auf der Basis des Mikrozensus-Panels hat, können die Ergebnisse für die räumlich immobilen Personen mit den Ergebnissen für alle Personen verglichen werden. Dabei zeigt sich, dass selbst bei denjenigen Wechselarten, bei denen sich die räumlich mobilen deutlich von den räumlich immobilen Personen unterscheiden, die Wechselquote insgesamt nah an derjenigen der räumlich immobilen Personen liegt. Für Längsschnittanalysen auf der Basis des Mikrozensus ist daraus zu schließen, dass die Analyse von Entwicklungen, die sich zwischen zwei Erhebungsjahren vollziehen, durch die systematischen Ausfälle infolge der Umzüge selbst dann kaum verzerrt sein dürften, wenn sich die Wechselquoten der räumlich mobilen und der räumlich immobilen Personen deutlich unterscheiden.

Zurückzuführen sind diese geringen Verzerrungen darauf, dass die Gruppe der räumlich mobilen Personen relativ klein ist.

Da bislang nur solche Veränderungen untersucht wurden, die sich zwischen zwei Erhebungszeitpunkten vollzogen haben, ist im Rahmen weiterführender Analysen zum einen geplant, die Zusammenhänge zwischen räumlicher Mobilität und Arbeitsmarktprozessen für längere Untersuchungszeiträume zu analysieren. Zum anderen sollen neben den Arbeitsmarktprozessen auch die Zusammenhänge zwischen räumlicher Mobilität und familiären Veränderungen untersucht werden.

Literaturverzeichnis

Frick, Joachim (1996): Lebenslagen im Wandel: Determinanten klein-räumlicher Mobilität in Westdeutschland. Frankfurt/New York

Frick, Joachim/Rendtel, Ulrich/Wagner, Gert (1993): Eine Strategie zur Kontrolle von Längsschnittgewichtungen am Beispiel des Sozio-oekonomischen Panels. Diskussionspapier Nr. 80, Deutsches Institut für Wirtschaftsforschung, Berlin

Haisken-DeNew, John P./Frick, Joachim R. (Eds.) (2000): Desktop Companion to the German Socio-Economic Panel Study. Unpublished paper. Deutsches Institut für Wirtschaftsforschung, Berlin

Hanefeld, Ute (1987): Das Sozio-oekonomische Panel: Grundlagen und Konzeption. Sozio-oekonomische Daten und Analysen für die Bundesrepublik Deutschland, Band 1. Frankfurt am Main/New York

Heidenreich, Joachim (1994): Hochrechnung des Mikrozensus ab 1990.

In: Gabler, Siegfried/Hoffmeyer-Zlotnik, Jürgen H. P./Krebs, Dagmar (Hrsg.): Gewichtung in der Umfragepraxis. Opladen, 112 – 123

Pannenberg, Markus (2000): Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (GSOEP) (1984 until 1998). Diskussionspapier Nr. 196. Deutsches Institut für Wirtschaftsforschung, Berlin

Pischner, Rainer (1994): Quer- und Längsschnittgewichtung des Sozio-oekonomischen Panels (SOEP). In: Gabler, Siegfried/Hoffmeyer-Zlotnik, Jürgen H. P./Krebs, Dagmar (Hrsg.): Gewichtung in der Umfragepraxis. Opladen

Pischner, Rainer (2000): Überarbeitete Querschnittshochrechnung der Wellen G-N (1990 bis 1997) des Sozio-oekonomischen Panels (SOEP) unter Einbeziehung der Ergänzungstichprobe E (Welle O). Unveröffentlichtes Manuskript. Deutsches Institut für Wirtschaftsforschung, Berlin

Rendtel, Ulrich (1995): Lebenslagen im Wandel: Panelausfälle und Panelrepräsentativität. Frankfurt am Main/New York

Rendtel, Ulrich/Wagner, Gert/Frick, Joachim (1995): Eine Strategie zur Kontrolle von Längsschnittgewichtungen in Panelerhebungen: Das Beispiel des Sozio-Oekonomischen Panels. In: Allgemeines Statistisches Archiv 79, 252 – 277

Rendtel, Ulrich/Pannenberg, Markus/Daschke, S. (1997): Die Gewichtung der Zuwanderer-Stichprobe des Sozio-oekonomischen Panels (SOEP). In: Vierteljahreshefte zur Wirtschaftsforschung, 2, 271 – 286

Zühlke, Sylvia (2001): Längsschnittanalysen auf der Basis des Mikrozensus: Methodische Probleme und Lösungsansätze. In: Statistische Analysen und Studien Nordrhein-Westfalen 4, 3 – 13

Sicherung persönlicher Angaben in Tabellendaten

Dipl.-Physiker Rüdiger Dietz Repsilber

1. Motivation der statistischen Geheimhaltung

Die Vermeidung der Offenlegung persönlicher Angaben in aggregierten Veröffentlichungstabellen, die statistische Geheimhaltung, ist eine fundamentale Aufgabe jeder Statistiken erhebenden und verbreitenden Institution, weil damit die für die Aussagefähigkeit der zu veröffentlichenden Daten unabdingbare Vertrauensbasis geschaffen und erhalten wird. Andererseits ist mit dem Schutz persönlicher Daten gegen ihre Offenlegung untrennbar ein Informationsverlust verbunden, der die Aussagefähigkeit der Statistik – wenn auch auf kontrollierbare Weise – einschränkt. Die Maxime muss daher sein: so viel Offenlegung wie möglich und nur so viel Geheimhaltung wie unbedingt nötig. Danach zu verfahren ist umso wichtiger, als diejenigen, die zu diesen Statistiken berichten, häufig auch zum Kreis der diese Statistiken Nachfragenden gehören, so dass ein wechselseitiges Interesse an einer optimalen Datensicherung besteht.

Als weiterer vertrauensbildender Maßnahme zur Förderung der Akzeptanz von Statistikerhebungen kommt der Offenlegung der angewendeten Geheimhaltungsverfahren selbst eine große Bedeutung zu. Das gilt auch für eine umfassende Unterrichtung über die Weiterentwicklung von Geheimhaltungsverfahren, die dem veröffentlichenden Statistiker aufgrund von immer effizienter arbeitender so genannter Attackersoftware aufgezwungen wird. Wegen der dadurch inzwischen erreichten Komplexität der Verfahren (Sicherung von unterschiedlichen Tabellen mit gemeinsamen Aggregaten, Berücksichtigung von Nutzerinformationen, die dieser schon vor dem Erscheinen der zu sichernden Tabelle haben kann) ist ein wirksamer Schutz von Tabellen gegen die Preisgabe von persönlichen Daten ohne ausgefeilte EDV-Unterstützung kaum mehr zu haben.

Jeder Statistiken verbreitenden Institution und der ihr zugeordneten EDV-Abteilung muss es ein besonderes Anliegen sein, trotz zunehmender Komplexität des Geheimhaltungsproblems möglichst einfache Verfahren zu entwickeln, die sich – zumindest im Prinzip – allgemein verständlich vermitteln lassen und die in bedienungsfreundliche EDV-Programme

umgesetzt werden können. Im Blickfeld steht dabei nicht allein der Nutzer der hinsichtlich der statistischen Geheimhaltung zu sichernden Tabellen, sondern auch der die Statistiken verbreitende Fachstatistiker, weil er die nötigen Sicherungsmaßnahmen gegenüber dem Nutzer überzeugend vertreten können muss, was wiederum der allgemeinen Akzeptanz von Statistikerhebungen zugute kommt.

Gegenstand der statistischen Geheimhaltung sind nach mehreren Kriterien gegliederte oft mehrfach durch Zwischensummen unterteilte Statistiktabellen. In solchen fein gegliederten Tabellen treten viele Werte auf, die einzelnen Berichtenden zugeordnet werden können und die z. B. durch Sperren dieser Werte geheim gehalten werden müssen. Die Unterdrückung solcher sensiblen Werte in der Veröffentlichungstabelle bezeichnet man als primäre Geheimhaltung.

Das Sperren von sensiblen Werten allein genügt jedoch nicht, um sie vor zu genauer Berechnung mit Hilfe noch offener Tabellenwerte mittels der Tabellen-Summen-Beziehungen und eines gewissen, bei den Tabellennutzern zu unterstellenden Vorwissens über die Tabellenwerte zu schützen. Unter Vorwissen sind diejenigen Kenntnisse zu verstehen, die

der Nutzer noch vor der Veröffentlichung der Tabelle haben kann, mit dem er die Tabellenwerte eingrenzen kann. Die Verhinderung der zu genauen Rückrechnung primär geheimer Werte aus noch offenen Tabellenwerten – ggf. unter Zuhilfenahme von zu unterstellendem Vorwissen – wird als sekundäre Geheimhaltung bezeichnet.

Die sekundäre Geheimhaltung hat also ihre Ursache in der Veröffentlichung von Summen- und Zwischensummenwerten. Ohne Angabe solcher Summenwerte wäre die primäre Geheimhaltung alleine bereits hinreichend für den Schutz von Einzelangaben. Andererseits kann der Tabellennutzer im Nachhinein keine Summen oder Zwischensummen berechnen, wenn zu diesen gesperrte Werte beitragen. Gerade die höheren Aggregationsniveaus bis hinauf in die Landessummen könnten dann nicht bestimmt werden. Erst eine sekundäre Geheimhaltung, die Sperrungen bevorzugt in den niedrigsten Aggregationsniveaus setzt, macht die Veröffentlichung solcher Summenwerte überhaupt möglich; sie leistet also einen wesentlichen Beitrag zur Erschließung von zu veröffentlichender Information.

Ein besonders einfaches Verfahren zur sekundären Geheimhaltung, das all diesen Anforderungen gerecht wird, ist das Quaderverfahren, das seit vielen Jahren im LDS NRW und auch in anderen statistischen Ämtern in Form von EDV-Programmen auf aggregierte Tabellendaten angewendet wird. Das Quaderverfahren bildet auch den Schwerpunkt in dieser Darstellung.

2. Zur primären Geheimhaltung

Die primäre Geheimhaltung wird durch Regeln festgelegt, die angeben, welche Tabellenwerte primär geheim gehalten werden sollen, weil für die zu diesen Werten beitragenden Einzelangaben ein erhöhtes Risiko der Offenlegung oder der näherungsweisen Offenlegung besteht (vgl. „Statistische Geheimhaltung in Tabellen“, Sarah Giessing, Forum der Bundesstatistik 31/1999). Demgemäß unterscheiden sich die Regeln zur primären Geheimhaltung zum einen durch das bei den Tabellennutzern zu unterstellende Vorwissen, zum anderen aber auch darin, ob eine exakte oder eine näherungsweise Offenlegung von Einzelangaben verhindert werden soll. Wenn im Rahmen der primären Geheimhaltung von der Vermeidung einer Offenlegung die Rede ist, so bezieht sich diese Aussage immer nur auf die Bestimmung von Einzelangaben durch die Kenntnis des jeweiligen Tabellenwertes und schließt die Zuhilfenahme anderer Werte der Tabelle aus.

2.1 Vermeidung einer exakten Offenlegung von Einzelangaben

Regeln, die eine exakte Offenlegung von Einzelangaben in Tabellenaggregaten verhindern sollen, sind als so genannte Fallzahlregeln eingeführt. Bezeichnet n die Anzahl der Berichtenden, die mit ihren Angaben zu einem Tabellenwert beitragen, so werden folgende Regeln in der amtlichen Statistik angewendet:

$n = 1$: Es trägt nur ein Berichtender zum Tabellenwert bei; als Einzelangabe ist dieser Tabellenwert primär geheim zu halten.

$n = 2$: Es tragen nur zwei Berichtende zum Tabellenwert bei, von denen jeder durch Subtraktion seines Wertes vom Tabellenwert den jeweils anderen Wert exakt berechnen kann; dieser Tabellenwert ist demgemäß ebenfalls primär geheim zu halten.

$n = m$

$m > 2$: Tabellenwerte mit weniger als $m + 1$ Meldeeinheiten müssen primär geheimgehalten werden, wenn unterstellt werden kann, dass ein Vorwissen über den exakten Summenwert von $m - 1$ Meldeeinheiten besteht, mit dem die verbleibende Einzelangabe der m Berichtenden aufgedeckt werden kann.

2.2 Vermeidung einer näherungsweisen Offenlegung von Einzelangaben

Die Fallzahlregeln verhindern zwar, dass eine Einzelangabe aus ihrem Tabellenwert exakt berechnet werden kann, sie schützen aber nicht vor näherungsweiser Offenlegung von Einzelangaben mit Hilfe von Vorwissen. Verfügt der Tabellennutzer beispielsweise über das Vorwissen, dass n Einzelwerte eines aus mehr als n Einzelangaben aggregierten Tabellenwertes den weitaus größten Beitrag zu diesem Tabellenwert leisten, so kann er mit weiterem Vorwissen, der Kenntnis eines Summenwertes aus $n - 1$ dieser dominierenden Einzelangaben, den noch verbleibenden großen Einzelwert bis auf den aus den restlichen nicht dominierenden Einzelwerten bestehenden Summenwert genau abschätzen.

Wenn z. B. eine von zwei dominierenden Einzelangaben bekannt ist – was man bei jedem Berichtenden eines der dominierenden Werte immer unterstellen muss –, so kann die andere dominierende Angabe aus dem Tabellenwert bis auf die Summe der restlichen nicht dominierenden Angaben genau berechnet werden.

Das Dominanzproblem hat also seine Entsprechung im Fallzahlproblem, wobei der Unterschied zwischen beiden in der Restsumme aus den nicht dominierenden Einzelangaben liegt, die im Dominanzfall nur zu einer näherungsweisen Offenlegung führen kann. In der amtlichen Statistik werden die $(n; k)$ -Dominanzregel und die $p\%$ -Regel zur Vermeidung einer zu genauen Offenlegung von Einzelangaben angewendet.

2.2.1 Die $(n; k)$ -Dominanzregel

Ein Tabellenwert aus mehr als n Einzelangaben ist primär geheim zu halten, wenn der Anteil der Summe der n größten Einzelwerte am Gesamttabellenwert größer ist als der vorgegebene Schrankenwert k . Diese Regel ist sinnvoll, wenn über $n - 1$ der dominierenden Einzelwerte ein Vorwissen über den exakten Wert ihrer Summe unterstellt werden kann. Von praktischer Bedeutung für die amtliche Statistik sind bisher die $(1; k)$ - und die $(2; k)$ -Regel mit k -Werten zwischen 50 % und 95 %, je nachdem, welche Sensibilität den zu schützenden Angaben zugedacht wird, wobei im $(2; k)$ -Fall in der Regel eine höhere Schranke k gesetzt wird als bei der $(1; k)$ -Regel.

2.2.2 Die $p\%$ -Regel

Ein Tabellenwert ist nach der $p\%$ -Regel primär geheim zu halten, wenn er nach Abzug des zweitgrößten Einzelbeitrages den größten um weniger als $p\%$ übersteigt, d. h. wenn das Restaggregat, der Tabellenwert abzüglich der beiden größten Einzelwerte, bezogen auf den größten Einzelwert kleiner als $p\%$ ausfällt.

Die Intention dieser Regel besteht in erster Linie im Schutz der am stärksten gefährdeten größten Einzelangabe gegen zu genaue Offenlegung durch die zweitgrößte, während der zweite Teil der Formulierung auf einen gewissen Zusammenhang mit der $(2; k)$ -Dominanzregel hindeutet. Man darf jedoch nicht verkennen, dass der Bezugswert in der $p\%$ -Regel der größte Einzelwert ist, in der Dominanzregel das schützende Restaggregat aber immer als Teil des Gesamtaggregates betrachtet wird.

Die Anwendung der $p\%$ -Regel auf aggregierte Daten führt demgemäß auch zu anderen Ergebnissen als die $(2; k)$ -Dominanzregel mit geeignetem Parameter k . Wie die Auswertungen einiger statistischer Landesämter unter Verwendung von Realdaten aber gezeigt haben, sind die Unterschiede zwischen beiden Re-

geln gering. Für die statistische Praxis werden p%-Werte zwischen 4 % und 8 % empfohlen, wenn man die p%-Regel anstelle der (1; k)- und der (2; k)-Dominanzregel einsetzen will. Welcher der genannten Regeln der Vorzug zu geben ist, scheint Gegenstand einer immer währenden Diskussion zu sein und hängt wohl auch von der in Betracht kommenden Statistik ab.

3. Sekundäre Geheimhaltung

Das Problem der sekundären Geheimhaltung ist durch die Forderung, so wenig Information wie möglich zu unterdrücken, als lineares Optimierungsproblem exakt lösbar. Doch sind solche Lösungsverfahren auf die in der „statistischen Praxis“ zu bearbeitenden umfangreichen Tabellen von größenordnungsmäßig einer Million Tabellenfeldern nicht anwendbar, weil deren Rechenzeiten mit dem Tabellenumfang exponentiell zunehmen und entsprechende EDV-Programme unzumutbar lange Computer-Rechenzeiten beanspruchen würden; man ist daher gezwungen, auf heuristische Verfahren auszuweichen. Das einfachste heuristische Verfahren, das bisher am schnellsten arbeitet und das anderen Heuristiken in Bezug auf die Datensicherung um nichts nachsteht, ist das in der Landesdatenverarbeitungszentrale des LDS NRW entwickelte iterative Quaderverfahren. Im Folgenden wird eine kurze Einführung in die sekundäre Geheimhaltung mit dem Quaderverfahren gegeben; eine ausführliche Darstellung nebst Literaturangaben findet man in der Schriftenreihe „Statistische Analysen und Studien NRW“ in der Ausgabe 3/2000.

Das Verfahren sichert nach mehreren Merkmalen gegliederte, auch mehrfach durch Zwischensummen unterteilte Statistiktabellen gegen zu genaue Rückrechnung ihrer primär geheimen, bereits gesperrten Werte durch zusätzliche Sperrungen, die Sekundärsperrungen von Tabellenfeldern. Es bietet Intervallschutz für die primär geheimen Werte, d. h. es verhindert, dass ein primär gesperrter

Wert genauer schätzbar ist, als es ein vom Anwender des Geheimhaltungsverfahrens vorgegebenes Schutzintervall um den geheimen Wert erlaubt. Darüber hinaus sichert das Verfahren einander überlappende Einzeltabellen, wobei diese in einem iterativen Prozess so lange aneinander abgeglichen werden, bis alle in mehreren Tabellen gemeinsamen Werte den selben Geheimhaltungsstatus haben.

Als Eingabedaten erwartet das Verfahren Tabellendaten, die hinsichtlich der primären Geheimhaltung bereits

Abb. 1

Kreise \ Wirtschaftsgruppen	A	B	C	D	Σ
Kreis 1	10,0	100,0 p s	20,0	100,0 s p	230,0
Kreis 2	0,8	50,1 s	0	0,1 p	51,0
Kreis 3	0,2	0,9	0	0,9	2,0
Reg.-Bez.	11,0	151,0	20,0	101,0	283,0

bearbeitet worden sind. Man spricht von n-dimensionalen (Einzel-) Tabellen, wenn diese nach n verschiedenen Merkmalen (auch hierarchisch) gegliedert sind, mit jeweils nur einer Randsumme für jedes Gliederungskriterium, und von überlappenden Tabellen, wenn diese gewisse Tabellenwerte gemeinsam haben. Als Beispiel für ein hierarchisches Gliederungskriterium kann die regionale Gliederung in Nordrhein-Westfalen dienen, wo in der untersten Hierarchiestufe (niedrigste Aggregation der Werte bezüglich dieser Gliederung) die Gemeinden zu ihren Kreisen beitragen, in der nächsthöheren Stufe die Kreise und kreisfreien Städte zu ihren Regierungsbezirken und wo die Regierungsbezirke schließlich zum Land zusammengefasst werden.

3.1 Quaderverfahren im Falle zwischensummenfreier Einzeltabellen

Kern des Verfahrens zur sekundären Geheimhaltung ist das „reine“ Quaderverfahren. Es sichert n-dimensionale Statistiktabellen, die nicht durch Zwischensummen unterteilt sind, hinreichend gegen zu genaue Rückrechnung der primär geheimen Werte.

3.1.1 Karree-Sicherung in zweidimensionalen Tabellen

Das Quaderverfahren lässt sich am anschaulichsten anhand kleiner zwischensummenfreier zweidimensionaler Tabellen erläutern. Die Abbildung 1 zeigt eine Tabelle dieses Typs: Als Zeilengliederung wurde eine regionale Gliederung angenommen, die Spalten kennzeichnen Wirtschaftsgruppen. In die Tabellenfelder sind die Werte und die Sperrvermerke, p für primär, s für sekundär geheim, eingetragen.

Als erstes erhebt sich für den die Tabelle veröffentlichenden Statistiker die Frage, ob mit diesem Muster von Sperrvermerken ein hinreichender Schutz für die primär geheimen Werte zu garantieren ist. Zur Beantwortung stellt er sich folgendes Vorgehen eines externen Tabellennutzers vor, nach dem dieser die Eingrenzung der geheimen Werte betreibt:

Zur Vereinfachung des Problems wird der Tabellennutzer die gegebene Tabelle in die auf die geheimen Werte reduzierte Form der Abbildung 2 bringen, wobei er die ihm unbekannten geheimen Werte durch die Symbole X_1 , X_2 , X_3 und X_4 ersetzt.

Eine Gesamtheit von Werten einer zweidimensionalen Tabelle, die wie X_1 , X_2 , X_3 , X_4 in den Eckpunkten eines geometrischen Karrees liegen, wird im Folgenden als Karree bzw. auch allgemeiner als zweidimensionaler Quader bezeichnet.

Um dem unbekannten Wert X_1 „näher zu kommen“, eliminiert der Tabellennutzer die restlichen drei Unbekannten X_2 , X_3 , X_4 mit Hilfe der Summenbeziehungen der Tabelle und erhält die Abbildung 3:

Abb. 2

Kreise \ Wirtschaftsgruppen	B	D	Σ reduz
Kreis 1	X_1	X_2	200,0
Kreis 2	X_3	X_4	50,2
Reg.-Bez. reduziert	150,1	100,1	250,2

Abb. 3

Kreise \ Wirtschaftsgruppen	B	D	Σ reduz
Kreis 1	X_1	$200,0 - X_1$	200,0
Kreis 2	$150,1 - X_1$	$50,2 - (150,1 - X_1)$	50,2
Reg.-Bez. reduziert	150,1	100,1	250,2

Wenn dem externen Tabellennutzer keine weiteren Informationen über die Tabellenwerte zugänglich sind, muss er sich mit obiger Tabelle (Abb. 3) zufrieden geben, die geheimen Werte sind dann hinreichend gesichert: Die Lösung des Gleichungssystems zur Bestimmung der unbekannten Werte des Karrees X_1, X_2, X_3, X_4 enthält einen frei wählbaren Parameter – bei dieser „Auflösung“ den Parameter X_1 .

Verfügt der Tabellennutzer aber beispielsweise über das Vorwissen, dass es sich bei dieser Tabelle um eine so genannte positive Tabelle handelt, dass also keine negativen Tabellenwerte zu erwarten sind, so kann er mit diesem Wissen den freien Parameter X_1 weiter eingrenzen:

1. Zeile (Kreis 1):
 $0 \leq X_1 \leq 200 \wedge 0 \leq 200 - X_1 \leq 200$
d. h. $0 \leq X_1 \leq 200$
2. Zeile (Kreis 2):
 $0 \leq 150,1 - X_1 \leq 50,2 \wedge 0 \leq 50,2 - (150,1 - X_1) \leq 50,2$
d. h. $99,9 \leq X_1 \leq 150,1$

Die Eingrenzungen sind die Folge der Positivität der Tabelle, wonach keiner der Tabellenwerte die jeweilige Randsumme übersteigt und außerdem jeder Wert größer oder höchstens gleich Null sein muss.

Die Spaltengleichungen liefern keine weitere Eingrenzung von X_1 . Wenn dem Tabellennutzer keine weiteren Informationen über die Tabellenwerte vorliegen, ist der Parameter X_1 der Lösung des o. g. Gleichungssystems zur Berechnung der Unbekannten X_1, X_2, X_3, X_4 mit dem Schutzintervall

$$99,9 \leq X_1 \leq 150,1$$

„hinreichend ungenau“ festgelegt, die geheimen Tabellenwerte sind hinreichend gesichert.

Um sich davon zu überzeugen, trägt man die Schutzintervallgrenzen des Lösungsparameters in Abbildung 3 ein und erhält:

Abb. 4

Kreise \ Wirtschaftsgruppen	B	D	Σ reduz
Kreis 1	[99,9; 150,1]	[49,9; 100,1]	200,0
Kreis 2	[0,0; 50,2]	[0,0; 50,2]	50,2
Reg.-Bez. reduziert	150,1	100,1	250,2

Eine Folge der Einparametrigkeit der Lösungen für die Karree-Unbekannten ist, dass die Spannweite der Wertebereiche für alle vier Unbekannten die gleiche ist. Die Spannweite der Schutzintervalle in einem Karree ist somit eine Karree-Eigenschaft. Man kann daher die Spannweite, die ein zu schützender geheimer Wert überstreichen soll, als Kriterium für die Auswahl geeigneter Karrees mit bereits gesperrten und/oder noch zu sperrenden Werten zum Schutze des geheimen Wertes benutzen.

Dabei zeigt sich, dass nicht nur Karrees mit Werten von der gleichen Größenordnung wie der zu schützende geheime Wert als Sicherungskarrees in Betracht kommen, sondern auch solche mit sehr viel kleineren Werten. Ja, sogar Nullen können als Sicherungswerte dienen, wenn die Karree-Spannweite nicht zu klein ausfällt.

Ein anderes sehr wichtiges Merkmal, das das Karree in zweidimensionalen Tabellen als Gesamtheit von Tabellenfeldern für die Sicherung geheimer Werte so attraktiv macht, ist seine – ebenfalls mit der Einparametrigkeit der Karree-Lösungen zusammen-

hängende – Minimaleigenschaft: Das Karree als Gesamtheit aller Tabellenwerte in den vier Ecken eines geometrischen Karrees ist die kleinste Gesamtheit von Werten, die als geheime Werte zum Schutze eines zu sichernden Wertes ausreicht.

Diese Minimaleigenschaft lässt sich wie folgt begründen: Als offensichtlichste Schutzmaßnahme muss der Tabellenersteller in der Zeile des zu sichernden Wertes (Pivots) und auch in der Pivot-Spalte einen zusätzlichen geheimen oder noch zu sperrenden Tabellenwert auswählen und ggf. sperren, damit der zu schützende Wert nicht einfach durch Differenzbildung aus der betreffenden Zeilen- bzw. Spaltensumme und den ande-

ren nicht geheimen Tabellenwerten in dieser Zeile bzw. Spalte berechnet werden kann. Damit ist für den Pivot-Wert aber noch kein hinreichender Schutz garantiert, weil u. U. die beiden ausgewählten „Sicherungswerte“ selbst mit Hilfe noch offener Werte und den entsprechenden Summen, zu denen sie beitragen, berechnet werden können. Beide Sicherungswerte lassen sich aber gemeinsam durch Hinzunahme nur eines geheimen oder noch zu sperrenden Wertes hinreichend schützen (wenn sich die Karree-Spannweite als groß genug erweist), das ist die kleinste Anzahl zusätzlicher Sicherungswerte; das Sperrmuster bildet ein Karree.

Als Beispiel betrachte man den primär geheimen Wert 100,0 des Tabellenfeldes (Kreis 1; B) in der Abbildung 1 als Pivot-Element, das durch die Werte 100,0 und 50,1 gegen Rückrechnung mit Hilfe der Zeilen- bzw. Spaltensumme 230,0 bzw. 151,0 und den noch offenen Werten in dieser Zeile bzw. Spalte geschützt wird. Diese beiden „Zusatzsperrungen“ werden durch nur einen, hier bereits gesperrten Wert 0,1 hinreichend gesichert (wenn die Spannweite des Karrees groß genug ist).

Wenn die Sicherheit der primär geheimen Werte gewährleistet ist, hat der veröffentlichende Statistiker zu prüfen, ob die vorgenommene Sicherung die günstigste ist. Dazu benötigt er handliche Kriterien. In der amtlichen Statistik wird angestrebt, in erster Linie so wenig Sperrungen wie möglich vorzunehmen und wenn dann noch Auswahlmöglichkeiten für die Sekundärsperrungen bestehen, diese so zu setzen, dass die Summe der sekundär gesperrten Werte minimal wird.

In der Beispieltabelle wurde die kleinstmögliche Anzahl von Sekundärsperrungen zum einen dadurch erreicht, dass als Sperrmuster ein Karree gewählt wurde und zum anderen durch die Einbeziehung eines bereits primär gesperrten Wertes (0,1) in das Sicherungskarree. Danach gibt es keine weiteren Auswahlmöglichkeiten von Sperrmustern, die ebenfalls nur zwei Sekundärsperrungen umfasst hätten; der die Tabelle der Abbildung 1 veröffentlichende Statistiker kann zufrieden sein.

Aus dieser ganzen Betrachtung lässt sich folgendes Vorgehen bei der Sicherung zweidimensionaler nicht durch Zwischensummen unterteilter Tabellen ableiten:

1. Für jeden primär geheimen Wert wird ein Sicherungskarree so ausgewählt, dass dieses möglichst viele bereits gesperrte Werte einbezieht; stehen dann noch mehrere Quader zur Auswahl, ist derjenige mit der kleinsten Summe noch zu sperrender Werte als Sicherungsquader zu bevorzugen. – Solches Vorgehen ist bereits ein heuristisches, weil in Einzelsicherungsprozessen Primärsperrung für Primärsperrung abgearbeitet wird und nicht alle Primärsperrungen gemeinsam, was zu Übersperrungen führen kann. Das Quaderverfahren ist eben nur bei voll besetzten n -dimensionalen Tabellen mit höchstens zwei zu sichernden Werten in Bezug auf die Anzahl der Sekundärsperrungen wirklich optimal. – Sind keine Vorinformationen zu berücksichtigen, ist die Tabelle nach diesem Arbeitsgang bereits hinreichend gesichert.

2. Muss beim Tabellennutzer ein Vorwissen über die Tabellenwerte unterstellt werden, mit dem er die geheimen Werte eingrenzen kann, wie z. B. das Vorwissen, dass es sich bei der vorliegenden Tabelle um eine positive Tabelle handelt, so ist obige Karree-Auswahl auf Karrees mit einer genügend großen Spannweite beschränkt. Um die richtige Auswahl treffen zu können, gibt sich der für die Sicherung der Tabellendaten Verantwortliche eine relative Mindestspannweite vor. Sicherungs-Karrees sind dann so auszuwählen, dass ihre auf den zu schützenden geheimen Wert bezogene Spannweite größer als die vorgegebene relative Mindestspannweite ist. – Wenn gemäß Punkt 1 mehrere, u. U. auch einander sich überdeckende Karrees in einer Tabelle vorkommen, werden die Karree-Spannweiten immer unabhängig von den anderen Sicherungskarrees bestimmt. Dieses Vorgehen ist legitim, weil eine als Karree-Spannweite ermittelte Spannweite eines geheimen Wertes durch die Sicherung anderer geheimer Werte derselben Tabelle unter Beibehaltung des Vorwissens niemals weiter eingengt werden kann.

3.1.2 Anmerkungen zur Berücksichtigung von Vorwissen

Für den Schutz primär geheimer Werte mittels sekundärer Geheimhaltung ist es – unabhängig vom eingesetzten Sperrverfahren – äußerst wichtig, dass die für die Auswahl des Sperrmusters festgelegte relative Mindestspannweite dem Tabellennutzer keinesfalls bekannt wird! Die Kenntnis der relativen Mindestspannweite ist als eine spezielle Art des Vorwissens zu werten, das, wie jedes Vorwissen über die Tabellenwerte, das Problem der sekundären Geheimhaltung wesentlich verschärft.

Jedes Sekundärsperrverfahren, auch die oben beschriebene Heuristik, ist so angelegt, dass der externe Tabellennutzer die zu schützenden Tabellenwerte nur bis auf ein Schutzinter-

vall genau eingrenzen kann. Der für den Schutz geheimer Werte Verantwortliche muss also unterstellen, dass die unteren und die oberen Schutzintervallgrenzen allgemein bekannt sind. Ob der betreffende gesperrte Wert primär geheim ist, wird meistens schon durch die veröffentlichte Anzahl der Berichtenden ($n \leq 2$) deutlich.

Kennt der Tabellennutzer darüber hinaus noch den Auswahlparameter relative Mindestspannweite, weiß er also, dass die ihm über die Schutzintervallgrenzen bekannte Spannweite eines primär geheimen Wertes größer als das Produkt aus bekannter relativer Mindestspannweite und unbekanntem geheimen Wert ist, so weiß er dadurch auch, dass der geheime Wert kleiner als die gegebene Spannweite dividiert durch die ebenfalls gegebene relative Mindestspannweite sein muss. Wenn dann diese obere Schranke des betreffenden geheimen Wertes kleiner als seine obere Schutzintervallgrenze ist, hat der Tabellennutzer das Schutzintervall u. U. stärker eingegrenzt, als es von dem die Tabelle veröffentlichenden Statistiker verantwortet werden kann.

Die Gefahr der Offenlegung bzw. näherungsweise Offenlegung wird bereits durch das Beispiel des Sicherungskarrees der Abbildung 1 sehr deutlich, wenn man hier unterstellt, dass die maßgeblich zur Auswahl gerade dieses Karrees beitragende relative Mindestspannweite 50 % beträgt. Mit diesem Vorwissen kann ein externer Tabellennutzer beispielsweise den primär geheimen, durch das Schutzintervall [99,9; 150,1] gesicherten Wert X_1 (vgl. linkes oberes Tabellenfeld in Abbildung 3 und 4) bis auf das Intervall [99,9; 100,4] eingrenzen: Die obere Schranke dieses neuen Intervalls erhält er durch Division der Schutzintervalllänge $150,1 - 99,9 = 50,2$ durch die relative Mindestspannweite 0,5. Das ursprüngliche Schutzintervall [99,9; 150,1] lässt sich mit Kenntnis der relativen Mindestspannweite hier soweit eingrenzen, dass die relative Spannweite des primär geheimen Wertes weit unter der zu tolerierenden relativen Mindestspannweite zu liegen kommt (sie ist kleiner als $(100,4 - 99,9)/100,0 = 0,5 \%$); der ge-

heime Wert ist damit näherungsweise offengelegt, was deutlich zeigt, dass man die relative Mindestspannweite geheim halten muss.

Alle bisher besprochenen Sicherungsmaßnahmen bleiben höchst unbefriedigend, wenn man bedenkt, dass die Nutzer von Statistiktabellen aufgrund ihrer Erfahrung von Berufs wegen sehr viel mehr über „ihre“ Tabellen wissen als beispielsweise nur, dass sie keine negativen Werte enthalten. In der Regel kann man davon ausgehen, dass die Tabellenwerte zumindest auf plus minus 50 % genau bekannt sind. Dann reicht aber ein nur für positive Tabellen hergeleiteter Intervallschutz nicht mehr aus, wie es die Beispieltabelle der Abbildung 1 belegt:

Zur Eingrenzung der geheimen Werte der Abbildung 1 geht der externe Tabellennutzer wieder über Abbildung 2 zur Abbildung 3 über, genau nach demselben Muster wie unter Punkt 3.1.1 beschrieben. Da der Tabellennutzer nun außerdem noch Schätzintervalle kennt, die die geheimen Werte überdecken, ersetzt er in Abb. 2 die unbekannten Werte durch ihre Schätzintervalle (Abb. 5) und stellt diese Tabelle der in Abb. 3 gegenüber.

Abb. 5

Kreise \ Wirtschaftsgruppen	B	D	Σ reduz
Kreis 1	[50; 150]	[50; 150]	200,0
Kreis 2	[25; 75]	[0,05; 0,15]	50,2
Reg.-Bez. reduziert	150,1	100,1	250,2

Betrachtet man Abb. 1 wieder als positive Tabelle, so gelten folgende Abschätzungen:

1. Zeile:

$$50 \leq X_1 \leq 150 \quad \wedge \\ 50 \leq 200,0 - X_1 \leq 150 \\ \text{d. h. } 50 \leq X_1 \leq 150$$

2. Zeile:

$$25 \leq 150,1 - X_1 \leq 50,2 \quad \wedge \\ 0,05 \leq 50,2 - (150,1 - X_1) \leq 0,15 \\ \text{d. h. } 99,95 \leq X_1 \leq 100,05$$

Der primär geheime Wert X_1 kann vom externen Tabellennutzer demnach bis auf eine relative Spannweite von 0,1 % genau eingegrenzt werden, er ist also zu genau bestimmbar.

Die Ursache für die sehr genaue Offenlegung des geheimen Wertes X_1 liegt in der Kleinheit des Sicherungspartners $X_4 = 0,1$ (vgl. Abb. 1 und Abb. 2) begründet und darin, dass die Schätzfehlergrenzen aus dem einheitlichen relativen Fehler (von 50 %) abgeleitet wurden. Dennoch kommen auch sehr kleine Tabellenwerte und sogar Nullen als Sperrpartner zur Sicherung geheimer Werte in Betracht, wenn nur ihre geschätzten Spannweiten nicht wesentlich kleiner als die der zu schützenden Werte sind. Ganz allgemein müssen nur zu unterstellende Schätzintervalle in die Berechnung der Quaderspannweite einbezogen werden, um auch sie gemäß 3.1.1, Regel 2 als Vorwissen bei der Quaderauswahl einzubringen (siehe Anhang).

Eine ganz andere Qualität des Vorwissens wird durch die so genannten Einzelangaben in die sekundäre Geheimhaltung eingetragen. Hier hat man zu berücksichtigen, dass der die Einzelangabe Berichtende selbst zum Kreis der Tabellennutzer gehört und somit seine Einzelangabe in der Tabelle wiedererkennen kann. Eine Einzelangabe darf daher nicht als Sicherungspartner in einem Karree zum Schutze eines geheimen Wertes dienen, es sei denn,

es wird noch ein weiteres Karree zur Sicherung desselben Pivot-Wertes herangezogen, dessen Einzelangaben mit Ausnahme des Pivot's selbst im ersten Karree nicht vorkommen.

Diese als Doppelquadersicherung bezeichnete Sicherung von geheimen Werten wird man immer dann vornehmen, wenn dadurch Sperrungen in die Randsummen vermieden werden können. Das ist insbesondere dann der Fall, wenn alle „Aggregate“ zur selben Randsumme Einzelangaben sind, d. h. wenn z. B. eine Zeile oder Spalte in einer zweidimensionalen Tabelle mit Ausnahme der jeweiligen Randsumme nur mit Einzelangaben belegt ist.

Eine Ausnahme von der Doppelquadersicherung tritt bei schwach besetzten Tabellen auf, wenn sich kein zweiter (z. B. zweidimensionaler) Quader finden lässt, dessen Einzelangaben im ersten nicht vorkommen, das Pivot ausgenommen. Dies ist immer dann der Fall, wenn der zu schützende Tabellenwert eine Einzelangabe ist, die gleichzeitig auch in einer der Randsummen auftritt (weil sonst nur leere Tabellenfelder zu dieser Randsumme beitragen). In diesem Falle kann man die Einzelangabe im Rand aber wie eine geheime Angabe von mehr als einem Berichtenden behandeln, ohne dabei eine Einbuße an Geheimhaltung zu riskieren: Diejenige Gliederung bei der eine (geheime) Einzelangabe in der Randsumme liegt, trägt nichts zur Offenlegung beider Einzelangaben bei.

3.1.3 Verallgemeinerung des Quaderverfahrens

Die gemäß 3.1.1 mit der Karreesicherung zu behandelnden zweidimensionalen Tabellen stellen einen zwar häufig anzutreffenden Anwendungsfall für die Sicherung aggregierter Daten mit dem Quaderverfahren dar, oft werden aber auch Tabellen veröffentlicht, die nach mehr als zwei qualitativen Merkmalen gegliedert sind, z. B. nach regionaler und nach wirtschaftlicher Systematik und gleichzeitig nach Betriebsgrößenklassen und nach Rechtsformen der Betriebe. Das Karree-Verfahren ist daher für Anwendungen auf n-dimensionale Tabellen, d. h. zu einem n-dimensionalen Quaderverfahren zu verallgemeinern. – In diesem Abschnitt werden ausschließlich Tabellen betrachtet, die bezüglich jeder Gliederung nur eine Randsumme und keine Zwischensummen aufweisen, es werden also zunächst nur zwischensummenfreie n-dimensionale Tabellen bearbeitet.

Die für die Sicherung von aggregierten Daten n-dimensionaler Tabellen mit dem Quaderverfahren fundamentale Definition ist die des n-dimensionalen Quaders:

Ein n-dimensionaler Quader ist die Gesamtheit aller Tabellenwer-

te, die durch zwei zueinander diametrale Tabellenwerte so festgelegt ist, dass die Ausprägung jedes Gliederungsmerkmals dieser Tabellenwerte die eines der beiden Diametralwerte ist. Zwei Tabellenwerte heißen zueinander diametral, wenn sich die Ausprägungen der Gliederungsmerkmale dieser Werte in jeder Gliederung voneinander unterscheiden.

Da jede Gliederungsausprägung eines Quaderwertes somit einen von zwei Werten annimmt, die Gliederungsausprägung eines der beiden Diametralwerte, umfasst ein n-dimensionaler Quader 2^n Tabellenwerte. Das entspricht genau der Erwartung, wenn man Tabellen betrachtet, die nach bis zu drei Merkmalen gegliedert sind:

Im Falle einer eindimensionalen Tabelle wird nur ein Partnerwert zur Sicherung eines geheimen Wertes benötigt; der eindimensionale Sicherungsquader umfasst zwei = 2^1 Tabellenwerte. In zweidimensionalen Tabellen kommen zu jedem der beiden Werte, Pivot plus Partner zur Sicherung gegen Rückrechnung in der ersten Gliederung, noch zwei Werte zur Sicherung bezüglich des zweiten Gliederungsmerkmals hinzu; das Karree umfasst vier = 2^2 Tabellenwerte. Bei drei Gliederungsmerkmalen muss man das Karree in einer der zweidimensionalen Tabellen bezüglich der ersten beiden Gliederungen in eine andere nach den ersten beiden Gliederungsmerkmalen gegliederte Tabelle projizieren, deren drittes Gliederungsmerkmal sich von dem der erstgenannten Tabelle unterscheidet; es kommen also vier weitere Tabellenwerte hinzu. Der dreidimensionale Quader besteht demnach aus $8 = 2^3$ Tabellenwerten, was genau der Anschauung entspricht.

Man erkennt daraus das bei fortlaufender Erhöhung der Tabellendimension wirkende Bildungsgesetz: Bei jeder Hinzunahme eines weiteren Gliederungsmerkmals verdoppelt sich die Anzahl der Quaderwerte, so dass bei Hinzufügen nur einer weiteren Gliederung zu einer n-dimensionalen Tabelle noch ein weiterer n-dimensio-

naler Quader aufgebaut werden muss, der sich vom ersten nur in der Indizierung durch das neu hinzugefügte Gliederungsmerkmal unterscheidet. Das ist zugleich auch die kleinste Anzahl von neu hinzuzunehmenden Tabellenwerten, weil für jeden Quaderwert der n-dimensionalen Tabelle mindestens ein weiterer geheimer Wert der n+1-dimensionalen Tabelle allein schon zum Schutz gegen Rückrechnung mit Hilfe der neu hinzugenommenen Gliederung eingebracht werden muss. Der Quader der n+1-dimensionalen Tabelle umfasst daher $2^n + 2^n = 2^{n+1}$ Quaderwerte.

Da Sperrmuster mit Quaderstrukturen für geheime zu sichernde Tabellenwerte einen hinreichenden Schutz gegen zu genaues Rückrechnen bieten (vgl. „Wahrung der Geheimhaltung sensibler Daten in mehrdimensionalen Tabellen mit dem Quaderverfahren“, R. D. Reipsilber, Statistische Analysen und Studien Nordrhein-Westfalen 3/2000), hat man mit der Quadersicherung zugleich auch nur die kleinste Anzahl von Partnerwerten auszuwählen, die als gesperrte Werte jeden zu schützenden geheimen Tabellenwert hinreichend sichern. Damit stellt sich das Quaderverfahren als das Sicherungsverfahren heraus, das von allen Verfahren zur sekundären Geheimhaltung aufgrund der kleinsten Anzahl auszuwählender Sicherungspartner das schnellste ist, d. h. dass das Quaderverfahren bereits prinzipiell die kürzeste Rechenzeit benötigt.

Um noch die für die Festlegung eines n-dimensionalen Quaders ausschlaggebende Indizierung der Quaderwerte durch die Ausprägungen der Gliederungsmerkmale zueinander diametraler Tabellenwerte zu veranschaulichen, sei auf die Beispieltabelle der Abbildung 1 verwiesen: Darin ist ein zu schützendes Pivotelement durch die Gliederungsausprägungen Kreis 1 als Zeilenindex und die Wirtschaftsgruppe B als Spaltenindex festgelegt. Ein zu diesem Pivot diametraler Tabellenwert ist der ebenfalls primär geheime Wert 0,1, der durch Kreis 2 und Wirtschaftsgruppe D indiziert ist. Lt. obiger Definition erkennt man die Diametralität daran, dass

sich sowohl die Gliederungsausprägung der Zeile als auch die der Spalte von den entsprechenden Gliederungsausprägungen des Pivots unterscheiden.

Die Gesamtheit der Quaderwerte findet man nun, indem man alle 2^2 Paare von Gliederungsausprägungen aufsucht, die bezüglich der beiden Tabellengliederungen entweder das primär geheime Pivot 100,0 oder den dazu diametralen Tabellenwert 0,1 indizieren. Das sind genau die Tabellenwerte mit den Indexpaaren (Kreis 1; Wgr. B), (Kreis 2; Wgr. B), (Kreis 1; Wgr. D), (Kreis 2; Wgr. D), wie es auch die anschauliche Karree-Definition nahe legt.

Mit der Definition eines n-dimensionalen Quaders ist es nunmehr möglich, das Sperrmuster zur Sicherung jedes einzelnen geheimen Wertes in einer n-dimensionalen, nicht durch Zwischensummen unterteilten Tabelle durch folgende Regelung festzulegen:

Ein durch n Gliederungsmerkmale indizierter geheimer Tabellenwert wird gegen zu genaues Rückrechnen mit Hilfe der Tabellensummenbeziehungen und beim Nutzer zu unterstellendem Vorwissen gesichert, indem ein n-dimensionaler Quader ausgewählt wird mit folgenden Eigenschaften (siehe auch Anhang),

- **der zu sichernde Tabellenwert (Pivot) ist keine Einzelangabe einer Randsumme (solche Einzelangaben werden durch ihre Einzelangaben im Inneren der Tabelle mittels Quader geschützt); das Pivot ist einer der beiden diesen Quader fixierenden diametralen Werte;**
- **die auf den Pivot-Wert bezogene Quader-Spannweite, ist größer als eine vorgegebene relative Mindestspannweite;**
- **der Quader ohne Pivot umfasst, außer im Falle von Randsummenwerten, keine Einzelangaben oder es wird zu demselben Pivot noch ein zweiter Quader ausgewählt, der – mit Ausnahme des Pivots selbst – keine Einzelangaben des ersten enthält;**

- **der Quader umfasst möglichst viele bereits gesperrte Tabellenwerte;**
 - **die noch offenen zu sperrenden Werte liegen nach Möglichkeit im Inneren der Tabelle, d. h. sie gehören keiner Randsumme an;**
 - **die Summe der noch offenen, ggf. auch gewichteten Tabellenwerte soll so klein wie möglich sein;**
- Anschließend werden alle noch offenen Quaderwerte gesperrt.**

Mit dieser Regelung werden alle im Punkt 3.1.1 besprochenen Sicherungsmaßnahmen auf n-dimensionale Tabellen übertragen. Die einzige noch störende Beschränkung der allgemeinen Anwendbarkeit des Quaderverfahrens betrifft die in diesem Absatz zugrunde gelegte Voraussetzung der zwischensummenfreien Tabellen. Wie man die in der „statistischen Praxis“ zu behandelnden oft mehrfach durch Zwischensummen unterteilten Tabellen in zwischensummenfreie überführen kann, zeigt der nachfolgende Abschnitt.

3.2 Das Quaderverfahren in der Anwendung

3.2.1 Aufteilung einer gegebenen Tabelle in Untertabellen

Tabellen, deren Werte und Merkmals-trägerzahlen bezüglich jeder Gliederung zu nur einer Randsumme aufaddiert werden, sind in (mehrfach) durch Zwischensummen unterteilten so genannten hierarchisch gegliederten Tabellen nur als Teilgesamtheiten realisiert, sie werden im Folgenden als Untertabellen bezeichnet. Untertabellen sind also Teile der gegebenen Statistiktabelle, die bezüglich jedes Gliederungsmerkmals genau eine Randsumme und sonst keine weiteren Zwischensummen aufweisen. Jede Untertabelle, für sich alleine betrachtet, erfüllt demnach genau die Voraussetzungen für die Anwendbarkeit des Quaderverfahrens.

Zur Veranschaulichung der Untertabellen-Aufteilung und deren Gliederungen genügt zunächst die Betrachtung

nur eines Gliederungsmerkmals, z. B. die regionale Gliederung Nordrhein-Westfalens, die sich von den Gemeinden über die Kreise und kreisfreien Städte zu den Regierungsbezirken bis zum Land erstreckt: Bezüglich dieser Gliederung lassen sich die Untertabellen des untersten Aggregationsniveaus aussondern, indem alle Gemeinden zusammengefasst werden, die zu einem bestimmten Kreis gehören und indem man dazu noch die Kreisergebnisse als Randsummen hinzufügt. So können für jeden Kreis alle Untertabellen aufgesucht werden, die sich aufgrund der anderen Gliederungen ergeben.

Die Untertabellen der bezüglich der regionalen Gliederung nächst höheren, der zweiten Aggregationsstufe findet man durch Aussonderung aller Ergebnisse, die die zu einem speziellen Regierungsbezirk gehörigen Kreise und kreisfreien Städte beitragen zuzüglich der Regierungsbezirksergebnisse als Randsummen in dieser Gliederung. Schließlich können noch alle Regierungsbezirksergebnisse und die Landesergebnisse von der gegebenen Statistiktabelle als Untertabellen abgetrennt werden. Auch hier können mehrere Untertabellen entstehen, weil die anderen Gliederungen möglicherweise auch hierarchisch strukturiert sind.

Bei der Abtrennung der Untertabellen von der Gesamttabelle wird nach dem Muster der regionalen Gliederung bei allen Gliederungen immer nur eine Aggregationsstufe zusammen mit den zugehörigen Aggregaten der nächst höheren Stufe als Randsummen aussortiert. Die so erhaltene Gesamtheit von Untertabellen bildet bezüglich der Aggregationsniveaus eine Hierarchie voneinander abhängiger Untertabellen. Diese Hierarchie wird zweckmäßig Untertabelle für Untertabelle nach absteigenden Hierarchiestufen mit dem Quaderverfahren behandelt, bis alle Untertabellen erfasst worden sind.

Eine Bearbeitung nach absteigenden Aggregationsstufen bietet sich insofern an, als jede Sperrung in einer Untertabelle höherer Hierarchiestufe immer auch Sperrungen in den zugehörigen

Untertabellen niedrigerer Aggregation bedeuten, die dann in diesen Untertabellen mit dem Quaderverfahren zu sichern sind.

Bei schwach besetzten Tabellen genügt jedoch die einmalige Abarbeitung aller Untertabellen der Gesamtheit nach absteigenden Hierarchiestufen meist nicht, weil in den Untertabellen der niedrigeren Hierarchiestufen oft Sperrungen in den Randsummen vorgenommen werden müssen, die als Sperrungen im Inneren der zugehörigen Untertabelle höherer Hierarchie auftreten und dann dort zu sichern sind. Die betroffenen höher aggregierten Untertabellen werden aber beim Abarbeiten der Untertabellengesamtheit nach abnehmenden Aggregationsniveaus nicht mehr erreicht, so dass ein iteratives Vorgehen angezeigt ist, bei dem die Untertabellengesamtheit so lange immer wieder vollständig durchlaufen wird, bis bei einem vollen Durchlauf der Gesamtheit keine neuen Sperreintragungen mehr vorgenommen wurden.

3.2.2 Aufstockung der Tabellendimension

Mit der Aufteilung der gegebenen hierarchisch gegliederten Statistiktabelle in eine Gesamtheit von Untertabellen, die dann iterativ abgearbeitet wird, wobei das Quaderverfahren alle Untertabellen der Gesamtheit hinreichend gegen zu genaues Rückrechnen sichert, ist eine hinreichende Sicherung der Gesamttabelle nicht zu machen. Dies widerlegen hierarchisch gegliederte Gegenbeispiel-Tabellen, bei denen jede Untertabelle für sich betrachtet, mit irgendeinem Geheimhaltungsverfahren zwar hinreichend gesichert werden kann und bei denen auch alle Untertabellen nach obigem iterativen Verfahren aneinander abgeglichen worden sind, die dennoch Rückrechnungen über mehrere Untertabellen hinweg erlauben.

Die Ursache für diese Geheimhaltungslücke liegt in der Aufteilung des Geheimhaltungsproblems in viele Teilprobleme. Diese Sicherungsmaßnahmen dürfen eben nicht unabhängig voneinander durchgeführt werden

und zwar selbst dann nicht, wenn alle Untertabellen so aneinander abgeglichen werden, dass jeder in mehreren Untertabellen auftretende Tabellenwert in all diesen Untertabellen den gleichen Geheimhaltungsstatus besitzt.

Nun gibt es aber noch eine andere Möglichkeit, die durch Zwischensummen unterteilte Statistiktafel zum Zwecke der Anwendung des Quaderverfahrens in eine zwischensummenfreie Tabelle zu überführen, die Aufstockung der Tabellendimension. Dazu führt man für jedes Aggregationsniveau jedes Gliederungskriteriums eine eigene Dimension ein, die zur Unterscheidung von den ursprünglichen Gliederungen als Aggregatdimension bzw. als Aggregatgliederung bezeichnet werden soll. Die nach Aggregatdimensionen gegliederte Tabelle enthält keine Zwischensummen mehr, derentwegen man eine Zerlegung in Untertabellen vornehmen müsste, sie kann daher mit dem Quaderverfahren hinreichend gesichert werden.

Die regionale Gliederung Nordrhein-Westfalens beispielsweise umfasst die Aggregatgliederung der Kreise nach ihren Gemeinden, die Aggregatgliederung der Regierungsbezirke nach ihren Kreisen und kreisfreien Städten sowie die Aggregatgliederung des Landes nach den Regierungsbezirken. Allein für diese eine hinsichtlich der Aggregation dreistufige regionale Gliederung werden also drei neue Gliederungen eingebracht. Die Aufstockung der Tabellendimension ergibt i. d. R. hochdimensionale Tabellen, die so genannten vollständigen Tabellen, die außerdem noch durch strukturelle Tabellenfelder und zusätzliche Summenfelder zu ergänzen sind. Um in vollständigen Tabellen zu einem zu sichernden geheimen Wert einen geeigneten Sicherungsquader zu finden, muss demnach ein sehr großer und vor allem hochdimensionaler Tabellenraum abgetastet werden, mit der Konsequenz sehr großer Rechenzeiten für die Sicherung der Tabelle.

Einen Ausweg aus dieser „Rechenzeitfalle“ bieten Zwischensummen in der Originaltafel, die keine geheimen Quaderwerte zur Sicherung eines geheimen Wertes enthalten; sie

koppeln das Geheimhaltungsproblem des gerade zu bearbeitenden Pivots von dem der Sicherung der Gesamttabelle gewissermaßen ab. Zur Sicherung des Pivots genügt die Bearbeitung desjenigen Tabellenteils, der durch Zwischensummen von der gegebenen Tabelle abgegrenzt wird, die nicht mit Sicherungsquaderwerten dieses Pivots belegt werden müssen.

Da man aber von vorneherein nicht wissen kann, auf welche Zwischensummen sich die Quader eines zu sichernden Pivots noch erstrecken müssen, wird bei der Quaderauswahl mit einem Tabellenteil niedrigster Aggregationsstufen begonnen und weitere über Zwischensummen anzukoppelnde Tabellenteile erst hinzugenommen, wenn Sicherungsquader nicht ohne diese Zwischensummen auskommen (z. B. zu geringe Besetzung). Diese Tabellenteile sind dann in Bezug auf ihre Dimensionen temporär aufzustocken und in diesem so erweiterten Tabellenraum wird nach einem geeigneten Quader gesucht, wodurch der abzutastende Raum erheblich eingegrenzt werden kann.

Derzeit entwickelt das LDS NRW das EDV-Programm QUIT (**Q**uaderverfahren **i**terativ), das die Abgrenzung von Teiltabellen zur Reduzierung der Rechenzeit ausnutzt, um so auch umfangreichere Tabellen noch hinreichend sichern zu können. Erste Probelaufe haben dabei ergeben, dass die Quaderauswahlmöglichkeiten noch erheblich weiter eingegrenzt werden müssen. Ein ganz wesentlicher Rechenzeitgewinn konnte bei höher aggregierten Pivot-Elementen durch Rückverfolgung der „Spuren“ des Pivots über die zu ihm beitragenden niedriger aggregierten primär geheimen Werte erreicht werden, indem die Auswahl nur auf Quader beschränkt wurde, die alle zum Pivot beitragenden primär geheimen Werte enthalten. Damit ist es bereits gelungen, die sehr umfangreiche Statistik des steuerbaren Umsatzes (etwa 700 000 Tabellenfelder) zu bearbeiten.

Nach ersten Erfahrungen mit diesem Konzept trifft man beim Aufsuchen von solchen hochdimensionalen Quadern immer wieder auf offene nicht

sperrbare Tabellenwerte wie strukturelle Nullen, Tabellenwerte mit zu kleinen Schätzintervallen etc., was letzten Endes zu Sperrungen in hochaggregierten Randsummen führt. Es konnte aber inzwischen gezeigt werden (vergleiche dazu auch o. g. Aufsatz in Statistische Analysen und Studien, S. 56), dass die Starrheit der Quaderstruktur zu Gunsten einer optimaleren Auswahl von Sperrkandidaten teilweise aufgegeben werden kann, ohne dabei auf Sicherheit und auf für die Quaderauswahl wichtige Symmetrieeigenschaften verzichten zu müssen. Solche erlaubten Quaderdeformationen sind im EDV-Programm QUIT bereits realisiert und führen durch die Reduktion der Quaderdimension zu erheblichen Rechenzeiteinsparungen.

Es sei erwähnt, dass die Aufstockung der Tabellendimension durch Umstrukturierung der Gliederungen (Umschlüsselung) auch mit Hilfe vorhandener Dienstprogramme möglich ist. Eine damit erhaltene vollständige Tabelle kann bereits jetzt schon mit dem Quaderverfahren hinreichend gesichert werden. Dieses Vorgehen bleibt aber wegen der mit der Dimensionsaufstockung verbundenen enormen Aufblähung der Tabelle bisher auf nicht zu große Tabellen (ca. 50 000 Felder in der gegebenen Tabelle) beschränkt.

3.2.3 Überlappende Statistiktabellen

In der „statistischen Praxis“ hat man es häufig mit mehreren – auch mehrfach durch Zwischensummen unterteilten – Einzeltabellen zu tun, die einander überlappen, d. h. die gewisse Aggregate gemeinsam haben. So kann beispielsweise der regional gegliederte steuerbare Umsatz einmal nach Rechtsformen der Betriebe, ein anderes Mal nach Beschäftigtengrößenklassen oder auch nach wirtschaftlicher Systematik „heruntergebrochen“ werden. Gemeinsam haben diese Tabellen die nur regional gegliederte Summentabelle.

Bei der Tabellensicherung träte hier kein neues Problem auf, wenn es ge-

länge, die Überlappungsbereiche beim Sperren von Werten zu meiden. Untersuchungen im Zusammenhang mit der Geheimhaltung der Handwerkszählung 1955 haben aber gezeigt, dass Sperrungen im Überlappungsbereich prinzipiell nicht auszuschließen sind. Daraus ergibt sich als zwingende Notwendigkeit, dafür zu sorgen, dass mehreren Einzeltabellen angehörende Aggregate in allen diesen Einzeltabellen den gleichen Geheimhaltungsstatus haben. Für obiges Beispiel bedeutet das, dass die nur noch regional gegliederten Gesamtsummenwerte in jeder der drei Einzeltabellen, in der nach Rechtsformen, der nach Beschäftigtengrößenklassen und der nach wirtschaftlicher Systematik gegliederten Tabelle, zumindest den gleichen Geheimhaltungsvermerk tragen.

Für die Sicherung solcher voneinander abhängiger Tabellen kommt daher nur eine gemeinsame Bearbeitung durch gegenseitigen Abgleich in Frage, ganz analog zum Abgleich der Untertabellen. Das bedeutet, dass alle zu einem Pool aneinander abzugleichender Tabellen gehörenden Einzeltabellen i. d. R. gleichzeitig veröffentlicht werden. Eine später erstellte Veröffentlichungstabelle, die Überlappungen mit vorhergehenden, bereits veröffentlichten Tabellen hat, kann nur dann gesichert werden, wenn der Abgleich mit allen in Frage kommenden „Vorgängertabellen“ ausschließlich Sperrungen in der „Nachzüglertabelle“ hervorbringt, sonst nicht.

Auch für den Abgleich von zu einem Pool zusammengefassten hierarchisch gegliederten Einzeltabellen sind im LDS NRW EDV-Programme erstellt worden. Das modernste davon ist GHMITER.22 (**G**eh**ei**m**h**altung **i**terativ); es ist sowohl für den Einsatz auf dem Großrechner von IBM unter VS wie auch für PC unter Windows geeignet und steht allen Anwendern der statistischen Ämter des Bundes und der Länder zur Verfügung. In diesem EDV-Programm ist die Überführung der hierarchisch gegliederten Einzeltabellen noch durch die Aufteilung in Untertabellen realisiert. Die Sicherung der primär gehei-

men Werte erfolgt daher in zwei iterativen Prozessen, der „inneren Iteration“ zum Abgleich der Untertabellen und der „äußeren Iteration“ zur Erreichung eines einheitlichen Geheimhaltungsstatus aller mehreren Einzeltabellen gemeinsamen Tabellenwerte.

Das derzeit in Entwicklung befindliche EDV-Programm QUIT (**Q**uaderverfahren zum **i**terativen Abgleich überlappender Tabellen), verzichtet auf eine innere Iteration und arbeitet mit der Aufstockung der Tabellendimension. Dieses Verfahren ist sehr rechenintensiv und wird dazu führen, dass man bei großen Tabellen Zerlegungen in größere Teiltabellen vornehmen muss, die dann wieder als Einzeltabellen in einer akzeptablen Zeit bearbeitet und in einem äußeren Iterationsprozess aneinander abgeglichen werden können. Dazu wird QUIT mit den gleichen Abgleichsmechanismen arbeiten, wie das von ihm wegen des damit erreichbaren höheren Sicherheitsstandards zu ersetzende GHMITER. QUIT steht voraussichtlich Ende 2002 der amtlichen Statistik zur Verfügung.

4. Erweiterung der Anwendung

4.1 Ausgabe von Schutzintervallgrenzen

Im Falle der in der amtlichen Statistik häufig auftretenden sehr fein gegliederten Tabellen werden aufgrund der mit der Feinheit der Tabellengliederung einhergehenden dünnen Besetzung sehr viele Sperrungen eingetragen, die in der Veröffentlichungstabelle dann beispielsweise als Sternchen erscheinen. Eine derart von Schutzsternchen perforierte Veröffentlichungstabelle mindert die Akzeptanz solcher Statistikerhebungen, wobei man allerdings auch nicht bereit ist, zwecks Einsparung von Sperrungen die Tabellengliederung zu vergrößern.

Einer inzwischen häufiger geäußerten Anregung folgend, kann man anstelle der Schutzsternchen für jeden geheimen Tabellenwert seine Schutz-

intervallgrenzen eintragen, die jeder Tabellennutzer mit Hilfe von linearen Optimierungsprogrammen selbst errechnen könnte, wenn diese nicht so aufwändig in der Handhabung wären. Um den Tabellennutzer zu unterstützen, hat das Statistische Bundesamt ein EDV-Programm in Auftrag gegeben, das solche Schutzintervallgrenzen auf der Basis der linearen Optimierung berechnet.

Solche linearen Optimierungen sind sehr rechenzeitaufwändig – woraus sich ja gerade die Notwendigkeit zur Verwendung von Näherungsverfahren in der sekundären Geheimhaltung ergibt –, so dass sich auch hier ein Ausweichen auf das Quaderverfahren anbietet. Wie in Abschnitt 3 beschrieben, wird für die Quaderauswahl die Quaderspannweite berechnet. Außerdem werden dabei gleichzeitig auch die Schutzintervallgrenzen jedes Quaderwertes bestimmt. Durch diese Intervallgrenzen werden für jeden Quaderwert Schutzintervalle beschrieben, die ein externer Tabellennutzer durch den Einsatz von linearen Optimierungsverfahren nicht weiter eingrenzen kann. Andererseits erfolgt die Quaderauswahl so, dass die Schutzintervalllänge, die Quaderspannweite, für den Schutz der geheimen Werte gegen zu genaues Rückrechnen ausreicht.

Es erscheint somit zweckmäßig, die Schutzintervallgrenzen abzuspeichern und nach der Bearbeitung mit dem Geheimhaltungsprogramm auszugeben; sie können dann anstelle der bisher für die gesperrten Werte eingesetzten Schutzsternchen ausgedruckt werden. Der Tabellennutzer erhält dadurch eine zusätzliche Information über den Bereich, den ein gesperrter Tabellenwert überdecken kann, die die Datensicherung akzeptabler macht, ohne dabei den erforderlichen Datenschutz in Frage zu stellen. Dieser Komfort kann allerdings erst durch QUIT geboten werden und nicht schon mit GHMITER, wenn letzteres noch mit Untertabellenabgleich arbeitet (Geheimhaltungslücke, siehe 3.2.2).

4.2 Quaderverfahren zur Sicherung von Einzelangaben

Die Beantwortung der Frage, was denn das Primäre der Geheimhaltung von Tabellendaten eigentlich sei, nämlich der Schutz von Einzelangaben gegen zu genaues Rückrechnen, d. h. der Intervallschutz der Einzelwerte, führt zu der Forderung, bei der Sicherung von Tabellendaten bereits das Einzelmateriale in den Sperrprozess des Quaderverfahrens mit einzubeziehen. Mit Hilfe der Quadersicherung der Einzelwerte lässt sich erreichen, dass kein Einzelwert genauer berechnet werden kann, als es seine durch den Sicherungsquader festgelegte relative Spannweite zulässt.

Das bedeutet, dass die primäre Geheimhaltung auf die Sperrung von Einzelangaben im Einzeldatenmaterial zurückgeführt wird; alle Primärsperrungen von Aggregaten werden an das Quaderverfahren übertragen. Damit erübrigt sich insbesondere die Diskussion um die Konzentrationsmaße der primären Geheimhaltung (vgl. Pkt. 2.2). Vom Tabellenerzeuger bleibt lediglich festzulegen, welche Genauigkeit der Rückrechenbarkeit er denn noch tolerieren kann, d. h. welche relative Mindestspannweite einem Einzelwert zuerkannt werden muss, damit dieser noch als geschützt gelten darf.

Die Sicherung besonders großer Werte vor Dominanz in ihren Aggregaten erledigt sich dabei von selbst: Durch die bei besonders großen Einzelwerten auch besonders großen Spannweiten aufgrund ihrer durch die Quadersicherung garantierten relativen Mindestspannweite bleibt immer ein hinreichender Mindestabstand von seinem kleinsten zu seinem größten Schätzwert, z. B. seinem Aggregatwert, wenn dieser veröffentlicht werden darf; anderenfalls wäre der betreffende Aggregatwert durch das Quaderverfahren gesperrt worden.

Um den Quaderschutz auf das Einzeldatenmaterial zu erweitern, genügt es, der gegebenen Statistiktafel noch eine Gliederung nach Einzeldaten anzufügen. Erste Untersuchun-

gen auch mit Realdaten haben ergeben, dass diese Gliederung in Bezug auf die unterste Aggregation der gegebenen n -dimensionalen Tafel den größten und den zweitgrößten Einzelwert sowie das Restaggregat, den Tabellenwert abzüglich der beiden größten Einzelwerte, enthalten sollte. – Das Hinzufügen weiterer Einzelangaben anstelle des Restaggregats würde zu Übersperrungen führen, weil zur Auswahl eines Doppelquaders zum Schutze des größten Einzelwertes der drittgrößte Einzelwert zum selben Aggregat oft zu klein ist, während das Restaggregat noch einen ausreichenden Schutz bieten kann.

Mit dieser neuen Gliederung erhält man eine $n+1$ -dimensionale Gesamttabelle, die drei neu hinzutretenden Tabellen des „Einzelmateriale“, von denen jede genau so gegliedert ist wie die gegebene n -dimensionale Tafel und die gegebene Tafel als Summentabelle in Bezug auf diese neu eingefügte Einzeldatengliederung. Für den Dominanzschutz ist es unerlässlich, außerdem noch die Schätzfehler anzugeben, die der Tabellenutzer aufgrund des zu unterstellenden Vorwissens bestimmen kann.

Um diese $n+1$ -dimensionale Gesamttabelle mit dem Quaderverfahren sichern zu können, müssen noch alle Einzelangaben der drei Einzelmateriale tabellen primär gesperrt werden. Die Anwendung des Quaderverfahrens mit einem die Schätzintervalle der Tabellenwerte berücksichtigenden Intervallschutz liefert dann außer dem Schutz gegen zu genaues Rückrechnen auch einen hinreichenden Dominanzschutz:

Unterstellt man, dass der Tabellenutzer „seine“ Tabellenwerte bis zu plus minus 100 % eingrenzen kann, so werden alle Werte der (n -dimensionalen) Veröffentlichungstabelle gesperrt, die nach der $p\%$ -Regel geheim zuhalten gewesen wären. Dabei ergibt sich der Prozentsatz p als Quotient aus der relativen Mindestspannweite, nach der die Sicherungsquader auszuwählen sind, und der relativen Schätzintervalllänge, dem zweifachen relativen Schätzfehler. Außer

diesen Sperreintragungen setzt das Quaderverfahren noch weitere Sperrungen, die als Sicherung der $p\%$ -Regel anzusehen sind. Auf diese Weise erklärt sich die $p\%$ -Regel ganz von selbst, als diejenige Schutzvorschrift, die einen hinreichenden Intervallschutz für alle Einzelangaben garantiert.

Kann man relative Schätzfehler annehmen, die größer als 100 % sind, so wird die $p\%$ -Regel mit der Anwendung des Quaderverfahrens auf die durch das Einzelmateriale erweiterte Tafel nur noch bedingt realisiert. Wenn der relative Schätzfehler schließlich als sehr groß gegenüber 100 % anzunehmen ist, bleibt nur noch der Intervallschutz, der sich aus der Positivität der Tafel ergibt.

Ein Dominanzschutz ist im Falle sehr großer Schätzfehler wegen der zu ungenauen Abschätzbarkeit der Einzelangaben aber auch nicht erforderlich, weil mit den großen Schätzintervallen keine ausreichenden Kenntnisse über den Anteil vorhanden sind, den die Einzelwerte an ihren Aggregaten haben. Das Verfahren der Quadersicherung in durch das Einzeldatenmaterial erweiterten Tabellen ist also allgemeingültiger als die $p\%$ -Regel mit anschließender Quadersicherung, die es andererseits aber bei zu unterstellenden relativen Schätzfehlern von bis zu 100 % auf eindrucksvolle Weise bestätigt.

Anhang

Grundlagen des Quaderverfahrens

Gegeben sei eine n -dimensionale Tafel T mit Randsummen bezüglich jedes der n Gliederungen, aber ohne Zwischensummen. Die Ausprägungen jeder Gliederung seien durchnummeriert, die Tabellenwerte mit diesen Nummern indiziert.

Definition 1:

Zwei Tabellenwerte $T_a, T_b \in T$ heißen zueinander diametral, symbolisch $T_a \# T_b$, wenn für ihre Indizes gilt

$a = (a_1, a_2, \dots, a_n)$, $b = (b_1, b_2, \dots, b_n)$
 $a_i \neq b_i$ für $i = 1, 2, \dots, n$.

Dadurch wird der Schätzfehler ε eingegrenzt gemäß

$$-\varepsilon_{\text{unten}} \leq \varepsilon \leq \varepsilon_{\text{oben}} \quad (4),$$

wobei sich die Beträge der Intervallgrenzen aus (4) ergeben:

$$\varepsilon_{\text{oben}} = \min [\min(X_{\text{oben}} - X), \min(X' - X'_{\text{unten}})] \quad (5)$$

$$\varepsilon_{\text{unten}} = \min [\min(X'_{\text{oben}} - X'), \min(X - X_{\text{unten}})] .$$

Definition 3:

$T_q \in Q(T_a, T_b)$ heißt bezüglich T_b gerade indiziert, wenn $\sum_i B_{ik} + \sum A_j$ gerade ist, anderenfalls heißt T_q ungerade indiziert; dabei bezeichnet A_j die Aggregationsstufe von T_q bezüglich der j -ten Gliederung, $j = 1, 2, \dots, n$, Aggregationsstufen mit 1 beginnend fortlaufend durchnummeriert ($T_a = \text{Pivot}$).

Jeder Quaderwert X bzw. $X' \in Q(T_a, T_b)$ kann vom Tabellennutzer somit höchstens bis auf sein **Schutzintervall** genau eingegrenzt werden:

$$X \in [X - \varepsilon_{\text{unten}}, X + \varepsilon_{\text{oben}}], \\ X' \in [X' - \varepsilon_{\text{oben}}, X' + \varepsilon_{\text{unten}}] \quad (6).$$

Die linearen Gleichungen zur Berechnung gesperrter Werte $X, Y, X', Y' \in Q(T_a, T_b)$, wo X, Y bezüglich T_b gerade, X', Y' bezüglich T_b ungerade indiziert sind, haben die Gestalt:

$$X + X' = \Sigma, \\ X - Y = \Sigma, \quad X' - Y' = \Sigma, \quad (1)$$

wobei Σ die Quadersumme bezeichnet: Σ ist die Randsumme abzüglich aller nicht zu $Q(T_a, T_b)$ gehörenden Tabellenwerte bezüglich der Summations-Gliederung. Tragen keine Randsummenwerte zu $Q(T_a, T_b)$ bei, gilt nur die erste Gleichung von (1).

Falls alle Tabellenwerte $X, X' \in Q(T_a, T_b)$ gesperrt sind, kann der externe Tabellennutzer nur Schätzwerte für X, X' angeben:

$$\hat{X} = X + \varepsilon; \quad \hat{X}' = X' - \varepsilon \quad (2);$$

mit nur einem unbekannten Parameter $\varepsilon \forall X, X' \in Q(T_a, T_b)$, der – falls kein Vorwissen unterstellt werden muss – beliebige Werte annehmen kann.

Bei zu unterstellendem Vorwissen kennt der Tabellennutzer **Schätzintervalle** mit Intervallgrenzen $T_{\text{oben}}, T_{\text{unten}} \forall T_t \in T$, so dass auch für die Schätzwerte (2) gilt,

$$X_{\text{unten}} \leq X + \varepsilon \leq X_{\text{oben}}, \\ X'_{\text{unten}} \leq X' - \varepsilon \leq X'_{\text{oben}} \quad (3).$$

Die Spannweite des Quaders $Q(T_a, T_b)$, range, ist demgemäß

$$\text{range} = \varepsilon_{\text{unten}} + \varepsilon_{\text{oben}} \quad (7).$$

Damit ist ein Quaderauswahlkriterium gegeben, das in den derzeit verfügbaren EDV-Programmen GHQUAR.44 und GHMITER.22 angewendet wird und das auch in QUIT (z. Zt. noch in der Entwicklung) die Quaderauswahl steuert: Es kommen nur solche Quader als Sicherungsquader in Betracht, deren Spannweite, range, bezogen auf den Betrag des zu schützenden Pivotwerts größer als eine vorgegebene relative Mindestspannweite ist.

Das Personal des Landes am 30. Juni 2000

Regionalisierte Ergebnisse der Personalstandstatistik

Dipl.-Volkswirt Heinz-Peter Emmerich

Entscheidungen im Rahmen der regionalen Wirtschaftspolitik und Wirtschaftsförderung haben im Hinblick auf die Nutzung und Entwicklung des Wirtschaftspotenzials eines Landes einen bedeutenden politischen Stellenwert. Angesichts der strukturellen wirtschaftlichen Unterschiede in NRW sind u. a. Fragen der Raumordnungspolitik sowie der Landes- und Regionalplanung von großer Bedeutung. Dabei erstreckt sich der für politische Entscheidungen notwendige Informationsbedarf auf nahezu alle Bereiche der (amtlichen) Statistik.

Darüber hinaus sind statistische Informationen über regionale Raumeinheiten auch für Zwecke der Regionalforschung von Interesse, um ökonomisch relevante Sachverhalte zu analysieren, die zur Beschreibung regionaler Disparitäten im Rahmen (regional-)wissenschaftlicher Untersuchungen herangezogen werden können.

Da auch von der regionalen Verteilung der Standorte von Behörden und Einrichtungen des Landes – und somit der dort Beschäftigten, aber auch der Personen, die das Dienstleistungsangebot der öffentlichen Einrichtungen in Anspruch nehmen und am Ort wohnen (z. B. Studenten, Seminarteilnehmer bzw. Teilnehmer an Fortbildungsveranstaltungen etc.) – ein positiver Effekt u. a. auf die regionale Wertschöpfung, Kaufkraft zu erwarten ist, soll diese Untersuchung einen kleinen Beitrag zu dem weiten Spektrum der wirtschaftspolitischen und raumordnungspolitischen Aufgabenstellung liefern.

Rechtsgrundlage, Berichtskreis und Erhebungsmerkmale

Die gesetzliche Grundlage zur Durchführung u. a. der Personalstandstatistiken ist in dem Gesetz über die Statistiken der öffentlichen Finanzen und des Personals im öffentlichen Dienst (Finanz- und Personalstatistikgesetz – FPStatG –) vom 08. März 2000 (BGBl. I S. 206) gere-

gelt. Danach sind bundesweit die Beschäftigten des öffentlichen Dienstes jährlich zum 30. Juni zu erfassen.

Inhaltlich erstreckt sich diese Statistik auf den zum o. g. Stichtag in einem unmittelbaren grundsätzlich entgeltspflichtigen Dienst- bzw. Arbeitsvertragsverhältnis (einschl. Ausbildungsverhältnis) zur Berichtsstelle stehenden Personenkreis. Dabei sind – von den auskunftspflichtigen Stellen – für jeden Beschäftigten persönliche, regionale sowie bezahlungsrelevante Merkmale in z. T. verschlüsselter Form anzugeben.

Der gesamte Merkmalskatalog ist inzwischen bei dem überwiegenden Teil der Berichtsstellen durch Nutzung der dort für eigene Zwecke vorhandenen Verwaltungsdateien für das Bezahlungswesen (d. h. für die Zahlbarmachung der Bezüge, Vergütungen und Löhne) maschinell abrufbar. So erhält das LDS NRW die statistisch relevanten Daten des Landespersonals – bis auf wenige Einzelfälle – aus den Dateien des Landesamtes für Besoldung und Versorgung (LBV NRW).

In diesem Beitrag wird die regionale Verteilung der Beschäftigten, die im (unmittelbaren) öffentlichen Dienst des Landes NRW tätig sind, nach ausgewählten Gesichtspunkten und Erhebungsmerkmalen dargestellt.

Zum Verfahren der Regionalisierung

Grundlage für die Regionalisierung der Landesbeschäftigten bilden die in der LBV-Datei vorhandene Dienststellennummer sowie weitere gespeicherte regionale Merkmale (z. B. Schulnummer, Betriebsnummer). Da als „kleinste“ regionale Einheit die Gemeinde zugrunde zu legen ist, sind die staatlichen Dienststellen jeweils daraufhin zu überprüfen, ob sich das dort eingesetzte Personal über mehr als eine kreisfreie Stadt oder kreisangehörige Gemeinde verteilt, d. h. die Frage möglicher Nebenstellen muss jeweils (vorab) geklärt werden. Die dazu notwendigen Informationen können zum großem Teil aus den Unterlagen des LBV sowie aus der im Landesorganisationsgesetz NRW getroffenen Einteilung der Behörden und Einrichtungen des Landes und den dazu ergangenen Bekanntmachungen und Verordnungen über die Veränderung oder Auflösung von Landesdienststellen und/oder deren Einzugsbereiche u. a. m. entnommen werden. Mit diesen Hinweisen sowie den intern gespeicherten regionalen Daten konnten die Beschäftigten der einzelnen Dienststellen entweder direkt über eine eigene Referenzdatei oder – im Falle einer Aufteilung von Dienststellen auf mehrere Nebenstellen – über die vorhandenen regionalen Informationen eindeutig einer Stadt/Gemeinde zugeordnet werden.

Gesamtergebnisse

Im Landesdienst waren zum Erhebungsstichtag 308 255 Vollzeit beschäftigte Beamtinnen/Beamte, Angestellte und Arbeiterinnen/Arbeiter tätig; davon entfielen 69,1 % (213 114) auf die in einem öffentlich rechtlichen Dienstverhältnis stehenden Beschäftigten, 27,1 % (83 499)

auf die Angestellten und 3,8 % (11 642) auf die Arbeiterinnen/Arbeiter. Die o. g. Gesamtzahl aller Vollzeitbeschäftigten enthielt auch 91 Personen mit einem Dienstort außerhalb von Nordrhein-Westfalen. Bei diesem Personenkreis handelte es sich u. a. um Bedienstete des Polizeipräsidenten der Wasserschutzpolizei NRW in Duisburg, die in der Stadt Bramsche (Niedersachsen) ihren Dienstort hatten, um Beschäftigte der Vertretung des Landes NRW beim Bund und bei der Europäischen Union sowie um Personen des Schul- und Hochschulbereiches, die in einem anderen Bundesland (z. B. Rheinland-Pfalz) bzw. im Ausland einer Beschäftigung nachgingen.

Die regionale Verteilung der Landesbeschäftigten in der Gliederung nach Voll- und Teilzeitarbeit (T1-Beschäftigte)¹⁾ sowie dem Dienstverhältnis zeigte, dass 55 % (117 476) aller Vollzeit beschäftigten Beamtinnen/Beamten und 77,1 % (64 378) der Angestellten in Dienststellen tätig waren, die ihren Sitz (einschl. Nebenstellen) in einer kreisfreien Stadt hatten.

Von diesen waren 64,6 % (75 862) Beamtinnen/Beamte bzw. 69,6 % (44 791) Angestellte in Dienstorten tätig, die in einer kreisfreien Stadt des Regierungsbezirks Düsseldorf oder des Regierungsbezirks Köln lagen. Im einzelnen entfielen 47 506 Beamtinnen/Beamte und 22 233 Angestellte auf die kreisfreien Städte des Regierungsbezirks Düsseldorf und 28 356 Beamtinnen/Beamte bzw. 22 558 Angestellte auf die des Regierungsbezirks Köln.

Stellt man die in diesen zwei Regionen eingesetzten Beamtinnen/Beamte und Angestellte in Relation zur jeweiligen Gesamtzahl der o. g. Dienstverhältnisse im Landesdienst, ist anzumerken, dass 53,7 % (114 459) aller Vollzeit beschäftigten Beamtinnen/Beamten und 61,7 % (51 500) aller Vollzeit beschäftigten Angestellten in Dienststellen bzw. Einrichtungen des Landes ihren

Dienst verrichteten, die im Regierungsbezirk Düsseldorf oder im Regierungsbezirk Köln lagen.

Personen) aller Vollzeitbeschäftigten im unmittelbaren Dienst des Landes NRW.

Vollzeit beschäftigtes Personal des Landes in NRW am 30. Juni 2000 nach dem Dienstort von Haupt- und Nebenstellen einer Dienststelle und Gemeindegrößenklassen		
Sitz der Dienststelle Gemeindegrößenklasse	Bevölkerung ¹⁾	Vollzeitbeschäftigte des Landes
Nordrhein-Westfalen	17 996 153	308 164
Kreisfreie Städte zusammen	7 432 511	191 468
davon mit ... Einwohnern		
500 000 und mehr	3 233 968	89 006
200 000 bis unter 500 000	3 101 088	89 923
unter 200 000	1 097 455	12 539
Kreisangehörige Gemeinden zusammen	10 563 642	116 696
davon mit ... Einwohnern		
100 000 und mehr	837 461	16 691
50 000 bis unter 100 000	3 141 720	40 555
20 000 bis unter 50 000	4 244 669	46 438
10 000 bis unter 20 000	1 879 769	10 828
5 000 bis unter 10 000	446 281	2 113
unter 5 000	13 742	71

1) am 30. Juni 2000

Bei den Vollzeit beschäftigten Arbeiterinnen/Arbeitern war der Anteil in diesen Regierungsbezirken zusammen mit 64,8 % bzw. 7 540 Personen – gemessen an ihrer Gesamtzahl – ebenfalls entsprechend hoch.

Bei den Teilzeit beschäftigten Personen (T1) änderte sich diese regionale Konzentration auf die Regierungsbezirke Düsseldorf und Köln – quantitativ gesehen – nur unwesentlich. Zum Erhebungstichtag hatten 53,4 % (28 065) aller T1-Beamtinnen/Beamten und 57,5 % (19 192) aller T1-Angestellten ihren Dienstort in einem dieser zwei (Regierungs-)Bezirke.

Im einzelnen ist die regionale Gliederung des o. g. Beschäftigtenkreises in den kreisfreien Städten und Kreisen nach dem Dienstverhältnis der Tabelle auf Seite 38ff. zu entnehmen.

Ausgehend von der Gesamtzahl aller Vollzeitbeschäftigten im Landesdienst waren gut dreiünftel (62,1 % bzw. 191 468 Personen) in Dienststellen tätig, die ihren Sitz in einer kreisfreien Stadt hatten; hier lebten mehr als zweiünftel der Gesamtbevölkerung (41,3 % bzw. 7,4 Mill. Einwohner). Demnach entfielen auf die kreisangehörigen Gemeinden, die einen Bevölkerungsanteil von 58,7 % (10,6 Mill. Einwohner) auf sich vereinigten, 37,9 % (116 696

Im Einzelnen war Folgendes festzustellen:

Die 308 164 Vollzeittätigen im Landesdienst mit einem Dienstort in NRW verteilten sich zu 4,2 % auf Gemeinden mit einer Einwohnerzahl bis 20 000, in denen 13,0 % der Gesamtbevölkerung lebten, während in Gemeinden mit 20 000 bis 100 000 Einwohner, die einen Bevölkerungsanteil von 41 % verzeichneten, ein Vollzeitbeschäftigtenanteil von 28,2 % festzustellen war.

Die kreisangehörigen und kreisfreien Städte mit über 100 000 Einwohnern, die insgesamt knapp die Hälfte (45,9 %) der Gesamtbevölkerung auf sich vereinigten, konnten 67,5 % aller Vollzeitbeschäftigten des Landes in ihren „administrativen“ Grenzen verzeichnen.

Allein in den 16 kreisfreien Städten mit einer Einwohnerzahl von über 200 000, auf die gut ein Drittel (35,2 %) der Gesamtbevölkerung entfiel, waren mehr als die Hälfte (58,1 %) aller Vollzeitkräfte beschäftigt. Auf die jeweilige Einwohnerzahl bezogen bedeutet dies, dass in den kreisfreien Städten 26 und in den Kreisgebieten 11 Vollzeit beschäftigte Landesbedienstete je 1 000 Einwohner tätig waren bzw. für Dienstleistungen dem Bürger zur Verfügung standen. Im Landesdurch-

1) Beschäftigte mit der Hälfte oder mehr als der Hälfte der regelmäßigen Wochenarbeitszeit eines Vollzeitbeschäftigten

Personal des Landes am 30. Juni 2000 nach dem Dienort von Haupt- und Nebenstellen einer Dienststelle

Lfd. Nr.	Sitz der Dienststelle	Vollzeitbeschäftigte				
		insgesamt	darunter Auszubildende	Beamten/innen/Beamte	Angestellte	Arbeiter/-innen
	Kreisfreie Städte					
1	Düsseldorf	27 688	2 362	15 016	10 988	1 684
2	Duisburg	9 136	1 238	6 747	2 199	190
3	Essen	13 050	1 355	7 780	4 666	604
4	Krefeld	3 984	337	2 964	930	90
5	Mönchengladbach	4 415	576	3 638	729	48
6	Mülheim an der Ruhr	1 619	17	1 371	221	27
7	Oberhausen	2 556	241	2 084	443	29
8	Remscheid	1 514	25	1 278	232	4
9	Solingen	1 729	194	1 451	269	9
10	Wuppertal	6 908	760	5 177	1 556	175
	Kreise					
11	Kleve	3 959	458	3 416	501	42
12	Mettmann	4 085	34	3 494	542	49
13	Neuss	4 441	185	3 725	677	39
14	Viersen	2 941	77	2 531	396	14
15	Wesel	4 851	67	4 099	678	74
16	Reg.-Bez. Düsseldorf	92 876	7 926	64 771	25 027	3 078
	davon					
17	kreisfreie Städte	72 599	7 105	47 506	22 233	2 860
18	Kreise	20 277	821	17 265	2 794	218
	Kreisfreie Städte					
19	Aachen	15 035	1 932	5 533	7 881	1 621
20	Bonn	11 506	960	5 280	5 249	977
21	Köln	26 558	2 677	16 003	9 158	1 397
22	Leverkusen	1 820	177	1 540	270	10
	Kreise					
23	Aachen	2 414	265	2 119	252	43
24	Düren	3 370	601	2 732	545	93
25	Erftkreis	4 271	256	3 562	602	107
26	Euskirchen	2 138	44	1 736	354	48
27	Heinsberg	2 618	51	2 171	430	17
28	Oberbergischer	3 182	312	2 587	544	51
29	Rheinisch-Bergischer Kreis	2 239	19	1 900	321	18
30	Rhein-Sieg-Kreis	5 472	513	4 525	867	80
31	Reg.-Bez. Köln	80 623	7 807	49 688	26 473	4 462
	davon					
32	kreisfreie Städte	54 919	5 746	28 356	22 558	4 005
33	Kreise	25 704	2 061	21 332	3 915	457
	Kreisfreie Städte					
34	Bottrop	1 038	3	889	149	–
35	Gelsenkirchen	5 262	824	4 167	1 034	61
36	Münster	16 530	1 900	7 888	7 249	1 393

1) mit der Hälfte oder mehr als der Hälfte der regelmäßigen Wochenarbeitszeit eines Vollzeitbeschäftigten – 2) einschl. Altersteilzeitbeschäftigten – 3) mit weniger

Teilzeitbeschäftigte (T1) ¹⁾²⁾				Teilzeitbeschäftigte (T2) ³⁾				Lfd. Nr.
insgesamt	Beamtinnen/ Beamte	Angestellte	Arbeiter/-innen	insgesamt	Beamtinnen/ Beamte	Angestellte	Arbeiter/-innen	
5 146	2 226	2 769	151	1 110	76	1 028	6	1
2 216	1 308	882	26	443	34	407	2	2
3 674	1 662	1 523	489	700	23	675	2	3
1 125	783	331	11	97	13	84	–	4
1 077	755	308	14	72	11	59	2	5
594	483	106	5	29	4	23	2	6
839	597	229	13	42	11	30	1	7
474	347	127	–	32	16	16	–	8
625	470	155	–	29	13	16	–	9
2 075	1 239	813	23	613	26	585	2	10
1 053	788	251	14	41	4	36	1	11
1 850	1 470	372	8	79	21	58	–	12
1 748	1 351	392	5	53	13	39	1	13
1 054	831	217	6	29	9	20	–	14
1 737	1 338	387	12	75	29	41	5	15
25 287	15 648	8 862	777	3 444	303	3 117	24	16
17 845	9 870	7 243	732	3 167	227	2 923	17	17
7 442	5 778	1 619	45	277	76	194	7	18
3 591	984	2 473	134	2 434	33	2 388	13	19
3 453	1 085	2 256	112	1 490	19	1 468	3	20
6 217	2 922	3 163	132	2 247	78	2 158	11	21
717	536	178	3	29	4	25	–	22
831	652	170	9	34	10	23	1	23
1 016	699	310	7	72	11	61	–	24
1 571	1 207	344	20	81	20	59	2	25
736	480	227	29	33	8	22	3	26
930	673	250	7	47	23	22	2	27
978	746	227	5	62	9	53	–	28
1 074	853	218	3	41	8	33	–	29
2 107	1 580	514	13	87	23	62	2	30
23 221	12 417	10 330	474	6 657	246	6 374	37	31
13 978	5 527	8 070	381	6 200	134	6 039	27	32
9 243	6 890	2 260	93	457	112	335	10	33
362	256	104	2	10	4	5	1	34
1 046	690	347	9	83	14	66	3	35
4 086	1 129	2 780	177	1 301	40	1 246	15	36

als der Hälfte der regelmäßigen Wochenarbeitszeit eines Vollzeitbeschäftigten

Noch: Personal des Landes am 30. Juni 2000 nach dem Dienort von Haupt- und Nebenstellen einer Dienststelle						
Lfd. Nr.	Sitz der Dienststelle	Vollzeitbeschäftigte				
		insgesamt	darunter Auszubildende	Beamten/innen/Beamte	Angestellte	Arbeiter/-innen
	Kreise					
37	Borken	3 798	306	3 334	452	12
38	Coesfeld	2 276	48	1 840	361	75
39	Recklinghausen	7 567	273	6 295	1 202	70
40	Steinfurt	4 455	365	3 795	622	38
41	Warendorf	2 475	47	2 120	313	42
42	Reg.-Bez. Münster	43 401	3 766	30 328	11 382	1 691
	davon					
43	kreisfreie Städte	22 830	2 727	12 944	8 432	1 454
44	Kreise	20 571	1 039	17 384	2 950	237
	Kreisfreie Stadt					
45	Bielefeld	8 565	1 207	6 190	2 167	208
	Kreise					
47	Gütersloh	3 349	242	2 882	431	36
48	Herford	2 785	53	2 379	395	11
49	Höxter	1 754	33	1 403	299	52
50	Lippe	5 313	599	4 131	1 081	101
51	Minden-Lübbecke	3 768	291	3 001	654	113
52	Paderborn	5 271	558	3 677	1 402	192
53	Reg.-Bez. Detmold	30 805	2 983	23 663	6 429	713
	davon					
54	kreisfreie Stadt	8 565	1 207	6 190	2 167	208
55	Kreise	22 240	1 776	17 473	4 262	505
	Kreisfreie Städte					
56	Bochum	9 944	930	6 149	3 206	589
57	Dortmund	12 574	1 412	8 640	3 601	333
58	Hagen	5 218	770	3 723	1 358	137
59	Hamm	3 176	325	2 593	565	18
60	Herne	1 643	9	1 375	258	10
	Kreise					
58	Ennepe-Ruhr-Kreis	2 845	30	2 429	396	20
59	Hochsauerlandkreis	4 662	467	3 642	842	178
60	Märkischer	5 162	304	4 253	861	48
61	Olpe	1 476	34	1 219	226	31
62	Siegen-Wittgenstein	4 719	444	3 443	1 151	125
63	Soest	4 010	95	3 109	841	60
64	Unna	5 030	399	4 039	851	140
65	Reg.-Bez. Arnsberg	60 459	5 219	44 614	14 156	1 689
	davon					
66	kreisfreie Städte	32 555	3 446	22 480	8 988	1 087
67	Kreise	27 904	1 773	22 134	5 168	602
68	Nordrhein-Westfalen	308 164	27 701	213 064	83 467	11 633
	davon					
69	kreisfreie Städte	191 468	20 231	117 476	64 378	9 614
70	Kreise	116 696	7 470	95 588	19 089	2 019
71	Dienstorte außerhalb NRWs	91	–	50	32	9
72	Personal des Landes NRW insgesamt	308 255	27 701	213 114	83 499	11 642

Anmerkungen Seite 38

Teilzeitbeschäftigte (T1) ¹⁾²⁾				Teilzeitbeschäftigte (T2) ³⁾				Lfd. Nr.
insgesamt	Beamtinnen/ Beamte	Angestellte	Arbeiter/-innen	insgesamt	Beamtinnen/ Beamte	Angestellte	Arbeiter/-innen	
1 268	975	284	9	69	17	49	3	37
1 026	755	246	25	53	14	36	3	38
2 422	1 806	578	38	131	29	89	13	39
1 734	1 329	393	12	81	14	59	8	40
1 103	878	217	8	42	19	17	6	41
13 047	7 818	4 949	280	1 770	151	1 567	52	42
5 494	2 075	3 231	188	1 394	58	1 317	19	43
7 553	5 743	1 718	92	376	93	250	33	44
2 398	1 192	1 173	33	694	17	673	4	45
1 324	1 052	262	10	59	19	40	–	47
1 089	818	267	4	45	10	33	2	48
616	422	182	12	40	1	28	11	49
1 671	1 140	498	33	158	21	134	3	50
1 213	844	336	33	44	17	24	3	51
1 585	946	621	18	905	24	879	2	52
9 896	6 414	3 339	143	1 945	109	1 811	25	53
2 398	1 192	1 173	33	694	17	673	4	54
7 498	5 222	2 166	110	1 251	92	1 138	21	55
2 612	1 106	1 444	62	888	24	860	4	56
3 063	1 684	1 335	44	942	41	900	1	57
1 151	584	492	75	337	9	324	4	58
801	544	242	15	34	11	20	3	59
481	346	126	9	27	5	22	–	60
1 159	895	253	11	49	15	32	2	58
1 198	816	356	26	69	16	47	6	59
1 567	1 182	369	16	83	26	52	5	60
452	339	109	4	18	4	10	4	61
1 285	756	509	20	375	13	358	4	62
1 327	940	369	18	68	14	51	3	63
1 422	1 102	268	52	59	10	43	6	64
16 518	10 294	5 872	352	2 949	188	2 719	42	65
8 108	4 264	3 639	205	2 228	90	2 126	12	66
8 410	6 030	2 233	147	721	98	593	30	67
87 969	52 591	33 352	2 026	16 765	997	15 588	180	68
47 823	22 928	23 356	1 539	13 683	526	13 078	79	69
40 146	29 663	9 996	487	3 082	471	2 510	101	70
9	2	4	3	–	–	–	–	71
87 978	52 593	33 356	2 029	16 765	997	15 588	180	72

schnitt entfielen 17 Vollzeit beschäftigte Personen auf je 1 000 Einwohner.

Aus der vorliegenden Übersicht kann man auch schließen, dass sich die Gemeinden – gemessen an der Einwohnerzahl – in ihrer Bedeutung als Behördenstandorte wesentlich unterscheiden. In den (einwohner-) großen Städten konzentriert sich die Zahl der Landesbediensteten – und damit u. U. die Anzahl der Landesbehörden und -einrichtungen – gemessen an der Bevölkerungsverteilung überproportional stark.

Hinsichtlich der „Versorgung“ mit Vollzeit beschäftigten Landesbediensteten wies die kreisfreie (Landeshaupt-) Stadt Düsseldorf mit 27 688 die größte Zahl auf, gefolgt von Köln (26 558), Münster (16 530), Aachen (15 035), Essen (13 050) und Dortmund (12 574).

Die Verteilung der Vollzeitbeschäftigten im Landesdienst je 1 000 Einwohner zeigte dagegen eine andere regionale Struktur. Danach hatten die kreisfreien Städte Münster und

Herne die Besatzziffer in den hier genannten Städten z. T. erheblich über dem vergleichbaren Durchschnittswert (26) aller kreisfreien Städte.

Allgemein war die Streuung dieses Datenwertes in der o. g. Gruppe sehr groß. Die Spannweite reichte von 9 bis 62 Vollzeitkräften je 1 000 Einwohner.

So wiesen die kreisfreien Städte Bottrop, Mülheim an der Ruhr und Herne mit 9 Vollzeitbeschäftigten pro 1 000 Einwohner den kleinsten Wert auf, gefolgt von den kreisfreien Städten Solingen (10), Leverkusen (11), Oberhausen (12) und Remscheid (13). Die kreisfreien Ruhrgebietsstädte Gelsenkirchen, Dortmund, Essen und Bochum lagen mit 19, 21, 22 bzw. 25 Vollzeitkräften je 1 000 Einwohner über dem Landesdurchschnittswert von 17 (vgl. nachfolgende Grafik).

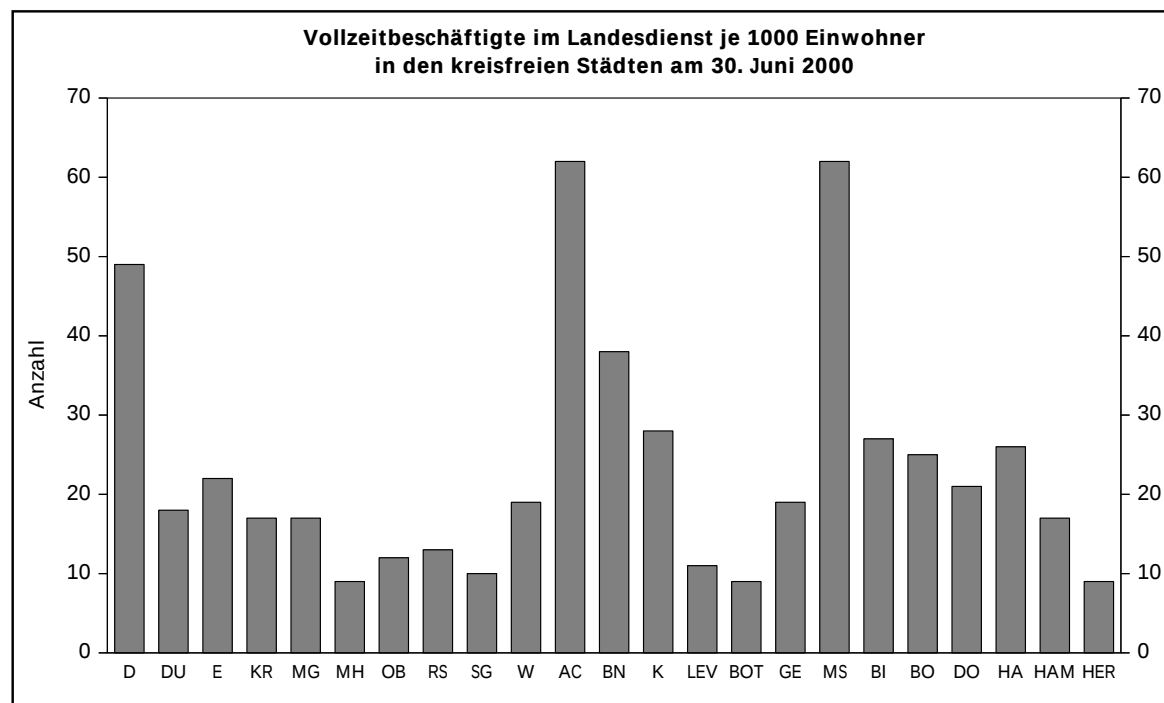
Innerhalb der Kreisgebiete war die Spannweite dieser Quote um den entsprechenden Durchschnittswert (11) nicht so stark ausgeprägt. Sie lag – auf dieser regionalen Bezugs-

Bezogen auf den Landesdurchschnitt (17) lag die Zahl der Vollzeitbeschäftigten je 1 000 Einwohner jedoch bei fast allen Kreisen (30 Kreise) darunter; nur der Kreis Paderborn konnte mit einer „Dienstleistungsquote“ von 18 einen über dem Landesdurchschnitt liegenden Wert erreichen, während der Hochsauerlandkreis und der Kreis Siegen-Wittgenstein mit jeweils 16 Vollzeitpersonen je 1 000 Einwohner diesen knapp verfehlten. Die Abweichungen in diesen 30 Kreisen waren mit + 7 bzw. –3 Vollzeitbeschäftigten je 1 000 Einwohnern vom Durchschnittswert aller Kreise – im Gegensatz zu den kreisfreien Städten – relativ gering.

Interessant ist auch ein Blick auf die „regionale“ Rangfolgeziffer in der Differenzierung nach dem Indikator „Beschäftigte insgesamt je 1 000 Einwohner“.

Während sich die Stadt Aachen mit 86 Beschäftigten je 1 000 Einwohner vor Münster (83), Düsseldorf (60), Bonn (55) sowie Bielefeld und Köln (jeweils 36) platzieren konnte, belegten die Städte Bottrop und Herne (jeweils 12) – wie

schon bei der Quote „Vollzeittätige je 1 000 Einwohner“ – gemeinsam den letzten Platz. Bei den Kreisen lagen bei dieser Quote Paderborn (27), Siegen-Wittgenstein und der Hochsauerlandkreis (mit jeweils 21) sowie Lippe (20) und Soest (18) vorne. Schlusslicht war der Kreis Aachen (11), der auch bei



Aachen mit 62 Beschäftigten je 1 000 Einwohner die größte Versorgungsquote. Es folgten die (kreisfreien) Städte Düsseldorf, Bonn, Köln und Bielefeld mit jeweils 49, 38, 28 bzw. 27 Vollzeitkräften je 1 000 Einwoh-

ebene – zwischen 8 (Kreis Mettmann, Kreis Aachen, Rheinisch-Bergischer- und Ennepe-Ruhr-Kreis) und 18 (Kreis Paderborn) Vollzeitkräften je 1 000 Einwohner.

der Vollzeitbeschäftigtenquote – neben anderen Gebietskörperschaften – am Ende der „Rankingskala“ lag. (Siehe Tabelle Seite 43f.)

**Personal des Landes in NRW am 30. Juni 2000 nach dem Dienstort
von Haupt- und Nebenstellen einer Dienststelle**

Sitz der Dienststelle	Beschäftigte je 1 000 Einwohner				Rangfolgeziffer	
	insgesamt	Vollzeitbeschäftigte	Teilzeit- beschäftigte (T1)1,3	Teilzeit- beschäftigte (T2)2	Beschäftigte insgesamt	Vollzeit- beschäftigte
Reg.-Bez. Düsseldorf						
Kreisfreie Städte						
Düsseldorf	60	49	9	2	3	3
Duisburg	23	18	4	1	13	14
Essen	29	22	6	1	9	9
Krefeld	22	17	5	0	16	17
Mönchengladbach	21	17	4	0	18	16
Mülheim an der Ruhr	13	9	3	0	48	48
Oberhausen	15	12	4	0	29	27
Remscheid	17	13	4	0	23	23
Solingen	14	10	4	0	38	38
Wuppertal	26	19	6	2	12	11
Kreise						
Kleve	17	13	4	0	22	21
Mettmann	12	8	4	0	51	53
Neuss	14	10	4	0	40	42
Viersen	13	1 0	4	0	44	43
Wesel	14	10	4	0	41	41
Reg.-Bez. Köln						
Kreisfreie Städte						
Aachen	86	62	15	1 0	1	2
Bonn	55	38	11	5	4	4
Köln	36	28	6	2	5	5
Leverkusen	16	11	4	0	25	30
Kreise						
Aachen	11	8	3	0	54	54
Düren	17	13	4	0	24	24
Erftkreis	13	9	3	0	46	46
Euskirchen	15	11	4	0	30	29
Heinsberg	14	11	4	0	37	37
Oberbergischer Kreis	15	11	3	0	35	33
Rheinisch-Bergischer Kreis	12	8	4	0	50	51
Rhein-Sieg-Kreis	13	1 0	4	0	45	45
Reg.-Bez. Münster						
Kreisfreie Städte						
Bottrop	12	9	3	0	52	50
Gelsenkirchen	23	19	4	0	14	12
Münster	83	62	15	5	2	1

1) mit der Hälfte oder mehr als der Hälfte der regelmäßigen Wochenarbeitszeit eines Vollzeitbeschäftigten – 2) einschl. Altersteilzeitbeschäftigten – 3) mit weniger als der Hälfte der regelmäßigen Wochenarbeitszeit eines Vollzeitbeschäftigten

Noch: Personal des Landes in NRW am 30. Juni 2000 nach dem Dienstort von Haupt- und Nebenstellen einer Dienststelle						
Sitz der Dienststelle	Beschäftigte je 1 000 Einwohner				Rangfolgeziffer	
	insgesamt	Vollzeitbeschäftigte	Teilzeit- beschäftigte (T1)1,3)	Teilzeit- beschäftigte (T2)2)	Beschäftigte insgesamt	Vollzeit- beschäftigte
noch: Reg.-Bez. Münster						
Kreise						
Borken	14	11	4	0	39	36
Coesfeld	16	11	5	0	26	35
Recklinghausen	15	11	4	0	32	28
Steinfurt	14	10	4	0	36	40
Warendorf	13	9	4	0	47	49
Reg.-Bez. Detmold						
Kreisfreie Stadt						
Bielefeld	36	27	7	2	6	6
Kreise						
Gütersloh	14	1 0	4	0	43	44
Herford	15	11	4	0	31	34
Höxter	15	11	4	0	28	31
Lippe	20	15	5	0	20	20
Minden-Lübbecke	16	12	4	0	27	25
Paderborn	27	18	5	3	11	13
Reg.-Bez. Arnsberg						
Kreisfreie Städte						
Bochum	34	25	7	2	7	8
Dortmund	28	21	5	2	10	10
Hagen	33	26	6	2	8	7
Hamm	22	17	4	0	15	15
Herne	12	9	3	0	49	47
Kreise						
Ennepe-Ruhr-Kreis	12	8	3	0	53	52
Hochsauerlandkreis	21	16	4	0	19	18
Märkischer	15	11	3	0	34	32
Olpe	14	10	3	0	42	39
Siegen-Wittgenstein	21	16	4	1	17	19
Soest	18	13	4	0	21	22
Unna	15	12	3	0	33	26
Nordrhein-Westfalen	23	17	5	1	x	x
davon						
kreisfreie Städte	34	26	6	2	x	x
Kreise	15	11	4	0	x	x

Regionalisierung nach Zonen

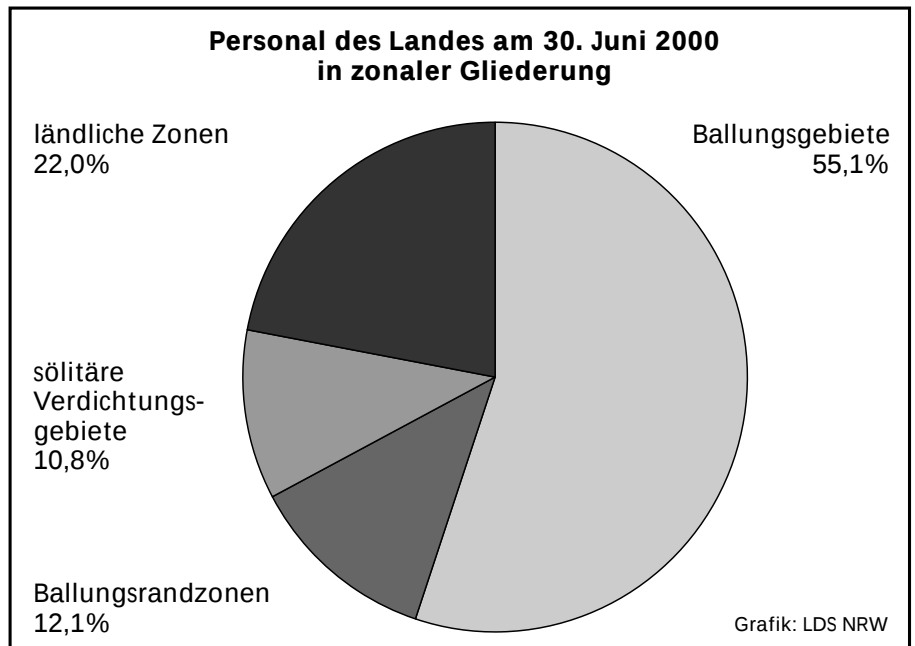
Im Landesentwicklungsprogramm (LEPro) sind auch aufgrund raumordnungspolitischer Untersuchungen die Regionen des Landes u. a. in räumliche Zonen unterteilt. Hierzu ist folgende Untergliederung nach den durch § 21 LEPro festgelegten Abgrenzungen vorgegeben:

Ballungskerne: Dabei handelt es sich um Verdichtungsgebiete, deren durchschnittliche Bevölkerungsdichte 2 000 Einwohner je km² übersteigt oder in absehbarer Zeit übersteigen wird und deren Flächengröße mindestens 50 km² beträgt. Dazu zählen nahezu alle kreisfreien Städte.

Ballungsrandzonen: Dies sind an Ballungskerne angrenzende Verdichtungsgebiete, die eine durchschnittliche Bevölkerungsdichte von 1 000 bis 2 000 Einwohnern je km² aufweisen oder in absehbarer Zeit aufweisen werden. Als Beispiele seien die Städte Ratingen, Velbert, Moers, Bergisch-Gladbach, Stolberg, Marl, Dorsten, Hamm und Iserlohn genannt.

Solitäre Verdichtungsgebiete: Hierzu gehören Städte, die außerhalb von Ballungskernen und Ballungsrandzonen liegen, aber Erscheinungsformen siedlungsmäßiger Verdichtung aufweisen, die denen der Ballungskerne und Ballungsrandzonen vergleichbar sind. Dies sind die Städte Münster, Bielefeld, Paderborn und Siegen.

Ländliche Zonen sind Gebiete mit überwiegend ländlicher Raumstruktur, die eine durchschnittliche Bevölkerungsdichte von weniger als 1 000 Einwohnern je km² aufweisen und durch eine aufgelockerte Verteilung städtischer und dörflicher Siedlungen gekennzeichnet sind. Dazu gehören u. a. aus dem Regierungsbezirk Düsseldorf die Städte und Gemeinden des Kreises Kleve und aus dem Regierungsbezirk Köln die der Kreise Düren, Euskirchen und des Oberbergischen Kreises. Aus dem Regierungsbezirk Münster gehören u. a. die Städte und Gemeinden der



Kreise Borken, Coesfeld, Steinfurt und Warendorf in diese (LEP-) Zone. Der Regierungsbezirk Detmold ist mit allen kreisangehörigen Gemeinden (außer Paderborn, Stadt) vertreten, während aus dem Regierungsbezirk Arnsberg die Kreise Hochsauerland, Olpe, Siegen-Wittgenstein (außer Siegen, Stadt) und Soest zu dieser zonalen Kategorie zählen.

Eine regionale Auswertung unter dem Gesichtspunkt dieser räumlichen Gliederung ergibt u. a. folgendes Ergebnis: Mehr als die Hälfte (55,1 %) aller Landesbeschäftigten (227 573) sowie 56 % aller Vollzeitbeschäftigten (172 450) waren demnach mit ihrem Dienstort den Ballungsgebieten und 12,1 % (49 986) aller bzw. 11,8 % (36 540) der Vollzeitbeschäftigten den Ballungsrandzonen zuzuordnen, während 21,6 % bzw. 66 421 Vollzeitbeschäftigte und 22,0 % (90 931) aller Beschäftigten des Landes ihren Dienstort in ländlichen Zonen hatten. Auf die (Städte der) solitären Verdichtungsgebiete entfielen 32 753 (10,6 %) Vollzeiträfte und 44 408 (10,8 %) aller Landesbeschäftigten des unmittelbaren Dienstes.

Ein ähnliches Bild trifft auch für die Teilzeitbeschäftigungsverhältnisse zu. So hatten 52,6 % und damit 55 123

2) Beschäftigte mit weniger als der Hälfte der regelmäßigen Wochenarbeitszeit eines Vollzeitbeschäftigten

Personen ihren Dienstort in Ballungsgebieten, 13 446 bzw. 12,8 % in Ballungsrandzonen; 24 510 Personen (23,4 %) gingen ihrer Teilzeitbeschäftigung in einem Dienstort in ländlichen Zonen nach; 11 655 Teilzeitbeschäftigte bzw. 11,1 % hatten ihre Arbeitsstätte in den „Solitärstädten“.

Altersstruktur

Eine Auswertung der vorliegenden Daten zum Durchschnittsalter der Vollzeitbeschäftigten in der Gliederung nach dem Dienstort in kreisfreien Städten und Kreisen des Landes NRW, die naturgemäß geprägt wird durch die (personelle) Struktur und Aufgabe der regional ansässigen Behörden und Einrichtungen des Landes, ergibt folgendes Bild:

Während das Durchschnittsalter der Vollzeitlandesbeschäftigten bei 43 Jahren lag, betrug es bei Beschäftigten mit Dienstort „Kreisfreie Stadt“ 42 und bei denen mit Dienstort in einer „Stadt/Gemeinde“ eines Kreises 44 Jahre.

Die Spannweite des Durchschnittsalters der Vollzeiträfte mit Dienstort „kreisfreie Stadt“ reichte von 39 bis 46 Jahren. Die durchschnittlich jüngsten Vollzeitbeschäftigten waren in Aachen (39 Jahre) tätig, gefolgt von Bonn, Köln, Gelsenkirchen

**Personal des Landes in NRW am 30. Juni 2000 nach dem Dienstort
von Haupt- und Nebenstellen einer Dienststelle**

Sitz der Dienststelle	Vollzeitbeschäftigte		Rangfolgeziffer zum Durchschnittsalter
	Anzahl	Durchschnittsalter in Jahren	
Reg.-Bez. Düsseldorf			
Kreisfreie Städte			
Düsseldorf	27 688	42	40
Duisburg	9 136	42	40
Essen	13 050	42	40
Krefeld	3 984	43	30
Mönchengladbach	4 415	43	30
Mülheim an der Ruhr	1 619	45	4
Oberhausen	2 556	44	15
Remscheid	1 514	44	15
Solingen	1 729	43	30
Wuppertal	6 908	42	40
Kreise			
Kleve	3 959	43	30
Mettmann	4 085	45	4
Neuss	4 441	44	15
Viersen	2 941	44	15
Wesel	4 851	44	15
Reg.-Bez. Köln			
Kreisfreie Städte			
Aachen	15 035	39	54
Bonn	11 506	41	49
Köln	26 558	41	49
Leverkusen	1 820	44	15
Kreise			
Aachen	2 414	45	4
Düren	3 370	41	49
Erftkreis	4 271	44	15
Euskirchen	2 138	45	4
Heinsberg	2 618	45	4
Oberbergischer Kreis	3 182	44	15
Rheinisch-Bergischer Kreis	2 239	45	4
Rhein-Sieg-Kreis	5 472	44	15
Reg.-Bez. Münster			
Kreisfreie Städte			
Bottrop	1 038	45	4
Gelsenkirchen	5 262	41	49
Münster	16 530	41	49

Noch: **Personal des Landes in NRW am 30. Juni 2000 nach dem Dienstort von Haupt- und Nebenstellen einer Dienststelle**

Sitz der Dienststelle	Vollzeitbeschäftigte		Rangfolgeziffer zum Durchschnittsalter
	Anzahl	Durchschnittsalter in Jahren	
noch: Reg.-Bez. Münster			
Kreise			
Borken	3 798	44	15
Coesfeld	2 276	46	1
Recklinghausen	7 567	44	15
Steinfurt	4 455	45	4
Warendorf	2 475	45	4
Reg.-Bez. Detmold			
Kreisfreie Stadt			
Bielefeld	8 565	42	40
Kreise			
Gütersloh	3 349	43	30
Herford	2 785	44	15
Höxter	1 754	46	1
Lippe	5 313	43	30
Minden-Lübbecke	3 768	44	15
Paderborn	5 271	42	40
Reg.-Bez. Arnsberg			
Kreisfreie Städte			
Bochum	9 944	42	40
Dortmund	12 574	42	40
Hagen	5 218	42	40
Hamm	3 176	43	30
Herne	1 643	46	1
Kreise			
Ennepe-Ruhr-Kreis	2 845	45	4
Hochsauerlandkreis	4 662	43	30
Märkischer	5 162	43	30
Olpe	1 476	44	15
Siegen-Wittgenstein	4 719	44	15
Soest	4 010	45	4
Unna	5 030	43	30
Nordrhein-Westfalen	308 164	43	x
davon			
kreisfreie Städte	191 468	42	x
Kreise	116 696	44	x

und Münster mit 41 Jahren sowie u. a. Düsseldorf, Duisburg, Essen und Wuppertal mit 42 Jahren. Die durchschnittlich ältesten Vollzeitkräfte arbeiteten in Herne (46 Jahre), gefolgt von Bottrop und Mülheim an der Ruhr (45 Jahre) sowie Leverkusen, Remscheid und Oberhausen mit 44 Jahren.

Bei den Vollzeitbeschäftigten mit Dienort in der Gemeinde eines Kreises ist festzustellen, dass das Durchschnittsalter auf dieser Regionalebene zwischen 41 und 46 Jahren lag, also knapp über der Spannweite in den kreisfreien Städten. Beschäftigte des Landes in Gemeinden des Kreises Düren waren im Durchschnitt 41 Jahre alt und damit die „Jüngsten“. Ihr Durchschnittsalter lag um 3 Jahre unter dem Durchschnitt dieser Körperschaftsgruppe. Es folgten die Beschäftigten mit Dienort in Gemeinden des Kreises Paderborn (42 Jahre) sowie der Kreise Kleve, Gütersloh, Lippe, Hochsauerland, Märkischer Kreis und Unna mit jeweils 43 Jahren. Die durchschnittlich ältesten Beschäftigten (46 Jahre) arbeiteten in kreisangehörigen Gemeinden der Kreise Höxter und Coesfeld. (Vgl. Tabelle Seite 46f.)

Die Frauenquote und die regionale Verteilung der Beschäftigten nach dem Geschlecht

Der Anteil der weiblichen Arbeitnehmer an der Gesamtzahl aller Beschäftigten im unmittelbaren Dienst des Landes NRW konnte sich gegenüber Mitte 1999 leicht um 0,6 Prozentpunkte auf 49,5 % und somit auf einen Stand von 204 563 Personen erhöhen.

Während die Frauenquote im Vollzeitbereich von 38,7 in 1999 auf 39,1 in 2000 leicht um 0,4 Punkte anstieg, konnte auch das traditionell hohe Niveau im Teilzeitbereich mit einer Quote über 80 gehalten werden. Zwar ging diese bei den T1-Arbeitnehmerinnen um einen Punkt auf 86 im Jahr 2000 zurück, gleichzeitig stieg aber die der weiblichen T2-Beschäftigten²⁾ um 3,3 Punkte auf 49,7 an.

Bezüglich der räumlichen Verteilung der weiblichen Beschäftigten auf die Dienorte der Behörden und Einrichtungen des Landes NRW können u. a. folgende Aussagen getroffen werden:

Von den 120 585 Vollzeit beschäftigten weiblichen Arbeitnehmern mit Dienort in NRW arbeiteten 76 744 (63,6 %) Personen in einer kreisfreien Stadt und 43 841 (36,4 %) in einer kreisangehörigen Gemeinde. Bei den weiblichen T1-Beschäftigten relativiert sich die regionale Verteilung erheblich: Von den insgesamt 75 616 Personen gingen 39 420 (52,1 %) ihrer Arbeit in einer Behörde/Einrichtung des Landes in einer kreisfreien Stadt und 36 196 (47,9 %) – also nahezu die gleiche Anzahl – in einer Gemeinde eines Kreises in NRW nach.

Auf konkrete regionale Einheiten bezogen ist festzustellen, dass die weiblichen Vollzeitkräfte in den kreisfreien Städten Köln (11 542), Düsseldorf (11 506) und Münster (10 401) am stärksten vertreten waren. Mit 415 Personen wies Bottrop den kleinsten Wert auf.

Unter den Kreisen lagen die höchsten Zahlen in Recklinghausen (2 717), Rhein-Sieg-Kreis (2 453) und im Märkischen Kreis (2 054); der kleinste Wert (472) war im Kreis Olpe festzustellen.

Im Einzelnen war bei den weiblichen T1-Beschäftigten u. a. folgende regionale Verteilung zu beobachten: Mit 5 097 Personen war diese Beschäftigtengruppe am Dienort Köln am stärksten vertreten, gefolgt von Düsseldorf (4 252) und Essen (3 201).

Bei den Kreisen belegte Recklinghausen den ersten Platz (2160). Es folgten die Kreise Mettmann (1697), Neuss (1606), Wesel (1584) und der Erftkreis (1464). Schlusslicht war – wie auch bei den weiblichen Vollzeitkräften – der Kreis Olpe (414).

Betrachtet man die Gesamtzahl aller Beschäftigten – also die Summe der Vollzeit- und Teilzeitkräfte – in der Gliederung nach dem Dienort und

Geschlecht, so ergab sich u. a. folgendes Ergebnis:

In Düsseldorf stand den 16 381 weiblichen Personen eine um ca. 1 180 höhere Zahl männlicher Beschäftigter in den Behörden und Einrichtungen des Landes gegenüber; in Aachen waren es rd. 2 400 mehr Männer als Frauen, die es auf insgesamt 9 317 Personen brachten. In Köln hatten die weiblichen Arbeitnehmer (17 795) mit einem Plus von rd. 570 und in Leverkusen mit einem „Überschuss“ von 406 Personen die Oberhand, während es im westfälischen Münster rd. 1 100 weibliche Personen (insgesamt 10 401) weniger als männliche waren.

In 11 der 31 Kreise war eine „Männerdominanz“ vorhanden (z. B. die Kreise Siegen-Wittgenstein mit +1 113, Paderborn +805, Hochsauerland +705, Soest +495 und Kleve +303), während in 18 Kreisen das weibliche Geschlecht dominierte (z. B. die Kreise Rhein-Sieg-Kreis mit +1 324, Mettmann +1 162, Erftkreis +895, Neuss +746 und Rheinisch-Bergischer Kreis +620). In zwei Kreisen (Lippe mit 3 567 männlichen und 3 575 weiblichen Arbeitnehmern und Borken mit 2 559 männlichen und 2 576 weiblichen Personen) war (fast) ein geschlechtsspezifisches Patt festzustellen. (Vgl. Tabelle Seite 50f.)

Beschäftigte nach Dienorten und Aufgabenbereichen

Eine Aufteilung des Personals nach kreisfreien Städten und Kreisen des Landes NRW und in der Gliederung nach den Hauptfunktionen entsprechend dem staatlichen Funktionsplan macht u. a. deutlich, dass „naturgemäß“ die Hauptfunktion 0 (Allgemeine Dienste) in allen hier dargestellten Regionaleinheiten mit Landesbediensteten besetzt war (insgesamt 122 501 Vollzeitbeschäftigte). Hinter diesem Aufgabenbereich verbergen sich Dienststellen wie z. B. der Landtag, der Ministerpräsident

3) Zwischenzeitlich aufgelöst; die durchgeführten Sprachkursmaßnahmen wurden nach Unna-Massen verlagert.

und die Staatskanzlei, die einzelnen Ministerien, die Bezirksregierungen, Polizeibehörden und -einrichtungen einschließlich der Polizeiführungsakademie in Münster, Gerichte der ordentlichen Gerichtsbarkeit, Staatsanwaltschaften, Gerichte der allgemeinen Verwaltungsgerichtsbarkeit sowie die Finanzgerichte Düsseldorf, Köln und Münster, Justizvollzugseinrichtungen, Oberfinanzdirektionen und Finanzämter, aber auch die Fortbildungsakademie des Innenministeriums, das Institut für öffentliche Verwaltung NRW und das Landesprüfungsamt für Verwaltungslaufbahnen in Hilden, das Institut der Feuerwehr in Münster u. a. m.

Ebenso verhielt es sich mit der Hauptfunktion 1 „Bildungswesen, Wissenschaft, Forschung, kulturelle Angelegenheiten“. Für die regionale Verteilung der Beschäftigten sind hier natürlich die Standorte der Dienststellen bzw. Einrichtungen der (wissenschaftlichen) Bildung (Schul- und Hochschulwesen, medizinische Einrichtungen etc.) ausschlaggebend. Universitätsstädte wie Köln mit einer Vollzeitbeschäftigtenzahl von 14 557, Aachen (11 143), Münster (9 983), Düsseldorf (9 937), Essen (7 989) und Dortmund (7 275) sind u. a. in diesem Zusammenhang zu nennen.

Insgesamt arbeitete in diesem Hauptaufgabenbereich – der neben den genannten Teilbereichen auch das Materialprüfungsamt und die Zentralstelle für die Vergabe von Studienplätzen (ZVS) in Dortmund, die staatlichen Bibliotheken, Archive und Büchereistellen sowie weitere Einrichtungen aus dem Bereich der Wissenschaft und Forschung außerhalb von Hochschulen beinhaltet – mit 57,4 % der zahlenmäßig größte Teil aller Vollzeitbeschäftigten 177 003 Personen.

Danach waren in den beiden o. g. Hauptfunktionen 299 504 Personen oder 97,2 % aller Vollzeitkräfte des unmittelbaren Landesdienstes tätig.

Für die „Soziale Sicherung, soziale Kriegsfolgeaufgaben und Wiedergutmachung“, die in der Hauptfunk-

tion 2 zusammengefasst sind, waren 3 563 (1,2 %) Vollzeitbeschäftigte eingesetzt. Hinsichtlich der regionalen Verteilung kann gesagt werden, dass z. B. in Düsseldorf 423 Bedienstete u. a. in der Landesanstalt für Arbeitsschutz (167) und im Versorgungsamt (198), in Köln 385 Personen u. a. im Versorgungsamt (253) und Staatlichen Amt für Arbeitsschutz (114), in Münster 418 Beschäftigte u. a. im Landesversorgungsamt NRW (258) und Versorgungsamt (157) und in Dortmund 335 Personen im Versorgungsamt (245) sowie im Staatlichen Amt für Arbeitsschutz (90) einer Vollzeitbeschäftigung nachgingen. Darüber hinaus waren 182 Personen in der Landesstelle für Aussiedler, Zuwanderer und ausländische Flüchtlinge in Unna-Massen und 5 Personen in der Außenstelle Waldbrö³⁾ tätig. Daneben arbeiteten zum Erhebungsstichtag 12 Personen im Landeszentrum für Zuwanderung NRW, das als eine eigenständige Abteilung der o. g. Landesstelle mit Sitz in Solingen geführt wird.

Ferner waren aus diesem Aufgabenbereich 55 Personen im Landesversicherungsamt, 143 im Versorgungsamt und 79 im Staatlichen Amt für Arbeitsschutz in der Ruhrgebietsstadt Essen als Vollzeitkraft tätig. Die übrigen 1 538 Vollzeitbeschäftigten verteilten sich auf die hier nicht im einzelnen genannten Einrichtungen und Ämter der Kriegsoferversorgung, die vornehmlich ihren Sitz in einer kreisfreien Stadt haben.

In der Hauptfunktion 3 „Gesundheit, Sport und Erholung“ waren 2 376 Personen Vollzeit beschäftigt. In Bielefeld, wo z. B. das Landesinstitut für den öffentlichen Gesundheitsdienst ansässig ist, versahen insgesamt 159 Vollzeitbeschäftigte (davon 87 Vollzeitkräfte im Staatlichen Umweltamt Bielefeld) ihre Arbeit. In Düsseldorf waren es 392 (u. a. 234 in der Zweigstelle des Landesumweltamtes Essen und 15 in der des Landesinstituts für den öffentlichen Gesundheitsdienst NRW Bielefeld sowie 125 im Staatlichen Umweltamt Düsseldorf), in

Münster 190 (u. a. 145 beim Staatlichen Umweltamt), in Essen 354 (Landesumweltamt) und in Hagen 164 Personen (Staatliches Umweltamt).

Im Sektor „Wohnungswesen, Raumordnung und kommunale Gemeinschaftsdienste“ (Hauptfunktion 4) gingen 364 Personen einer Vollzeitbeschäftigung nach. Unter regionalen Aspekten waren diese Beschäftigte ausschließlich dem Landesvermessungsamt in Bonn zuzuordnen.

Der Bereich „Ernährung, Landwirtschaft und Forsten“ (Hauptfunktion 5) war mit 589 Vollzeitkräften belegt. Davon waren 195 Mitarbeiterinnen/Mitarbeiter in Münster (u. a. 126 Personen des Chemischen Landes- und Staatlichen Veterinäruntersuchungsamtes), 68 in Düsseldorf (u. a. 64 Personen des Landesamtes für Ernährungswirtschaft und Jagd), 84 in Krefeld (Staatliches Veterinäruntersuchungsamt) und 78 in Warendorf (Nordrhein-Westfälisches Landgestüt) tätig. Bei den übrigen 159 Vollzeitkräften, deren Betätigungsfeld vornehmlich im Veterinär-, Agrar- und Forstbereich lag, wären als zuzuordnende Dienststellen neben den Bezirksregierungen auch die Direktoren der Landwirtschaftskammern Rheinland mit Sitz in Bonn und Westfalen mit Sitz in Münster zu nennen.

Von den insgesamt in dieser Hauptfunktion Vollzeittätigen waren 25,8 % (152) in Dienststellen im Regierungsbezirk Düsseldorf und 46,3 % (273) in solchen im Regierungsbezirk Münster beschäftigt. Die übrigen 164 Personen verteilten sich relativ gleichmäßig auf Dienststellen/Dienstorte in den übrigen Regierungsbezirken.

Bezüglich der regionalen Konzentration nach Körperschaftsgruppen kann festgestellt werden, dass 64,3 % der Beschäftigten dieses Hauptaufgabenbereiches in Dienststellen tätig waren, die ihren Sitz in einer kreisfreien Stadt hatten, während für gut ein Drittel (35,7 %) die Dienststelle in einem Kreisgebiet lag.

4) Ab 2001 erfolgt der Nachweis beim „Landesbetrieb Straßenbau NRW“.

Personal des Landes am 30. Juni 2000 nach dem Dienort von Haupt- und Nebenstellen einer Dienststelle								
Sitz der Dienststelle	Männliche Beschäftigte				Weibliche Beschäftigte			
	insgesamt	Vollzeit- beschäftigte	Teilzeit- beschäftigte (T1) ¹⁾²⁾	Teilzeit- beschäftigte (T2) ³⁾	insgesamt	Vollzeit- beschäftigte	Teilzeit- beschäftigte (T1) ¹⁾²⁾	Teilzeit- beschäftigte (T2) ³⁾
Reg.-Bez. Düsseldorf								
Kreisfreie Städte								
Düsseldorf	17 563	16 182	894	487	16 381	11 506	4 252	623
Duisburg	5 977	5 362	378	237	5 818	3 774	1 838	206
Essen	8 705	7 855	473	377	8 719	5 195	3 201	323
Krefeld	2 477	2 343	99	35	2 729	1 641	1 026	62
Mönchengladbach	2 763	2 624	118	21	2 801	1 791	959	51
Mülheim an der Ruhr	1 017	968	46	3	1 225	651	548	26
Oberhausen	1 541	1 449	82	10	1 896	1 107	757	32
Remscheid	981	928	49	4	1 039	586	425	28
Solingen	900	839	52	9	1 483	890	573	20
Wuppertal	5 050	4 379	339	332	4 546	2 529	1 736	281
Kreise								
Kleve	2 678	2 569	99	10	2 375	1 390	954	31
Mettmann	2 426	2 261	153	12	3 588	1 824	1 697	67
Neuss	2 748	2 595	142	11	3 494	1 846	1 606	42
Viersen	1 909	1 807	97	5	2 115	1 134	957	24
Wesel	3 160	2 996	153	11	3 503	1 855	1 584	64
Reg.-Bez. Köln								
Kreisfreie Städte								
Aachen	11 743	9 357	826	1 560	9 317	5 678	2 765	874
Bonn	8 430	6 846	767	817	8 019	4 660	2 686	673
Köln	17 227	15 016	1 120	1 091	17 795	11 542	5 097	1 156
Leverkusen	1 080	993	80	7	1 486	827	637	22
Kreise								
Aachen	1 453	1 362	81	10	1 826	1 052	750	24
Düren	2 109	1 990	95	24	2 349	1 380	921	48
Erftkreis	2 514	2 391	107	16	3 409	1 880	1 464	65
Euskirchen	1 371	1 304	61	6	1 536	834	675	27
Heinsberg	1 856	1 772	76	8	1 739	846	854	39
Oberbergischer	1 906	1 782	105	19	2 316	1 400	873	43
Rheinisch-Bergischer Kreis	1 367	1 272	82	13	1 987	967	992	28
Rhein-Sieg-Kreis	3 171	3 019	130	22	4 495	2 453	1 977	65
Reg.-Bez. Münster								
Kreisfreie Städte								
Bottrop	651	623	27	1	759	415	335	9
Gelsenkirchen	3 158	3 023	107	28	3 233	2 239	939	55
Münster	11 516	9 894	963	659	10 401	6 636	3 123	642

1) mit der Hälfte und mehr als der Hälfte der regelmäßigen Wochenarbeitszeit eines Vollzeitbeschäftigten – 2) einschl. Altersteilzeitbeschäftigten – 3) mit weniger als der Hälfte der regelmäßigen Wochenarbeitszeit eines Vollzeitbeschäftigten

Noch: Personal des Landes am 30. Juni 2000 nach dem Dienstort von Haupt- und Nebenstellen einer Dienststelle								
Sitz der Dienststelle	Männliche Beschäftigte				Weibliche Beschäftigte			
	insgesamt	Vollzeit- beschäftigte	Teilzeit- beschäftigte (T1) ¹⁾²⁾	Teilzeit- beschäftigte (T2) ³⁾	insgesamt	Vollzeit- beschäftigte	Teilzeit- beschäftigte (T1) ¹⁾²⁾	Teilzeit- beschäftigte (T2) ³⁾
noch: Reg.-Bez. Münster								
Kreise								
Borken	2 559	2 442	104	13	2 576	1 356	1 164	56
Coesfeld	1 617	1 504	105	8	1 738	772	921	45
Recklinghausen	5 131	4 850	262	19	4 989	2 717	2 160	112
Steinfurt	3 051	2 852	186	13	3 219	1 603	1 548	68
Warendorf	1 611	1 524	85	2	2 009	951	1 018	40
Reg.-Bez. Detmold								
Kreisfreie Stadt								
Bielefeld	6 427	5 543	518	366	5 230	3 022	1 880	328
Kreise								
Gütersloh	2 301	2 125	171	5	2 431	1 224	1 153	54
Herford	1 994	1 854	129	11	1 925	931	960	34
Höxter	1 259	1 185	60	14	1 151	569	556	26
Lippe	3 567	3 318	187	62	3 575	1 995	1 484	96
Minden-Lübbecke	2 496	2 370	116	10	2 529	1 398	1 097	34
Paderborn	4 283	3 519	200	564	3 478	1 752	1 385	341
Reg.-Bez. Arnsberg								
Kreisfreie Städte								
Bochum	7 638	6 522	631	485	5 806	3 422	1 981	403
Dortmund	8 981	7 849	566	566	7 598	4 725	2 497	376
Hagen	3 562	3 220	164	178	3 144	1 998	987	159
Hamm	1 994	1 934	58	2	2 017	1 242	743	32
Herne	1 029	975	46	8	1 122	668	435	19
Kreise								
Ennepe-Ruhr-Kreis	1 771	1 658	105	8	2 282	1 187	1 054	41
Hochsauerlandkreis	3 317	3 170	142	5	2 612	1 492	1 056	64
Märkischer	3 279	3 108	159	12	3 533	2 054	1 408	71
Olpe	1 045	1 004	38	3	901	472	414	15
Siegen-Wittgenstein	3 746	3 267	259	220	2 633	1 452	1 026	155
Soest	2 950	2 805	127	18	2 455	1 205	1 200	50
Unna	3 317	3 180	134	3	3 194	1 850	1 288	56
Nordrhein-Westfalen	208 372	187 579	12 353	8 440	204 526	120 585	75 616	8 325
davon								
kreisfreie Städte	130 410	114 724	8 403	7 283	122 564	76 744	39 420	6 400
Kreise	77 962	72 855	3 950	1 157	81 962	43 841	36 196	1 925
Dienstorte außerhalb NRWs	63	62	1	-	37	29	8	-
Personal des Landes NRW insgesamt	208 435	187 641	12 354	8 440	204 563	120 614	75 624	8 325

Anmerkungen Seite 50

**Personal des Landes in NRW am 30. Juni 2000 nach dem Dienstort
von Haupt- und Nebenstellen einer Dienststelle**

Lfd. Nr.	Sitz der Dienststelle	insgesamt			
			0	1	2
			allgemeine Dienste	Bildungswesen, Wissenschaft, Forschung, kulturelle Angelegenheiten	soziale Sicherung, soziale Kriegs- folgeaufgaben, Wiedergutmachung
	Kreisfreie Städte				
1	Düsseldorf	27 688	16 823	9 937	423
2	Duisburg	9 136	3 640	5 205	137
3	Essen	13 050	4 430	7 989	277
4	Krefeld	3 984	1 622	2 185	–
5	Mönchengladbach	4 415	2 018	2 334	59
6	Mülheim an der Ruhr	1 619	723	881	–
7	Oberhausen	2 556	1 055	1 501	–
8	Remscheid	1 514	682	832	–
9	Solingen	1 729	573	1 142	12
10	Wuppertal	6 908	3 273	3 465	164
	Kreise				
11	Kleve	3 959	1 794	2 113	–
12	Mettmann	4 085	1 671	2 408	–
13	Neuss	4 441	2 061	2 373	–
14	Viersen	2 941	1 324	1 615	–
15	Wesel	4 851	1 904	2 865	–
16	Reg.-Bez. Düsseldorf	92 876	43 593	46 845	1 072
	davon				
17	kreisfreie Städte	72 599	34 839	35 471	1 072
18	Kreise	20 277	8 754	11 374	–
	Kreisfreie Städte				
19	Aachen	15 035	3 539	11 143	220
20	Bonn	11 506	3 222	7 735	9
21	Köln	26 558	11 401	14 557	385
22	Leverkusen	1 820	662	1 157	–
	Kreise				
23	Aachen	2 414	514	1 856	–
24	Düren	3 370	1 318	1 973	3
25	Erftkreis	4 271	1 902	2 360	–
26	Euskirchen	2 138	854	1 238	–
27	Heinsberg	2 618	1 028	1 586	–
28	Oberbergischer	3 182	824	2 321	5
29	Rheinisch-Bergischer Kreis	2 239	809	1 409	–
30	Rhein-Sieg-Kreis	5 472	2 039	3 389	–
31	Reg.-Bez. Köln	80 623	28 112	50 724	622
	davon				
32	kreisfreie Städte	54 919	18 824	34 592	614
33	Kreise	25 704	9 288	16 132	8
	Kreisfreie Städte				
34	Bottrop	1 038	332	706	–
35	Gelsenkirchen	5 262	2 148	2 867	224
36	Münster	16 530	5 713	9 983	418

Vollzeitbeschäftigte						Lfd. Nr.
davon im Aufgabenbereich						
3	4	5	6	7	8	
Gesundheit, Sport und Erholung	Wohnungswesen, Raumordnung und kommunale Gemeinschaftsdienste	Ernährung, Landwirtschaft und Forsten	Energie- und Wasserwirtschaft, Gewerbe, Dienstleistungen	Verkehrs- und Nachrichtenwesen	Wirtschafts- unternehmen	
392	–	68	42	–	3	1
129	–	–	25	–	–	2
354	–	–	–	–	–	3
93	–	84	–	–	–	4
–	–	–	–	–	4	5
15	–	–	–	–	–	6
–	–	–	–	–	–	7
–	–	–	–	–	–	8
–	–	–	–	–	2	9
–	–	–	–	–	6	10
7	–	–	–	–	45	11
–	–	–	–	–	6	12
–	–	–	–	–	7	13
–	–	–	–	–	2	14
–	–	–	20	–	62	15
990	–	152	87	–	137	16
983	–	152	67	–	15	17
7	–	–	20	–	122	18
121	–	–	12	–	–	19
138	364	14	–	–	24	20
108	–	18	82	–	7	21
–	–	–	–	–	1	22
–	–	–	–	–	44	23
–	–	2	25	–	49	24
–	–	–	–	–	9	25
–	–	4	–	–	42	26
–	–	–	–	–	4	27
–	–	–	–	–	32	28
–	–	–	–	–	21	29
5	–	–	–	–	39	30
372	364	38	119	–	272	31
367	364	32	94	–	32	32
5	–	6	25	–	240	33
–	–	–	–	–	–	34
–	–	–	20	–	3	35
190	–	195	16	–	15	36

Noch: **Personal des Landes in NRW am 30. Juni 2000 nach dem Dienstort von Haupt- und Nebenstellen einer Dienststelle**

Lfd. Nr.	Sitz der Dienststelle	insgesamt			
			0	1	2
			allgemeine Dienste	Bildungswesen, Wissenschaft, Forschung, kulturelle Angelegenheiten	soziale Sicherung, soziale Kriegsfolgeaufgaben, Wiedergutmachung
37	Kreise Borken	3 798	1 204	2 583	–
38	Coesfeld	2 276	774	1 424	64
39	Recklinghausen	7 567	2 790	4 482	58
40	Steinfurt	4 455	1 391	3 048	–
41	Warendorf	2 475	908	1 477	–
42	Reg.-Bez. Münster	43 401	15 260	26 570	764
	davon				
43	kreisfreie Städte	22 830	8 193	13 556	642
44	Kreise	20 571	7 067	13 014	122
45	Kreisfreie Stadt Bielefeld	8 565	3 640	4 555	179
46	Kreise Gütersloh	3 349	1 485	1 860	–
47	Herford	2 785	1 086	1 697	–
48	Höxter	1 754	527	1 149	54
49	Lippe	5 313	2 203	2 987	47
50	Minden-Lübbecke	3 768	1 161	2 283	–
51	Paderborn	5 271	1 709	3 402	44
52	Reg.-Bez. Detmold	30 805	11 811	17 933	324
	davon				
53	kreisfreie Stadt	8 565	3 640	4 555	179
54	Kreise	22 240	8 171	13 378	145
55	Kreisfreie Städte Bochum	9 944	3 809	6 120	15
56	Dortmund	12 574	4 774	7 275	335
57	Hagen	5 218	2 315	2 715	–
58	Hamm	3 176	1 674	1 502	–
59	Herne	1 643	590	1 053	–
60	Kreise Ennepe-Ruhr-Kreis	2 845	1 046	1 784	–
61	Hochsauerlandkreis	4 662	2 234	2 130	59
62	Märkischer	5 162	1 742	3 369	–
63	Olpe	1 476	537	882	–
64	Siegen-Wittgenstein	4 719	1 333	3 265	24
65	Soest	4 010	1 471	2 210	166
66	Unna	5 030	2 200	2 626	182
57	Reg.-Bez. Arnsberg	60 459	23 725	34 931	781
	davon				
58	kreisfreie Städte	32 555	13 162	18 665	350
59	Kreise	27 904	10 563	16 266	431
60	Nordrhein-Westfalen	308 164	122 501	177 003	3 563
	davon				
61	kreisfreie Städte	191 468	78 658	106 839	2 857
62	Kreise	116 696	43 843	70 164	706

Vollzeitbeschäftigte						Lfd. Nr.
davon im Aufgabenbereich						
3	4	5	6	7	8	
Gesundheit, Sport und Erholung	Wohnungswesen, Raumordnung und kommunale Gemeinschaftsdienste	Ernährung, Landwirtschaft und Forsten	Energie- und Wasserwirtschaft, Gewerbe, Dienstleistungen	Verkehrs- und Nachrichtenwesen	Wirtschafts- unternehmen	
-	-	-	-	-	11	37
-	-	-	-	-	14	38
171	-	-	59	-	7	39
-	-	-	-	-	16	40
2	-	78	-	-	10	41
363	-	273	95	-	76	42
190	-	195	36	-	18	43
173	-	78	59	-	58	44
159	-	-	20	-	12	45
-	-	-	-	-	4	46
-	-	-	-	-	2	47
-	-	-	-	-	24	48
4	-	60	-	-	12	49
124	-	-	-	-	200	50
3	-	3	11	-	99	51
290	-	63	31	-	353	52
159	-	-	20	-	12	53
131	-	63	11	-	341	54
-	-	-	-	-	-	55
-	-	-	190	-	-	56
166	-	-	20	-	2	57
-	-	-	-	-	-	58
-	-	-	-	-	-	59
7	-	-	-	-	8	60
4	-	63	11	-	161	61
-	-	-	-	-	51	62
-	-	-	-	-	57	63
43	-	-	-	-	54	64
141	-	-	-	-	22	65
-	-	-	20	-	2	66
361	-	63	241	-	357	57
166	-	-	210	-	2	58
195	-	63	31	-	355	59
2 376	364	589	573	-	1 195	60
1 865	364	379	427	-	79	61
511	-	210	146	-	1 116	62

Im Bereich „Energie- und Wasserwirtschaft, Gewerbe, Dienstleistungen“ – dies ist aus haushaltssystematischer Sicht die Hauptfunktion 6 – waren insgesamt 573 Vollzeitbeschäftigte in den Dienststellen der Berg- und Eichverwaltung eingesetzt. Hinsichtlich der regionalen Verteilung wären z. B. die Städte Dortmund (190 Beschäftigte im Landesoberbergamt NRW und Eichamt), Köln (82 Beschäftigte in der Landeseichdirektion NRW und Eichamt Köln), Düsseldorf (42 Beschäftigte im Eichamt) sowie der Kreis Recklinghausen (59 Beschäftigte im Bergamt und Eichamt Recklinghausen sowie im Bergamt Marl) zu nennen.

In der Hauptfunktion 7 – Verkehrs- und Nachrichtenwesen – sind keine Beschäftigten des Landes ausgewiesen. Der Grund liegt darin, dass die dort beschriebenen Aufgaben und deren personalstatistischer Nachweis z. B. für den Bereich „Bundesfernstraßen/Bundesautobahnen“ in NRW von den Landschaftsverbänden (im Auftrag des Landes) wahrgenommen und somit im kommunalen Bereich⁴⁾ (im Abschnitt 66) erfolgt; des Weiteren sind in dieser Hauptfunktion z. B. auch Aufgaben des Bundes (z. B. Flugsicherung) aufgeführt, die – funktional – beim Bundespersonal nachgewiesen werden.

Der Sektor „Wirtschaftsunternehmen, allgemeines Grund- und Kapitalvermögen, Sondervermögen“ (Hauptfunktion 8), denen die Beschäftigten der Forstämter, des Staatsbads Oeynhausen sowie der Sondervermögen Haus Büren'scher Fonds, Bergischer Schulfonds Düsseldorf und Münster'scher Studiefonds zuzuordnen sind, umfasste zum Erhebungstichtag 1 195 Vollzeitbeschäftigte, die sich zu über 90 % auf die Kreise verteilten.

So konzentrierte sich im Regierungsbezirk Detmold, in dem 353 Personen (29,5 %) in Dienststellen des o.g. Aufgabenbereichs tätig waren, die Zahl der Beschäftigten auf die Kreise Minden-Lübbecke (200 Personen, davon 180 im Staatsbad Oeynhausen und 20 im Forstamt Minden) und Paderborn (99 Personen in den Staatlichen Forstämtern Neuenheerse und Paderborn sowie im Haus Büren'scher Fonds in Büren). Im Regierungsbezirk Arnsberg waren zum Erhebungstichtag 357 Personen bzw. 29,9 % aller Vollzeitbeschäftigten der o. g. Hauptfunktion tätig; darunter 161 Mitarbeiterinnen/Mitarbeiter (45,1 %) im Hochsauerlandkreis (u. a. im Staatlichen Forstamt Arnsberg und Brilon sowie bei den Leitern der Forstämter Olsberg und Schmallenberg der Landwirtschaftskammer (LwK) Westfalen-Lippe), 54 (15,1 %)

im Kreis Siegen-Wittgenstein (im Staatlichen Forstamt Hilchenbach und beim Leiter des Forstamtes Siegen der LwK Westfalen-Lippe), 57 (16 %) im Kreis Olpe (u. a. Leiter des Forstamtes Altenhundem der LwK Westfalen-Lippe in Lennestadt und im Staatlichen Forstamt Attendorn) sowie 51 (14,3 %) im Märkischen Kreis (u. a. Staatliches Forstamt Attendorn und Leiter des Forstamtes Lüdenscheid der LwK Westfalen-Lippe). (Vgl. Tabelle Seite 52ff.)

Neben diesen Grunddaten lassen sich auf der Basis der räumlichen Zuordnungseinheit „Gemeinde“ viele – auch nichtadministrative – Aggregate (wie z. B. zentralörtliche und zonale Gliederung, Mittel- und Oberbereiche, Wirtschaftsräume, Raumordnungsregionen und Größenklassen, aber auch frei wählbare Regionen in gemeindescharfer Abgrenzung) in Kombination mit dem erhobenen personalstatistischen Merkmalspektrum darstellen. Diesbezügliche Auswertungen und Datenwünsche können aus den im Hause geführten Fachinformationssystemen – standardisiert oder nutzerspezifisch – in angemessener Zeit vorgenommen werden.