

## 8 Der Mikrozensus und die Möglichkeiten seiner Regionalisierung

*Robert Herter-Eschweiler*

### 8.1 Einleitung

Der Mikrozensus ist eine gesetzlich angeordnete Erhebung der amtlichen Statistik mit dem Zweck, grundlegende statistische Daten über die Bevölkerungsstruktur, die wirtschaftliche und soziale Lage der Bevölkerung, der Familien, Lebensgemeinschaften und Haushalte, die Erwerbstätigkeit, Arbeitsuche, Aus- und Weiterbildung, Wohnverhältnisse und Gesundheit in tiefer fachlicher und regionaler Gliederung bereitzustellen. Die Ergebnisse hierüber gehen ein in Regierungsberichte, in das Jahresgutachten des Sachverständigenrates zur Begutachtung der gesamtwirtschaftlichen Entwicklung, bilden die Grundlage für die laufende Arbeitsmarkt- und Berufsforschung, den jährlichen Rentenversicherungsbericht der Bundesregierung und vieles andere mehr. Des Weiteren dient der Mikrozensus für eine Reihe kleinerer Erhebungen der empirischen Sozial- und Meinungsforschung sowie der amtlichen Statistik als Hochrechnungs-, Adjustierungs- und Kontrollinstrument. Durchgeführt wird der Mikrozensus jährlich seit 1957, in den neuen Ländern (einschließlich Berlin-Ost) seit 1991. Lediglich in den Jahren 1983 und 1984 wurde der Mikrozensus wegen der Rechtsunsicherheit der für 1983 geplanten Volkszählung ausgesetzt.

Da die Merkmale des Mikrozensus weitgehend identisch sind mit den Merkmalen der Arbeitskräfteerhebung der Europäischen Union, wurde diese 1968 in den Mikrozensus integriert. Die Ergebnisse der Arbeitskräfteerhebung werden unter anderem herangezogen für die Ermittlung der EU-Indikatoren zur Beschäftigungspolitik und zur nachhaltigen Entwicklung sowie für die Verteilung der Mittel aus den Regional- und Sozialfonds der Europäischen Union. Zudem bildet die Arbeitskräfteerhebung die Grundlage für die monatliche Berichterstattung zur Erwerbslosigkeit in den Mitgliedsstaaten der Europäischen Union.

Da das Erheben, Speichern, Verwenden und Weitergeben personenbezogener Daten durch öffentliche Stellen, als Eingriff in das grundrechtlich geschützte Recht jeder Person auf informationelle Selbstbestimmung, einer gesetzlichen Grundlage bedarf, die dem rechtstaatlichen Gebot der Normenklarheit entsprechen muss<sup>1</sup>, wurde bis zum Gesetz zur Durchführung einer Repräsentativstatistik über die Bevölkerung und die Arbeitsmarktbeteiligung sowie die Wohnsituation der Haushalte (Mikrozensusgesetz – MZG) vom 07.12.2016, die gesetzliche Grundlage stets zeitlich befristet. Mit dem Mikrozensusgesetz vom 07.12.2016 wird die bisherige Befristung aufgehoben. Als Grund hierfür wird die Integration der Haushaltstatistiken der Gemeinschaftsstatistik über Einkommen und Lebensbedingungen sowie der euro-

---

1 Siehe hierzu: Bundesverfassungsgericht, Urteil vom 15.12.1983 – 1 BvR 209/83 (Volkszählungsurteil).

päischen Erhebung über die Informations- und Kommunikationstechnologien in den Mikrozensus und deren unbefristeten Datenlieferungsverpflichtungen genannt (siehe Bundestagsdrucksache 18/9418: 22).

Die Gemeinschaftsstatistik über Einkommen und Lebensbedingungen (EU-SILC) wird seit 2005 in allen Mitgliedstaaten der Europäischen Union sowie mittlerweile auch in anderen europäischen Ländern (Schweiz, Norwegen, Island, Nordmazedonien, Türkei) erhoben. Sie dient der Erhebung aktueller und vergleichbarer multidimensionaler Quer- und Längsschnitt-Mikrodaten über Einkommen, Armut, soziale Ausgrenzung und Lebensbedingungen und ist im Europäischen Statistischen System (ESS) verankert. Weitere Themen der Befragung sind neben den verschiedenen Bestandteilen des Einkommens weitere wichtige Lebensbereiche wie etwa die soziale Ausgrenzung, die Wohnsituation, die Arbeitsbedingungen, die Bildung oder die Gesundheit. Besondere Bedeutung erhält EU-SILC mit der 2010 beginnenden Strategie Europa 2020. Die EU-SILC-Daten werden für die Überwachung der Armut und der sozialen Eingliederung in der EU verwendet. Auf der Grundlage der EU-SILC wurde das Ziel festgelegt, bis 2020 die Zahl der von Armut und sozialer Ausgrenzung Betroffenen um 20 Millionen zu verringern.

Die Informations- und Kommunikationstechnologien (IKT)-Erhebung ist eine europäische Erhebung, die jährlich seit dem Jahr 2002 durchgeführt wird. Mit der Befragung über die Nutzung von Informations- und Kommunikationstechnologien (IKT) in privaten Haushalten werden Daten zur Ausstattung und zur Nutzung moderner Informations- und Kommunikationstechnologien, insbesondere Computer und Internet, erhoben. Die Schwerpunkte liegen auf Fragen zu Art, Häufigkeit und ausgewählten Zwecken der Internetnutzung (zum Beispiel E-Commerce, E-Government, oder E-Learning). Zusätzlich werden Informationen darüber erhoben, welche Bedenken und Hindernisse Menschen von der Nutzung neuer Informations- und Kommunikationstechnologien abhalten.

Da die Integration tiefgreifende methodische und organisatorische Änderungen erfordert, wird die Integration der beiden Statistiken stufenweise umgesetzt. Für die Jahre 2017 bis 2019 wird der bisherige Mikrozensus mit nur geringen Anpassungen der Methoden und Erhebungsmerkmale weitergeführt. Als größter methodischer Eingriff ist hier die Fokussierung auf die Befragten, die in Privathaushalten leben, zu nennen. Für Personen in Gemeinschaftsunterkünften wird ab 2017 nur ein verkürztes Frageprogramm (Wohnrechtsstatus, Alter, Geschlecht, Familienstand, Staatsangehörigkeit und Hauptstatus) erhoben. Ab 2020 wird dann die Statistik über Einkommen und Lebensbedingungen, deren Erhebungsmerkmale weitgehend mit Auskunftspflicht in das Mikrozensusgesetz aufgenommen wurden, in den Mikrozensus als Unterstichprobe integriert. Der Auswahlatz der Unterstichprobe beträgt höchstens 12% und beschränkt sich auf Personen, die im Auswahlbezirk ihren Hauptwohnsitz haben.

Des Weiteren wird ab der Erhebung 2020 die Arbeitkräfteerhebung wiederum mit einer Unterstichprobe von höchstens 45% erhoben und es findet für diese Unterstichprobe eine schnellere Rotation statt. Zwar verbleiben die Auswahlseinheiten der Arbeitkräfteerhebung wie bisher viermal in der Erhebung, jedoch in einem 2-(2)-2-

Schema. Die Bewohner einer Auswahleinheit werden zu Beginn in zwei aufeinanderfolgenden Quartalen befragt, setzen dann zwei Quartale aus und werden anschließend wiederum in zwei aufeinanderfolgenden Quartalen befragt. Weitere Neuerungen, die ab der Erhebung 2020 umgesetzt werden, sind (a) die Einführung einer festen Berichtswoche und dass (b) mit Ausnahme der Zusatzprogramme „Wohnsituation“, „vertraglich vereinbarte Datenübertragungsrate“ und „Zahl der lebend geborene Kinder für Frauen im Alter von 15 bis 75 Jahren“, die mit dem vollen Stichprobenumfang von 1% erhoben werden, alle anderen Zusatzprogramme (Schichtarbeit, Gesundheitszustand, Krankenversicherungsschutz, weitere Eigenschaften der Haupttätigkeit für Erwerbstätige und Pendlereigenschaften) innerhalb der Unterstichprobe der Arbeitskräfteerhebung erfragt werden. Infolge der unterjährigen Rotation der Auswahlbezirke in der Arbeitskräfteerhebung werden die Zusatzprogramme allerdings nur bei der jeweils ersten Befragung innerhalb eines Jahres erhoben.

Um die Möglichkeiten einer Regionalisierung mit dem Mikrozensus zu verstehen, ist es wichtig, das Erhebungsdesign und die Hochrechnung in ihren Grundzügen zu kennen. Bevor also die regionalen Gliederungen, die im Mikrozensus enthalten sind, (Abschnitt 8.5) und in welcher Form eine Georeferenzierung mit dem Mikrozensus möglich ist (Abschnitt 8.6), kurz beschrieben werden, werden ebenfalls kurz das Erhebungsdesign des Mikrozensus (Abschnitt 8.2), die Hochrechnung (Abschnitt 8.3) und der Stichprobenfehler im Mikrozensus skizziert. Da für die, ab 2020 bzw. 2021 geltenden Neuerungen der Unterstichproben noch keine endgültigen Festlegungen zur Hochrechnung vorliegen, gilt weiterhin das seit 2005 eingeführte Hochrechnungsverfahren.

## 8.2 Das Erhebungsdesign des Mikrozensus

Auswahleinheiten sind im Mikrozensus Zusammenfassungen von Anschriften, einzelne Anschriften oder Teile von Anschriften, die aus den Volkszählungen (genauer den Gebäude- und Wohnungszählungen) ermittelt werden. In der Folge der Durchführung des Zensus 2011 wurde daher die Stichprobenziehung im Mikrozensus ab der Erhebung 2016 auf eine neue Auswahlgrundlage umgestellt. Das 1990 eingeführte Erhebungsdesign (vgl. Meyer 1994) wurde mit wenigen Modifikationen beibehalten. In diesem Beitrag wird nur eine kurze Übersicht über das Erhebungsdesign gegeben, eine detaillierte Beschreibung geben Bihler und Zimmermann (2016).

### 8.2.1 Regionale Schichtung

In regionalstatistischer Hinsicht sollte das Erhebungsdesign des Mikrozensus eine möglichst tiefgehende Ergebnisdarstellung mit genügender Genauigkeit ermöglichen. Zu diesem Zweck wurden, mit Ausnahme von Hamburg, Berlin und München, bei denen eine innerstädtische Unterteilung vorgenommen wurde, auf der Ebene der kreisfreien Städte und der Landkreise regionale Schichten gebildet, die in der Regel 200.000 oder mehr Einwohner umfassen. Auch hier ist die Schichtbildung von zwei Ausnahmen geprägt. In der Region Hannover bilden die Stadt Hannover und

das Umland Hannover jeweils eigene regionale Schichten. Gleiches gilt für die Städtereion Aachen. Die Stadt Aachen und das Umland Aachen bilden jeweils eigene regionale Schichten. Gegenwärtig bestehen 243 regionale Schichten mit durchschnittlich rund 340.000 Einwohner (Stand: 31.12.2016).

Innerhalb der regionalen Schichten konnte noch eine weitere Untergliederung vorgenommen werden, die ebenfalls mit wenigen Ausnahmen kreistrennscharf gebildet wurde. Derzeit bestehen 365 solcher sogenannter regionaler Schichtuntergruppen mit durchschnittlich 226 00 Einwohnern (Stand: 31.12.2016). Auf dieser Ebene wird zwar eine Hochrechnung vorgenommen (siehe Abschnitt 8.3.3), es werden jedoch nur einige ausgewählte Grunddaten ausgewiesen, da der wichtigste Nutzen der regionalen Schichtuntergruppen in dem Regionalisierungseffekt der Stichprobenziehung gesehen wird.

Für die Hochrechnung (siehe Abschnitt 8.3) wurden die regionalen Schichten soweit zu sogenannten regionalen Anpassungsschichten zusammengefasst, dass diese durchschnittlich mindestens 500.000 Einwohner umfassen. Diese Zusammenfassung ist notwendig, um größere Verzerrungen der Ergebnisgenauigkeit auf Bundes- und Landesebene bzw. der Ebene von Regierungsbezirken oder NUTS 2-Regionen aufgrund des Hochrechnungsverfahrens zu vermeiden. Daher orientieren sich die regionalen Anpassungsschichten mit einer Ausnahme an den Regierungsbezirken bzw. NUTS 2-Regionen, d.h. es werden in der Regel keine bezirksübergreifenden regionale Anpassungsschichten gebildet. Aktuell gibt es 145 regionaler Anpassungsschichten mit durchschnittlich 570.000 Einwohnern (Stand: 31.12.2016).

## 8.2.2 Klumpung – Fachliche Schichtung

Neben der regionalen Schichtung ist eine weitere, fachliche Schichtung im Erhebungsdesign enthalten, mit der wie bisher die „unterschiedlichen Siedlungsstrukturen und Wohnformen vom Ein- oder Zweifamilienhaus über das Mehrfamilienhaus bis zur Gemeinschaftsunterkunft möglichst repräsentativ einbezogen werden“ soll (Müller 1992: 180; siehe Fußnote 2). Hierzu wurden vier Schichten nach Anschriftengrößenklassen auf der Basis der in der Auswahlgrundlage enthaltenen Hausnummern und der in einem Gebäude lebenden Personen gebildet. Anders als bei der Erstellung der Auswahlgrundlage auf Basis der Volkszählung 1987 durften gemäß § 23 Zensusgesetz 2011 für die Planung und Ziehung der Stichprobe auf Anschriftenebene die Zahl der Wohnungen und die Zahl der dort wohnenden Personen, auch in Sonderbereichen, verwendet werden. Das Schätzverfahren der Zuordnung eines Gebäudes zu einer fachlichen Schicht musste nicht mehr angewendet werden. Abweichungen sind aber in der Definition eines Sonderbereichs im Zensus und von Personen ohne eigene Haushaltsführung im Mikrozensus (den sogenannten Gemeinschaftsunterkünften) gegeben. So zählen im Zensus 2011 Studentenwohnheime zu den Sonderbereichen. Da die dort lebenden Personen normalerweise eine eigene Haushaltsführung aufweisen, werden sie im Mikrozensus nicht den Gemeinschaftsunterkünften zugeordnet. Anschriften von See- und Binnenschiffern, die im Zensus als Sonderfälle typisiert wurden, werden im Mikrozensus als Gemeinschaftsunter-

kunft ausgewiesen. Ebenso werden im Mikrozensus Anschriften ohne Wohnungen, an denen Personen wohnen, als Gemeinschaftsunterkünfte gekennzeichnet. Im Zensus hingegen wurden diese Fälle nicht als Sonderbereich deklariert.

Wie bisher wurden vier Anschriftengrößenklassen gebildet:

- Anschriftengrößenklasse 1: Anschriften mit 1 bis 4 Wohnungen werden mit einer Richtgröße von 12 Wohnungen zusammengefasst. Um die Streuung der Größe der Auswahleinheiten gering zu halten, wurde für die Zahl der Wohnungen nur ein Streubereich von – in der Regel – 10 bis 13 Wohnungen zugelassen. Da die Anschriftengrößenklasse 1 maximal 70 Personen umfassen soll, sind in Ausnahmefällen auch Zusammenfassungen von weniger als 10 Wohnungen zugelassen.
- Anschriftengrößenklasse 2: Anschriften mit 5 bis 10 Wohnungen. Sie bilden in der Regel eine eigene Auswahleinheit.
- Anschriftengrößenklasse 3: Anschriften mit 11 oder mehr Wohnungen werden mit einer Richtgröße von 6 Wohnungen geteilt. Jeder Zerlegungsteil bildet eine Auswahleinheit. Infolgedessen, dass es keine Wohnungsidentifikatoren gibt, kann die Lage der Zerlegungsteile in der Anschriftengrößenklasse 3 nicht vorab festgelegt werden. Sie kann nur in der späteren Feldarbeit erfolgen.
- Anschriftengrößenklasse 4: Anschriften ohne Wohnungen und Sonderbereiche mit den zuvor beschriebenen Abweichungen. Je nach Anzahl der gemeldeten Personen wurden die Anschriften geteilt oder zusammengefasst mit dem Ziel, durchschnittlich 15 Personen zu befragen.

### 8.2.3 Auswahltechnik

Die in Abschnitt 8.2.2 beschriebenen Klumpen bildeten die Auswahleinheiten, aus denen 20 1%-Stichproben gezogen wurden. Zur Verbesserung der regionalen Repräsentation erfolgte vor der Auswahl eine regionale Anordnung der Auswahleinheiten. Die Auswahl erfolgte nach folgendem Verfahren:

- Die Auswahleinheiten wurden nach regionaler Schichtuntergruppe, Kreis, Gemeindegrößenklasse, Gemeinde, Ortsteil, Postleitzahl, Straße, Hausnummer und Hausnummerzusatz sortiert.
- Je 100 in der Reihenfolge der Sortierung aufeinanderfolgende Auswahleinheiten bildeten eine sogenannte „Zone“. Den Auswahleinheiten innerhalb einer Zone wurden per Pseudo-Zufallsgenerator eine Zahl von 0 bis 99 zugeordnet. Um zyklisch wiederkehrenden Regelmäßigkeiten entgegenzuwirken<sup>2</sup>, wurde für jede Zone eine neue Permutation der Zahlen 0 bis 99 ermittelt.
- Auswahleinheiten mit gleicher Nummer, d.h. gleicher „Stichprobennummer“, wurden zu einer 1%-Stichprobe zusammengefasst. Damit waren 100 1%-Stichproben gegeben.

2 Bei systematischen Auswahlen, gezogen wird jedes n-te Element, kann es zu Verzerrungen kommen, wenn der Ziehungsabstand, mit dem ein Element ausgewählt wird, in einer ganzzahligen Beziehung zu den wiederkehrenden Merkmalsausprägungen steht.

- Je vier in der Reihenfolge der Sortierung aufeinanderfolgende Zonen bilden einen sogenannten „Block“. Analog zu den Zonen wurden den Blöcken per Pseudo-Zufallsgenerator die Zahlen 1 bis 4 zugeordnet. Auswahlseinheiten mit gleicher „Blocknummer“ bildeten ein sogenanntes „Rotationsviertel“. Jedes „Rotationsviertel“ wird nach viermaliger Befragung gegen ein neues ausgetauscht.
- Für die Generierung von Substichproben (Arbeitskräfteerhebung und deren Ad-hoc-Module, Einkommen und Lebensbedingungen sowie Informations- und Kommunikationstechnologie) wurden weitere Zerlegungen vorgenommen.
- Ab dem Erhebungsjahr 2020 wird die Befragung zu einer festgelegten Berichtswoche durchgeführt. Um den Befragungszeitpunkt bestimmen zu können, wurden jeweils vier aufeinanderfolgende „Blöcke“ mit einer Zufallspermutation der Zahlen 1 bis 4 versehen, die das Quartal kennzeichnen. Des Weiteren wurden für ein Wochenkennzeichen schichtübergreifend die Auswahlseinheiten mit einer Zufallspermutation der Ziffern 1 bis 13 belegt<sup>3</sup>. Der Befragungszeitpunkt für eine Auswahlseinheit „ergibt sich [...] dann aus dem Zusammenspiel von Stichproben- und Rotationsviertelnummer (Jahr) sowie der Kombination aus Quartals- und Wochenkennzeichen (Kalenderwoche)“ (Bihler, Zimmermann 2016: 26).

In einem solchen Auswahlverfahren, einer einfachen, geschichteten Klumpenauswahl, sollten die Klumpen heterogen und die Schichten homogen sein. Denn tendenziell vergrößert eine Klumpenauswahl den Stichprobenfehler im Vergleich zu einer einfachen Zufallsauswahl. Generell gilt: Je homogener die Klumpen hinsichtlich eines interessierenden Merkmals, je größer die Klumpen und je unterschiedlicher die Klumpengrößen sind, desto größer ist der Stichprobenfehler. Im Gegensatz zur Klumpung führt die Schichtung zu einer Verringerung des Stichprobenfehlers, wenn die Schichten hinsichtlich eines interessierenden Merkmals homogen sind bzw. es eine hohe Korrelation zwischen einem interessierenden Merkmal und dem oder den Schichtungsmerkmal(en) vorliegt. In Anbetracht des Mikrozensus als Mehrthemenstichprobe sind dem Schichtungseffekt Grenzen gesetzt. So ist der Hauptzweck der Schichtung weniger in der Reduktion des Stichprobenfehlers zu sehen, als vielmehr in der regionalen Repräsentation der Mikrozensusergebnisse.

#### 8.2.4 Partielle Rotation

Um die Ergebnisgenauigkeit von Veränderungen zu erhöhen und um Kosten zu sparen, wurde bereits im Oktober 1962 eine sogenannte partielle Rotation der Auswahlseinheiten eingeführt. Danach wird nach vier Befragungen in einer Auswahlseinheit die Auswahlseinheit gegen eine neue ausgetauscht. Dieser Grundsatz bleibt auch dann erhalten, wenn die Umstellung der Arbeitskräfteerhebung auf eine unterjährige Rotation ab dem Erhebungsjahr 2020 eingeführt wird. Damit sind für einzelne Unterstichproben im Mikrozensus unterschiedliche Rotationsgeschwindigkeiten geben.

---

3 Die Möglichkeit von Quartalen mit 14 Wochen wurde dabei mit berücksichtigt.



Wie in der Einleitung beschrieben, verbleiben die Auswahleinheiten der Arbeitskräfteerhebung wie bisher viermal in der Erhebung, jedoch in einem 2-(2)-2- Schema. Die Bewohner einer Auswahleinheit werden zu Beginn in zwei aufeinanderfolgenden Quartalen befragt, setzen dann zwei Quartale aus und werden anschließend wiederum in zwei aufeinanderfolgenden Quartalen befragt. Für die übrigen Unterstichproben (Einkommen und Lebensbedingungen sowie Informations- und Kommunikationstechnologie) und die Auswahleinheiten, die zu keiner Unterstichprobe gehören, bleibt es bei der bisherigen Rotationsgeschwindigkeit. Zwischen den einzelnen Befragungszeitpunkten liegen jeweils 12 Monate.

### 8.2.5 Aktualisierung

Zieleinheit des Mikrozensus ist eine Anschrift. Damit jede Anschrift mit einer angebbaren Wahrscheinlichkeit ausgewählt werden kann, bedarf es einer Aktualisierung neu gebauter Wohnungen. Diese Aktualisierung wird jährlich durchgeführt und erfolgt über die Meldungen zur Bautätigkeitsstatistik. Die gemeldeten Anschriften werden in die bereits erwähnten Größenklassen eingeteilt und aus ihnen, ebenfalls wie bereits geschildert, Auswahleinheiten gebildet. Gegenüber der Auswahl auf der Basis des Zensus 2011 ergeben sich die folgenden Modifikationen: Die Auswahleinheiten der ersten Anschriftengrößenklasse, die mit 1 bis 4 Wohnungen gebildet werden, haben als Richtwert 6 Wohnungen (nicht 12 wie in der Grundausswahl). Die dritte Anschriftengrößenklasse beginnt bereits ab 9 Wohnungen pro Gebäude. Schließlich können Gemeinschaftsunterkünfte aus den Meldungen zur Bautätigkeitsstatistik direkt zugeordnet werden. Damit sind die Auswahleinheiten aus allen Anschriftengrößenklassen annähernd gleich groß und werden pro regionaler Schicht in nur einer Schicht, der sogenannten Neubauschicht zusammengefasst. Die systematische Auswahl mit Zufallsstart erfolgt kontinuierlich über die Jahre hinweg, wobei die Auswahleinheiten primär nach Aktualisierungsjahr und sekundär nach regionalen Gesichtspunkten sortiert werden. Rotationsviertel und weitere Teilstichproben werden ebenfalls systematisch mit Zufallsstart – analog zur Grundausswahl – ermittelt.

## 8.3 Die Hochrechnung des Mikrozensus

Der Mikrozensus ist zwar eine Stichprobe mit Auskunftspflicht, dennoch kann auch im Mikrozensus der Stichprobenplan nicht exakt eingehalten werden. So werden beispielsweise nicht alle Interviews in dem Quartal durchgeführt, in dem sie nach dem Stichprobenplan hätten durchgeführt werden sollen. Zudem dürften auch systematische Fehler vorkommen. Um diese systematischen Fehler durch unbekannte Antwortausfälle zu verringern, wird im Mikrozensus eine gebundene Hochrechnung mit Anpassungen an ausgewählte Bezugsmerkmale der laufenden Bevölkerungsstichprobe und weiteren Datenquellen durchgeführt. Mit der Anpassung werden des Weiteren die Zufallsfehler der Erhebungsmerkmale verringert, die mit den Bezugsmerkmalen hoch korrelieren. Auch wird eine gewisse Konsistenz mit den in der amtlichen Statistik verfügbaren Bevölkerungsdaten erreicht.

Das eine Antwort von einer befragten Person vorliegt, wird im Mikrozensus als zweistufiger, unabhängiger Prozess aufgefasst. Die Wahrscheinlichkeit der Antwortgewährung ist demnach das Produkt aus der Wahrscheinlichkeit in die Auswahl des Mikrozensus zu gelangen und der Antwortwahrscheinlichkeit. Dadurch kann die Hochrechnung in zwei getrennten Schritten erfolgen:

1. Schritt: gebundene Hochrechnung von der Netto- auf die Bruttostichprobe,
2. Schritt: gebundene Hochrechnung der Bruttostichprobe an ausgewählte Bezugsmerkmale der laufenden Bevölkerungsfortschreibung und weitere Datenquellen.

Unter den Rahmenbedingungen, dass die Hochrechnung – getrennt nach den einzelnen Bundesländern – quartalsweise und die Anpassung auf der Ebene der regionalen Anpassungsschichten erfolgen soll und es einen gleichen Hochrechnungsfaktor für alle Personen eines Haushalts geben soll, lässt sich ein einfaches Hochrechnungsverfahren auf der Basis des Verhältnisses von Soll/Ist disjunkter Bezugsmerkmale nicht realisieren. Bei einem solchen Verfahren würde es zu sehr kleinen Zellenbesetzungen in der Anpassungsmatrix kommen und dazu führen, dass die Schätzungen instabil würden. Für die Hochrechnung beider Schritte wird ab den Mikrozensen 2005 daher eine verallgemeinerte Regressionsschätzung eingesetzt, die eine Anpassung an getrennte Anpassungsmatrizen ermöglicht. Der Totalwert einer interessierenden Variablen Y lässt sich unter Einbezug von Bezugsmerkmalen und unter der Annahme eines linearen Zusammenhangs zwischen der interessierenden Variable Y und den Bezugsmerkmalen X, wie folgt in Matrixnotation angeben:

$$\begin{aligned} \hat{t}_y &= \hat{t}_{y,HT} + \hat{\mathbf{B}}(\mathbf{t}_x - \hat{\mathbf{t}}_{x,HT}) \\ &= \sum_{k=1}^n \left( 1 + (\mathbf{t}_x - \hat{\mathbf{t}}_{x,HT})' \left( \sum_{k=1}^n d_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \mathbf{x}_k \right) d_k y_k \\ &= \sum_{k=1}^n w_k y_k \end{aligned}$$

wobei:

$\hat{t}_y$  = geschätzter Totalwert der interessierenden Variablen Y

$\hat{t}_{y,HT}$  = geschätzter Totalwert der interessierenden Variablen Y in freier Hochrechnung

$$\hat{t}_{y,HT} = \sum_{k=1}^n d_k y_k$$

$d_k$  = Kehrwert der Wahrscheinlichkeit für Person k, in die Stichprobe zu gelangen und zu antworten

$$d_k = \frac{1}{\pi_k \hat{\Theta}_k}$$

$y_k$  = Wert der Variablen Y für Person k

$n$  = Stichprobenumfang der Nettostichprobe



$\pi_k$  = Ziehungswahrscheinlichkeit für Person k

$\hat{\Theta}_k$  = geschätzte Antwortwahrscheinlichkeit für Person k

$\mathbf{t}_x$  = Vektor der bekannten Bezugsmerkmale, an die eine Anpassung erfolgen soll

$\hat{\mathbf{t}}_{x,HT}$  = Vektor der geschätzten Bezugsmerkmale in freier Hochrechnung

$\hat{\mathbf{B}}$  = geschätzte Regressionskoeffizienten:

$$\hat{\mathbf{B}} = \left( \sum_{k=1}^n d_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \left( \sum_{k=1}^n d_k \mathbf{x}_k y_k \right)$$

$\mathbf{x}_k$  = Vektor der Bezugsmerkmale für Person k

$w_k$  = gebundener Hochrechnungsfaktor für Person k

Der verallgemeinerte Regressionsschätzer lässt sich also darstellen als Horvitz-Thompson-Schätzer (HT-Schätzer), der um die gewichtete Abweichung zwischen den bekannten Totalwerten der Bezugsmerkmale und den korrespondierenden HT-Schätzungen korrigiert wird. Für Details der verallgemeinerten Regressionsschätzung im Mikrozensus siehe: Afentakis, Bihler 2005: 1042 und folgende Seite.

### 8.3.1 Die gebundene Hochrechnung von der Netto- zur Bruttostichprobe

Bei der gebundenen Hochrechnung von der Netto- auf die Bruttostichprobe, der sogenannten Kompensation, werden die befragten Haushalte an die insgesamt zu befragenden Haushalte (= befragte Haushalte + ausgefallene Haushalte) unter Bezugnahme von Hilfsvariablen angepasst. Von den Hilfsvariablen wird angenommen, dass sie mit der Antwortwahrscheinlichkeit eines Haushalts korrelieren. Infolge zeitlicher Verschiebungen in der Realisierung der zu einem bestimmten Zeitpunkt vorgesehenen Interviews, liegt zwar keine Übereinstimmung zwischen Auswahl- und Befragungsquartal vor, „es wird jedoch davon ausgegangen, dass sich die ausgefallenen Haushalte zwischen den Berichtsquartalen in ihrer Struktur nur zufällig voneinander unterscheiden und daher die zeitlich verschobene Berechnung der Kompensationsfaktoren keine systematischen Auswirkungen auf die Mikrozensusergebnisse hat“ (Afentakis, Bihler 2005: 1043).

Die Auswahl der Hilfsvariablen erfolgte unter zwei Gesichtspunkten:

- die Hilfsvariablen sollen eine möglichst hohe Korrelation mit der Ausfallwahrscheinlichkeit aufweisen und
- die Hilfsvariablen müssen für die ausgefallenen Haushalte bekannt sein.

Insbesondere der letzte Aspekt schränkt, da über die ausgefallenen Haushalte keine Fremdeinkünfte eingeholt werden dürfen, die Auswahl der Hilfsvariablen ein. In der nachfolgenden Tabelle sind die Hilfsvariablen mit ihren Ausprägungen aufgeführt. Des Weiteren kann der Tabelle entnommen werden auf welcher regionalen Einheit die Hilfsvariablen berücksichtigt werden.

Tabelle 8.3.1-01 Grundmodell für die Kompensation der Antwortausfälle

	Regionale Ebene	Kompensationsterme
	<i>Privathaushalte</i>	
	Bundesland	- Rotationsviertel - Neubauschicht (Ja/Nein)
In Kompensationsklassen mit weniger als zehn Haushalten wird die Anpassung auf der nächst höheren regionalen Einheit vorgenommen.	Regionale Anpassungsschicht	- Haushaltsgröße (Ein-, Zwei-, Drei- und Mehrpersonenhaushalte) - Staatsangehörigkeit der Haushaltsbezugs-person (deutsch/nicht deutsch)
	evtl.: Regierungsbezirk	- Wohnsitz der Haushaltsbezugs-person (Hauptwohnung/Nebenwohnung)
	Bundesland	
		Zusätzlich für Einpersonenhaushalte: - Geschlecht - Alter (unter 60 Jahre/60 Jahre oder älter)
	Regionale Schichtuntergruppe	- Privathaushalte insgesamt
	<i>Gemeinschaftsunterkünfte</i>	
	Regierungsbezirk	- Zahl der Personen in Gemeinschaftsunterkünfthen

Quelle: Afentakis, Bihler 2005: 1044, modifizierte und um die Gemeinschaftsunterkünfte erweiterte Darstellung.

Bei der vorstehenden Auflistung handelt es sich um das sogenannte Grundmodell der Kompensation, das nur dann zur Anwendung kommt, wenn jede Anpassungszelle mit mindestens zehn Haushalten besetzt ist. Liegt eine kleinere Besetzungszahl in einer Anpassungszelle vor, wird die betreffende Hilfsvariable auf der nächst höheren, regionalen Einheit angepasst. Weist eine Anpassungszelle auch auf der Ebene der Bundesländer noch eine zu geringe Besetzungszahl auf, wird die betreffende Hilfsvariable aus dem Modell entfernt. Das konkrete Kompensationsmodell variiert in den einzelnen Bundesländern dementsprechend in Abhängigkeit der Zellenbesetzungen in den Anpassungszellen.

Antwortausfälle von Personen aus Gemeinschaftsunterkünften werden bei der Kompensation besonders behandelt. Hier erfolgt lediglich eine Anpassung der befragten Personen an die insgesamt zu befragenden Personen aus Gemeinschaftsunterkünften.

Die Kehrwerte der im ersten Schritt ermittelten Faktoren bilden dann die Eingangsgewichte für den zweiten Schritt der Hochrechnung, der gebundenen Hochrechnung der Bruttostichprobe an Bezugsmerkmale der laufenden Bevölkerungsfortschreibung und weitere Datenquellen.

### 8.3.2 Die gebundene Hochrechnung der Bruttostichprobe an bekannte Bezugsmerkmale der Grundgesamtheit

Analog zu dem Verfahren der gebundenen Hochrechnung von der Netto- auf die Bruttostichprobe, erfolgt die gebundene Hochrechnung der Bruttostichprobe über ausgewählte Bezugsmerkmale der laufenden Bevölkerungsfortschreibung sowie weiterer Datenquellen. Auch hier erfolgt die Hochrechnung über getrennte Randverteilungen unter Heranziehung einer verallgemeinerten Regressionsschätzung. Ebenso wird eine Modellreduktion vorgenommen, wenn eine Anpassungsklasse mit weniger als zehn befragten Einheiten, hier Personen, besetzt ist. Das Grundmodell sowie die Datenquellen sind in der Tabelle 8.3.2-01 dargestellt.

Tabelle 8.3.2-01 Anpassungsgrundmodell für die Quartals- und Jahreshochrechnung

Regionale Ebene	Hochrechnungsterme	Quelle
Bundesland <sup>a)</sup>	- Alter (unter 15, 15 bis 44, 45 Jahre oder älter) differenziert nach Geschlecht	Laufende Bevölkerungsfortschreibung
	- Staatsangehörigkeit (deutsch, türkisch, EU, nicht EU) differenziert nach Geschlecht <sup>b)</sup>	Ausländerzentralregister
	- Zeit- und Berufssoldaten einschließlich Bundes- und Bereitschaftspolizei <sup>c)</sup>	Bundesministerium der Verteidigung, Bundesministerium des Innern, für Bau und Heimat; Bundespolizei
	- Grundwehr-/Zivildienstleistende <sup>d)</sup>	
	- Bevölkerung insgesamt je Monat	Laufende Bevölkerungsfortschreibung
Regierungsbezirk	- Staatsangehörigkeit (deutsch, nicht deutsch) differenziert nach Geschlecht	Laufende Bevölkerungsfortschreibung
Regionale Anpassungsschicht	- Bevölkerung insgesamt	Laufende Bevölkerungsfortschreibung

a) Anpassung an die Bevölkerung insgesamt, wenn in den Anpassungsklassen auf der Ebene der Bundesländer weniger als zehn befragte Personen vorliegen.

b) Zusammenfassung der Staatsangehörigkeit (deutsch/nicht deutsch) mit der Differenzierung nach dem Geschlecht, wenn in einer Anpassungsklasse weniger als zehn Personen befragt wurden.

c) Zusammenfassung von „Zeit- und Berufssoldaten einschließlich Bundes- und Bereitschaftspolizei“ wenn in einer Anpassungsklasse weniger als zehn befragte Haushalte vorliegen.

d) Die Grund- und Zivildienstleistenden sind nicht mehr Bestandteil der Hochrechnungsterme.

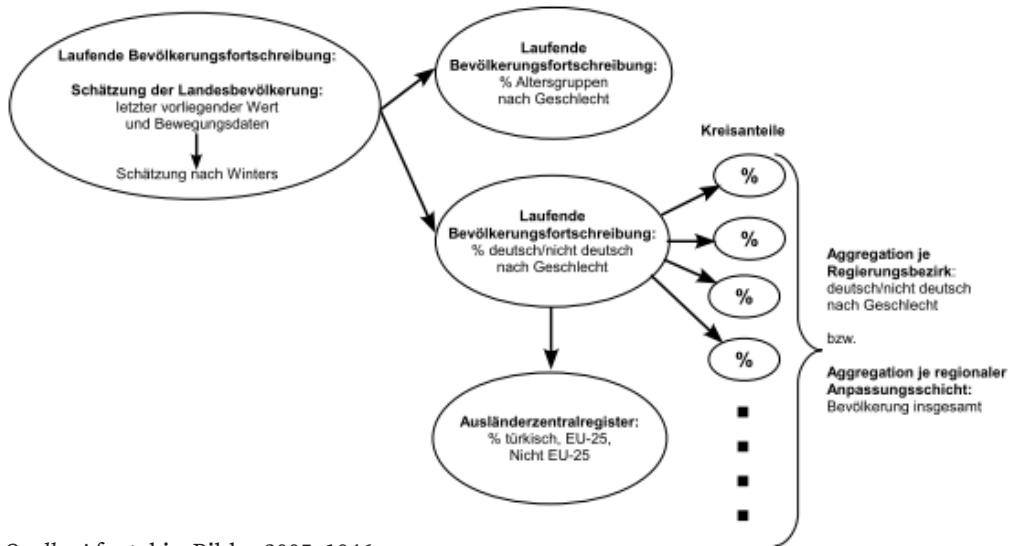
Quelle: Afentakis, Bihler 2005: 1045, leicht modifizierte Darstellung.

Bei der Anpassung nach der Staatsangehörigkeit beziehungsweise nach Gruppen von Staatsangehörigkeiten erfolgt diese nicht an die Eckwerte des Ausländerzentralregisters, sondern an dessen Struktur. Das heißt die Anteilswerte der Personen

mit türkischer Staatsangehörigkeit, mit einer Staatsangehörigkeit aus einem Land der Europäischen Union beziehungsweise mit einer sonstigen Staatsangehörigkeit werden auf die Ergebnisse der laufenden Bevölkerungsfortschreibung übertragen. Damit soll einerseits die Struktur des Ausländerzentralregisters genutzt werden und andererseits das Niveau der laufenden Bevölkerungsfortschreibung erhalten bleiben. Als notwendig wurde dieses Verfahren erachtet, da es zum Teil größere Abweichungen der Ausländerzahlen zwischen Ausländerzentralregister und laufender Bevölkerungsfortschreibung gibt, die Konsistenz der Eckwerte jedoch gewahrt bleiben soll.

Da alle Hochrechnungsterme nur auf der Personenebene zur Verfügung stehen, aber alle Personen eines Haushalts den gleichen Hochrechnungsfaktor erhalten sollen, werden für Personen in Privathaushalten „anstelle der Ausprägungen auf der Personenebene die Haushaltsmittelwerte genommen. [... Die Bezugsmerkmale sind somit] für alle Personen eines Haushalts gleich, was zu gleichen Hochrechnungsfaktoren innerhalb des Haushalts führt. [... Dennoch werden] durch die Mittelwertbildung [...] die Eckwerte weiterhin getroffen, wenn sie aus der Stichprobe hochgerechnet werden“ (Afentakis, Bihler 2005: 1042 und folgende Seite). Würde dieses Verfahren nicht gewählt, ergäben sich unterschiedliche Ergebnisse für einzelne Haushaltsmitglieder (beispielsweise die Anzahl verheiratet und zusammenlebender Ehepartner), die per Definition gleich sein müssten, da im Mikrozensus alle Personen eines Haushalts befragt werden. Des Weiteren werden, um große Schwankungen der Hochrechnungsfaktoren zu vermeiden, die Hochrechnungsfaktoren (HRF) auf das Intervall  $0,01 * EG \leq HRF \leq 5 * EG$  beschränkt (EG = Eingangsgewichte).

Wie eingangs erwähnt, erfolgt die Hochrechnung auf der Quartalsebene. Die Ergebnisse der laufenden Bevölkerungsfortschreibung und auch die Daten aus den anderen Datenquellen liegen zu diesem Zeitpunkt aber noch nicht vor und müssen daher geschätzt werden. Ausgehend von den letzten verfügbaren Werten für die Bundeslandbevölkerung aus der laufenden Bevölkerungsfortschreibung sowie Zeitreihen aus der Bevölkerungsbewegung (Geburten, Sterbefälle, Zu- und Fortzüge), werden aus den relativen Anteilen der entsprechenden Bevölkerungsgruppen die Bezugsmerkmale, an die der Mikrozensus angepasst werden soll, gebildet. Die Zeitreihen der Bevölkerungsbewegung umfassen etwa 24 Monate und die Anteilwerte werden mithilfe gleitender Mittelwerte ermittelt, um Sprünge am Jahresende zu vermeiden. Diese Sprünge „können infolge einer einmaligen Verbuchung der Einbürgerungen am Jahresende bzw. beim Übergang in eine neue Altersgruppe entstehen“ (Afentakis, Bihler 2005: 1046). Das Modell der Schätzung (zu Details siehe Afentakis, Bihler 2005: 1045 und folgende Seiten) der Bezugsmerkmale ist wie folgt gegeben:



Quelle: Afentakis, Bihler 2005: 1046.

Abbildung 8.3.2-01 Schätzmodell der Bezugsmerkmale für die Hochrechnung des Mikrozensus ab 2005

Die einzelnen Quartale werden als voneinander unabhängige Stichproben aufgefasst, sodass der Jahreshochrechnungsfaktor als arithmetischer Mittelwert der einzelnen Quartalshochrechnungsfaktoren gebildet wird.

### 8.3.3 Die gebundene Hochrechnung für tief regionalisierte Ergebnisse

Für tief regionalisierte Ergebnisse wird eine gesonderte Hochrechnung auf der Ebene der regionalen Schichtuntergruppen vorgenommen. Aufgrund der geringeren Anzahl befragter Personen auf dieser Ebene, können die Ergebnisse nicht in der fachlichen Tiefe dargestellt werden, wie dies für die Anpassungsschichten oder regionale Einheiten einer höheren Ebene gilt.

Die Hochrechnung erfolgt ebenfalls mit einem verallgemeinerten Regressionsmodell wie es für die Kompensation und die Hochrechnung der Quartalergebnisse herangezogen wird. Um Inkonsistenzen mit den Ergebnissen der „Standardhochrechnung“ möglichst klein zu halten, werden die Bezugsmerkmale der Anpassung aus dem hochgerechneten Mikrozensus genommen. Lediglich auf der Ebene der Schichtuntergruppen wird die laufende Bevölkerungsfortschreibung als externe Datenquelle herangezogen.

Ebenso wie bei der „Standardhochrechnung“ wird auch bei der tief regionalisierten Hochrechnung eine Mindestbesetzungszahl von zehn befragten Personen in einer Anpassungsklasse gefordert. Bei Nichterfüllung erfolgt dann gleichfalls eine Modellreduktion, sodass die nachfolgend aufgeführten Hochrechnungsterme wiederum das Grundmodell darstellen, deren konkrete Ausgestaltung in den einzelnen Bundesländern von der Besetzungszahl der Anpassungsklassen abhängig ist.

Tabelle 8.3.2-02 Anpassungsgrundmodell für die tief regionalisierte Hochrechnung

Regionale Ebene	Hochrechnungsterme	Quelle
Bundesland	<ul style="list-style-type: none"> <li>- Alter (unter 15, 15 bis 44, 45 Jahre oder älter) differenziert nach Geschlecht</li> <li>- Staatsangehörigkeit (deutsch, türkisch, EU, nicht EU) differenziert nach Geschlecht</li> <li>- Zeit- und Berufssoldaten einschließlich Bundes- und Bereitschaftspolizei, Grundwehrdienstleistende, Zivilbevölkerung</li> <li>- Bevölkerung insgesamt je Monat</li> <li>- Erwerbsstatus (erwerbstätig, erwerbslos, Nichterwerbsperson)</li> </ul>	Mikrozensus
Regionale Anpassungsschicht	<ul style="list-style-type: none"> <li>- Staatsangehörigkeit (deutsch, nicht deutsch) differenziert nach Geschlecht</li> </ul>	Mikrozensus
Regionale Schichtuntergruppe	<ul style="list-style-type: none"> <li>- Bevölkerung insgesamt</li> </ul>	Laufende Bevölkerungsfortschreibung

## 8.4 Der Stichprobenfehler des Mikrozensus

Die Hochrechnung im Mikrozensus erfolgt mit Hilfe eines verallgemeinerten Regressionschätzers. Unter Berücksichtigung des Erhebungsdesign als einstufige Klumpenauswahl und unter Berücksichtigung der Bezugsmerkmale, an die der Mikrozensus angepasst wird, kann die Varianz einer derart hochgerechneten, interessierenden Variablen Y wie folgt geschätzt werden:

$$\hat{V}(\hat{t}_y) = \sum_{h=1}^H \frac{N_h^2}{n_h} \left( 1 - \frac{n_h}{N_h} \right) \frac{1}{n_h - 1} \left( \sum_{i \in S_h} z_{hi}^2 - \frac{\sum_{i \in S_h} z_{hi}}{n_h} \right)$$

wobei:

$\hat{V}(\hat{t}_y)$  = Varianz des hochgerechneten Totalwerts  $\hat{t}_y$

$N_h$  = Anzahl der Auswahleinheiten in der h-ten Schicht in der Grundgesamtheit

$n_h$  = Anzahl der Auswahleinheiten in der h-ten Schicht in der Stichprobe

$S_h$  = Menge der Stichprobenauswahleinheiten in der h-ten Schicht

$z_{hi}$  = gewichtete Residuen der Summe aller Personen der i-ten Auswahleinheit in der h-ten Schicht:

$$z_{hi} = \sum_{i \in S_h} \frac{w_k}{d_k} (y_k - \hat{\mathbf{B}}' \mathbf{x}_k)$$



- $w_k$  = gebundener Hochrechnungsfaktor für Person k  
 $d_k$  = Kehrwert der Wahrscheinlichkeit für Person k, in die Stichprobe zu gelangen und zu antworten  
 $y_k$  = Wert der Variablen Y für Person k  
 $\hat{\mathbf{B}}$  = geschätzte Regressionskoeffizienten:  

$$\hat{\mathbf{B}} = \left( \sum_{k=1}^n d_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \left( \sum_{k=1}^n d_k \mathbf{x}_k y_k \right)$$
 $\mathbf{x}_k$  = Vektor der Bezugsmerkmale für Person k

Die Varianzschätzung kann nur annähernd bestimmt werden, da nicht alle Bedingungen zur Berechnung von Varianzen geschichteter Klumpenstichproben erfüllt sind. So müssen für eine korrekte Varianzschätzung geschichteter Klumpenstichproben jede Schicht mindestens zwei Klumpen (Auswahleinheiten) enthalten und innerhalb jeden Klumpens müssen mindestens zwei Haushalte befragt worden sein. In der Aufsplittung auf die einzelnen Quartale treten jedoch Schichten mit nur einem Klumpen und Klumpen mit nur einem befragten Haushalt auf. Um die betroffenen Schichten und Klumpen dennoch zu berücksichtigen werden Schichten mit nur einem Klumpen mit benachbarten Schichten innerhalb einer regionalen Anpassungsschicht oder eines Regierungsbezirkszusammengefasst. Ebenso erfolgt eine Zusammenfassung der benachbarten Klumpen.

Methodisch korrekt wäre des Weiteren die Varianzschätzung für die Schichten auf der Basis des Auswahlplans vorzunehmen. Da dies die Anzahl der Schichten mit nur einem Klumpen erhöhen würde, werden für die Varianzschätzung die Schichten auf der Basis der realisierten Interviews herangezogen. Es wird also angenommen, dass sich die Verschiebungen der Interviews zwischen den einzelnen Quartalen nicht systematisch voneinander unterscheiden.

Weiterhin müssten aus methodischer Sicht die jährlichen Neubausauswahlen als gesonderte Schicht behandelt werden. Aber auch diese Differenzierung würde zu sehr vielen kleinen Schichten führen. Die jährlichen Neubausauswahlen werden daher gemäß ihrer Gebäudegrößenklasse mit den Gebäudegrößenklassen der Grundausswahl zusammengelegt.

Die Varianzen werden je Land und Quartal berechnet. Aufgrund des geringen Auswahlprozents können die Quartalsstichproben in guter Annäherung als unabhängig betrachtet werden. Ebenso die Differenzierung nach den Bundesländern, da nach diesen geschichtet wurde. Die entsprechenden Bundesländer- und Quartalsvarianzen können also für Bundes- und Jahresschätzwerte addiert werden.

Da die Fehlerrechnung relativ aufwendig ist und nicht für jede Tabelle durchgeführt werden kann, wurde „durch eine Regressionsanalyse untersucht, ob der Zufallsfehler (abhängige Variable) mit dem Wert des hochgerechneten Tabellenfelds (unabhängige Variable) zusammenhängt. Bei einem ausreichend guten Zusammenhang lässt sich dann für andere Ergebnisse, für die keine Fehlerrechnung durchgeführt wurde, der Zufallsfehler abschätzen“ (Statistisches Bundesamt 2011: 229). In

der Abschätzung des quadrierten relativen Standardfehlers ( $v_g^2$ ) wird angenommen, dass  $v_g^2$  näherungsweise umgekehrt proportional zum hochgerechneten Wert  $\hat{n}$  des Tabellenfeldes  $g$  ist:

$$v_g^2 = a + \frac{b}{\hat{n}_g} + e_g$$

$a$  und  $b$  sind Konstanten,  $e_g$  das Residuum des hochgerechneten Wertes  $\hat{n}$  des Tabellenfeldes  $g$ .

Um auszuschließen, dass die Konstanten negative Werte annehmen können, wurde für die hochgerechnete Bevölkerungszahl in der Grundgesamtheit ( $bev$ ) der Zufallsfehler gleich 0 gesetzt. Dadurch ergibt sich für die Konstante  $a$ :  $a = -b/bev$  und es folgt für die Abschätzung des relativen Standardfehlers:

$$v_g = b \sqrt{\left( \frac{1}{\hat{n}_g} - \frac{1}{bev} \right)} + e_g$$

Die darauf abgeleitete Regression wurde für mehrere Gruppen durchgeführt. Die deutlichsten Unterschiede zeigten sich in den folgenden beiden Gruppen:

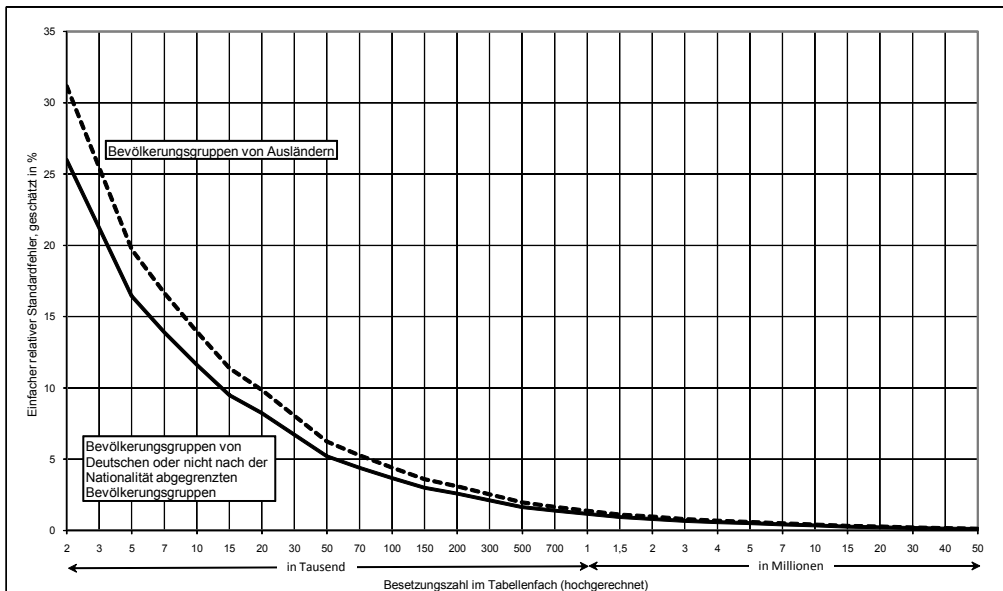
- Tabellenfelder für Deutsche oder Tabellenfelder für Bevölkerungsgruppen ohne Differenzierung nach der Nationalität, mit

$$v_g = 11,62531 \sqrt{\left( \frac{1}{\hat{n}_g} - \frac{1}{82,461Mill} \right)}$$

- Tabellenfelder von Ausländern, mit

$$v_g = 13,92822 \sqrt{\left( \frac{1}{\hat{n}_g} - \frac{1}{82,461Mill} \right)}$$

Basierend auf diesen Ergebnissen wurde die bisherige Fehlergrafik aktualisiert (siehe Abbildung 8.4-01). Es ist zu beachten, dass die Abbildung 8.4-01 im Gegensatz zu den, in den Fachserien publizierten Fehlergrafiken, keine logarithmierte Skalierung der Skalenachsen enthält.



Bei den ausgewählten regionalen Einheiten aus GV-ISys kann nochmals zwischen administrativen und nicht-administrativen Gebietseinheiten unterschieden werden. Die nachfolgende Tabelle gibt einen Überblick über die im Mikrozensus enthaltenen regionalen Einheiten.

Tabelle 8.5-01 Regionale Einheiten im Mikrozensus

Administrative Raumeinheiten	Mikrozensuspezifische Raumeinheiten	Nichtadministrative Raumeinheiten
Bundesländer	Regionale Anpassungsschichten	Planungsregionen
Regierungsbezirke	Regionale Schichten	Raumordnungsregionen
Kreise/kreisfreie Städte	Regionale Schichtuntergruppen	Gemeindegößenklassen
Gemeinden	Gebäudeschichten	Arbeitsmarktregionen
		Arbeitsamtsbezirke
		Siedlungsstrukturelle Gebietstypen der BBSR
		- Regionsgrundtypen der BBSR
		- Differenzierte Regionstypen der BBSR
		- Siedlungsstrukturelle Kreistypen der BBSR
		- Siedlungsstrukturelle Gemeindetypen der BBSR
		BIK-Regionen
		- BIK-Regionstypen
		- BIK-Strukturtypen
		- BIK-Gemeindegößenklassen
		NUTS 2-Regionen
		Stadt-Land-Typologie von Eurostat

Es kann mit dem Mikrozensus also zunächst auf bereits im Datensatz enthaltene regionale Einheiten zurückgegriffen werden. Darüber hinaus können, aufbauend auf diesen Einheiten, eigenständig neue regionale Einheiten gebildet werden, beispielsweise Großstadtregionen, Verdichtungsräume (siehe hierzu Abschnitt 5.3) oder die Arbeitsmarktregionen von Kropp und Schwengler (2011).

Der Regionalisierung sind jedoch Grenzen gesetzt, die einerseits durch die Anzahl befragter Personen in einer regionalen Einheit und andererseits durch die Hochrechnung gegeben sind. So stimmen für regionale Einheiten, für die eine Anpassung im Rahmen der Hochrechnung durchgeführt wird, die Verteilungen der Anpassungsmerkmale im Mikrozensus mit den entsprechenden Verteilungen der externen Bevölkerungsdaten überein. Die Übereinstimmung gilt ebenso für übergeordnete regionale Einheiten, die sich additiv aus den regionalen Einheiten, auf denen die Anpassung der Hochrechnung durchgeführt wurde, zusammensetzen. Erfolgt bei-

spielsweise die Hochrechnung auf der Ebene der regionalen Anpassungsschichten, so werden diese sowie die übergeordneten Einheiten, die sich additiv aus den regionalen Anpassungsschichten zusammensetzen (Regierungsbezirke – soweit vorhanden –, Bundesländer und das Bundesgebiet), verteilungstreu bezüglich der verwendeten Anpassungsmerkmale dargestellt. Für regionale Einheiten, die eine Untermenge der regionalen Einheiten sind, auf denen die Anpassung der Hochrechnung erfolgte, ist eine Übereinstimmung der Verteilungen der Anpassungsmerkmale in den untergeordneten regionalen Einheiten nicht zwingend gegeben. Die Verteilung einer übergeordneten regionalen Einheit kann das Resultat verschiedener Verteilungen der untergeordneten regionalen Einheiten sein. Beispielhaft sei dies an einem fiktiven Beispiel verdeutlicht, siehe Tabelle 8.5-02. In der Tabelle wird die Altersverteilung dreier Altersklassen in einer Anpassungsschicht dargestellt, die sich aus zwei untergeordneten Raumeinheiten zusammensetzt. Für die untergeordneten Raumeinheiten werden zwei mögliche Versionen der Altersverteilung ausgewiesen, die zu der übergeordneten Altersverteilung führen.

*Tabelle 8.5-02* Fiktives Beispiel einer Altersklassenverteilung in einer Anpassungsregion, die sich aus zwei untergeordneten regionalen Einheiten zusammensetzt

Altersklasse	Anpassungsregion									
	davon in der untergeordneten regionalen Einheit									
	Insgesamt		Altersverteilung: Version 1				Altersverteilung: Version 2			
			A		B		A		B	
n	%	n	%	n	%	n	%	n	%	
< 15 Jahre	50	14,29	25	14,29	25	14,29	20	10,53	30	18,75
15 – 44 Jahre	150	42,86	75	42,86	75	42,86	90	47,37	60	37,50
≥ 45 Jahre	150	52,86	75	52,86	75	52,86	80	42,11	70	43,75
Insgesamt	350	100	175	100	175	100	190	100	160	100

Wie der Tabelle 8.5-02 entnommen werden kann, können unterschiedliche Verteilungen in untergeordneten Raumeinheiten zu einer gleichen Verteilung auf der nächsthöheren Raumeinheit führen.

Unter der Annahme, dass die Abweichungen in den Untermengen zufällig streuen, können jedoch Auswertungen auf der Basis von Gebietstypen dieser Untermengen vorgenommen werden. Als Beispiel solcher Gebietstypen seien die politischen Gemeindegrößenklassen, die BIK-Gemeindegrößenklassen, die siedlungsstrukturellen Gebietstypen des BBSR oder die Stadt-Land-Typologie von Eurostat genannt.

Das Beispiel verdeutlicht auch, dass den tief regionalisierten Ergebnissen des Mikrozensus Grenzen gesetzt sind. Mit Ausnahme der Bevölkerung auf der Ebene der regionalen Schichtuntergruppen, die an die Ergebnisse der laufenden Bevölkerungsfortschreibung angepasst werden, erfolgt die Anpassung aller anderen

Anpassungsmerkmale an den hochgerechneten Ergebnissen des Mikrozensus. Die Verteilungen übergeordneter regionaler Einheiten werden also auf untergeordnete Raumeinheiten übertragen. Die Güte der tief regionalisierten Ergebnisse ist dementsprechend unter anderem auch abhängig von der Homogenität der Verteilungen der Anpassungsmerkmale der regionalen Schichtuntergruppen innerhalb der Raumeinheit, auf der die Anpassung in der Hochrechnung erfolgte.

Da der Mikrozensus dezentral durchgeführt wird, liegt die Verantwortung für die Daten bei den einzelnen Statistischen Landesämtern. Dementsprechend obliegt die Auswertung der Mikrozensusergebnisse unterhalb der Ebene der Bundesländer auch den einzelnen Statistischen Landesämtern.

## 8.6 Die Georeferenzierung im Mikrozensus

Der Statistische Verbund hat sich zum Ziel gesetzt, bis Ende 2019 alle georeferenzierbaren Statistiken zu georeferenzieren. Für den Mikrozensus werden bereits ab 2018 die Wohnanschriften mit Gitterzellen mit einem Raster von 100m x 100m versehen. Diese Informationen werden für die an einer Anschrift verfügbare maximale Datenübertragungsrate genutzt (siehe § 6 Absatz 1 Nummer 10 Buchstabe b des Mikrozensusgesetzes). Über weitere Nutzungsmöglichkeiten im Rahmen der amtlichen Statistik – insbesondere für Stichprobenerhebungen – wird derzeit diskutiert. In einer Machbarkeitsstudie wurde für die 6- bis 10-jährigen Kinder in Hessen die Entfernung zur nächst gelegenen Grundschule analysiert. Der Fokus der Analyse bestand weniger auf der inhaltlichen Interpretation der Ergebnisse als vielmehr auf den technischen Umsetzungsmöglichkeiten (siehe Statistisches Bundesamt 2018). Es sind aber nicht nur die technischen Umsetzungsmöglichkeiten zu klären, sondern auch die rechtlichen Rahmenbedingungen und wie die Veröffentlichung in der amtlichen Statistik gestaltet sein soll. Eine zentrale Frage, die unter rechtlichen Gesichtspunkten zu beantworten ist, lautet: Dürfen auf Ebene der Gitterzellen Informationen aus verschiedenen Statistiken miteinander kombiniert werden? Bezüglich der Art der Veröffentlichung werden in der amtlichen Statistik bisher kleinräumige kartographische Darstellungen als Rasterkarten eingesetzt (Atlas Agrarstatistik, Zensus-Atlas 2011), um die Lage eines Objektes in einem Einzugsgebiet oder, wie in der zuvor zitierten Machbarkeitsstudie, Entfernungszonen zwischen Geoobjekten darzustellen. Für stichprobenbasierte Statistiken stoßen solche Darstellungen allerdings an ihre Grenzen. Ob es darüber hinaus noch weitere Anwendungsmöglichkeiten in der amtlichen Statistik gibt, ist noch zu diskutieren. So hat Pforr (2018) erste Versuche unternommen, den Mikrozensus für Schätzungen von räumlichen Kontexten zu nutzen.

Eine weitere Option der Nutzung der Georeferenzierung im Mikrozensus kann in der Bildung innerstädtischer Raumtypen gesehen werden. In Abhängigkeit der Lage ließen sich beispielsweise in Bezug zum (Haupt-)Zentrum einer Stadt gemäß einer klassischen Zentrenkonzeption die Lagetypen „City“, „Cityrand“, „Innenstadtrand“ und „Stadtrand“ bilden (siehe Kapitel 6.2.2). Zwar dürfte infolge der Auswahltechnik eine entsprechende Verteilung der Auswahleinheiten in einer Stadt gegeben sein, aufgrund der Stichprobekonzeption und der Hochrechnung des Mikrozensus müsste



aber geprüft werden, ob solche Raumeinheiten für eine konkrete Stadt machbar sind. Unter der Annahme, dass die Abweichungen in den einzelnen Lagetypen zufällig streuen, wäre jedoch zu prüfen, ob für Stadttypen (siehe hierzu beispielsweise die Einteilung der Städte in der innerstädtischen Raumbeobachtung) Anwendungen möglich sind.

