



Dr. Nilgun Massih-Tehrani

ist Diplom-Soziologin und seit 2018 als Referentin im Sachgebiet „Zensus“ des Bayerischen Landesamts für Statistik tätig. Sie ist Mitarbeiterin im Teilprojekt „Haushaltegenerierung“ des Zensus 2021.



Adrian Reichert

hat in München Volkswirtschaftslehre und Statistik studiert. Im Rahmen des Zensus 2021 leitet er den Projektbereich „Gebäude- und Wohnungszählung, Haushaltegenerierung“ im Bayerischen Landesamt für Statistik.

DIE HAUSHALTEGENERIERUNG IM ZENSUS 2021

Dr. Nilgun Massih-Tehrani, Adrian Reichert

↘ **Schlüsselwörter:** Zensus 2021 – Haushaltegenerierung – registergestützter Zensus – Einwohnerzahl – Wohnhaushalt

ZUSAMMENFASSUNG

Ziel des Zensus 2021 ist neben der Ermittlung der Einwohnerzahl auch die Bereitstellung von Daten zu Anzahl und Struktur von Haushalten und deren Wohnsituation. Hierfür erstellt das Teilprojekt Haushaltegenerierung einen bundesweiten Einzeldatenbestand auf Personenebene, den sogenannten zensustypischen Datensatz. Dieser ermöglicht erhebungsteilübergreifende Auswertungen zu demografischen, haushaltsstatistischen und erwerbsstatistischen Daten sowie zu Wohnungs- und Gebäudedaten auf kleinräumiger Ebene. Auch im Zensus 2021 wird ein registergestütztes Verfahren umgesetzt. Für die Erstellung des zensustypischen Datensatzes ist daher die Integration unterschiedlicher Register- und Erhebungsdaten sowie die Korrektur der Über- und Untererfassungen der Melderegister auf Einzelpersonenebene erforderlich. Die für den Zensus 2011 entwickelten Verfahren der Haushaltegenerierung wurden für den Zensus 2021 maßgeblich weiterentwickelt.

↘ **Keywords:** 2021 Census – household generating procedure – register-based census – number of inhabitants – residential household

ABSTRACT

The objective of the 2021 Census is not only to determine the number of inhabitants, but also to provide data on the number and structure of households and their housing situation. To this end, a nationwide database at the level of individuals – the so-called census-typical data set – is created in the “household generating procedure” subproject. This census-typical data set permits the intertwined analysis of demographic, household and employment statistics, and of dwelling and building data, all at a small-area level. The 2021 Census, too, uses a register-based method. This means that the creation of the census-typical data set requires the integration of different register and survey data as well as the correction of overcoverage and undercoverage in the population register. The household generating procedure developed for the 2011 Census has been significantly refined for the 2021 Census.

1

Einleitung

Der folgende Beitrag gibt einen Einblick in die geplanten methodischen Weiterentwicklungen der Haushaltegenerierung im Rahmen des Zensus 2021. Dazu erfolgt zunächst eine kurze Einführung in die Problemstellung und die grundsätzliche Arbeitsweise der Haushaltegenerierung. Im Anschluss daran werden die wichtigsten methodischen Neuerungen in der Haushaltegenerierung skizziert.

Der letzte Zensus in Deutschland wurde zum Stichtag 9. Mai 2011 durchgeführt. Zur Vermeidung hoher Kosten, zur Entlastung der Befragten und im Hinblick auf die Aspekte Datenschutz und Datensparsamkeit entschied sich der Gesetzgeber dazu, die Volkszählung 2011 erstmals nicht mehr als Befragung aller Einwohnerinnen und Einwohner durchzuführen. Vielmehr wurden die Daten registergestützt – das heißt durch Auswertungen der Melderegister und anderer Verwaltungsregister – ermittelt. Befragungen sollten nur noch ergänzend erfolgen.¹ Auch der Zensus 2021 wird registergestützt durchgeführt werden.²

Der Zensus dient neben der Ermittlung der Einwohnerzahl auch der Bereitstellung fachlich und regional tief gegliederter Ergebnisse. Damit differenzierte Auswertungen auf kleinräumiger Ebene möglich werden, bedarf es der Zusammenführung der verschiedenen Datenquellen des Zensus. Die Datenintegration ist unter anderem Aufgabe der Haushaltegenerierung. Erst durch dieses Verfahren kann ein bundesweiter Einzeldatenbestand auf Personenebene zur Verfügung gestellt werden, der erhebungsteilübergreifende Auswertungen zu demografischen, haushaltsstatistischen und erwerbsstatistischen Daten sowie zu Wohnungs- und Gebäudedaten ermöglicht. So können beispielsweise nicht nur reine Haushaltsergebnisse (wie die Anzahl der Einpersonnen- und Mehrpersonenhaushalte), sondern auch erhebungs-

teilübergreifende Ergebnisse (wie die durchschnittliche Wohnungsgröße von Haushalten mit Kindern) ausgewiesen werden (Kreuzmair/Reisch, 2013).

Zunächst wird auf die allgemeine Zielsetzung und Arbeitsweise der Haushaltegenerierung eingegangen, bevor die methodischen Neuerungen der Haushaltegenerierung im Zensus 2021 erläutert werden.

2

Zielsetzung der Haushaltegenerierung

Zur Erstellung fachlich und regional tief gegliederter Ergebnisse führt die Haushaltegenerierung zum einen eine Integration verschiedener Datenquellen und zum anderen eine Synchronisation der integrierten Daten mit der berechneten Einwohnerzahl durch. Die Integration der Melderegister und der Gebäude- und Wohnungszählung ermöglicht es, Informationen über die Zahl und Struktur von Haushalten und deren Wohnsituation zu gewinnen. Zudem führt die Haushaltegenerierung die Korrektur der Über- und Untererfassungen (Karteileichen und Fehlbestände) der Melderegister auf Einzeldatenbasis durch. Datenintegration und -synchronisation unterliegen unterschiedlichen Herausforderungen:

Mit einem registergestützten Verfahren werden Daten zur Struktur und Anzahl von Wohnhaushalten nicht mehr primärstatistisch erhoben, sondern durch die Integration verschiedener Datenquellen (insbesondere Melderegister und Gebäude- und Wohnungszählung) ermittelt. Die Information zu Haushaltszusammenhängen liegt in den Melderegistern jedoch nicht vollständig vor. Gebäude- und wohnungsstatistische Informationen werden zwar primärstatistisch erhoben, liefern jedoch keine Information zu den konkreten Wohnverhältnissen von Haushalten in Deutschland. Kleinräumig gegliederte Haushalts- und Wohnungsbelegungsdaten können demnach nur durch die anschriftenweise Zusammenführung der Personendaten aus den Melderegistern zu Haushalten und durch die Verknüpfung mit den Gebäude- und Wohnungsdaten der Gebäude- und Wohnungszählung erreicht werden (Vorndran, 2004, hier: Seite 1245). Im Mittelpunkt der Haushaltegenerierung steht die statistische Einheit des Wohnhaushalts, der jedoch weder im Melderegister noch in der Gebäude- und Wohnungszählung direkt erfasst wird. Deshalb werden Verfahren

1 Entwurf eines Gesetzes zur Vorbereitung eines registergestützten Zensus einschließlich einer Gebäude- und Wohnungszählung 2011 (Zensusvorbereitungsgesetz 2011 – ZensVorbG 2011) vom 30. Mai 2007 (BT-Drucksache 16/5525); Gesetz über den registergestützten Zensus im Jahre 2011 (Zensusgesetz 2011 – ZensG 2011) vom 8. Juli 2009 (BGBl. I Seite 1781).

2 Entwurf eines Gesetzes zur Durchführung des Zensus im Jahr 2021 (Zensusgesetz 2021 – ZensG 2021) vom 25. März 2019 (BT-Drucksache 19/8693).

eingesetzt, mit denen Haushaltszusammenhänge auf Basis von Registerangaben nachgebildet und mit den primärstatistisch erhobenen Daten der Gebäude- und Wohnungszählung verknüpft werden.

Eine besondere Herausforderung hierbei liegt darin, mit der Unvollständigkeit sowie Mehrdeutigkeit der verschiedenen Datenquellen umzugehen. Daher ist es nicht ausreichend, eine reine Datenzusammenführung im Sinne eines Record Linkage durchzuführen, sondern es werden zusätzlich statistische Schätzverfahren eingesetzt, um Haushalte zu bilden. Außerdem lassen sich die Datenquellen nicht immer eindeutig zusammenführen – beispielsweise entspricht die aus den Melderegistern generierte Anzahl der Haushalte nicht immer der durch Befragung erhobenen Anzahl der Wohnungen einer Anschrift. Hierfür wurde eine Vorgehensweise entwickelt, die eine eindeutige Lösung findet, zugleich aber den primärstatistisch erhobenen Angaben Vorrang gewährt.

Auch für die Datensynchronisation zwischen der durch Hochrechnung festgelegten Einwohnerzahl und dem Melderegisterbestand auf Einzeldatenebene werden in der Haushaltegenerierung neue Verfahren entwickelt. An Stichprobenanschriften kann die Korrektur von Karteileichen und Fehlbeständen auf Basis der primärstatistisch festgestellten Informationen unmittelbar durchgeführt werden. An allen anderen Anschriften wird hingegen ein statistisches Korrekturverfahren angewandt, um einen auf Einzelpersonenebene auswertbaren Datensatz zu erstellen und die Personendaten mit den durch Hochrechnung festgelegten Einwohnerzahlen abzustimmen. Der Umfang des zu korrigierenden Personenbestands wird aus der Stichprobe geschätzt. Zudem werden weitere Stichprobeninformationen zu Über- und Untererfassungen genutzt, um im Personenbestand des Melderegisters geeignete Personen zu bestimmen, die als Karteileiche gelöscht oder als Fehlbestand imputiert werden. Auch hierbei handelt es sich nicht um eine reine Datenzusammenführung, sondern um den Einsatz statistischer Schätzverfahren, um die Über- und Untererfassungen im Einzeldatenbestand möglichst plausibel zu korrigieren.

Auf Basis der Erfahrungen aus dem Zensus 2011 konnte das Verfahren der Haushaltegenerierung weiterentwickelt werden. Die Haushaltegenerierung im Zensus 2021 soll eine höhere Ausschöpfung der Registerinformationen sowie der Gebäude- und Wohnungsdaten erreichen.

Außerdem ermöglichen die Weiterentwicklungen, stärker auf die Plausibilität der Ergebnisse auf Personen-, Haushalts- und Anschriftenebene zu achten. Hierfür wird in manchen Bearbeitungsschritten der Haushaltegenerierung auf zu detaillierte Hochrechnungen verzichtet und stärker auf modernere statistische Schätzverfahren zurückgegriffen. Dabei werden diverse Merkmale aus dem Melderegister sowie primärstatistisch erhobene Informationen einbezogen, die dabei helfen, plausiblere Haushaltszusammenhänge auf kleinräumiger Ebene zu erreichen. Einerseits wird darauf geachtet, dass die Gesamtzahl der Haushalte, Karteileichen und Fehlbestände auf der Gemeindeebene plausibel ist. Andererseits wird auch stärker berücksichtigt, an welchen Anschriften bestimmte Haushaltstypen zu bilden sind oder an welchen Anschriften und bei welchen Personen die Korrektur der Karteileichen und Fehlbestände vorzunehmen ist. Weiterhin werden im Vergleich zur Haushaltegenerierung 2011 vermehrt globale Lösungsansätze (zum Beispiel durch Einsatz von Optimierungsverfahren) umgesetzt. Das zeitgleiche Bearbeiten aller Haushalte und Wohnungen einer Anschrift (oder im Falle des Korrekturverfahrens sogar ein zeitgleiches Bearbeiten aller Personen einer Gemeinde) erzielt verbesserte Ergebnisse. Außerdem werden Unplausibilitäten vermieden, die im Zuge einer schrittweisen Bearbeitung einzelner Haushalte und Wohnungen nicht zu verhindern wären.

Die Änderungen werden in den Kapiteln 4 und 5 ausführlicher behandelt, zunächst erfolgt jedoch eine kurze Übersicht über die Arbeitsweise der Haushaltegenerierung im Zensus 2021.

3

Die Haushaltegenerierung im Überblick

Die Haushaltegenerierung im Zensus 2021 ist ein Verfahren, mit dessen Hilfe Daten, die aus verschiedenen Quellen stammen und sich auf verschiedene, für den Zensus relevante statistische Einheiten beziehen, zusammengeführt, miteinander verknüpft und bereinigt werden. Im Zentrum des Verfahrens steht die Konstruktion einer im Zensus nicht direkt beobachteten statistischen Einheit, nämlich des Wohnhaushalts. Darüber hinaus dient die Haushaltegenerierung dazu, die ein-

fließenden amtlichen Registerdaten durch Abgleich mit Daten aus statistischen Primärerhebungen zu korrigieren. Ziel ist die Erstellung des zensustypischen Datensatzes, der auf Personenebene und in fachlich sowie regional tiefer Gliederung ausgewertet werden kann (Kreuzmair/Reisch, 2013, hier: Seite 14).

Ihren Ursprung hat die Haushaltegenerierung in einem bereits in den 1980er-Jahren etablierten Verfahren aus der Städtestatistik, mit dessen Hilfe Haushaltszusammenhänge aus Melderegisterdaten konstruiert werden. Im Rahmen des Zensustests 2001 wurde die Tauglichkeit eines solchen Verfahrens für die Durchführung eines Zensus überprüft. Zusätzlich zu den Angaben des Melderegisters wurden hierfür auch gebäude- und wohnungsstatistische Informationen herangezogen (Vorndran, 2004, hier: Seite 1246). Schließlich wurde das basale Verfahren der Haushaltegenerierung im Zensus 2011 gegenüber dem Zensustest erheblich ausgebaut und weiterentwickelt, insbesondere im Hinblick auf die Verarbeitung und den gegenseitigen Abgleich verschiedener Datenbestände.³ Die Haushaltegenerierung im Zensus 2021 baut auf den Erfahrungen aus dem Zensus 2011 auf, steuert an geeigneten Stellen methodisch nach und setzt modernere statistische Verfahren ein.

Die grundlegende Operation der Haushaltegenerierung besteht darin, Beziehungen zwischen statistischen Einheiten zu erzeugen. Dabei sind viele verschiedene Arten von Beziehungen zu berücksichtigen: Hierarchische Beziehungen verknüpfen über- und untergeordnete statistische Einheiten, etwa Haushalte mit Wohnungen und Personen mit Haushalten. Alle Beziehungen dieser Art sind eindeutig, das heißt jede Person ist eindeutig einer Anschrift und einem Haushalt zugeordnet. Da die Anzahl und Struktur von Wohnhaushalten erfasst werden soll, ist jeder Haushalt mit nur einer Wohnung verknüpft. Weitere, nicht hierarchische Beziehungen verknüpfen Personen mit Personen (beispielsweise Eltern-Kind-Beziehungen, Ehen, Lebenspartnerschaften), um Beziehungsstrukturen innerhalb von Wohnhaushalten abzubilden.

³ Für detaillierte Informationen zur Entwicklung der Haushaltegenerierung im Zensustest und zum Verfahren im Zensus 2011 siehe Braun, 2004; Statistische Ämter des Bundes und der Länder, 2004; Vorndran, 2004; Kreuzmair/Reisch, 2013; Hofmeister/Fürnrohr, 2014.

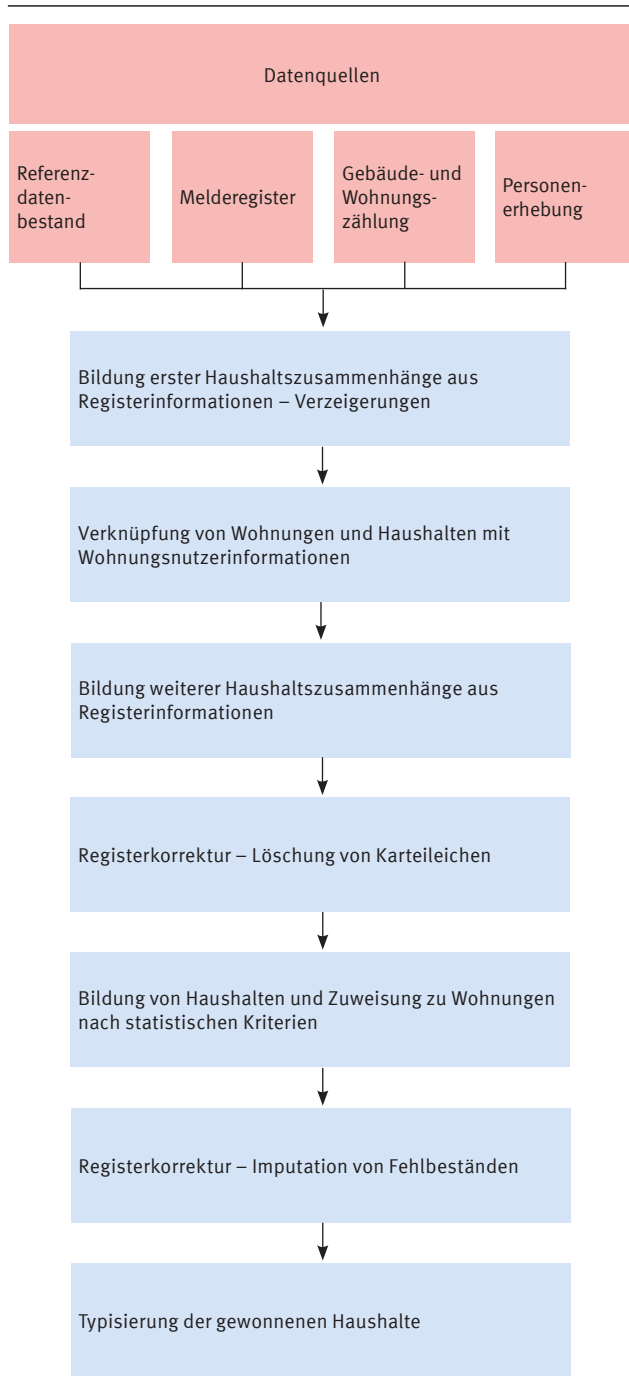
Die in der Haushaltegenerierung verarbeiteten Daten beziehen sich auf Gemeinden, Anschriften, Gebäude, Wohnungen, Haushalte, Beziehungen und Personen. Die Daten stammen zum Teil aus amtlichen Quellen, etwa aus den Registern der Einwohnermeldeämter, zum Teil auch aus primärstatistischen Erhebungen, etwa der Gebäude- und Wohnungszählung und den Personenerhebungen an Haushaltsstichproben- und Sonderbereichsanschriften. [↪ Grafik 1 auf Seite 40](#)

Die wichtigsten amtlichen Datenquellen sind die Register der Einwohnermeldeämter in den Kommunen. Aus diesen werden Daten gewonnen, mit deren Hilfe sich Personen bestimmten Gemeinden und Anschriften zuordnen lassen und die zudem Informationen über gesetzliche Beziehungen zwischen Personen enthalten, etwa Ehen, Lebenspartnerschaften oder Eltern-Kind-Beziehungen. Die Gebäude- und Wohnungszählung liefert Daten, die eine Beschreibung der Anschriften im Hinblick auf die dort vorliegenden Gebäude und Wohnungen ermöglichen. Sie erlauben zudem eine Zuordnung von Personen und Haushalten zu Wohnungen. Die Personenerhebungen an Anschriften der Haushaltsstichprobe und an Sonderbereichsanschriften liefern für einen Teil der Anschriften Daten über die dort tatsächlich wohnenden Personen und deren Haushaltszusammenhänge. Darüber hinaus werden auf der Grundlage der Haushaltsstichprobe Hochrechnungen erstellt, die primär dazu dienen, die Einwohnerzahl der Gemeinden zu bestimmen. In der Haushaltegenerierung werden weitere Hochrechnungen aus der Haushaltsstichprobe für die Bereinigung der amtlichen Registerdaten verwendet, die Fehler in Form von Karteileichen und Fehlbeständen aufweisen.

Die Haushaltegenerierung führt in einem mehrstufigen Verfahren Daten aus all diesen Quellen zusammen, konstruiert und beschreibt Haushaltszusammenhänge, bereinigt die Registerdaten mit Rückgriff auf primärstatistische Erkenntnisse und erstellt einen zensustypischen Datensatz.

Wie in Grafik 1 abgebildet, dient der erste Bearbeitungsschritt der Bildung von Haushaltszusammenhängen über Verzeigerungsinformationen aus den Melderegistern. Verzeigerungen sind in den Melderegistern enthaltene Informationen zu gesetzlichen Beziehungen zwischen zwei gemeldeten Personen, zum Beispiel Ehen oder Eltern-Kind-Beziehungen. In der Haushaltegenerierung wird eine Beziehung auf Basis einer Verzeigerung daher

Grafik 1
Datenquellen und Ablauf der Haushaltegenerierung



immer dann zwischen zwei Personen einer Anschrift gebildet, wenn deren im Melderegister geführte Ordnungsnummern oder andere eindeutige personenbezogene Merkmale gegenseitig aufeinander verweisen.

Im Bearbeitungsschritt „Verknüpfung von Wohnungen und Haushalten mit Wohnungsnutzerinformationen“ werden erste Verknüpfungen von Wohnungen und Haushalten durchgeführt. Mithilfe eines maschinellen Namensabgleichs werden an einer Anschrift gemeldete Personen mit Wohnungsnutzerinformationen der Gebäude- und Wohnungszählung derselben Anschrift abgeglichen. Die im Melderegister gefundenen Wohnungsnutzer und ihre Haushalte werden mit der entsprechenden Wohnung verknüpft und ihre Haushalte zusammengelegt. Darüber lässt sich beispielsweise der Haushaltszusammenhang von nichtehelichen Lebensgemeinschaften, über die keine Angaben in den Melderegistern vorliegen, rekonstruieren. Bei Nicht-Eindeutigkeit der Zuordnungen werden Haushaltstrennungen oder Wohnungszusammenlegungen durchgeführt, da jeder Wohnhaushalt grundsätzlich nur mit einer Wohnung verknüpft sein darf.

Im Schritt „Bildung weiterer Haushaltszusammenhänge aus Registerinformationen“ werden Registerinformationen genutzt, um weitere (weniger eindeutige) Haushaltszusammenhänge zu bilden. Anders als bei der Verarbeitung von Verzeigerungsinformationen werden in diesem Schritt auch nicht gesetzliche Beziehungen gebildet, wie etwa nicht eingetragene Lebensgemeinschaften oder Großeltern-Kind-Beziehungen.

Nachdem die Informationen aus den Melderegistern durch die vorherigen Schritte weitestgehend ausgeschöpft sind, erfolgt anschließend mit der Bereinigung des Einzeldatenbestands um hochgerechnete Übererfassungen der erste Teil der Registerkorrektur.¹⁴ Grundprinzip der Korrektur von Karteileichen ist das Herunterbrechen der im Zuge der Einwohnerzahlhochrechnung festgestellten Gesamtkarteileichenzahl einer Gemeinde auf einzelne Anschriften und Personen (Datensynchronisation). Ziel ist dabei, durch die Auswahl der Karteileichen nach statistischen Kriterien keine räumlichen und demografischen Verzerrungen im Einzeldatensatz zu erzeugen. Eine detailliertere Darstellung dieses Verfahrens enthält Kapitel 5 dieses Aufsatzes.

Danach erfolgt der Bearbeitungsschritt „Bildung von Haushalten und Zuweisung zu Wohnungen nach statistischen Kriterien“. Ein Großteil der Haushalts- und

4 Der zweite Teil der Registerkorrektur erfolgt erst zu einem späteren Zeitpunkt im Bearbeitungsschritt „Registerkorrektur – Imputation von Untererfassungen des Melderegisters“.

Wohnungszusammenhänge kann in den vorhergehenden Schritten der Haushaltegenerierung bereits gebildet werden. Wurden alle Informationen aus dem Melderegister ausgeschöpft, werden noch unverknüpfte Personen und Wohnungen in diesem Bearbeitungsschritt nach statistischen Kriterien verknüpft. Bestehen an einer Anschrift mindestens genauso viele Wohnungen wie Haushalte, werden die Haushalte unter Berücksichtigung ihres Wohnflächenbedarfs mit Wohnungen der Anschrift verknüpft. An Anschriften, an denen mehr unverknüpfte Haushalte als Wohnungen bestehen, werden unverknüpfte Haushalte mithilfe eines statistischen Schätzverfahrens und eines Optimierungsverfahrens auf bereits verknüpfte Haushalte sowie unverknüpfte Wohnungen verteilt.

Als letzter Schritt vor der Haushaltstypisierung erfolgt der zweite Teil der Registerkorrektur, die Imputation der hochgerechneten Untererfassungen in den Melderegistern. Dazu werden bereits bestehende Personen aus dem Melderegister, bei denen davon ausgegangen werden kann, dass sie potenziellen Fehlbeständen möglichst ähnlich sind, kopiert und als neue und eigenständige Personen wieder in den Datensatz eingefügt. Welche Personen zur Kopie ausgewählt werden, wird auf Basis statistischer Schätzverfahren entschieden. Anschließend müssen die kopierten Personen konkreten Anschriften zugeordnet und mit den dort bestehenden Haushalten und Wohnungen verknüpft werden. Dieser zweite Teil der Registerkorrektur wird als Abschlusschritt der Bildung und Verknüpfung von Haushalten durchgeführt. Damit können bewohnte, aber bisher unverknüpfte Wohnungen vorrangig mit Fehlbeständen verknüpft werden. Diesem Verfahren liegt die Annahme zugrunde, dass eine in der Gebäude- und Wohnungszählung gemeldete bewohnte Wohnung tatsächlich besteht. Sie ist insbesondere deshalb zu diesem Zeitpunkt der Haushaltegenerierung noch unverknüpft, weil der Wohnungsnutzer der Wohnung ein Fehlbestand ist. Auch zur Korrektur um die hochgerechneten Fehlbestände erfolgt in Kapitel 5 eine detailliertere Darstellung.

Im letzten Schritt der Haushaltegenerierung wird die Haushaltstypisierung vorgenommen, die Informationen zur Anzahl und Struktur von Wohnhaushalten liefert. Die Ermittlung privater Haushalte nach dem „Konzept des gemeinsamen Wohnens“ sowie die Typisierung von Haushalten und Familien sind Vorgabe der Europäischen Union (EU) und in der Durchführungsverordnung

2017/543¹⁵ genauer geregelt. Die Typisierung der in der Haushaltegenerierung gewonnenen Haushalte erfolgt nach drei verschiedenen Typisierungen: die EU-Typisierung von Hauptwohnsitzpersonen, die sogenannte nationale EU-Typisierung von Haupt- und Nebenwohnsitzpersonen sowie die nationale Typisierung von Haupt- und Nebenwohnsitzpersonen, die nah an den Kriterien des Mikrozensus angelegt ist.

4

Integration verschiedener Datenquellen im zensustypischen Datensatz

Welche Haushaltszusammenhänge bestehen und in welchen Wohnungen und Gebäuden diese Haushalte leben, wird durch die Zusammenführung von Registerangaben mit wohnungsstatistischen Informationen und den Angaben zu Wohnungsnutzern der Gebäude- und Wohnungszählung ermittelt. Ein Großteil der Haushaltszusammenhänge kann über Verzeigerungen aus den Melderegistern eindeutig zugewiesen werden. Ein weiterer großer Anteil der Haushalte kann über die Angabe von bis zu zwei Wohnungsnutzern aus der Gebäude- und Wohnungszählung zusammengeführt werden. Die Registerinformationen liefern jedoch keine vollständigen Haushaltsinformationen und es werden auch nur bis zu zwei Wohnungsnutzer je Wohnung erfragt. Daher können nicht alle Personen in den ersten beiden Bearbeitungsschritten Haushalten und Wohnungen zugeordnet werden. Eine eindeutige Zuordnung ist insbesondere meist dann nicht möglich, wenn in einem Haushalt mehrere Personen leben, die nicht verzeigert sind (zum Beispiel Wohngemeinschaften oder nichteheliche Lebensgemeinschaften).

Im Verfahren der Haushaltegenerierung 2011 erfolgte in diesen Fällen auf Basis von weiteren (weniger sicheren) Registerinformationen die Bildung weiterer Haushaltszusammenhänge aus Registerinformationen. Hierbei wurden Registerangaben jedoch nur für die Suche wei-

5 Durchführungsverordnung (EU) 2017/543 der Kommission vom 22. März 2017 zur Festlegung der Regeln für die Anwendung der Verordnung (EG) Nr. 763/2008 des Europäischen Parlaments und des Rates über Volks- und Wohnungszählungen in Bezug auf die technischen Spezifikationen für die Themen sowie für deren Unterliederungen (Amtsblatt der EU Nr. L 78, Seite 13).

terer Verwandtschafts- und Partnerschaftsbeziehungen genutzt. Konnten keine weiteren Haushaltszusammenhänge an einer Anschrift gefunden werden, dann erfolgte die Haushaltgenerierung nach statistischen Kriterien mithilfe einer gemeinde- oder kreisweiten Haushaltshochrechnung von 14 verschiedenen Haushaltstypen. Ziel der statistischen Generierung war es, durch Haushaltszusammenlegungen an einer Anschrift die Anzahl der hochgerechneten Haushaltstypen auf Gemeinde- oder Kreisebene zu erreichen (Kreuzmair/Reisch, 2013).

In der Haushaltgenerierung 2021 werden hingegen zunächst die Informationen der Melderegister stärker ausgeschöpft, um weitere Haushaltszusammenhänge zu bilden. Es wird nicht nur nach Verwandtschafts- und Partnerschaftsbeziehungen, sondern auch nach sonstigen Haushaltszusammenhängen gesucht. Beispielsweise werden Merkmale wie die gleiche vorherige Anschrift, das gleiche Einzugsdatum, der Familienname oder auch die Staatsangehörigkeit in unterschiedlichen Kombinationen einbezogen, um Personen an einer Anschrift zu Haushalten zusammenzuführen.

Sind alle Möglichkeiten der Haushaltsbildung auf Basis von Registerangaben erschöpft, erfolgt die statistische Generierung von Haushalten. Im Gegensatz zum Vorgehen der Haushaltgenerierung 2011 wird auf die detaillierte gemeinde- und kreisweite Haushaltshochrechnung verzichtet. Die Erfahrung zeigte, dass zu detaillierte Zielvorgaben auf Gemeindeebene nur schwer zu erreichen waren und der plausiblen Verteilung von Personen auf Haushalte und Wohnungen einer Anschrift zu wenig Augenmerk geschenkt wurde. Deshalb soll im neuen Verfahren dazu übergegangen werden, die Zielvorgaben der statistischen Generierung zu vergrößern (3 Haushaltsgrößenklassen statt 14 Haushaltstypen) und diese auf Anschriften- und nicht auf Gemeindeebene zu beziehen. Jede Anschrift erhält damit eine Vorgabe, welcher Haushaltsgrößenverteilung sich nach der statistischen Generierung möglichst anzunähern ist. Die Bestimmung dieser Haushaltsgrößenklassen erfolgt für jede Anschrift auf Basis eines statistischen Schätzverfahrens, das sich auf unterschiedliche Personen-, Wohnungs-, Gebäude- und Anschriftenmerkmale bezieht.

Auf Basis der geschätzten Zielvorgaben soll an Anschriften, an denen zu diesem Zeitpunkt noch mehr unverknüpfte Haushalte als unverknüpfte Wohnungen

bestehen, eine plausible Verteilung der unverknüpften Personen auf verknüpfte Haushalte und noch unverknüpfte Wohnungen gefunden werden. Hierbei wird anstatt eines schrittweisen Vorgehens der Haushaltgenerierung wie beim Zensus 2011 ein Optimierungsverfahren eingesetzt, um ein zeitgleiches Bearbeiten aller Haushalte und Wohnungen an einer Anschrift zu gewährleisten. Unplausibilitäten, die durch eine schrittweise Bearbeitung entstehen, können dadurch verhindert werden. Im Zensus 2011 konnte die schrittweise Bearbeitung in manchen Fällen nämlich dazu führen, dass große Haushalte mit Wohnungen mit kleiner Wohnfläche verknüpft wurden und kleine Haushalte mit Wohnungen mit großer Wohnfläche. Die zeitgleiche Bearbeitung aller Haushalte und Wohnungen einer Anschrift im Zensus 2021 ermöglicht hingegen die bestmögliche Verteilung der unverknüpften Haushalte auf Wohnungen.

Aufgrund des Verzichts auf sehr detaillierte Hochrechnungen auf Gemeinde- oder Kreisebene, der stärkeren Ausschöpfung der Personen-, Wohnungs- und Anschriftenmerkmale sowie der Umsetzung eines globalen Lösungsansatzes, kann besser gewährleistet werden, dass nicht nur die Gesamtzahl der gebildeten Haushalte auf Gemeindeebene plausibel ist, sondern auch eine plausible Verteilung der Personen auf Haushalte und Wohnungen an einer Anschrift erreicht wird. Dadurch verbessert sich die Qualität erhebungsteilübergreifender Auswertungen zu haushalts- und wohnungsstatistischen Informationen auf kleinräumiger Ebene.

5

Synchronisation der integrierten Daten mit der hochgerechneten Einwohnerzahl: das Korrekturverfahren

Aufgrund der Konzeption des Zensus 2021 als registergestützter Zensus erfolgt die Haushaltebefragung – und damit die Existenzfeststellung der im Melderegister verzeichneten Personen – nur auf Stichprobenbasis. Damit ist nur an Anschriften, die für die Stichprobe ausgewählt wurden, bekannt, wie viele und welche Personen dort tatsächlich leben. Für die Ermittlung der Gesamteinwohnerzahl einer jeden Gemeinde ist daher eine Hochrechnung aus den Ergebnissen der Haushaltsstichprobe erforderlich. Diese Hochrechnung bestimmt nicht nur die Gesamteinwohnerzahl einer Gemeinde, sondern auch die nicht durch die Befragung aufgedeckten Übererfassungen (Karteileichen) und Untererfassungen (Fehlbestände) sowie deren Verteilungen hinsichtlich demografischer Merkmale.

Wie in den vorherigen Kapiteln dargelegt, besteht eine der Hauptaufgaben der Haushaltegenerierung darin, einen auf Einzeldatensatzebene auswertbaren Datenbestand zu erzeugen. Um dieses Ziel zu erreichen, müssen daher die auf Gemeindeebene hochgerechneten Karteileichen und Fehlbestände auf Einzelpersonenebene heruntergebrochen werden. Das bedeutet, dass konkrete Melderegistereinträge als potenzielle Karteileichen und zu imputierende Fehlbestände zu markieren sind. Diese Aufgabe der Datensynchronisation übernimmt das Korrekturverfahren in der Haushaltegenerierung. Maßgabe ist dabei, die demografische Struktur des Personenbestands durch die Korrekturen möglichst wenig zu verfälschen. Gegenüber dem Korrekturverfahren, das in der Haushaltegenerierung des Zensus 2011 zum Einsatz kam (Hofmeister/Fürnrohr, 2014), werden für den Zensus 2021 drei Änderungen grundsätzlicher Art vorgenommen:

1. Anders als im Zensus 2011 kommt das Korrekturverfahren in allen Gemeinden zum Einsatz. Durch die Ausweitung der Haushaltsbefragungen auf Stichprobenbasis auf Gemeinden mit weniger als 10 000 Einwohnerinnen und Einwohnern kann im Zensus 2021 auch in diesen Gemeinden eine Korrektur um hochgerech-

nete Karteileichen und Fehlbestände im Rahmen des Korrekturverfahrens erfolgen. Da jedoch mit sinkender Gemeindegröße auch der Detailgrad möglicher Hochrechnungen sinkt, erfolgt bei kleineren Gemeinden eine Verringerung des Umfangs der berücksichtigten demografischen Merkmale.

2. Bei der Auswahl konkreter Melderegistereinträge zur Löschung von Karteileichen und zur Kopie für die Imputation wird stärker auf eine plausible Verteilung hinsichtlich der wichtigsten demografischen Merkmale geachtet, um Verfälschungen durch das Korrekturverfahren zu minimieren. Eine Hochrechnung von Karteileichen und Fehlbeständen hinsichtlich der wichtigsten demografischen Merkmale (Geschlecht, Nationalität und Alter) ist nur univariat, also für jedes Merkmal einzeln möglich. Daher muss durch das Korrekturverfahren vor der Auswahl konkreter Personen zunächst die gemeinsame multivariate Verteilung dieser Merkmale, jeweils getrennt nach Karteileichen und Fehlbeständen, bestimmt werden. Im Zensus 2011 erfolgte die Bestimmung der Verteilung mithilfe einer Monte-Carlo-Simulation unter Annahme von Unabhängigkeit zwischen den demografischen Merkmalen. Durch eine Formulierung als Optimierungsproblem nach der Vorgehensweise des IPF-Verfahrens (Agresti, 2013, hier: Seite 364 ff.) kann auf diese nur hilfswise verwendete Unabhängigkeitsannahme verzichtet werden. Die hochgerechneten Karteileichen- und Fehlbestandszahlen je Gesamtgemeinde und je demografischer Merkmalsausprägung gehen dabei als zwingend einzuhaltende Randbedingungen ein. Gleichzeitig kann eine Referenzstruktur definiert werden, der sich das Ergebnis, also die multivariaten Verteilungen von Karteileichen und Fehlbeständen, so gut wie möglich annähern soll. Als Referenz wird hier die demografische Bevölkerungsstruktur aus den Melderegistern verwendet. Durch das Optimierungsverfahren wird dann unter allen zulässigen Lösungen, die die Nebenbedingungen einhalten, diejenige Lösung mit der geringsten Distanz zur Referenzstruktur gewählt. Zusätzlich zur Bestimmung der multivariaten Verteilungen von Karteileichen und Fehlbeständen auf Gemeindeebene erfolgt eine Schätzung der individuellen Karteileichen- und Fehlbestandswahrscheinlichkeiten auf Einzelpersonenebene. Dazu werden statistische Modelle aus dem Bereich des Machine Learning (maschinelles Lernen) eingesetzt. Sie sollen anhand der Muster in den Ergebnissen der Haushaltsstichprobe präzisieren, wie hoch


eben jene Wahrscheinlichkeiten für jede nicht befragte Person ausfallen. Die konkrete Auswahl von Melderegistereinträgen zur Löschung beziehungsweise Imputation erfolgt dann durch Kombination der Informationen auf Gemeindeebene und auf Personenebene. Dadurch ist im Vergleich zum Zensus 2011 eine gezieltere Auswahl von Karteileichen und Fehlbeständen möglich, durch die eine geringere demografische und regionale Verzerrung im Ergebnisdatensatz zu erwarten ist.

3. Bei der Löschung von Karteileichen und der Imputation von Fehlbeständen wird im Zensus 2021 auf die Berücksichtigung von haushaltsstatistischen Hochrechnungen verzichtet. Für das Korrekturverfahren im Zensus 2011 wurde aus der Stichprobe die jeweilige Anzahl von Karteileichen- und Fehlbestandshaushalten untergliedert nach Größe und Typ je Gemeinde hochgerechnet. Diese gingen neben den demografischen Randbedingungen als zusätzliche Nebenbedingungen bei der Auswahl von konkreten Personen zur Löschung und Imputation ein. Auf derartige haushaltsstatistische Hochrechnungen wird im Zensus 2021 verzichtet: Einerseits hat sich gezeigt, dass die zu starke Untergliederung zu Qualitätsproblemen in der Hochrechnung geführt hat; andererseits war in der Regel keine gleichzeitige Erfüllung der demografischen und haushaltsstatistischen Nebenbedingungen möglich. Haushaltszusammenhänge von Karteileichen und Fehlbeständen sollen daher im Korrekturverfahren im Zensus 2021 unmittelbar in die statistischen Modelle einbezogen werden; sie finden so einen noch direkteren Eingang in die Auswahl konkreter Personen. Damit werden bevorzugt Personen als Karteileichen markiert, die in Haushaltskonstellationen leben, in denen auch in der Stichprobe gehäuft Karteileichen identifiziert werden. Gleichermaßen werden bevorzugt Personen zur Kopie als Fehlbestände ausgewählt, die hinsichtlich ihres Haushaltszusammenhangs Ähnlichkeiten zu typischen Fehlbeständen aus der Stichprobe aufweisen.

6

Fazit

Auch im Zensus 2021 kommt der Haushaltegenerierung wieder eine zentrale Bedeutung zu. Durch Datenintegration verschiedener Register- und Erhebungsdaten wird die Anzahl und Struktur von Haushalten und deren Wohnsituation ermittelt. Zudem wird der Personenbestand auf Einzeldatenbasis an die hochgerechnete Einwohnerzahl (und damit an die ermittelte Anzahl der Karteileichen und Fehlbestände) einer Gemeinde angepasst. Dadurch kann ein bundesweiter Einzeldatenbestand auf Personenebene zur Verfügung gestellt werden, der erhebungsteilübergreifende Auswertungen zu demografischen, haushaltsstatistischen, erwerbsstatistischen sowie zu Wohnungs- und Gebäudedaten ermöglicht.

Das weiterentwickelte Verfahren der Haushaltegenerierung im Zensus 2021 stellt eine höhere Datenqualität durch den Einsatz modernerer Schätzverfahren und Optimierungsverfahren in Aussicht. Es erfolgt eine stärkere Ausschöpfung der Personen-, Wohnungs- und Anschriftenmerkmale zur Bildung von Haushaltszusammenhängen; im Fokus steht auch eine gezieltere Auswahl bei der Löschung von Karteileichen und der Imputation von Fehlbeständen. Durch diese Weiterentwicklungen ist im Vergleich zur Haushaltegenerierung im Zensus 2011 eine plausiblere Abbildung der tatsächlichen Haushalts- und Wohnungszusammenhänge auf Anschriftenebene sowie eine bessere Auswahl von Karteileichen und Fehlbeständen hinsichtlich demografischer und anschriftenbezogener Merkmale zu erwarten. 

LITERATURVERZEICHNIS

Agresti, Alan. *Categorical Data Analysis*. 3. Auflage. Hoboken, New Jersey 2013.

Braun, Ralph. *Zensus 2001: Ergebnisse der Testerhebungen und Ausblick auf ein zukünftiges Zensusmodell*. In: Statistisches Monatsheft Baden-Württemberg. Ausgabe 5/2004, Seite 3 ff.

Hofmeister, Katrin/Fürnrohr, Michael. *Das Korrekturverfahren beim Zensus 2011*. In: Bayern in Zahlen. Ausgabe 6/2014, Seite 310 ff.

Kreuzmair, Ingrid/Reisch, Marco. *ZENSUS 2011: Ablauf der Haushaltegenerierung*. In: Statistisches Monatsheft Baden-Württemberg. Ausgabe 4/2013, Seite 13 ff.

Statistische Ämter des Bundes und der Länder 2004. [Ergebnisse des Zensus 2001](#). In: Wirtschaft und Statistik. Ausgabe 8/2004, Seite 813 ff.

Vorndran, Ingeborg. [Zensus 2001 – Verfahren und Ergebnisse der Haushaltegenerierung](#). In: Wirtschaft und Statistik. Ausgabe 11/2004, Seite 1245 ff.

RECHTSGRUNDLAGEN

Durchführungsverordnung (EU) 2017/543 der Kommission vom 22. März 2017 zur Festlegung der Regeln für die Anwendung der Verordnung (EG) Nr. 763/2008 des Europäischen Parlaments und des Rates über Volks- und Wohnungszählungen in Bezug auf die technischen Spezifikationen für die Themen sowie für deren Untergliederungen (Amtsblatt der EU Nr. L 78, Seite 13).

Entwurf eines Gesetzes zur Vorbereitung eines registergestützten Zensus einschließlich einer Gebäude- und Wohnungszählung 2011 (Zensusvorbereitungsgesetz 2011 – ZensVorbG 2011) vom 30. Mai 2007 (Drucksache 16/5525).

Entwurf eines Gesetzes zur Durchführung des Zensus im Jahr 2021 (Zensusgesetz 2021 – ZensG 2021) vom 25. März 2019 (BT-Drucksache 19/8693).

Gesetz über den registergestützten Zensus im Jahre 2011 (Zensusgesetz 2011 – ZensG 2011) vom 8. Juli 2009 (BGBl. I Seite 1781).

Herausgeber

Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung

Dr. Daniel Vorgrimler

Redaktionsleitung: Juliane Gude

Redaktion: Ellen Römer

Ihr Kontakt zu uns

www.destatis.de/kontakt

Erschienen im August 2019

Das Archiv älterer Ausgaben finden Sie unter www.destatis.de

Print

Einzelpreis: EUR 19,- (zzgl. Versand)

Jahresbezugspreis: EUR 114,- (zzgl. Versand)

Bestellnummer: 1010200-19907-1

ISSN 0043-6143

ISBN 978-3-8246-1090-7

Download (PDF)

Artikelnummer: 1010200-19907-4, ISSN 1619-2907

Vertriebspartner

IBRo Versandservice GmbH

Bereich Statistisches Bundesamt

Kastanienweg 1

D-18184 Roggentin

Telefon: +49 (0) 382 04 / 6 65 43

Telefax: +49 (0) 382 04 / 6 69 19

destatis@ibro.de

Papier: Design Offset, FSC-zertifiziert

© Statistisches Bundesamt (Destatis), 2019

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.