

METHODISCHE WEITERENTWICKLUNGEN IN DER LEBENSFORMENBERICHTERSTATTUNG AUF GRUNDLAGE DES MIKROZENSUS AB DEM BERICHTSJAHR 2020

Tim Hochgürtel, Clarissa Wilke

↳ **Schlüsselwörter:** Lebensformen – Mikrozensus 2020 – Plausibilisierung – Imputation

ZUSAMMENFASSUNG

Mit dem Berichtsjahr 2020 wurde der Mikrozensus umfangreichen methodischen und konzeptionellen Neuerungen unterzogen, unter anderem wurden weitere Erhebungen in den Mikrozensus integriert. Dies erfordert eine Harmonisierung in der Berichterstattung und führt auch zu einer Anpassung des Lebensformenkonzepts.

Die Veröffentlichung der Ergebnisse soll künftig früher und häufiger erfolgen. Dazu sind in der Lebensformenberichterstattung in verstärktem Maße standardisierte und automatisierte Methoden zur Plausibilisierung von Einzeldaten sowie zur Erzeugung weiterer Merkmale im Zuge der Typisierung und Bandsatzerweiterung einzusetzen. Der Beitrag beschreibt den Ablauf dieser standardisierten und automatisierten Prozesse.

↳ **Keywords:** *living arrangements – microcensus 2020 – plausibility checks – imputation*

ABSTRACT

The microcensus underwent substantial methodological and conceptual changes in the reporting year 2020, among them the inclusion of further surveys into the microcensus. As a consequence, the reporting has to be harmonised and the living arrangements concept is adjusted.

Results will be published sooner and more frequently. To this end, more standardised and automated methods have to be used in living arrangements reporting to check the plausibility of microdata and generate further variables from the information obtained. The article describes these standardized and automated processes.



Tim Hochgürtel

studierte Soziologie in Mainz und ist Referent im Referat „Bevölkerungsstatistische Auswertungen und Analysen aus dem Mikrozensus“ des Statistischen Bundesamtes. Schwerpunkte seiner Arbeit sind Analysen sowie Publikationen zu privaten Haushalten und Lebensformen.



Clarissa Wilke

hat in Mainz Politik- und Wirtschaftswissenschaften studiert. Seit 2015 ist sie im Statistischen Bundesamt tätig und seit April 2018 im Referat „Bevölkerungsstatistische Auswertungen und Analysen aus dem Mikrozensus“. Schwerpunkte ihrer Arbeit sind Auswertungen und Publikationen zu privaten Haushalten und Lebensformen.

1

Berichterstattung zu Lebensformen auf Basis des Mikrozensus

Seit 1957 berichtet die amtliche Statistik auf Grundlage des Mikrozensus über familiäre Strukturen der deutschen Bevölkerung (Herberger, 1957). Seither ist die Berichterstattung einem stetigen Wandel unterworfen. Dieser Wandel betrifft hierbei sowohl Aspekte der Methodik in der Erhebung des Mikrozensus, als auch die konzeptuelle Grundlage, mit der über Familien und Lebensformen berichtet wird (Emmerling/Riede, 1997; Lotze/Breiholz, 2002; Nöthen, 2005; Hochgürtel/Rammelt, 2017).

Mit dem Mikrozensus 2020 wurden wiederum umfassende Neuregelungen umgesetzt. Unter anderem sind seit dem Erhebungsjahr 2020 die Erhebung LEBEN IN EUROPA (deutsche Bezeichnung der Gemeinschaftsstatistik über Einkommen und Lebensbedingungen – EU-SILC) und seit dem Erhebungsjahr 2021 die Erhebung „Informations- und Kommunikationstechnologien in privaten Haushalten“ (IKT) Bestandteil des Mikrozensus.

Der Mikrozensus wird nun als integrierte Erhebung durchgeführt, bei welcher bisher weitgehend separat durchgeführte Erhebungen als Module einer gemeinsamen Erhebung realisiert werden. Dieses Vorgehen soll unter anderem Synergieeffekte erzielen, um die in der Vergangenheit gestiegenen Anforderungen an die Haushaltserhebungen zu bewältigen. So nutzen die bisher separat durchgeführten Erhebungen unter anderem dieselbe Stichprobengrundlage (Bihler/Zimmermann, 2016) sowie dieselben Erhebungsinstrumente (Hochgürtel, 2013; Hundenborn/Enderer, 2019).

Die Integration der bisher separat durchgeführten Haushaltserhebungen spiegelt sich auch in der Vereinheitlichung von Auswertungskonzepten wider. Bis einschließlich 2019 bildete die Vereinigungsmenge aus Haushalten am Haupt- und Nebenwohnsitz die Grundlage der Haushaltsberichterstattung im Mikrozensus. Seit dem Berichtsjahr 2020 wird einheitlich, über alle Module hinweg, das Konzept des Hauptwohnsitzhaushalts als Grundlage verwendet. Ein Hauptwohnsitzhaushalt ist dabei definiert als Haushalt, in dem mindestens eine Person im Alter von 16 Jahren oder älter mit Hauptwohnsitz lebt (Hochgürtel/Weinmann, 2020).

Daneben ändern sich mit dem Berichtsjahr 2020 auch die Veröffentlichungszyklen. Erstergebnisse sollen künftig drei bis vier Monate nach Ende eines Berichtsjahrs zur Verfügung stehen. Damit erfolgt eine Bereitstellung rund vier Monate früher als dies in den Berichtsjahren bis einschließlich 2019 der Fall war. Eine Veröffentlichung der Endergebnisse ist rund zehn Monate nach den Erstergebnissen geplant (Hundenborn/Enderer, 2019; Schmidt/Stein, 2021).

Um die zweimal im Jahr geplante Veröffentlichung (Erst- und Endergebnisse) zu ermöglichen, müssen die Methoden der Plausibilisierung und Datenaufbereitung noch stärker als bisher automatisiert werden. Daraus resultiert eine Anpassung der entsprechenden Anwendungen.

Der vorliegende Beitrag beschreibt in Kapitel 2 die Plausibilisierung der für die Erfassung der Lebensformen zentralen Merkmale in den Rohdaten der Mikrozensus-Befragung. Danach benennt Kapitel 3 notwendige Modifikationen des Lebensformenkonzepts im Kontext des Mikrozensus 2020. Des Weiteren veranschaulicht der Beitrag in Kapitel 4 Bandsatzerweiterungen und Typisierungen neuer Merkmale und stellt in diesem Zusammenhang auch die Vorgehensweise zur Imputation fehlender Angaben vor. Der Beitrag schließt mit einem Ausblick.

2

Plausibilisieren von Rohdaten

Mit den Neuerungen des Mikrozensus ab dem Berichtsjahr 2020 ist beabsichtigt, Veröffentlichungen schneller zu realisieren. Dazu werden die Datenaufbereitungsschritte im Kontext der Lebensformenberichterstattung standardisiert und automatisiert. Im Folgenden ist das Vorgehen zur Plausibilisierung der Rohdaten beschrieben.

Im Zuge der Erstellung von Typisierungen und Bandsatzerweiterungen (siehe hierzu auch Kapitel 4) werden die Befragten eines Haushalts zu einer oder mehreren Lebensformen gruppiert. Jede befragte Person des Haushalts gehört genau einer Lebensform an. Für diese Gruppierung werden die Angaben der Befragten genutzt.

Für eine Bildung der Lebensformen (siehe hierzu auch Kapitel 3) sind die Informationen zu Wohnsitz, Familien-

Tabelle 1

Merkmale des Mikrozensus zur Erfassung von Eltern und Partnern

Befragungsmerkmal	Plausibilisiertes Merkmal	Merkmal	Korrekturbedürftige Ausprägungen
			Anteil an den Erstergebnissen 2020 in %
AB0701P	TL0701P	Lebt Ihre Mutter in diesem Haushalt?	0,3
AB0702P	TL0702P	Personen-Nummer der Mutter	0,8
AB0801P	TL0801P	Lebt Ihr Vater in diesem Haushalt?	0,3
AB0802P	TL0802P	Personen-Nummer des Vaters	0,7
AB0901P	TL0901P	Lebt Ihr/-e Ehepartner/-in in diesem Haushalt?	0,1
AB0902P	TL0902P	Personen-Nummer Ehepartner/-in	0,4
AB1001P	TL1001P	Lebt Ihr/-e Lebenspartner/-in in diesem Haushalt?	0,1
AB1002P	TL1002P	Personen-Nummer Lebenspartner/-in	0,1

stand, Geburtsjahr und Geschlecht relevant. Darüber hinaus wird berücksichtigt, ob die Befragten mit Eltern und einem Partner beziehungsweise einer Partnerin in einem Haushalt leben. Hierzu wird im Mikrozensus separat nach einem Ehepartner beziehungsweise einer Ehepartnerin und nach einem Lebenspartner beziehungsweise einer Lebenspartnerin gefragt. [↘ Tabelle 1](#) enthält in Spalte 3 die Merkmale des Mikrozensus zu Elternteilen und Partnern.

Die Qualität der realisierten Gruppierung hängt von kohärenten und widerspruchsfreien Informationen der Befragten ab. Daher ist vor der Durchführung der Gruppierung die Kohärenz der Angaben der Befragten zu prüfen. Bei widersprüchlichen Angaben muss eine Korrektur erfolgen.

Die Widerspruchsfreiheit muss für alle Angaben zu familiären Beziehungen von Haushaltsmitgliedern gegeben sein. Inkohärenzen können dadurch entstehen, dass einzelne Haushaltsmitglieder bei der Beantwortung von Fragen widersprüchliche Angaben machen. Eine Inkohärenz auf Individualebene liegt beispielsweise vor, wenn eine Person zwar benennt, dass sie mit einem Ehepartner beziehungsweise einer Ehepartnerin im Haushalt lebt, aber nicht angibt, welches Haushaltsmitglied der Ehepartner beziehungsweise die Ehepartnerin ist. Auch aus der Kombination von Angaben verschiedener Haushaltsmitglieder können sich Widersprüchlichkeiten ergeben. Dies ist beispielsweise der Fall, wenn in einem Haushalt eine Person A eine andere Person B als Ehepartner beziehungsweise Ehepartnerin benennt, die Person B hingegen keinen Ehepartner beziehungsweise keine Ehepartnerin angibt.

Da für die Bildung von Lebensformen eine Vielzahl von Merkmalen Berücksichtigung findet, gibt es sehr viele Möglichkeiten, Inkohärenzen im Antwortverhalten eines Haushalts zu erzeugen. Die Konzeption und Programmierung eines Plausibilisierungs-Algorithmus, der jede denkbare Inkohärenz lösen kann, scheitert daher an der ungemein großen Menge möglicher widersprüchlicher Angaben.

Aus diesem Grund wurde eine andere Methode entwickelt, um die Widerspruchsfreiheit der Angaben zu familiären Beziehungen durch eine Plausibilisierung herzustellen. Die Angaben zu Eltern und Partnern oder Partnerinnen im Haushalt gelten als plausibel, wenn der Haushalt 40 vorab definierte Kriterien erfüllt. Anhand der Kriterien wird bei der Plausibilisierung jeweils geprüft, ob innerhalb eines Haushalts die getätigten Angaben zusammenpassen. Dies wird sowohl für Angaben der Befragten zu ihren Eltern umgesetzt als auch zu Angaben, die sie zu ihrem Partner oder ihrer Partnerin machen. [↘ Tabelle 2](#)

Wird in einem Haushalt mindestens ein Kriterium verletzt, gilt dieser als inkohärent. Ein Plausibilisierungsalgorithmus bereinigt die durch die Kriterienliste festgestellten Inkohärenzen. Der Vorteil dieser Vorgehensweise besteht darin, dass der Plausibilisierungsalgorithmus ausschließlich empirisch auftretende Probleme löst und nicht dazu imstande sein muss, Lösungen für alle potenziellen Widersprüche anzubieten. Dies erfordert jedoch auch die Pflege des Plausibilisierungsalgorithmus; er ist zu erweitern, wenn neue, bisher noch nicht bekannte Inkohärenzen auftreten. Der Plausibilisierungsalgorithmus wurde bisher für sechs

Tabelle 2

Kriterien der Kohärenzprüfung

Kriterien zur Prüfung der Kohärenz von Eltern und Partnerbeziehungen in Haushalten		Verstöße ¹	
		vor Plausibilisierung	nach Plausibilisierung
1 Personen-Nummer			
1.1	Jede Person hat eine Personen-Nummer.	0	0
1.2	Jede Personen-Nummer gibt es nur einmal im Haushalt.	0	0
2 Angaben zu Elternteilen			
2.1	Angabe zur Frage „Lebt Ihre Mutter in Ihrem Haushalt?“ ist vorhanden.	0	0
2.2	Wenn Mutter im Haushalt lebt, ist Angabe zur Personen-Nummer der Mutter vorhanden.	38	0
2.3	Angabe der Personen-Nummer der Mutter ist ungleich der eigenen Personen-Nummer.	21	0
2.4	Es existiert eine Person mit Personen-Nummer der Mutter im Haushalt.	933	0
2.5	Die benannte Mutter ist mindestens 12 Jahre alt.	25	0
2.6	Wenn angegeben wird, dass eigene Mutter nicht im Haushalt lebt, ist keine Personen-Nummer der Mutter benannt.	0	0
2.7	Wenn angegeben wird, dass eigene Mutter im Haushalt lebt, ist die angegebene Personen-Nummer der Mutter ungleich der angegebenen Personen-Nummer des Vaters.	38	0
2.8	Angabe zur Frage „Lebt Ihr Vater in Ihrem Haushalt?“ ist vorhanden.	19	0
2.9	Wenn Vater im Haushalt lebt, ist Angabe zur Personen-Nummer des Vaters vorhanden.	38	0
2.10	Angabe der Personen-Nummer des Vaters ist ungleich der eigenen Personen-Nummer.	19	0
2.11	Es existiert eine Person mit Personen-Nummer des Vaters im Haushalt.	739	0
2.12	Benannter Vater ist mindestens 12 Jahre alt.	24	0
2.13	Wenn angegeben wird, dass eigener Vater nicht im Haushalt lebt, ist keine Personen-Nummer des Vaters benannt.	0	0
2.14	Im Haushalt existieren keine Personen A und B, die sich gegenseitig als Elternteil angeben.	216	0
2.15	Alle Personen unter 14 Jahren haben mindestens einen Elternteil benannt.	1 233	0
2.16	Im Haushalt existieren keine Personen A und B, welche beide eine Person C als Elternteil benennen und die jeweils andere Person als Elternteil angeben.	104	0
3 Angaben zur Partnerschaft			
3.1	Angabe zur Frage „Lebt Ihr Ehepartner in Ihrem Haushalt?“ ist vorhanden.	0	0
3.2	Wenn Ehepartner im Haushalt lebt, ist Angabe zur Personen-Nummer der Ehepartner vorhanden.	267	0
3.3	Angabe der Personen-Nummer des Ehepartners ist ungleich der eigenen Personen-Nummer.	272	0
3.4	Es existiert eine Person mit Personen-Nummer des Ehepartners im Haushalt.	1 633	0
3.5	Ehepartner sind mindestens 14 Jahre alt.	25	0
3.6	Wenn angegeben wird, dass kein Ehepartner im Haushalt lebt, ist keine Personen-Nummer des Ehepartners benannt.	0	0
3.7	Wenn ein Ehepartner benannt ist, wird das Vorhandenseins eines Lebenspartners verneint.	0	0
3.8	Wenn ein Ehepartner benannt ist, wird keine Personen-Nummer für einen Lebenspartner benannt.	0	0
3.9	Die Personen-Nummer des Ehepartners ist ungleich der Personen-Nummer der Mutter.	101	0
3.10	Die Personen-Nummer des Ehepartners ist ungleich der Personen-Nummer des Vaters.	134	0
3.11	Wenn Person A die Person B als Ehepartner benennt, so gibt Person B die Person A als Ehepartner an.	2 126	0
3.12	Angabe zur Frage „Lebt Ihr Lebenspartner in Ihrem Haushalt?“ ist vorhanden.	0	0
3.13	Wenn Lebenspartner im Haushalt lebt, ist Angabe zur Personen-Nummer des Lebenspartners vorhanden.	0	0
3.14	Angabe der Personen-Nummer des Lebenspartners ist ungleich der eigenen Personen-Nummer.	33	0
3.15	Es existiert eine Person mit Personen-Nummer des Lebenspartners im Haushalt.	442	0
3.16	Der Lebenspartner ist mindestens 14 Jahre alt.	29	0
3.17	Wenn angegeben wird, dass kein Lebenspartner im Haushalt lebt, ist keine Personen-Nummer des Lebenspartners benannt.	0	0
3.18	Wenn angegeben wird, dass ein Lebenspartner im Haushalt lebt, ist keine Personen-Nummer des Ehepartners benannt.	0	0
3.19	Die Personen-Nummer des Lebenspartners ist ungleich der Personen-Nummer der Mutter.	0	0
3.20	Die Personen-Nummer des Lebenspartners ist ungleich der Personen-Nummer des Vaters.	0	0
3.21	Wenn Person A die Person B als Lebenspartner benennt, so gibt Person B die Person A als Lebenspartner an.	1 040	0
3.22	Es ist nicht sowohl eine Personen-Nummer des Ehepartners und eine Personen-Nummer des Lebenspartners benannt.	0	0

¹ Grundlage der Prüfung bilden die 651 609 Sätze der Einzeldaten für die ersten Ergebnisse des Mikrozensus 2020.

Quartale genutzt. Es zeigt sich bereits jetzt, dass bisher unbekannte Probleme kaum noch auftreten.

Verletzt ein Haushalt eines der Kriterien, ist damit eine Widersprüchlichkeit in den Antworten identifiziert. Damit sind aber noch keine Hinweise verbunden, wie die Widersprüchlichkeit aufgelöst werden kann. Je nach Haushaltskontext ist das Problem adäquat zu lösen. Entsprechend ist im Plausibilisierungsalgorithmus eine Vielzahl von Fallunterscheidungen implementiert. Zwischen den Kriterien, mit welchen die Kohärenz der Angaben des Haushalts geprüft wird, und den Fallunterscheidungen des Plausibilisierungsalgorithmus besteht eine n:m-Beziehung. So sind derzeit rund 100 verschiedene Inkohärenzfälle zu unterscheiden. Für jeden Fall sieht der Plausibilisierungsalgorithmus eine entsprechende Korrektur vor.

Eine inkohärente Angabe eines Befragten kann dazu führen, dass gegen mehrere Kriterien verstoßen wird. Dies illustriert das folgende Beispiel: In einem Haushalt geben die Personen A und B beide an, mit einem Ehepartner beziehungsweise einer Ehepartnerin im Haushalt zu leben. Person A hat eine Person B als Ehepartner beziehungsweise Ehepartnerin benannt. Person B hat hingegen nicht angegeben, welches Haushaltsmitglied der Ehepartner oder die Ehepartnerin ist. Damit ist das Kriterium 3.2 nicht erfüllt: Eine Person, die angibt mit einem Ehepartner beziehungsweise einer Ehepartnerin im Haushalt zu leben, muss auch benennen, welche andere Person des Haushalts dieser Ehepartner oder die Ehepartnerin ist. Darüber hinaus ist auch das Kriterium 3.11 nicht erfüllt, nach dem Haushaltsangaben nur dann als kohärent gelten, wenn Person A und B sich gegenseitig als Ehepartner beziehungsweise Ehepartnerin benannt haben. In einigen Fällen bedarf es lediglich einer einzigen Korrektur, um mehrere Verstöße eines Haushalts gegen Plausibilisierungskriterien zu bereinigen.

Die Verletzung eines Kriteriums muss nicht immer identische Ursachen haben. Je nach Haushaltskontext können verschiedene Fälle eintreten. Entsprechend muss der Plausibilisierungsalgorithmus verschiedene Fälle unterscheiden können, um die jeweils angemessene Plausibilisierung umzusetzen. Für das obige Beispiel geht der Plausibilisierungsalgorithmus wie folgt vor: Hat Person B (1.) als Familienstand „verheiratet“ angegeben, (2.) Person A nicht als Elternteil benannt und (3.) keine weitere Person C als Ehe- oder Lebenspartner beziehungs-

weise Ehe- oder Lebenspartnerin angegeben, wird im Zuge der Plausibilisierung der Person B die Person A als Ehepartner beziehungsweise Ehepartnerin zugewiesen. Ist hingegen mindestens eine der drei genannten Bedingungen nicht erfüllt, wird durch die Plausibilisierung der Person A zugewiesen, dass diese keinen Ehepartner beziehungsweise Ehepartnerin im Haushalt hat.

Im Kontext der Plausibilisierung werden durch den Algorithmus neue Merkmale erstellt. Die in Spalte 1 der Tabelle 1 genannten Befragungsmerkmale bleiben als Originalmerkmale unverändert. Neu generiert werden Merkmale mit den in Spalte 2 der Tabelle 1 aufgeführten Bezeichnungen. Im Falle eines plausiblen Merkmalswerts entspricht der Merkmalswert des plausibilisierten Merkmals dem des Originalmerkmals.

Nur bei Merkmalswerten mit Korrekturbedarf weicht der Merkmalswert der plausibilisierten Merkmale von den Originalangaben ab. Tabelle 1 benennt in Spalte 4 den Anteil der Merkmalswerte je Merkmal, die durch die Plausibilisierung verändert wurden. Bezogen auf die Merkmale der Tabelle 1 sind durch die Plausibilisierung in den Erstergebnissen des Mikrozensus 2020 jeweils weniger als 1% der Merkmalswerte zu korrigieren. Insgesamt gab es in 2,6% der Haushalte einen Korrekturbedarf zu mindestens einer Angabe.

In Tabelle 2 ist in Spalte 3 ausgewiesen, wie viele Verstöße in den Mikrodaten vorliegen, welche die Grundlage der Erstergebnisse des Mikrozensus 2020 bilden. Insgesamt liegen 651 609 Sätze in den Mikrodaten zu den Erstergebnissen des Berichtsjahres 2020 vor. Gegen 24 Kriterien liegt mindestens ein Verstoß vor. Das Maximum mit 2 126 Verstößen ist beim oben erläuterten Kriterium 3.11 gegeben. Dies entspricht einem Anteil nicht plausibler Angaben von 0,3%.

Durch den Plausibilisierungsalgorithmus werden diese Inkohärenzen bereinigt. Die durch die Plausibilisierung erzeugten Merkmale weisen keine Inkohärenzen mehr für die Bevölkerung in Hauptwohnsitzhaushalten auf (siehe Tabelle 2, Spalte 4).

Der eingesetzte Plausibilisierungsalgorithmus erlaubt die automatisierte Plausibilisierung der Angaben zu Eltern und Partnern in den Einzeldaten des Mikrozensus. Damit ist eine zur Erstellung von Typisierungen und Bandsatzerweiterungen der Lebensformen vorausgesetzte hinreichende Datenqualität in den benannten

Merkmale gesichert. Dabei lässt sich der Algorithmus gleichermaßen für Quartals- wie Jahresdaten einsetzen; seine Nutzung reduziert den manuellen Aufwand zur Durchführung der Plausibilisierung auf ein Minimum. Damit leistet der automatisierte Plausibilisierungsalgorithmus einen Beitrag zur schnelleren Bereitstellung von Ergebnissen.

Zugleich lässt sich die Korrekturleistung des Algorithmus evaluieren. Zum einen ist es möglich, die Anzahl der Verstöße gegen die 40 Plausibilisierungskriterien zu messen. Zum anderen lässt sich für die plausibilisierten Merkmale der Anteil der Merkmalswerte bestimmen, die durch den Algorithmus modifiziert wurden.

3

Lebensformenkonzept

3.1 Lebensformen in der Berichterstattung bis einschließlich 2019

Seit 2005 wird auf Basis des Lebensformenkonzepts über Familien und Lebensformen berichtet (Nöthen, 2005). In diesem Kontext werden Personen, die zum Zeitpunkt der Befragung im Haushalt leben, zu Lebensformen gruppiert. Sofern Personen mit ihrem Partner oder ihrer Partnerin in einem Haushalt leben, bilden diese Personen mit ihrem Partner beziehungsweise ihrer Partnerin eine gemeinsame Lebensform. In eine Lebensform werden auch die Kinder der Personen integriert, sofern diese Kinder den Familienstand „ledig“ aufweisen und selbst weder mit Partner beziehungsweise Partnerin noch mit eigenen Kindern im Haushalt leben. Das Alter der Kinder bleibt im Lebensformenkonzept unberücksichtigt.

Das Lebensformenkonzept unterscheidet damit vier grundlegende Lebensformen: (1) Paare mit ledigen Kindern, (2) Paare ohne lediges Kind, (3) Alleinerziehende: Lebensformen mit einem Elternteil und ledigen Kindern und (4) Alleinstehende.

Als Familien gelten hierbei die Lebensformen mit ledigen Kindern, das heißt Paare mit ledigen Kindern oder Alleinerziehende. Damit bestehen Familien immer aus genau zwei Generationen. Personen, welche den Haushalt bereits verlassen haben, werden im Lebensformen-

konzept nicht berücksichtigt. Unter den Paaren ohne Kind können auch Paare sein, deren Kinder bereits ausgezogen sind.

Lebensformen weisen genau eine Bezugsperson auf. Seit dem Mikrozensus 2005 ist die Bezugsperson in gemischtgeschlechtlichen Partnerschaften der männliche Partner, bei gleichgeschlechtlichen Partnerschaften der ältere Partner beziehungsweise die ältere Partnerin, bei Alleinerziehenden der alleinerziehende Elternteil und bei Alleinstehenden die Person selbst. Bei gleichgeschlechtlichen Partnerschaften gleichaltriger Partner beziehungsweise Partnerinnen entscheidet die Reihenfolge, in der die Partner beziehungsweise Partnerinnen im Fragebogen eingetragen sind. Bezugsperson dieser Partnerschaft ist dann der Partner oder die Partnerin mit der niedrigeren Personenummer.

Die Bezugsperson erfüllt eine Repräsentanzfunktion für die Lebensform. Werden in der Berichterstattung Lebensformen nach Eigenschaften dargestellt, bei denen es sich um Individualeigenschaften und nicht um Eigenschaften der Lebensform als solche handelt, überträgt die Bezugsperson ihre Individualeigenschaft an die Lebensform. So werden Lebensformen nach dem Alter der Bezugsperson der Lebensformen ausgewiesen (zum Beispiel in Fachserie 1 Reihe 3, Tabelle 3.1; Statistisches Bundesamt, 2021).

In der Berichterstattung werden Lebensformen am Hauptwohnsitz ausgewiesen. Hierbei überträgt die Bezugsperson ihre individuelle Wohnsituation auf die Lebensform. Lebt die Bezugsperson am Hauptwohnsitz, so handelt es sich um eine Lebensform am Hauptwohnsitz. Lebensformen können hierbei nach weiteren Eigenschaften unterschieden werden. So weist die Berichterstattung unter anderem Paare danach aus, ob die Partner verheiratet oder gleichgeschlechtlich sind.

Die Berichterstattung umfasst dabei nicht nur die Darstellung der Lebensformen, sondern weist ebenfalls die Bevölkerung in Lebensformen aus. Die Bevölkerung in Lebensformen am Hauptwohnsitz entspricht der Menge der Personen, welche Mitglied einer Lebensform am Hauptwohnsitz sind.

3.2 Lebensformen in der Berichterstattung ab 2020

Die Realisierung des Mikrozensus als integrierte Haushaltserhebung erfordert auch Harmonisierungen in den Veröffentlichungskonzepten. Dies betrifft auch das Lebensformenkonzept.

Weiterhin werden zur Bildung von Lebensformen ausschließlich Mitglieder desselben Haushalts berücksichtigt. Nach wie vor stehen Partnerschafts- und Eltern/Kind-Beziehungen im Zentrum des Lebensformenkonzepts. Wie auch vor dem Berichtsjahr 2020 werden Befragte zu Lebensformen gruppiert. Haushaltsmitglieder bilden mit ihrem Partner oder ihrer Partnerin eine Lebensform, sofern ein solcher beziehungsweise eine solche im Haushalt vorhanden ist. Daneben bilden Kinder mit ihren Eltern eine Lebensform, sofern die Kinder nicht selbst mit eigenen Kindern oder einem Partner beziehungsweise einer Partnerin im Haushalt leben. Damit unterscheidet das Lebensformenkonzept noch immer nach Alleinstehenden, Alleinerziehenden sowie Paaren mit Kindern und Paaren ohne Kind zwischen vier grundlegenden Lebensformen. Nach wie vor werden Lebensformen, in welchen Kinder leben, als Familien ausgewiesen. Das Alter der Kinder bleibt weiterhin unberücksichtigt. Familien sind also auch künftig Alleinerziehende und Paare mit Kindern. Ebenso wird noch immer in der Lebensform eine Bezugsperson bestimmt, deren Individualeigenschaften an die Lebensform übergehen kann. Die Bestimmung der Bezugsperson bleibt ebenfalls unverändert.

Eine erste Modifikation des Lebensformenkonzepts bezieht sich auf den Familienstand der Kinder. In der Ausgestaltung des Lebensformenkonzepts bis zum Berichtsjahr 2019 mussten Kinder den Familienstand „ledig“ aufweisen. Dem Konzept entsprechend, welches in anderen Modulen des integrierten Mikrozensus verwendet wird, bleibt der Familienstand bei der Abgrenzung von Kindern im Lebensformenkonzept ab 2020 unberücksichtigt. Lebt eine geschiedene Person mit den eigenen Eltern in einem Dreipersonenhaushalt zusammen, wurden bis einschließlich 2019 zwei Lebensformen gebildet. Die geschiedene Person wurde als „alleinstehend“ klassiert, die Eltern bildeten ein Paar ohne Kind. Ab dem Jahr 2020 müssen Kinder nicht mehr den Familienstand ledig aufweisen. Entsprechend handelt

es sich ab dem Berichtsjahr 2020 in dem beschriebenen Fall um ein Paar mit Kind.

Eine weitere Konzeptmodifikation besteht darin, dass ab dem Berichtsjahr 2020 nicht mehr die Lebensformen am Hauptwohnsitz die Menge der Lebensformen in der Berichterstattung bilden, sondern die Lebensformen in Hauptwohnsitzhaushalten.

Jede Lebensform ist Teil eines Haushalts. In den meisten Fällen besteht ein Haushalt aus genau einer Lebensform. Als Hauptwohnsitzhaushalte werden solche Haushalte bezeichnet, in denen mindestens eine Person im Alter von mindestens 16 Jahren mit Hauptwohnsitz lebt (Hochgürtel/Weinmann, 2020).

Lebensformen in Hauptwohnsitzhaushalten sind Lebensformen, welche (1.) in Hauptwohnsitzhaushalten ansässig sind **und** (2.) in welchen mindestens eine Person mit Hauptwohnsitz lebt. Nebenwohnsitz-Lebensformen sind hingegen solche Lebensformen, (1.) deren Mitglieder alle mit Nebenwohnsitz im Haushalt ansässig sind **oder** (2.) welche sich in Nebenwohnsitz-Haushalten befinden.

Zur Bestimmung einer Lebensform in Hauptwohnsitzhaushalten spielt das Alter der Personen – anders als bei der Abgrenzung von Haushalten in Hauptwohnsitzhaushalten – keine Rolle. Die Vorgehensweise zeigt auch, dass im Unterschied zur Berichterstattung bis einschließlich 2019 jedes Mitglied der Lebensform und nicht mehr nur die Bezugsperson den individuellen Hauptwohnsitzstatus an die Lebensform übertragen kann.

In einem Hauptwohnsitzhaushalt mit genau einer Lebensform handelt es sich bei dieser immer um eine Lebensform in einem Hauptwohnsitzhaushalt. Befinden sich in einem Hauptwohnsitzhaushalt hingegen mehrere Lebensformen, muss es sich nicht bei allen um eine Lebensform im Hauptwohnsitzhaushalt handeln. Ein konkretes Beispiel hierfür: In einem Hauptwohnsitzhaushalt, der aus zwei alleinstehenden Personen A und B besteht, ist Person A 20 Jahre alt und in der Wohnung mit Hauptwohnsitz gemeldet. Die zweite Person B ist 18 Jahre alt und in der Wohnung mit Nebenwohnsitz gemeldet. Hier handelt es sich also um einen Hauptwohnsitzhaushalt mit zwei Lebensformen, wobei lediglich Person A eine Lebensform in Hauptwohnsitzhaushalten bildet. Die Lebensform der Person B gehört zwar einem Hauptwohnsitzhaushalt an, bildet aber

aufgrund des individuellen Nebenwohnsitz-Status eine Nebenwohnsitz-Lebensform.

Als Bevölkerung in Lebensformen in Hauptwohnsitzhaushalten wird die Menge der Mitglieder von Lebensformen in Hauptwohnsitzhaushalten ausgewiesen. Zur Bevölkerung in Hauptwohnsitzhaushalten zählen hingegen alle Haushaltsmitglieder von Hauptwohnsitzhaushalten. Das obige Beispiel zeigt, dass die Bevölkerung in Hauptwohnsitzhaushalten von der Bevölkerung in Lebensformen in Hauptwohnsitzhaushalten abweichen kann. Die Bevölkerung in Lebensformen in Hauptwohnsitzhaushalten ist eine Teilmenge der Bevölkerung in Hauptwohnsitzhaushalten.

4

Typisierungen und Bandsatzerweiterungen für die Lebensformenberichterstattung

Bei der Datenaufbereitung des Mikrozensus werden für die Lebensformenberichterstattung aus den Befragungsinformationen mit Typisierungen und Bandsatzerweiterungen weitere Merkmale gebildet. Typisierung bezeichnet dabei die Bildung neuer Merkmale, bei der aus den Individualangaben einer Person weitere redundante Merkmale erzeugt werden. So wird etwa auf Basis der individuellen Angabe, ob der Ehepartner beziehungsweise die Ehepartnerin oder der Lebenspartner beziehungsweise die Lebenspartnerin im Haushalt lebt, ein weiteres Merkmal „Partner im Haushalt“ angelegt. Bandsatzerweiterungen sind dabei die Bildung solcher Merkmale, bei denen Aggregatsinformationen erzeugt und der jeweiligen Einheit (hier: Lebensform) zugeordnet werden. Beispielsweise wird für Familien das Merkmal „Alter des jüngsten Kindes“ erstellt und an alle Personen der Familie in der Mikrodaten-Datei angespielt.

Für die Lebensformenberichterstattung werden auf diese Weise mittels Typisierungen und Bandsatzerweiterungen rund 100 neue Merkmale gebildet. Deren Bildung orientiert sich vorrangig an den Merkmalen, die auch bereits vor 2020 erzeugt wurden (Lengerer und andere, 2005). Bei einigen Merkmalen wurde deren Erstellung methodisch neu gestaltet. So wurde etwa die Methode

zur Abbildung von Generationenbeziehungen im Haushalt neu gefasst (Hochgürtel, 2021).

In den Typisierungen und Bandsatzerweiterungen werden aus gegebenen Informationen der Befragten weitere Informationen abgeleitet. In einem automatisierten Prozess werden auf Basis von Ausgangsmerkmalen neue Merkmale erzeugt. Im Mikrozensus besteht für die meisten Merkmale Auskunftspflicht. Sie gewährleistet, dass die Angaben zu den Ausgangsmerkmalen in den meisten Fällen vorliegen. Im Mikrozensus gibt es zu Merkmalen mit Auskunftspflicht in der Regel so gut wie keinen Item-Nonresponse. Als Item-Nonresponse wird hierbei der Fall beschrieben, dass ein Merkmalsträger zwar grundsätzlich an der Befragung teilnimmt, zu einem bestimmten Merkmal aber keine Angaben macht.

Dennoch fehlen vereinzelt Angaben zu Pflichtmerkmalen. So findet sich in den Mikrodaten eines Erhebungsjahrs meistens eine kleine Anzahl von Merkmalen, bei denen jeweils einige Merkmalsangaben fehlen. Dabei sind in jedem Jahr andere Merkmale von solchen Missings betroffen. Auch Ausgangsmerkmale für Typisierungen und Bandsatzerweiterungen können von Missings betroffen sein.

Damit auch für Merkmalsträger, die in den Ausgangsmerkmalen Missings vorliegen haben, eine Typisierung und Bandsatzerweiterung erfolgen kann, werden für ausgewählte Variablen fehlende Werte in Ausgangsmerkmalen imputiert. Der fehlende Wert wird so durch einen geschätzten Wert ersetzt.

An dieser Stelle kommt eine Hot-Deck-Imputation zum Einsatz (Schafer/Graham, 2002). Dabei werden die fehlenden Angaben eines Merkmals durch die Angaben von Merkmalsträgern imputiert, welche für das entsprechende Merkmal eine Angabe getätigt haben. Für jeden zu imputierenden Merkmalswert wird zufällig eine Angabe aus den gemessenen Merkmalswerten gezogen.

Bestimmt werden muss, für welche Merkmale im Falle von Item-Nonresponse eine Imputation durchgeführt werden soll. Diese Merkmale werden im Folgenden mit X bezeichnet. Daneben muss benannt werden, welche Teilmenge günstig im Sinne des jeweiligen Merkmals ist.

Hat ein Merkmalsträger j in einem Merkmal ein Item-Nonresponse, so wird bei dem im Zuge der Imputation erzeugten Merkmal \tilde{X}_i dem Merkmalsträger j zufällig ein Merkmalswert zugewiesen. Die Wahrscheinlichkeit,

dass ein Merkmalsträger j eine Merkmalsausprägung k erhält, ist mit der relativen Häufigkeit von k gegeben.

Jedem Merkmalsträger j ist eine univariat-verteilte Zufallszahl z_j aus dem Intervall von 0 bis 1 zugewiesen. Bei der Erzeugung der Zufallszahl sind die Merkmalsträger immer einheitlich sortiert. Daneben wird bei der Erzeugung von z ein fester Startwert (Seed) verwendet. Somit werden bei wiederholter Ausführung der Imputation immer die gleichen Zufallszahlen erzeugt und damit auch immer der gleiche Wert k imputiert.

Hat der Merkmalsträger j kein Item-Nonresponse, so gilt $\tilde{X}_i = X_i$.

Als Formel ist die Hot-Deck-Imputation unten dargestellt.

Dabei gilt Folgendes:

i indiziert die Merkmale X , für welche die Imputation durchgeführt wird

\tilde{X}_i Merkmal mit imputierten Werten für Missings in X_i

j indiziert die günstigen Merkmalsträger von X_i

k indiziert die Merkmalsausprägungen von X_i

n Anzahl der günstigen Merkmalsträger

hrf_j Hochrechnungsfaktor

z_j Zufallszahl des Merkmalsträgers j

Die Anwendung des Hot-Deck-Verfahrens ist aufgrund der folgenden Eigenschaften an dieser Stelle vorteilhaft:

- › Die beschriebene Hot-Deck-Imputation lässt sich vollständig automatisieren. Die Imputation ist in die automatisierte Erstellung von Typisierungen und Bandsatzerweiterungen integriert. Imputationen werden in vorübergehend angelegten Hilfsmerkmalen realisiert. Diese Hilfsmerkmale werden für weitere Schritte der Typisierungen und Bandsatzerweiterungen genutzt.

- › Die Methode ist für fast alle diskreten Merkmale gleichermaßen geeignet. Dabei muss vorab nicht bekannt sein, in welchen Merkmalen Missings auftreten. Einzelne Merkmale lassen sich aber auch aus der Methode ausschließen.

- › Nur für „echten“ Item-Nonresponse erfolgt eine Imputation. Die Methode kann unterscheiden, ob ein Merkmalsträger keine Angabe im Sinne eines Item-Nonresponse getätigt hat oder ob der Merkmalsträger nicht zum entsprechenden Merkmal befragt wurde. Nur wenn ein Item-Nonresponse vorliegt, wird eine Imputation vorgenommen.

- › Die imputierten Merkmalswerte sind reproduzierbar. Durch die Verwendung einer einheitlichen Sortierung sowie eines Startwertes (Seed) werden bei wiederholt durchgeführten Imputationen die gleichen Ergebnisse erzeugt.

- › Sofern der Item-Nonresponse durch MCAR verursacht ist (missing completely at random; das heißt die Wahrscheinlichkeit eines Item-Nonresponse ist unabhängig von der Merkmalsausprägung [Little/Rubin, 2002]), wird durch die Hot-Deck-Imputation die Erwartungstreue der Schätzer der univariaten Verteilung des entsprechenden Merkmals nicht verletzt.

- › Die Imputation lässt sich einem Monitoring unterziehen. Es wird dokumentiert, für welche Merkmale welche Anzahl von Imputationen vorgenommen wurde.

Die hier skizzierte Imputation wird angewandt, um den automatisierten Ablauf zur Erstellung von Typisierung und Bandsatzerweiterung nicht zu stören. Bisher wurde im Kontext der Typisierung und Bandsatzerweiterung in jedem Jahr nur für wenige Merkmale eine Imputation vorgenommen. Auch die Anzahl der vorgenommenen Imputationen je Merkmal ist niedrig. So hat die Imputation keinen signifikanten Einfluss auf die Ergebnisse der Typisierung und Bandsatzerweiterung.

Formeldarstellung der Hot-Deck-Imputation

$$\tilde{X}_{ij} = \begin{cases} X_{ij} & \text{wenn } X_{ij} \neq \text{missing} \\ k & \text{wenn } X_{ij} = \text{missing} \end{cases} \cap \frac{\sum_{j=1}^n hrf_j | X_{ij} \neq \text{missing} \cap X_{ij} < k}{\sum_{j=1}^n hrf_j | X_{ij} \neq \text{missing}} < z_j \leq \frac{\sum_{j=1}^n hrf_j | X_{ij} \neq \text{missing} \cap X_{ij} \leq k}{\sum_{j=1}^n hrf_j | X_{ij} \neq \text{missing}}$$

5

Ausblick

Mit der Weiterentwicklung des Mikrozensus zu einem integrierten System der Haushaltsstatistiken sind für die Lebensformenberichterstattung eine Reihe von Herausforderungen verbunden. Zum einen erfordert ein integriertes System auch eine harmonisierte Berichterstattung. Dieser wurde mit den beschriebenen Konzeptanpassungen Rechnung getragen.

Zum anderen erhöhen sich die Anforderungen an die Veröffentlichungstätigkeit. Einerseits werden Ergebnisse als Erstergebnisse rund vier Monate früher veröffentlicht als dies bisher der Fall war. Andererseits ist mit der Veröffentlichung von Endergebnissen eine weitere Veröffentlichungspflicht verbunden.


Um diese erhöhten Anforderungen zu erfüllen, setzt die Statistikproduktion der Lebensformenberichterstattung in stärkerem Maße als bisher auf Standardisierung und Automatisierung. Dies geht über die hier beschriebenen Prozessschritte der Plausibilisierung, Imputation sowie die Erstellung von Typisierungen und Bandsatzerweiterungen hinaus.

Auch in weiteren Schritten der Statistikproduktion werden verstärkt standardisierte und automatisierte Verfahren eingesetzt. Vor dem hier beschriebenen Prozess der Plausibilisierung werden ausgewählte Merkmale der Mikrodaten des Mikrozensus unter anderem automatisiert darauf geprüft, ob Missings vorliegen, gültige Wertebereiche verletzt werden und inkohärente Angaben bestehen.

Daneben werden quartals- und jahresweise automatisierte und standardisierte Qualitätsanalysen durchgeführt. So werden Responsequoten bestimmt und Zeitreihen zentraler Merkmale der Lebensformenberichterstattung gebildet.

Standardisierte und automatisierte Verfahren werden auch für die Produktion von Veröffentlichungsprodukten eingesetzt, sowohl für die weitgehend automatisierte Erstellung der Fachserie als auch für die Prüfung, beispielsweise von Datenbank-Tabellen sowie regelmäßig erstellten Auskunft- und Veröffentlichungstabellen. Für die Erstellung von Sonderauswertungen wurden eben-

falls Instrumente entwickelt, die den manuellen Aufwand deutlich reduzieren.

Auch in Zukunft wird der Ausbau von standardisierten und automatisierten Prozessen weiter fortgesetzt. 

LITERATURVERZEICHNIS

Bihler, Wolf/Zimmermann, Daniel. [Die neue Mikrozensusstichprobe ab 2016](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2016, Seite 20 ff.

Emmerling, Dieter/Riede, Thomas. [40 Jahre Mikrozensus](#). In: Wirtschaft und Statistik. Ausgabe 3/1997, Seite 160 ff.

Herberger, Lothar. [Der Mikrozensus als neues Instrument zur Erfassung sozial-ökonomischer Tatbestände](#). In: Wirtschaft und Statistik. Ausgabe 4/1957, Seite 209 ff.

Hochgürtel, Tim. [Das künftige System der amtlichen Haushaltsstatistiken](#). In: Wirtschaft und Statistik. Ausgabe 7/2013, Seite 457 ff.

Hochgürtel, Tim. [Haushalte mit drei Generationen](#). In: WISTA Wirtschaft und Statistik. Ausgabe 5/2021, Seite 114 ff.

Hochgürtel, Tim/Rammelt, Sabine. [Die auskunftspflichtige Erfassung von Lebensgemeinschaften im Mikrozensus ab 2017](#). In: WISTA Wirtschaft und Statistik. Ausgabe 4/2018, Seite 47 ff.

Hochgürtel, Tim/Weinmann, Julia. [Haushalte in der Berichterstattung des Mikrozensus ab 2020](#). In: WISTA Wirtschaft und Statistik. Ausgabe 3/2020, Seite 89 ff.

Hundenborn, Janina/Enderer, Jörg. [Die Neuregelung des Mikrozensus ab 2020](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2019, Seite 9 ff.

Lengerer, Andrea/Bohr, Jeanette/Janßen, Andrea. *Haushalte, Familien und Lebensformen im Mikrozensus: Konzepte und Typisierungen*. In: ZUMA (Zentrum für Umfragen, Methoden und Analysen)-Arbeitsbericht. Ausgabe 2005/05. 2005. [Zugriff am 20. Dezember 2021]. Verfügbar unter: nbn-resolving.org

Little, Roderick, A./Rubin, Donald, B. *Statistical analysis with missing data*. Zweite Auflage. New York 2002.

Lotze, Sabine/Breiholz, Holger. [Zum neuen Erhebungsdesign des Mikrozensus](#). In: Wirtschaft und Statistik. Ausgabe 5/2002, Seite 359 ff.

Nöthen, Manuela. [Von der „traditionellen Familie“ zu „neuen Lebensformen“](#). In: Wirtschaft und Statistik. Ausgabe 1/2005, Seite 25 ff.

Schafer, Joseph L./Graham, John W. *Missing data: our view of the state of the art*. In: Psychological methods. 2002. Ausgabe Juni 7(2), Seite 147 ff.

Schmidt, Marcus/Stein, Jana Lucia. [Die Hochrechnung im Mikrozensus ab 2020](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2021, Seite 54 ff.

Statistisches Bundesamt. *Haushalte und Familien. Ergebnisse des Mikrozensus 2020 (Erstergebnisse)*. [Zugriff am 5. Januar 2022]. Verfügbar unter: www.destatis.de

Herausgeber
Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung
Dr. Daniel Vorgrimler
Redaktion: Ellen Römer

Ihr Kontakt zu uns
www.destatis.de/kontakt

Erscheinungsfolge
zweimonatlich, erschienen im Februar 2022
Ältere Ausgaben finden Sie unter www.destatis.de sowie in der [Statistischen Bibliothek](#).

Artikelnummer: 1010200-22001-4, ISSN 1619-2907

© Statistisches Bundesamt (Destatis), 2022
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.