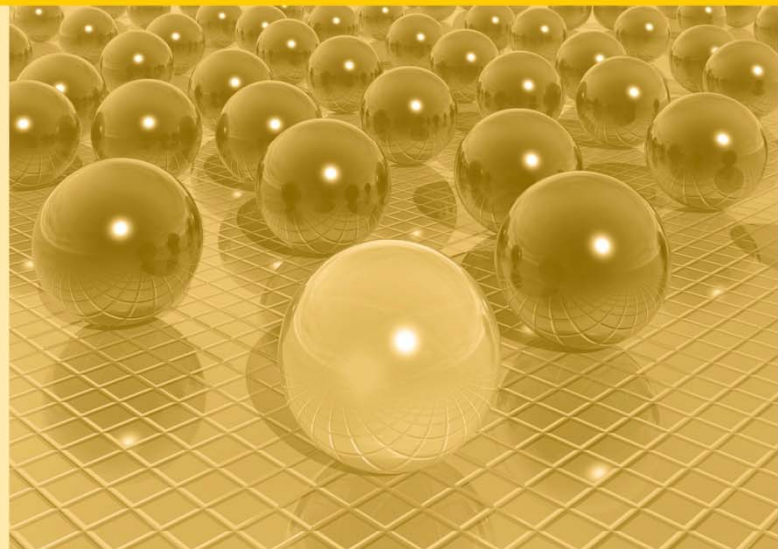


Metadatenreport



Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus 2014 am
Gastwissenschaftsarbetsplatz sowie per kontrollierter Datenfernverarbeitung
(sog. On-Site-Zugangswege/-Nutzungen)

DOI: 10.21242/12211.2014.00.00.1.1.1

Version 1

Impressum

Herausgeber: Statistische Ämter des Bundes und der Länder
Herstellung: Information und Technik Nordrhein-Westfalen
Telefon 0211 9449-01 • Telefax 0211 9449-8000
Internet: www.forschungsdatenzentrum.de
E-Mail: forschungsdatenzentrum@it.nrw.de

Fachliche Informationen

zu dieser Veröffentlichung:

Forschungsdatenzentrum der
Statistischen Ämter der Länder
– Düsseldorf –
Tel.: 0211 9449-2877
Fax: 0211 9449-8087
forschungsdatenzentrum@it.nrw.de

Informationen zum Datenangebot:

Statistisches Bundesamt
Forschungsdatenzentrum
Tel.: 0611 75-2420
Fax: 0611 75-3915
forschungsdatenzentrum@destatis.de

Forschungsdatenzentrum der
Statistischen Ämter der Länder
– Geschäftsstelle –
Tel.: 0211 9449-2883
Fax: 0211 9449-8087
forschungsdatenzentrum@it.nrw.de

Erscheinungsfolge: unregelmäßig
Erschienen im Dezember 2020

Diese Publikation wird kostenlos als PDF-Datei zum Download unter www.forschungsdatenzentrum.de angeboten.

© Information und Technik Nordrhein-Westfalen, Düsseldorf, 2020
(im Auftrag der Herausbergemeinschaft)

Vervielfältigung und Verbreitung, nur auszugsweise, mit Quellenangabe gestattet. Alle übrigen Rechte bleiben vorbehalten.

Fotorechte Umschlag: ©artSILENCE –Fotolia.com

Empfohlene Zitierung:

Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder: Metadatenreport. Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus 2014 am Gastwissenschafts Arbeitsplatz sowie per kontrollierter Datenfernverarbeitung (EVAS-Nummer: 12211, 12212, 12213). Version 1. DOI: 10.21242/12211.2014.00.00.1.1.1 Düsseldorf 2020.

Metadatenreport

Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus 2014 am Gastwissenschaftsarbetsplatz sowie per kontrollierter Datenfernverarbeitung (sog. On-Site-Zugangswege/-Nutzungen)

DOI: 10.21242/12211.2014.00.00.1.1.1

Version 1

Inhalt

1	Datenaufbereitung in den FDZ.....	2
1.1	Datenaufbereitung	2
1.2	Anonymisierungsmaßnahmen.....	3
1.3	Methodik der Verknüpfung.....	3
2	Produkt.....	4
2.1	Merkmale und Merkmalsbeschreibung.....	4
2.1.1	Merkmalsdefinitionen	4
2.1.2	Datensatzbeschreibung / Schlüsselverzeichnis.....	4
2.2	Vergleichbarkeit der Merkmale über die Zeit.....	5
2.3	Eckwerte relevanter Merkmale und Merkmalskombinationen	8
2.4	Auswertbare regionale Ebene.....	9
2.5	Produktversionen.....	10
3	Praktische Hinweise.....	11
3.1	Hinweise zur Geheimhaltung	11
3.1.1	Gesetzliche Grundlagen der statistischen Geheimhaltung.....	11
3.1.2	Geheimhaltung von Ergebnissen	11
3.1.3	Praktische Tipps zur Vermeidung von Geheimhaltungsfällen.	12
3.2	FAQ.....	12
3.3	Verfügbare Tools.....	14
	Anhang.....	15

1 Datenaufbereitung in den FDZ

1.1 Datenaufbereitung

Das Datenmaterial wird einer Vollzähligkeits- und Vollständigkeitskontrolle unterzogen, bei der geprüft wird, ob alle Erhebungsmerkmale und die dazugehörigen, typisierten Merkmale, die für das Mikrozensusjahr vorgesehen sind, belegt sind. Nicht belegte Merkmale werden aus dem Datenmaterial entfernt. Gleichzeitig wird darauf geachtet, dass die einzelnen Variablen die korrekten Formate aufweisen.

Im Zuge des Abgleichs mit den zentralen Veröffentlichungen¹ des Statistischen Bundesamts (siehe Metadatenreport Teil I, Kapitel 4 Zentrale Veröffentlichungen) werden die Hochrechnungsfaktoren so angeglichen, dass sie ohne zusätzliche Multiplikation auf die Gesamtbevölkerung hochrechnen². Fälle, die keinen positiven Hochrechnungsfaktor besitzen, werden aus dem Datenmaterial entfernt. Dies betrifft ca. 9% der Fälle aus dem Originaldatenmaterial. Für die betroffenen Haushalte liegt kein Interview vor, weil es sich beispielsweise um leerstehende Wohnungen oder weggefallene Haushalte handelt.

Die einzelnen Merkmale werden auf inhaltliche Konsistenz geprüft. Inkonsistenzen werden im Schlüsselverzeichnis (siehe auch Kapitel 2.1.2) dokumentiert und, sofern diese plausibel auflösbar sind, bereinigt. Dabei wird gewährleistet, dass die inhaltliche Aussage der Angaben nicht verändert wird.

Weiterhin wird geprüft, ob sich die Filterführung laut Fragebogen im Datenmaterial widerspiegelt. Für Personengruppen, denen gemäß Filterführung bestimmte Fragen nicht gestellt wurden, werden für die betroffenen Merkmale systemfehlende Werte vergeben (Ausprägung „leer“). Treten Inkonsistenzen bei der Filterführung auf, werden diese im Schlüsselverzeichnis (siehe auch Kapitel 2.1.2) dokumentiert und gegebenenfalls bereinigt, ohne die inhaltliche Aussage der Angaben zu verändern.

Zudem wird überprüft, ob unterschiedliche fachliche Ebenen, wie Haushalte, Familien und Lebensformen korrekt abgegrenzt und ausgewertet werden können.

Außerdem wird der in den Daten vorhandene Gebietsstand ermittelt. Die einzelnen Regionalkennungen (siehe Abschnitt 2.4 Auswertbare regionale Ebene) werden in diesem Zuge auf inhaltliche

¹ Exemplarisch finden Sie im Anhang dieses Reports Häufigkeitstabellen sowie die dazugehörige SPSS-Syntax mit Verweis auf die jeweilige Publikation des Statistischen Bundesamtes.

² Nähere Informationen zu den in diesem Produkt vorliegenden Hochrechnungsfaktoren finden Sie im Abschnitt 3.2 FAQ „Wann wird welcher Hochrechnungsfaktor verwendet?“.

Konsistenz geprüft. Treten Inkonsistenzen auf, werden diese im Schlüsselverzeichnis (siehe auch Kapitel 2.1.2) dokumentiert und gegebenenfalls bereinigt.

1.2 Anonymisierungsmaßnahmen

Direkte Identifikatoren und Hilfsmerkmale werden aus dem Datenmaterial entfernt.

Die Ordnungsnummern des Auswahlbezirks, des Haushalts im Auswahlbezirk und der Person im Haushalt werden systemfrei sortiert, so dass anhand der Position des einzelnen Falls im Datenmaterial kein Rückschluss auf einzelne Personen möglich ist.

Im Datenmaterial für die Gastwissenschaftsarbeitsplätze ist die Gemeindeebene für das Bundesland Bayern pseudonymisiert, so dass zwar die Information verfügbar ist, ob zwei Haushalte derselben Gemeinde angehören, nicht aber, um welche Gemeinde es sich handelt.

1.3 Methodik der Verknüpfung

Der Mikrozensus stellt kein bereits verknüpftes Produkt dar, allerdings besteht die Möglichkeit, mit Hilfe der EF1- EF4 und EF 63, auf Individualebene die Jahrgänge 2012-2015 miteinander zu verknüpfen³.

Da sich nach 2015 die Auswahlgrundlage der Stichprobe verändert hat, ist keine Verknüpfung der Jahre 2012-2015 mit nachfolgenden Erhebungsjahren möglich⁴.

³ Informationen zur Verknüpfungsmethode und Mustersyntaxen sind bei Herter-Eschweiler und Schimpel-Neimanns 2018 zu finden.

⁴ Genauere Informationen in Bihler/Zimmermann 2016 (<https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2016/06/die-neue-mikrozensusstichprobe-062016.pdf>).

2 Produkt

2.1 Merkmale und Merkmalsbeschreibung

2.1.1 Merkmalsdefinitionen

Folgende Klassifikationen finden in dem beschriebenen Produkt Anwendung:

- Klassifikation der Berufe, Ausgabe 2010 (KldB 2010)
<https://statistik.arbeitsagentur.de/DE/Navigation/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010/Systematik-Verzeichnisse/Systematik-Verzeichnisse-Nav.html>
- Internationale Standardklassifikation der Berufe, Ausgabe 2008 (ISCO-08)
<http://www.ilo.org/public/english/bureau/stat/isco/isco08/>
- Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008)
<https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/Downloads/klassifikation-wz-2008-3100100089004.pdf>
- Bildungsklassifikation International Standard Classification of Education, Ausgabe 2011 (ISCED-2011)
<http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf>
- Staats- und Gebietssystematik
https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/Staatsangehoerigkeitsgebietsschluessel_pdf.pdf
- Amtlicher Gemeindegchlüssel (AGS)/ Regionalschlüssel (RS)
<https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/inhalt.html>

2.1.2 Datensatzbeschreibung / Schlüsselverzeichnis

Der Mikrozensus ist eine Haushalts-, Personen- und Wohnungsstatistik (in Jahren mit Zusatzprogramm „Wohnen“, wie 2014). Die vollständige Liste der im Datensatz vorhandenen Variablen finden Sie im Schlüsselverzeichnis. Sofern Variablen das Ergebnis einer Frage des Fragebogens wiedergeben, finden Sie die zugehörige Fragennummer in der Spalte „Fragebogennummer“ des Schlüsselverzeichnisses. Das Zusatzprogramm beschäftigt sich 2014 mit der Wohnsituation. Den vollständigen Fragebogen mitsamt des Ad-Hoc-Moduls „Arbeitsmarktsituation von Zuwanderern und ihren direkten Nachkommen“ finden Sie bei den Erhebungsunterlagen.

Schlüsselverzeichnis

https://www.forschungsdatenzentrum.de/sites/default/files/mz_2014_on-site_svz_v1.pdf

Erhebungsunterlagen

https://www.forschungsdatenzentrum.de/sites/default/files/mz_2014_on-site_eu.pdf

2014 wurde das Zusatzprogramm „Wohnsituation der Haushalte“ sowie das Ad-Hoc-Modul „Arbeitsmarktsituation von Zuwanderern sowie ihren Nachkommen“ abgefragt.

2.2 Vergleichbarkeit der Merkmale über die Zeit

Folgende Merkmale weisen im Vergleich zum Vorjahr Veränderungen auf:

Gegenwärtige (Haupt-)Erwerbstätigkeit

- EF138 (Könnten Sie innerhalb der nächsten 2 Wochen mehr Stunden arbeiten?): Die Ausprägung „9 – Keine Angabe“ wurde entfernt.
- EF196, EF196U1 – EF196U2 (In welchem Kreis, Landkreis oder kreisfreien Stadt arbeiten Sie? (Kreis)): Die Variable ist neu hinzugefügt worden.
- EF196A, EF196AU1 – EF196AU3 (In welchem Kreis, Landkreis oder kreisfreien Stadt lag damals Ihr Wohnsitz? (freiwillige Beantwortung)): Die Variable ist neu hinzugefügt worden.

Weitere Erwerbstätigkeit

- EF197 (Ist Ihre weitere Tätigkeit eine geringfügige Beschäftigung?): Die Ausprägung „9- Keine Angabe“ entfällt.

Gegenwärtiger Besuch von Hochschule und Schule

- EF291 (Falls eine Hochschule/Fachhochschule besucht wird/ wurde: Wie ist die Bezeichnung Ihres Studienganges?): Die Variable ist neu hinzugefügt worden.

Schul-/ Ausbildungsabschluss

- EF312 (Höchster beruflicher Ausbildungs- oder Hochschul-/Fachhochschulabschluss): Die Ausprägungen „16- Ausbildungsstätten, Schulen für Gesundheits- und Sozialberufe: 3- jährig“ und „17- Ausbildungsstätten/Schulen für Erzieher/-innen“ wurden hinzugefügt, eine Vergleichbarkeit ist gegeben.
- EF316 (Welchen höchsten weiteren beruflichen Abschluss?): Die Ausprägungen „11- Ausbildungsstätten, Schulen für Gesundheits- und Sozialberufe: 3- jährig“ und „12- Ausbildungsstätten/Schulen für Erzieher/-innen“ wurden hinzugefügt, eine Vergleichbarkeit ist gegeben.

- EF318 (Höchster erreichter Grad der allgemeinen oder beruflichen Bildung): Die Variable entfällt.
- EF320 (Höchster beruflicher Ausbildungs- oder Hochschul-/Fachhochschulabschluss mit Grad): Die Ausprägungen „14- Ausbildungsstätten, Schulen für Gesundheits- und Sozialberufe: 3- jährig“ und „15- Ausbildungsstätten/Schulen für Erzieher/-innen“ wurden hinzugefügt, eine Vergleichbarkeit ist gegeben.

Wohnsituation des Haushaltes (Zusatzprogramm – im Vergleich mit 2010 zu betrachten)

- EF489 (In welcher Art von Gebäude wohnen Sie?)
- EF490 (Wie viele Wohnungen, einschließlich der leer stehenden Wohnungen, gibt es in dem Gebäude, in dem Sie wohnen?)
- EF491 (Bewohnen Sie Ihre Wohnung/Ihr Haus als...)
- EF492 (Wie groß ist die Wohnfläche der gesamten Wohnung/ des Hauses?)
- EF493 (Wann ist Ihr Haushalt in die Wohnung/ das Haus eingezogen?)
- EF500 (Enthält die monatliche Zahlung an Ihren Vermieter/Ihre Vermieterin, Ihre Hausverwaltung Nebenkosten (kalte Nebenkosten)?)
- EF501 (Enthält die monatliche Zahlung an Ihren Vermieter/Ihre Vermieterin, Ihre Hausverwaltung Nebenkosten (warme Nebenkosten)?)
- EF503 (Welche Energie wird überwiegend für Ihre Warmwasserversorgung verwendet?)

Datenfelder mit EU-Schlüsseln/ Typisierungen/ Hochrechnungsfaktoren

- EF517 (Höchster erreichter Grad der allgemeinen oder beruflichen Bildung (ISCED 2011)): Umstellung von ISCED 97 auf ISCED 2011.
- EF540 (Höchster erreichter Grad der allgemeinen oder beruflichen Bildung (ISCED 2011)): Umstellung von ISCED 97 auf ISCED 2011.

Angaben zum Haupteinkommensbezieher im Haushalt

- EF745 (Höchster schulischer und/oder beruflicher Abschluss des Haupteinkommensbeziehers nach ISCED): Umstellung von ISCED 97 auf ISCED 2011.

Angaben zur Bezugsperson in der Familie/Lebensform (Lebensformkonzept)

- EF833 (Höchster schulischer und/oder beruflicher Abschluss des Haupteinkommensbeziehers nach ISCED): Umstellung von ISCED 97 auf ISCED 2011.

Angaben zum Ehe- oder Lebenspartner der Bezugsperson (Lebensformenkonzept)

- EF857 (Höchster schulischer und/oder beruflicher Abschluss des Ehe- oder Lebenspartners nach ISCED): Umstellung von ISCED 97 auf ISCED 2011.

EU-Unterstichprobe

- EF955 (Hochrechnungsfaktor Jahr): Die Variable wurde entfernt.

Fortsetzung Wohnsituation der Haushalte 2014

- EF986 (Fläche für Mittelwertberechnung): Die Variable kann mit dem Zusatzprogramm aus dem Jahr 2010 verglichen werden.
- EF987 (Bruttokaltmiete für Mittelwertberechnung): Die Variable kann mit dem Zusatzprogramm aus dem Jahr 2010 verglichen werden.
- EF988 (Nettokaltmiete für Mittelwertberechnung): Die Variable kann mit dem Zusatzprogramm aus dem Jahr 2010 verglichen werden.
- EF989 (Nettokaltmiete je m²): Die Variable kann mit dem Zusatzprogramm aus dem Jahr 2010 verglichen werden.
- EF990 (Warmmiete für Mittelwertberechnung): Die Variable kann mit dem Zusatzprogramm aus dem Jahr 2010 verglichen werden.
- EF991 (Warmmiete je m²): Die Variable kann mit dem Zusatzprogramm aus dem Jahr 2010 verglichen werden.
- EF992 (kalte Nebenkosten für Mittelwertberechnung): Die Variable kann mit dem Zusatzprogramm aus dem Jahr 2010 verglichen werden.
- EF993 (warme Nebenkosten für Mittelwertberechnung): Die Variable kann mit dem Zusatzprogramm aus dem Jahr 2010 verglichen werden.
- EF994 (Anzahl verwendete Energiearten für Warmwasserversorgung): Die Variable kann mit dem Zusatzprogramm aus dem Jahr 2010 verglichen werden.
- EF995 (Anzahl verwendeter Energiearten für Beheizung der Wohnräume): Die Variable kann mit dem Zusatzprogramm aus dem Jahr 2010 verglichen werden.

Ad hoc Modul

- Das Ad hoc Modul ändert sich jährlich. Das Ad hoc Modul 2014 „Arbeitsmarktsituation von Zuwanderern sowie ihren Nachkommen“ umfasst die EFs ab EF1080.

Migrationstypisierungen

- EF 2001(Migrationsstatus) besitzt seit 2014 neue Ausprägungen: (16) definiert als „Zuwanderer: als Deutscher im Ausland geboren, beide Eltern haben keinen MHG“, (26) „Spätaussiedler: mit formaler Einbürgerung, Ehegatte oder Kind eines Spätaussiedlers, zeitgleich seit 1950 zugewandert, maschinell umgesetzt“, (51) „Zuwanderer: als Deutscher im Ausland geboren, jedoch kein Spätaussiedler, einseitiger Migrationshintergrund“

2.3 Eckwerte relevanter Merkmale und Merkmalskombinationen

Verteilung der Bevölkerung am Hauptwohnsitz

Tab. 1: EF46NEU Geschlecht

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	männlich	39.651.403	49,0	49,0	49,0
	weiblich	41.245.067	51,0	51,0	100,0
Gesamt		80.896.470	100,0	100,0	

Tab. 2: EF49NEU Familienstand

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	ledig	32.800.076	40,5	40,5	40,5
	verheiratet	36.620.556	45,3	45,3	85,8
	verwitwet	11.475.838	14,2	14,2	100,0
	Gesamt	80.896.470	100,0	100,0	

Privathaushalte am Haupt- und Nebenwohnsitz

Tab. 3: EF20NEU Privathaushalte kategorisiert

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	Einpersonenhaushalte	16.411.772	40,8	40,8	40,8
	Zweipersonenhaushalte	13.837.236	34,4	34,4	75,2
	Dreipersonenhaushalte	4.968.359	12,4	12,4	87,6
	Vierpersonenhaushalte	3.672.307	9,1	9,1	96,7
	Fünf- und mehr Personenhaus- halte	1.333.091	3,3	3,3	100,0
	Gesamt	40.222.765	100,0	100,0	

Tab. 4: EF1 Bundesland

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	Schleswig-Holstein	1.419.280	3,5	3,5	3,5
	Hamburg	976.707	2,4	2,4	6,0
	Niedersachsen	3.828.927	9,5	9,5	15,5
	Bremen	359.638	,9	,9	16,4
	Nordrhein-Westfalen	8.555.079	21,3	21,3	37,6
	Hessen	2.942.851	7,3	7,3	45,0
	Rheinland-Pfalz	1.901.361	4,7	4,7	49,7
	Baden-Württemberg	5.072.521	12,6	12,6	62,3
	Bayern	6.219.208	15,5	15,5	77,8
	Saarland	491.875	1,2	1,2	79,0
	Berlin	1.965.940	4,9	4,9	83,9
	Brandenburg	1.234.973	3,1	3,1	86,9
	Mecklenburg-Vorpommern	828.806	2,1	2,1	89,0
	Sachsen	2.156.546	5,4	5,4	94,4
	Sachsen-Anhalt	1.159.751	2,9	2,9	97,2
	Thüringen	1.109.304	2,8	2,8	100,0
	Gesamt	40.222.765	100,0	100,0	

Im Anhang finden Sie diese und weitere relevante Tabellen mit der Syntax für das Programm SPSS sowie allen Fundstellen in den Fachserien des Statistischen Bundesamtes.

2.4 Auswertbare regionale Ebene

- Bundesebene (NUTS 0)
- Landesebene (NUTS 1)
- Regierungsbezirksebene (NUTS 2)
- Kreisebene (ggf. eingeschränkte Aussagekraft durch geringe Zellbesetzungen) (NUTS 3)
- Gemeindeebene (ggf. eingeschränkte Aussagekraft durch geringe Zellbesetzungen) (LAU)

- Anpassungsschichten (regionale Schicht und regionale Untergruppe)⁵
- Raumordnungsregionen
- Arbeitsmarktregionen
- Nielseengebiete
- Arbeitsagenturbezirke

2.5 Produktversionen

1.0.0

- Erstveröffentlichung

2.0.0

- Neue Datenlieferung für den Mikrozensus. Enthalten sind Angaben zum Wohnmodul.

2.1.1

- Der Amtliche Gemeindeschlüssel (AGS) des Mikrozensus wurde an den AGS des Gemeindeverzeichnisses angepasst. Betroffen davon sind Berlin, Hamburg und Aachen.

2.1.1 (DOI 10-21242-12211-2014-00-00-1-1-0)

- Die Variablen EF 313, EF 313U1 und EF 313U2 wurden in den Datensatz aufgenommen.
- Es wurde eine eindeutige Benennung der Regionalen Anpassungsschicht (EF 564), der Regionalen Schicht (EF 566) und der Regionalen Untergruppe (EF 568) vorgenommen.
- Der Standard-File-Hinweis wurde aus dem Datensatz entfernt.
- Der Hinweis auf die Pseudonymisierung bayrischer Daten (EF 560UG1) wurden in das Schlüsselverzeichnis aufgenommen.

2.2.0 (DOI 10-21242-12211-2014-00-00-1-1-1)

- Verknüpfbarkeit zum Längsschnitt:
- Die Variablen EF 3, EF 4 und EF63 wurden zur Bildung einer jahresübergreifenden Personen-ID in den Datensatz aufgenommen.

⁵ In Berlin und Hamburg ist eine Auswertung auf Bezirksebene möglich. In München kann eine Aufteilung nach innerstädtischen Bezirken und nach Bezirken in der Randlage der Stadt München vorgenommen werden.

3 Praktische Hinweise

3.1 Hinweise zur Geheimhaltung

3.1.1 Gesetzliche Grundlagen der statistischen Geheimhaltung

Unter Geheimhaltung versteht man das Sicherstellen der absoluten Anonymität der Ergebnisse statistischer Analysen. Konkret bedeutet das, dass im Rahmen der Geheimhaltung sichergestellt wird, dass mit den veröffentlichten Ergebnissen keine Rückschlüsse auf einen Einzelfall (z. B. Person, Haushalt, Betrieb, Einrichtung) gezogen werden können. Statistische Geheimhaltung wird überall dort angewendet, wo statistische Ergebnisse oder Einzeldaten die geschützten Räume der amtlichen Statistik verlassen.

Die Geheimhaltung in der amtlichen Statistik ist in § 16 Bundesstatistikgesetz (BStatG) geregelt und beinhaltet, dass Einzelangaben über persönliche und sachliche Verhältnisse, die für eine Bundesstatistik angegeben werden, von den jeweils durchführenden statistischen Stellen geheim zu halten sind, soweit es keine anderslautenden Bestimmungen gibt. Dies wird auch als Statistikgeheimnis bezeichnet. Das Statistikgeheimnis verpflichtet die amtliche Statistik, die erhaltenen Informationen zu schützen, das heißt sie in einer Form zu anonymisieren, die keine Rückschlüsse mehr auf die betreffende Person und den dargelegten Sachverhalt enthält. Die Geheimhaltung ist auch im Hinblick auf die informationelle Selbstbestimmung von besonderem Interesse: Viele Erhebungen der amtlichen Statistik unterliegen der Auskunftspflicht, somit steht es den Befragten nicht frei selbst zu entscheiden, ob sie eine Information weitergeben möchten. Die amtliche Statistik muss deshalb sicherstellen, dass die erhobenen Daten keinem Befragten zugeordnet werden können.

Das BStatG sieht jedoch auch Fälle vor, in denen das Statistikgeheimnis nicht gilt. In § 16 BStatG sind die Ausnahmen von der Geheimhaltungspflicht dargelegt. Unter anderem wird dort festgelegt, unter welchen Umständen die Daten der amtlichen Statistik für die Wissenschaft zugänglich gemacht werden dürfen und welche Regeln dabei einzuhalten sind.

3.1.2 Geheimhaltung von Ergebnissen

Um die gesetzlich vorgeschriebene Geheimhaltung von Einzelfällen in den Daten sicherzustellen, müssen alle Ergebnisse, die am Gastwissenschaftsarbetsplatz oder per Kontrollierter Datenfernverarbeitung erzeugt werden, vor ihrer Freigabe an den Nutzer von den FDZ einer Geheimhaltungsprüfung unterzogen werden. Dabei stellen die FDZ sicher, dass die Ergebnisse absolut anonym sind und eine Reidentifikation einzelner Befragter nach menschlichem Ermessen ausgeschlossen werden kann. Entsprechend handeln auch die Fachabteilungen der Statistischen Ämter vor der Veröffentlichung von Ergebnissen.

Zur Sicherstellung der Geheimhaltung wenden die FDZ verschiedene Geheimhaltungsregeln an, die individuell auf die jeweilige Statistik zugeschnitten sind. In der Broschüre „Regelungen zur Auswertung von Mikrodaten in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder“ werden die gebräuchlichsten Regeln zur primären Geheimhaltung dargestellt. Diese Regeln werden in den FDZ im Grunde auf alle Statistiken angewendet. Die Anlage dieser Broschüre enthält Informationen darüber, welche Geheimhaltungsregeln auf welche Statistiken anzuwenden sind.

Die Broschüre finden Sie hier:

https://www.forschungsdatenzentrum.de/sites/default/files/fdz_broschuere_regelungen_2019.pdf

3.1.3 Praktische Tipps zur Vermeidung von Geheimhaltungsfällen

Treten in den erstellten Analysen Geheimhaltungsfälle auf, werden diese Werte von den FDZ zur Sicherstellung der Geheimhaltung durch ein Sperrmuster ersetzt. Gerade in Kreuztabellen entstehen so durch die notwendige Sekundärspernung schnell viele Lücken in den Auswertungen. Da eine einmal zur Sekundärspernung herangezogene Tabellenzelle auch in allen folgenden Analysen gesperrt werden muss (tabellenübergreifende Geheimhaltung) – auch, wenn es in der neu erstellten Tabelle nicht nötig wäre – ist es sinnvoll, bei jeder Ergebniserstellung darauf zu achten, dass möglichst keine Geheimhaltungsfälle erzeugt werden. Treten in einem Output Geheimhaltungsfälle auf, steht es dem betreuenden FDZ frei, die Prüfung und Freigabe des Outputs abzulehnen.

Um Geheimhaltungsfälle in den Analysen zu vermeiden, sollte immer darauf geachtet werden, dass die erstellten Analysen auf ausreichend großen Fallzahlen beruhen. Bei geringen Fallzahlen empfiehlt es sich, Variablenausprägungen zusammenzufassen und damit größere Fallzahlen zu erzielen.

3.2 FAQ

Wie werden in den Mikrozensusdaten Haushalte selektiert?

In den Veröffentlichungen der amtlichen Statistik wird eine Selektion nach allen Privathaushalten (EF31=1) und der Bezugsperson im Haushalt (EF37=1) vorgenommen. Alternativ dazu ist es möglich eine eindeutige Haushalts-ID über die Aneinanderreihung der Eingabefelder Bundesland (EF1), Auswahlbezirk (EF3), Haushaltsnummer (EF4) und Kennzeichnung von Überhanginterviews (EF5U2) zu generieren und das Datenmaterial anhand dieser ID über alle Personen des Haushalts zu aggregieren. Sofern es sich nicht um ein Überhanginterview handelt (EF5U2=missing), empfiehlt es sich der Haushalts-ID eine 0 anzuhängen, um Dubletten aufgrund der Verschiebung von Stellen zu vermeiden.

Wie können Mütter und Väter ausgewertet werden?

Zunächst muss eine Haushaltsidentifikationsnummer erstellt werden. Hierfür werden folgende Variablen verkettet:

- EF1 Bundesland
- EF3 Auswahlbezirk
- EF4 Haushaltsnummer
- EF5U2 Kennzeichnung der Jahresüberhänge

Sofern EF5U2 missing ist, handelt es sich nicht um ein Überhang Interview. Um Dubletten durch die Verschiebung von Stellen zu vermeiden, ist in diesen Fällen eine 0 anzuhängen. Über die Eingabefelder EF53 und EF55 besteht die Möglichkeit die Personnummer der Mutter bzw. des Vaters zu ermitteln. Bei Überhanginterviews ist diesen Nummern EF5U2 anzuhängen. Über EF5 können diese Nummern Personen nun im Haushalt identifiziert werden. EF5 ist die EF5U2 bereits angehängt.

Wieso sind in den Daten doppelte Personennummern enthalten?

Die Doppelung entsteht durch die in den Daten enthaltenen Jahresüberhänge aus dem Vorjahr. Über das Merkmal EF5U2 können diese von den tatsächlich in diesem Jahr befragten Fällen abgegrenzt werden.

Wann wird welcher Hochrechnungsfaktor verwendet?

Für Ergebnisse auf Ebene der Anpassungsschichten und höher (zum Beispiel Bundesländer) wird der Standardhochrechnungsfaktor (EF952) verwendet. Für Ergebnisse unterhalb der Anpassungsschichten (zum Beispiel Kreisebene) bis hinunter zur regionalen Untergruppe wird der Hochrechnungsfaktor für tiefregionalisierte Ergebnisse (EF953) empfohlen. Bei der quartalsweisen statistischen Analyse (die Ergebnisse basieren also nur auf der Anzahl der Personen, die in dem jeweiligen Quartal befragt wurden) findet der Standardhochrechnungsfaktor für die Quartale (EF951) Anwendung. Eine ausführliche Erläuterung der Modelle für die einzelnen Faktoren ist dem Metadatenreport Teil I in Abschnitt 2.6 zu entnehmen.

Beinhaltet eine Analyse auf Bundesebene nur Personen der zehnpromzentigen EU-Unterstichprobe des Ad-Hoc-Modules, ist der Hochrechnungsfaktor EF956 zu verwenden.

Im Rahmen des Zusatzprogramms wurde auf freiwilliger Basis das Thema „Arbeitsmarktsituation von Zuwanderern sowie ihren Nachkommen“ erfragt. Für die Schätzung der Antwortwahrscheinlichkeiten wird die Erhebungsmethode, der Erwerbstyp differenziert nach Geschlecht und Altersklassen

auf Bundesebene, der Erwerbstyp differenziert nach Geschlecht auf der Landesebene, Altersklassen ebenfalls auf der Landesebene sowie die zwei Nationalitätengruppen Deutsche/Ausländer in den Gebieten Ost/West eingesetzt. Die Methodik der Berechnung der Hochrechnungsfaktoren folgt der Methodik wie von Afentakis und Bihler (2005) beschrieben, wobei die Haushaltsebene ausgeblendet bleibt, da Auswertungen im Zusammenhang mit Haushalten und Familien von untergeordneter Bedeutung sind.

3.3 Verfügbare Tools

Für dieses Produkt werden seitens der FDZ keine weiterführenden Tools angeboten.

Anhang

Ausgewählte Häufigkeitstabellen und SPSS-Syntax (Anhang zu Abschnitt 2.3):

Bevölkerung (Personen am Hauptwohnsitz):

```
**** Bevölkerung am Hauptwohnsitz auswählen.
*****
IF (EF30=1 OR EF30=2) Hauptwohnsitz=1.
FILTER BY Hauptwohnsitz.
EXECUTE.
*** Hochrechnung: Hochrechnungsfaktor (Personenfaktor) in der regionalen.
**** Anpassungsschicht (EF952).
*****
WEIGHT BY EF952.
EXECUTE.
```

```
VARIABLE LABELS
EF46 'EF46NEU Geschlecht'.
EXECUTE.
```

```
ADD VALUE LABELS
EF46
      1 'männlich'
      2 'weiblich'.
EXECUTE.
```

```
**** Häufigkeiten für Geschlecht (EF46).
FREQUENCIES
VARIABLES= EF46
/ORDER=ANALYSIS.
```

EF46NEU Geschlecht

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	männlich	39.651.403	49,0	49,0	49,0
	weiblich	41.245.067	51,0	51,0	100,0
	Gesamt	80.896.470	100,0	100,0	

```
**** Familienstand entsprechend der Publikationen zusammenfassen.
*****
RECODE EF49 (4=3) (5=2) (6=3) (7=3) (ELSE=COPY) INTO EF49NEU.
EXECUTE.
FORMATS EF49NEU (F1.0).
EXECUTE.

VARIABLE LABELS
EF49NEU 'EF49NEU Familienstand'.
EXECUTE.
```

ADD VALUE LABELS
EF49NEU

- 1 'ledig'
- 2 'verheiratet'
- 3 'verwitwet'.

EXECUTE.

**** Häufigkeiten für Familienstand (EF49)

FREQUENCIES
VARIABLES= EF49NEU
/ORDER= ANALYSIS.

EF49NEU Familienstand

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	ledig	32.800.076	40,5	40,5	40,5
	verheiratet	36.620.556	45,3	45,3	85,8
	verwitwet	11.475.838	14,2	14,2	100,0
	Gesamt	80.896.470	100,0	100,0	

**** Überwiegender Lebensunterhalt EF401. Wird entsprechend der Publikation umkodiert.

RECODE EF401 (7=2) (5 THRU 9=3) (ELSE=COPY) INTO EF401NEU.

EXECUTE.

FORMATS EF401NEU (F1.0).

EXECUTE.

**** Labeln der Eingabefelder, die zur Prüfung der Daten benötigt werden.

VARIABLE LABELS

EF401NEU 'EF401NEU Überwiegender Lebensunterhalt'.

EXECUTE.

ADD VALUE LABELS

EF401NEU

- 1 'Erwerbs-/Berufstätigkeit'
- 2 'Arbeitslosengeld/-hilfe'
- 3 'Rente, Pension'
- 4 'Unterhalt durch Eltern, Ehemann/-frau oder andere Angehörige'
- 5 'Eigenes Vermögen, Vermietung, Zinsen, Altenteil'
- 6 'Sozialhilfe'
- 8 'Sonstige Unterstützungen (z.B. BAföG)'
- 9 'Elterngeld/Erziehungsgeld'.

EXECUTE.

**** Häufigkeiten für

**** Überwiegender Lebensunterhalt (EF401).

FREQUENCIES
VARIABLES= EF401NEU
/ORDER= ANALYSIS.

EF401NEU Überwiegender Lebensunterhalt

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	Erwerbs-/Berufstätigkeit	35.801.526	44,3	44,3	44,3
	Arbeitslosengeld/-hilfe	3.815.317	4,7	4,7	49,0
	Rente, Pension	20.842.893	25,8	25,8	74,7
	Unterhalt durch Eltern, Ehemann/- frau oder andere Angehörige	20.436.733	25,3	25,3	100,0
	Gesamt	80.896.470	100,0	100,0	

**** Um Erwerbsstatus zu erhalten, wird EF29 (Nichterwerbspersonen) in Anlehnung an die
**** Publikation umkodiert.

RECODE EF29 (3=4) (ELSE=COPY) INTO EF29NEU.
EXECUTE.

FORMATS EF29NEU (F1.0).
EXECUTE.

**** Labeln der Eingabefelder, die zur Prüfung der Daten benötigt werden.

VARIABLE LABELS
EF29NEU 'EF29NEU Erwerbstyp'.
EXECUTE.

ADD VALUE LABELS
EF29NEU

1 'Erwerbstätige'
2 'Erwerbslose'
4 'Nichterwerbspersonen'.

EXECUTE.

**** Häufigkeiten für Erwerbstyp (EF29)

FREQUENCIES
VARIABLES= EF29NEU
/ORDER= ANALYSIS.

EF29NEU Erwerbstyp

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	Erwerbstätige	39.941.548	49,4	49,4	49,4
	Erwerbslose	2.090.210	2,6	2,6	52,0
	Nichterwerbspersonen	38.864.712	48,0	48,0	100,0
	Gesamt	80.896.470	100,0	100,0	

**** Um die Stellung im Beruf nur von Erwerbstätigen zu.

**** erhalten, wird hier EF117 (Stellung im Beruf) umkodiert in EF117_NEU.

```

*****
DO IF (EF29=1).
RECODE EF117 (1=1) (2=1) (3=3) (4=4) (5=5) (6=6) (8=6) (9=4) (12=5) (14=4) (16=4) (17=5)
INTO EF117NEU.
END IF.
EXECUTE.
FORMATS EF117NEU (F2.0).
EXECUTE.

```

```

**** Labeln der Eingabefelder, die zur Prüfung der Daten benötigt werden.
*****
VARIABLE LABELS
EF117NEU 'EF117NEU Stellung im Beruf für Erwerbstätige'
EXECUTE.

```

```

ADD VALUE LABELS
EF117NEU
    1 'Selbständige mit und ohne Beschäftigte'
    3 'mithelfende Familienangehörige'
    4 'Beamte, Richterinnen und Richter'
    5 'Angestellte und kaufmännische/technische Auszubildende'
    6 'Arbeiterinnen und Arbeiter, Heimarbeiterinnen und Heimarbeiter und gewerbliche Auszubildende'.
EXECUTE.

```

```

**** **** Stellung im Beruf für Erwerbstätige (EF117_NEU).
FREQUENCIES
VARIABLES= EF117NEU
/ORDER= ANALYSIS.

```

EF117NEU Stellung im Beruf für Erwerbstätige

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	1 Selbständige mit und ohne Beschäftigte	4.192.447	5,2	10,9	10,9
	3 mithelfende Familienangehörige	176.401	,2	,5	11,4
	4 Beamte, Richterinnen und Richter	1.985.808	2,5	5,2	16,5
	5 Angestellte und kaufmännische/technische Auszubildende	24.105.098	29,8	62,7	79,3
	6 Arbeiterinnen und Arbeiter, Heimarbeiterinnen und Heimarbeiter und gewerbliche Auszubildende	7.973.761	9,9	20,7	100,0
	Gesamt	38.433.515	47,5	100,0	
Fehlend	System	42.462.954	52,5		
Gesamt		80.896.470	100,0		

Privathaushalte am Haupt- und Nebenwohnsitz⁶:

**** Alle Fälle auswählen, die Bezugsperson in Privathaushalten sind, um die.
 **** Privathaushalte auszählen zu können.
 IF (EF31=1 AND EF37=1) Privathaushalte=1.
 FILTER BY Privathaushalte.
 EXECUTE.
 WEIGHT BY EF952.
 EXECUTE.

**** Umkodieren von EF20 (Anzahl der Personen in Privathaushalten) in.
 **** EF20NEU, um die Haushaltsgröße zu kategorisieren.
 RECODE EF20
 (1=1) (2=2) (3=3) (4=4) (0=0) (ELSE=5) INTO EF20NEU.
 EXECUTE.
 FORMATS EF20NEU (F1.0).
 EXECUTE.

**** Labeln und Häufigkeiten.

VARIABLE LABELS

EF20NEU 'EF20NEU Privathaushalte kategorisiert'.

ADD VALUE LABELS

EF20NEU

0 'kein Privathaushalt'
 1 'Einpersonenhaushalte'
 2 'Zweipersonenhaushalte'
 3 'Dreipersonenhaushalte'
 4 'Vierpersonenhaushalte'
 5 'Fünf- und Mehrpersonenhaushalte'.

EXECUTE.

FREQUENCIES

VARIABLES= EF20NEU

/ORDER= ANALYSIS.

EF20NEU Privathaushalte kategorisiert

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	Einpersonenhaushalte	16.411.772	40,8	40,8	40,8
	Zweipersonenhaushalte	13.837.236	34,4	34,4	75,2
	Dreipersonenhaushalte	4.968.359	12,4	12,4	87,6
	Vierpersonenhaushalte	3.672.307	9,1	9,1	96,7
	Fünf- und Mehrpersonenhaushalte	1.333.091	3,3	3,3	100,0
	Gesamt	40.222.765	100,0	100,0	

⁶ Vgl. Zahlen in Destatis Fachserie 1, Reihe 3, 2014 https://www.destatis.de/GPStatistik/servlets/MCR-FileNodeServlet/DEHeft_derivate_00016323/2010300147004.pdf Tab. 1.1 und 1.4

**** Labeln und Häufigkeiten.

VARIABLE LABELS
EF1 'EF1 Bundesland'.

FREQUENCIES
VARIABLES= EF1
/ORDER= ANALYSIS.

EF1 Bundesland

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	Schleswig-Holstein	1.419.280	3,5	3,5	3,5
	Hamburg	976.707	2,4	2,4	6,0
	Niedersachsen	3.828.927	9,5	9,5	15,5
	Bremen	359.638	,9	,9	16,4
	Nordrhein-Westfalen	8.555.079	21,3	21,3	37,6
	Hessen	2.942.851	7,3	7,3	45,0
	Rheinland-Pfalz	1.901.361	4,7	4,7	49,7
	Baden-Württemberg	5.072.521	12,6	12,6	62,3
	Bayern	6.219.208	15,5	15,5	77,8
	Saarland	491.875	1,2	1,2	79,0
	Berlin	1.965.940	4,9	4,9	83,9
	Brandenburg	1.234.973	3,1	3,1	86,9
	Mecklenburg-Vorpommern	828.806	2,1	2,1	89,0
	Sachsen	2.156.546	5,4	5,4	94,4
	Sachsen-Anhalt	1.159.751	2,9	2,9	97,2
	Thüringen	1.109.304	2,8	2,8	100,0
	Gesamt	40.222.765	100,0	100,0	

**** Umkodieren von EF735 entsprechend der Publikation umkodiert.

RECODE EF735
(6=2) (2=2) (8=4) (9=5) (7=2) (ELSE=COPY) INTO EF735NEU.
EXECUTE.

FORMATS EF735NEU (F1.0).
EXECUTE.

**** Labeln und Häufigkeiten.

VARIABLE LABELS
EF735NEU 'EF735NEU Familienstand des Haupteinkommensbeziehers'.

ADD VALUE LABELS
EF735NEU

1 'ledig'
2 'verheiratet'
4 'geschieden'
5 'verwitwet'.

EXECUTE.

FREQUENCIES
VARIABLES= EF735NEU
/ORDER= ANALYSIS.

EF735NEU Familienstand des Haupteinkommensbeziehers

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	ledig	11.992.579	29,8	29,8	29,8
	verheiratet	18.593.607	46,2	46,2	76,0
	geschieden	4.707.766	11,7	11,7	87,7
	verwitwet	4.928.813	12,3	12,3	100,0
	Gesamt	40.222.765	100,0	100,0	

Statistische Ämter des Bundes und der Länder,
Metadatenreport – Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus 2014 am
Gastwissenschaftsarbetsplatz sowie per kontrollierter Datenfernverarbeitung (sog. On-Site-Zugangswege/
Nutzungen)

Fotorechte Umschlag: ©artSILENCEcom – Fotolia.com