

Wichtige Informationen zur Nutzung des Mikrozensus Scientific Use Files 2019

**Dokumentation und
Datenaufbereitung**

Herausgeber:

Statistisches Bundesamt
Referat F 36 Mikrozensus – Auswertung und
Analyse
Graurheindorfer Str. 198
53117 Bonn
E-Mail: MZ-ScientificUseFile@destatis.de

<http://www.destatis.de>

GESIS - Leibniz-Institut für Sozialwissenschaften
German Microdata Lab

B6, 4-5
68159 Mannheim
E-Mail: gml@gesis.org
Kontakt: Tobias Roth (Tel.: 0621 / 12 46 – 263)
<http://www.gesis.org/gml>

Stand: März 2022

Inhalt

Einleitung.....	4
1 Allgemeine Informationen zum Mikrozensus.....	5
1.1 Was ist der Mikrozensus?	5
1.2 Das Wichtigste über den Mikrozensus 2019 im Überblick	5
1.2.1 Grundgesamtheit.....	5
1.2.2 Berichtszeitraum	6
1.2.3 Stichprobe	6
1.2.4 Erhebungs- und Analyseeinheiten	7
1.2.5 Erhebungsprogramm und Erhebungsverfahren	7
1.2.6 Hochrechnung / GewichtungsvARIABLEN.....	9
2 Kurzbeschreibung des Scientific Use Files Mikrozensus 2019	11
2.1 Auswahl der Substichprobe.....	11
2.2 Hochrechnungs- / GewichtungsvARIABLEN im Scientific Use File.....	12
2.3 Anonymisierungsmaßnahmen, Regionalangaben.....	12
2.4 Die wichtigsten Veränderungen gegenüber 2018 in Stichworten	13
2.5 Möglichkeit der Längsschnittverknüpfung	14
2.6 Zur Abgrenzung der Bevölkerungsbegriffe und zur Hochrechnung.....	14
2.7 Zugang zum Mikrozensus Scientific Use File und Beratung	15
3 Literaturhinweise	16
3.1 Allgemein	16
3.2 Ergebnisse des Mikrozensus 2019	16
3.3 Mikrozensusgesetz, EU-Verordnungen.....	16
3.3.1 Mikrozensusgesetz 2017	16
3.3.2 EU-Verordnungen zur Arbeitkräfteerhebung	16
A Anhang.....	18
A1 Informationen zur Nutzung der Setups.....	18
A2 Verteilungen zur Prüfung der Systemfiles bzw. des Ablaufs der Setups.....	20

Einleitung

Für Forscherinnen und Forscher, die sich erstmals für die Daten interessieren sowie für jene, die bereits mit den Daten anderer Erhebungszeitpunkte Erfahrungen gesammelt haben, gibt dieser Bericht einen Überblick zum Erhebungsprogramm und den Erhebungsverfahren des Mikrozensus sowie zur Aufbereitung und Dokumentation des Scientific Use Files (SUF) 2019. Diese Informationen sollen dabei helfen, die Daten sachgerecht auszuwerten. Im Folgenden werden deshalb zunächst Informationen zum Mikrozensus allgemein sowie zum Erhebungs- und Stichprobendesign genannt. Anschließend werden das SUF 2019 und die Veränderungen gegenüber dem Vorjahr beschrieben. Nach den Literaturhinweisen finden sich im Anhang Informationen zur Nutzung der Setups zum Einlesen der Rohdaten und zur Erzeugung von Systemfiles, die ab dem Erhebungszeitpunkt 2007 zusätzlich zu den Rohdaten im ASCII-Format für die Statistikprogramme SPSS, SAS und Stata im dokumentierten Format bereitgestellt werden. Der Anhang enthält ebenfalls Verteilungen zur Prüfung der Systemfiles. Um die Ergebnisqualität zu prüfen, werden im Rahmen der Datenaufbereitung für ausgewählte Merkmalsverteilungen Ergebnisse des SUF mit Ergebnissen des Mikrozensus verglichen; siehe u. a. den [Vergleich der Randverteilungen](#).

Ausführliche Informationen über den Mikrozensus und die Mikrozensus SUF, u. a. Fragebogen, Klassifikationen und Definitionen, stehen auf den [Web-Seiten der GESIS](#) zur Verfügung.

Weitere Informationen sind auf den Seiten des Statistischen Bundesamtes und der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder abrufbar:

[Was ist der Mikrozensus?](#)

[Häufige Fragen zum Mikrozensus](#)

[Forschungsdatenzentren | Datenangebot | Mikrozensus](#)

Die Aufbereitung und Dokumentation des faktisch anonymisierten Einzelmaterials erfolgt durch das Statistische Bundesamt, Referat F 36 – Mikrozensus - Auswertung und Analyse, in Kooperation mit dem German Microdata Lab (GML) der GESIS.

1 Allgemeine Informationen zum Mikrozensus

1.1 Was ist der Mikrozensus?

Der Mikrozensus ist die amtliche Repräsentativstatistik über die Bevölkerung und den Arbeitsmarkt in Deutschland mit einem Auswahlsatz von 1 % der Haushalte und der Bevölkerung. Er wird seit 1957 in Westdeutschland und seit 1991 auch in den neuen Bundesländern durchgeführt. Die Arbeitskräftestichprobe der Europäischen Union (EU Labour Force Survey) ist in den Mikrozensus integriert. Im Jahr 2016 wurde die gesamte Stichprobe erneuert. Aufgrund der zeitlichen Begrenzung des Mikrozensusgesetzes von 2005 bis einschließlich des Jahres 2016 war zur Weiterführung des Mikrozensus ein Anschlussgesetz notwendig. In dem Anschlussgesetz (Mikrozensusgesetz 2017) wurden die Haushaltsstatistiken u. a. mit dem Ziel weiterentwickelt, die Belastung der Befragten zu reduzieren. Die Weiterentwicklung des Mikrozensus erfolgt in zwei Stufen: In der ersten Stufe von 2017 bis 2019 wird der Mikrozensus wie bisher mit einigen Anpassungen der Erhebungsmerkmale weitergeführt. Zudem werden von Personen aus Gemeinschaftsunterkünften nur noch wenige Basisdaten erhoben, wobei die Leitungen der Gemeinschaftsunterkünfte auskunftspflichtig sind. In der zweiten Stufe werden neben dem bereits im Mikrozensus integrierten EU Labour Force Survey auch zwei weitere Haushaltsstatistiken der Europäischen Union (EU) in den Mikrozensus integriert: Die Gemeinschaftsstatistik über Einkommen und Lebensbedingungen (ab 2020) sowie die Erhebung über die private Nutzung von Informations- und Kommunikationstechnologien (ab 2021). Zudem werden die Erhebungsmerkmale an die Anforderungen der jeweiligen Erhebung angepasst.

Siehe auch:

Die Informationsseite zum Mikrozensus im Internet¹: www.mikrozensus.de

[Informationen zum Mikrozensus \[Informationen für die Befragten Mikrozensus 2019\]](#) (Statistische Ämter des Bundes und der Länder, 2018/2019).

Statistisches Bundesamt (2020): [Qualitätsbericht Mikrozensus 2019](#). Wiesbaden.²

1.2 Das Wichtigste über den Mikrozensus 2019 im Überblick

1.2.1 Grundgesamtheit

Zur Grundgesamtheit gehört die gesamte wohnberechtigte Bevölkerung in Privathaushalten und Gemeinschaftsunterkünften am Haupt- und Nebenwohnsitz.

Nicht zur Erhebungsgesamtheit gehören Angehörige ausländischer Streitkräfte sowie ausländischer diplomatischer Vertretungen mit ihren Familienangehörigen. Personen ohne Wohnung (Obdachlose) haben im Mikrozensus aufgrund der Anlage als „Flächenstichprobe“³ keine Erfassungschance.

¹ Über diese Seite haben die Befragten auch die Möglichkeit den Fragebogen elektronisch auszufüllen.

² Weite Teile der hier dargestellten allgemeinen Informationen zum Mikrozensus orientieren sich am Qualitätsbericht Mikrozensus 2019 (Statistisches Bundesamt, 2020).

³ Vereinfachend wird der Mikrozensus zwar als „Flächenstichprobe“ bezeichnet, konkret sind es aber Anschriften mit Wohnraum.

1.2.2 Berichtszeitraum

Seit 2005 wird das Befragungsvolumen gleichmäßig über alle Kalenderwochen des Jahres verteilt (kontinuierliche Erhebung). Der Berichtszeitraum ist die letzte Kalenderwoche vor der Befragung (gleitende Berichtswoche).⁴ Das SUF enthält Angaben zum Berichtsquartal.

1.2.3 Stichprobe

Der Mikrozensus ist eine einstufige Klumpenstichprobe („Flächenstichprobe“).

Auswahlgrundlage

Während bis einschließlich 2015 die Grundausswahl für die alten Bundesländer auf Ergebnissen der Volkszählung 1987, für die neuen Bundesländer und Ost-Berlin auf Ergebnissen aus dem Bevölkerungsregister Statistik der DDR 1990 basierte, dient ab dem Berichtsjahr 2016 der Zensus 2011 als Auswahlgrundlage für den Mikrozensus. Durch die neue Auswahlgrundlage der Stichprobe ergeben sich Einschränkungen in der Vergleichbarkeit zu den Vorjahren. Die Aktualisierung der Grundausswahl erfolgt jährlich durch Berücksichtigung der Neubautätigkeit über die Meldungen zur Bautätigkeitsstatistik (abweichend hiervon wurden im Jahr 2016 die Neubausauswahlen der Jahre 2011 bis 2016 kumuliert).

Auswahleinheiten

Die Auswahleinheiten sind Auswahlbezirke (Anschriften mit Wohnraum, die nach der Anzahl der an der Anschrift befindlichen Wohnungen in sogenannte Anschriftengrößenklassen unterteilt werden. So kann ein Auswahlbezirk mehrere (benachbarte) Anschriften umfassen, alle Wohnungen einer Anschrift oder auch nur Teile einer Anschrift. Bei Anschriften ohne Wohnungen wurden die an der Anschrift gemeldeten Personen zur Größenbestimmung des Auswahlbezirks genutzt).

Ein Auswahlbezirk verbleibt vier Jahre in der Stichprobe, in jedem Jahr wird 1/4 der Auswahlbezirke durch neue Auswahlbezirke ersetzt (rotierende Panelstichprobe). Infolge der Änderung der Auswahlgrundlage wurde im Jahr 2016 abweichend hiervon die gesamte Stichprobe erneuert. Die aus dem Auswahlbezirk wegziehenden Haushalte und Personen werden nicht weiter befragt, sondern durch zuziehende Haushalte und Personen ersetzt. Der Mikrozensus ist somit eine Wiederholungsbefragung mit teilweiser Überlappung der Erhebungseinheiten (partielle Rotation). Die Möglichkeit der Zusammenführung der Querschnittsdaten zu einem [Mikrozensuspanel](#) ist rechtlich seit dem Mikrozensusgesetz 1996 gegeben. Aufgrund der Erneuerung der kompletten Stichprobe im Jahr 2016 sind Verknüpfungen der Querschnittsdaten ab dem Erhebungszeitpunkt 2012 nur bis einschließlich der Erhebung 2015 möglich. Zudem werden aufgrund der Neuregelung des Mikrozensus ab 2020 Verknüpfungen der Querschnittsdaten ab dem Erhebungszeitpunkt 2016 nur bis einschließlich der Erhebung 2019 möglich sein.

Schichtungsmerkmale

Neben der regionalen Schichtung (Bundesland und Regionalschicht) wird fachlich nach der Anschriftengrößenklasse geschichtet.

⁴ Es sei bereits jetzt darauf hingewiesen, dass ab der Erhebung 2020 die Befragung wieder mit einer festen Berichtswoche durchgeführt wird. Anders als bis zur Erhebung 2004 sind die Berichtswochen jedoch gleichmäßig über das Jahr verteilt und die ausgewählten Haushalte müssen ihre Angaben auch bei Verzögerungen auf eine bestimmte Berichtswoche beziehen.

Auswahlsatz

Der Auswahlsatz des Mikrozensus beträgt 1 %. Die Ad-hoc-Module des EU Labour Force Survey werden als Unterstichprobe mit einem Auswahlsatz von 0,1 % durchgeführt.

Auswahltechnik

Grundauswahl:

Vor der Auswahl wurden die Auswahlbezirke regional angeordnet (je Schicht in der Sortierung nach regionaler Schichtungsgruppe, Kreis, Gemeindegrößenklasse, Gemeinde und Auswahlbezirksnummer).

Zerlegung der Auswahlgesamtheit in 100 1 %-Stichproben: Je 100 aufeinanderfolgende Auswahlbezirke („Zonen“) werden mit einer zufälligen Permutation der Zahlen 00 bis 99 versehen. Vier aufeinanderfolgende Zonen bilden einen Block.

Zerlegung der Auswahlgesamtheit in vier Rotationsviertel zu je 0,25 %: Je vier aufeinanderfolgende Zonen werden mit einer zufälligen Permutation der Zahlen 1 bis 4 versehen.

Die geplante Befragungswoche eines Auswahlbezirks ergibt sich durch eine Kombination des Quartalskennzeichens (je vier aufeinanderfolgende Blöcke werden mit einer zufälligen Permutation der Zahlen 1 bis 4 versehen) und des Wochenkennzeichens (die Auswahlbezirke werden mit einer zufälligen Permutation der Zahlen 1 bis 13 versehen).

Neubauauswahl:

Bei der Neubauauswahl, die in einer eigenen Anschriftengrößenklasse abgegrenzt ist, gibt es gegenüber der Grundauswahl im Stichprobenplan einige Modifikationen. Während ein Auswahlbezirk der Grundauswahl (ohne Gemeinschaftsunterkünfte) durchschnittlich neun Wohnungen umfasst, ist die Richtgröße bei der Neubauauswahl sechs Wohnungen. Für Auswahlbezirke der Anschriftengrößenklasse mit Gemeinschaftsunterkünften beträgt die Richtgröße 15 Personen.

1.2.4 Erhebungs- und Analyseeinheiten

Erhebungseinheiten sind Personen (in Privathaushalten und Gemeinschaftsunterkünften), Haushalte und Wohnungen.

Analyseeinheiten sind Personen, Lebensgemeinschaften, Familien, Haushalte und Wohnungen.

Bei der Zuordnung von Personen und Haushalten ist zu beachten, dass Personen mit mehreren Wohnsitzen am Ort ihrer Haupt- und ihrer Nebenwohnung erfasst werden können. Um Mehrfachzählungen zu vermeiden, werden in Veröffentlichungen üblicherweise nur Personen am Hauptwohnsitz ausgewählt.

1.2.5 Erhebungsprogramm und Erhebungsverfahren

Erhebungsprogramm

Das Fragenprogramm des Mikrozensus besteht aus einem gleichbleibenden Grundprogramm, das sozioökonomische Grundinformationen liefert und die laufende Beobachtung des Arbeitsmarktes ermöglicht. In den vierjährigen Zusatzprogrammen, die seit 2005 mit dem vollen Auswahlsatz von 1 % durchgeführt werden, werden weitere Angaben erfragt.⁵ Daneben werden im Rahmen des EU Labour Force Survey jährlich wechselnde Ad-hoc-

⁵ Für die Zusatzprogramme, die 2020 oder später erhoben werden, ist ein Auswahlsatz von 1 % nur noch dann vorgesehen, wenn dies unbedingt erforderlich ist, um hinreichend zuverlässige Ergebnisse zu erhalten oder einen Nachweis in regionaler oder tiefer fachlicher Gliederung zu ermöglichen.

Module eingesetzt. Die Ad-hoc-Module werden mit einem Auswahlsatz von 0,1 % erhoben. Für Personen in Gemeinschaftsunterkünften werden nur noch wenige Basisdaten erhoben (neben der Art der Gemeinschaftsunterkunft sind dies der Wohnrechtsstatus, der Geburtsmonat und das Geburtsjahr, das Geschlecht, der Familienstand, die Staatsangehörigkeit und der Hauptstatus). Auskunftspflichtig hierüber ist nicht die Person selbst, sondern die Leitung der Gemeinschaftsunterkunft.

Grundprogramm:

Im Grundprogramm sind im Mikrozensus 2019 folgende Themenbereiche enthalten: Merkmale der Person, Familien- und Haushaltszusammenhang, Haupt- und Nebenwohnung, Aufenthaltsdauer, Staatsangehörigkeit und Migrationshintergrund, Wohn-/Lebensgemeinschaft, Fragen für eingebürgerte Personen, Fragen für Ausländer/-innen, Quellen des Lebensunterhalts, Höhe des Einkommens, Angaben zur Renten- und Pflegeversicherung, Besuch von Schule und Hochschule, allgemeiner Schul- und beruflicher Ausbildungsabschluss, allgemeine und berufliche Weiterbildung, Erwerbstätigkeit, Arbeitslosigkeit und Arbeitssuche, frühere Erwerbstätigkeit, Situation ein Jahr vor der Erhebung, Angaben zur Behinderung (freiwillig)⁶.

Zusatzprogramme:

Die Zusatzprogramme 2019 umfassen Angaben zum Krankenversicherungsschutz und Zusatzangaben zur gegenwärtigen Haupttätigkeit. Für Auskunftspflichtige, die gemäß Stichprobenplan 2018 hätten befragt werden sollen, aber erst 2019 mit dem Fragebogen 2018 befragt wurden, liegen keine Angaben zu den Zusatzprogrammen vor. Zudem liegen für diese Befragtengruppe keine Angaben für 2019 neu aufgenommene Fragen des Grundprogramms vor.

Ad-hoc-Modul:

Im Rahmen der in den Mikrozensus integrierten Arbeitskräfteerhebung (AKE) werden normalerweise jährlich Ad-hoc-Module durchgeführt, die dazu dienen, detaillierte Informationen zu politisch relevanten Themen der EU zu sammeln, welche nicht Teil des Standardfrageprogramms der Arbeitskräfteerhebung sind. Im Jahr 2019 wurde das Ad-hoc-Modul zur „Arbeitsorganisation und Arbeitszeitgestaltung“ durchgeführt. Zielpopulation des Ad-hoc-Moduls waren Erwerbstätige in Privathaushalten.⁷ Der Auswahlsatz betrug 0,1 %. Die Erteilung der Auskunft war freiwillig.

Erhebungsverfahren

In den meisten Fällen erfolgen persönlich-mündliche Befragungen durch Interviewer/-innen (face-to-face), welche einen elektronischen Fragebogen mittels eines Laptops befüllen (CAPI). Die Organisation der Erhebung erfolgt dezentral durch die jeweiligen Statistischen Ämter der Länder. Neben dem face-to-face Interview besteht die Möglichkeit, einen Papierfragebogen⁸ selbst auszufüllen oder per Telefon⁹ interviewt zu werden. In allen Fällen dürfen von einem erwachsenen Haushaltsmitglied stellvertretend Angaben für andere Haushaltsmitglieder getätigt werden (sog. Proxy-Interviews).

⁶ Ab der Erhebung 2020 wird dieses Merkmal in der Unterstichprobe zur Arbeitsmarktbeteiligung erhoben.

⁷ Für weitere Hinweise zu Design, Frageprogramm und methodischen Problemen siehe Eurostat ([2020](#)).

⁸ Die CAPI-Fragen unterscheiden sich teilweise von den Fragen im Selbstausfüllendenbogen.

⁹ Telefoninterviews werden allerdings nur auf ausdrücklichen Wunsch des zu befragenden Haushalts durchgeführt.

1.2.6 Hochrechnung / Gewichtungsvariablen

Die Hochrechnung des Mikrozensus basiert auf einem zweistufigen Verfahren:¹⁰

(1) Zunächst erfolgt ein Ausgleich der nicht erreichbaren Haushalte durch Kompensationsfaktoren anhand von Informationen über diese Haushalte. Der Unit-Nonresponse beträgt im Mikrozensus 2019 6,4 % der Haushalte (ohne Jahresüberhänge; s. u.). (Bei Pflichtfragen liegt die Ausfallquote (Item-Nonresponse) i. d. R. unter 3 %)

(2) Anschließend werden die mit dem Kompensationsfaktor für Unit-Nonresponse gewichteten Stichprobenwerte an Populationswerte aus der laufenden Bevölkungsfortschreibung und dem Ausländerzentralregister nach Altersklassen (bis 14 Jahre, 15-44, 45 Jahre oder älter), Staatsangehörigkeitsgruppen (deutsch, türkisch, EU-28 und nicht EU-28) und Geschlecht auf unterschiedlichen regionalen Ebenen (Bundesland, NUTS-2, regionalen Anpassungsschichten¹¹) an die Ergebnisse der laufenden Bevölkungsfortschreibung angepasst.¹² Um zu schwach besetzte Anpassungsklassen zu vermeiden, werden die Kompensations- und Hochrechnungsfaktoren (EF950 und EF951) durch ein Kalibrierungsverfahren (Generalized Regression Estimation) berechnet. Damit ist eine Anpassung an getrennte Randverteilungen möglich. Zudem gilt die Bedingung, dass innerhalb eines Haushalts alle Personen den gleichen Gewichtungsfaktor erhalten. Die Hochrechnungsfaktoren für die Jahresergebnisse (EF952) stellen das arithmetische Mittel der jeweiligen Quartalsfaktoren dar.

Mit einem gleichen Kalibrierungsverfahren wird der Hochrechnungsfaktor für das Ad-hoc-Modul (EF956) berechnet. Bezüglich der Kompensation wird zusätzlich die Antwortwahrscheinlichkeit unter Berücksichtigung der Art der Erhebung (CAPI/Telefon, Selbstausfüllung) berücksichtigt. Bei den Staatsangehörigkeitsgruppen wird nur zwischen Deutschen/Ausländern/-innen unterschieden und hinsichtlich der regionalen Ebenen erfolgt die Differenzierung nach Bund, Ost-West-Unterscheidung und Bundesland. Die Altersklassen und die Zielpopulation orientieren sich an der inhaltlichen Ausgestaltung des Ad-hoc-Moduls. In 2019 richtete sich das Ad-hoc-Modul an Erwerbstätige in Privathaushalten. Da im Ad-hoc-Modul Auswertungen im Haushaltszusammenhang von untergeordneter Bedeutung sind, entfällt die Bedingung, dass alle Personen eines Haushalts den gleichen Gewichtungsfaktor erhalten müssen. Der Hochrechnungsfaktor wird zudem nur für Jahresergebnisse berechnet. Die Eckwerte für die Anpassung werden aus den hochgerechneten Werten des Mikrozensus ermittelt.

Für Auswertungen der Angaben zur Behinderung (EF2700, EF2701) steht der Hochrechnungsfaktor EF957 zur Verfügung. Für Personen mit amtlich festgestellter Behinderung und einem Grad der Behinderung von 50 % oder mehr wurden die Daten des Mikrozensus auf Landesebene¹³ an Ergebnisse der Schwerbehindertenstatistik angepasst. Für Leichtbehinderte erfolgt lediglich eine Einschätzung der Antwortausfälle. Der Hochrechnungsrahmen beinhaltet die Merkmale Grad der Behinderung (bis 49 % / 50 % oder mehr), Altersgruppen (Behinderung bis 49 % (Leichtbehinderte): bis 24 Jahre / 25-49 / 50-59 / 60-79, 80 Jahre oder älter; Behinderung 50 % oder mehr (Schwerbehinderte): bis 14 Jahre / 15-29 / 30-44 / 45-59 / 60-79 / 80 Jahre oder älter) und Geschlecht. Die Erstellung des Hochrechnungsfaktors erfolgt durch einen einfachen Quotientenschätzer. Mit der Berücksichtigung nur gültiger

¹⁰ Zu einer ausführlichen Darstellung siehe Afentakis und Bihler (2005).

¹¹ Regionale Anpassungsschichten sind Kreise oder Zusammenfassungen von Kreisen, sodass diese Gebiete im Durchschnitt mindestens 500.000 Einwohner/-innen umfassen.

¹² Für Zeit- und Berufssoldaten/-innen einschließlich Bundes- und Bereitschaftspolizei sowie Personen im freiwilligen Wehrdienst stammen die Eckwerte aus Bestandsdaten des Bundesministeriums der Verteidigung, des Bundesministeriums des Innern, für Bau und Heimat bzw. der Bundespolizei.

¹³ In den Erhebungen vor 2017 erfolgte die Anpassung lediglich auf Bundesebene.

Angaben zum Grad der Behinderung sollen sowohl zufällige als auch systematische Unterschiede zur Schwerbehindertenstatistik ausgeglichen werden. Angaben zur Schwerbehindertenstatistik liegen nur für die Gesamtbevölkerung vor. Für Personen in Gemeinschaftsunterkünften liegen allerdings im Mikrozensus keine Informationen zu den Behinderteneigenschaften vor, sodass eine Korrektur der Eckwerte vorgenommen wurde. Aus den Erhebungen 2009 und 2013 des Mikrozensus wurde eine Schätzung der in Gemeinschaftsunterkünften lebenden Menschen mit Behinderung vorgenommen und die Eckwerte der Schwerbehindertenstatistik in 2019 entsprechend korrigiert. An diese korrigierten Eckwerte wurde dann die Anpassung für die Menschen mit Behinderung in Privathaushalten vorgenommen.

2 Kurzbeschreibung des Scientific Use Files Mikrozensus 2019

Das SUF ist eine faktisch anonymisierte 70 %-Substichprobe des Mikrozensus.

2.1 Auswahl der Substichprobe

Auswahleinheiten waren bis 2011 Haushalte bzw. Wohnungen des Mikrozensus, wobei alle Personen eines ausgewählten Haushalts bzw. einer Wohnung in der Substichprobe enthalten waren. Wohnungen wurden in den Erhebungsjahren gezogen, in denen das Zusatzprogramm zur Wohnsituation (1998, 2002, 2006, 2010) durchgeführt wurde. Ab 2012 werden als Auswahleinheiten für die Substichprobe die Auswahlbezirke innerhalb eines Rotationsviertels herangezogen. Damit wird die Möglichkeit geschaffen selbstständig mit den SUF Paneldatensätze zu erzeugen. Infolge der Erneuerung der gesamten Mikrozensusstichprobe im Jahr 2016, sind Verknüpfungen der Querschnittsdaten ab dem Erhebungszeitpunkt 2012 nur bis einschließlich der Erhebung 2015 möglich. Zudem werden aufgrund der Weiterentwicklung des Mikrozensus ab 2020 Verknüpfungen der Querschnittsdaten ab dem Erhebungszeitpunkt 2016 nur bis einschließlich der Erhebung 2019 möglich sein. Das für den SUF 2012 bis 2015 verwendete Verfahren kann weitestgehend auch für die neue Stichprobe des SUF ab 2016 übernommen werden. Es wird folgende Auswahltechnik angewendet:

Sortierung der Auswahlbezirke nach

1. Kennung des Rotationsviertels
2. Bundesland
3. Größenklasse (Dezile) der Auswahlbezirke
4. Kennung über Grundauswahl/Aktualisierung der Auswahlbezirke
5. Kennung über den Einbezug eines Auswahlbezirks in das Ad-hoc-Modul
6. Regierungsbezirk
7. Regionale Anpassungsschicht
8. Regionale Schicht
9. Regionale Untergruppe
10. Gemeindegrößenklasse
11. Nummer des Auswahlbezirks

Die Sortierung folgt damit weitgehend der Schichtung des Originalfiles. Ab 2012 wurden in die Sortierung neu aufgenommen die Größenklasse (Dezile) der Auswahlbezirke sowie die Kennung über den Einbezug eines Auswahlbezirks in das Ad-hoc-Modul.

Zwar wurde bei der Grundauswahl bzw. wird bei der Aktualisierung der Auswahlbezirke eine einheitliche Größe der jeweiligen Anschriftengrößenklassen angestrebt, sie unterscheiden sich jedoch fluktuationsbedingt erheblich. Um eine daraus resultierende Erhöhung der Fehlervarianz zu begrenzen, erfolgt die zusätzliche Aufnahme einer Sortierung nach der Anzahl der befragten Personen in einem Auswahlbezirk.

Mit der Aufnahme der Kennung eines Auswahlbezirks in das Ad-hoc-Modul soll die Unterstichprobeneigenschaft des Ad-hoc-Moduls in der Substichprobenziehung kontrolliert werden.

Beide Merkmale wurden für so wichtig gehalten, dass sie in der Sortierung weit oben, an dritter bzw. fünfter Stelle platziert werden.

Schichtung:

Jeweils zehn in der Reihenfolge der Sortierung aufeinanderfolgende Auswahlbezirke bilden eine Schicht. Eine heterogene Zusammensetzung der Schicht in Bezug auf die Sortiermerkmale ist ungünstig, da sie tendenziell mit einem höheren Varianzbeitrag verbunden ist. Im Hinblick auf die Analysemöglichkeit nach Bundesländern sowie die anzustrebende Homogenität der Auswahlbezirke innerhalb einer Schicht, die varianzreduzierend wirkt, werden bei Sortierwechseln von Rotationsviertel, Bundesland und Größenklassen (Dezile) der Auswahlbezirke unvollständige Schichten mit weniger als zehn Auswahlbezirken zugelassen. Die letzte Schicht kann ebenfalls weniger als zehn Auswahlbezirke umfassen.

Auswahlsatz:

Im „Startjahr“ 2016 wurden für vollständige Schichten mit Hilfe eines „Pseudozufallsgenerators“ sieben unterschiedliche Zahlen im Intervall α mit $(1 \leq \alpha \leq 10, \alpha \in \mathbb{N})$ gezogen. Für jede vollständige Schicht wurden die Auswahlbezirke nach einer zufälligen Permutation ihrer Ordnungsnummer in der Schicht sortiert. In den so angeordneten Daten gilt jeder Auswahlbezirk mit gezogener Zufallszahl als ausgewählt. Der Stichprobenumfang jeder vollständigen Schicht beträgt 7.

Für die Auswahlbezirke unvollständiger Schichten wurden gleichverteilte unabhängige Zufallszahlen z im Intervall $(0, 1)$ gezogen. Es gelten diejenigen Auswahlbezirke als gezogen, deren Zufallszahl im Intervall $(0 \leq z \leq 0,7)$ liegt. Die Anzahl gezogener Auswahlbezirke in unvollständigen Schichten ist zufällig, die Ziehungswahrscheinlichkeit von 7/10 bleibt aber erhalten.

Mit der so gezogenen einfachen Stichprobe ohne Zurücklegen wurden gleiche Auswahlwahrscheinlichkeiten für alle Auswahlbezirke des SUF erreicht.

In den Folgejahren werden zunächst die Auswahlbezirke der Substichprobe des vorangegangenen Jahres ($t-1$) übernommen, die auch in der Stichprobe des aktuellen Jahres (t) vertreten sind. Für die Auswahlbezirke des im aktuellen Jahr neu hinzugekommenen Rotationsviertels sowie der Neubausauswahl wird die Auswahltechnik gleichermaßen eingesetzt. Die entsprechend gezogenen Auswahlbezirke werden anschließend mit den Auswahlbezirken der Substichprobe des Vorjahres zusammengeführt und bilden das SUF des aktuellen Jahres.

2.2 Hochrechnungs- / GewichtungsvARIABLEN im Scientific Use File

Das SUF enthält Hochrechnungsvariablen, die auf die Population (in 1.000) hochrechnen. Um eine verbesserte Anpassung der Verteilungen des SUF an die veröffentlichten Ergebnisse zu erreichen, orientiert sich die Erstellung der Hochrechnungsvariablen im SUF an den Hochrechnungsfaktoren im Originalfile. EF951, EF952 und EF956 werden mit einem Kalibrierungsverfahren ermittelt. Jedoch wird nur zwischen Deutschen/Ausländern/-innen unterschieden und das Bundesland als einzige regionale Einheit berücksichtigt. Des Weiteren ist EF952 nicht das arithmetische Mittel der Quartalshochrechnungsfaktoren (EF951). Die Eckwerte für die Anpassung werden aus den hochgerechneten Werten des Mikrozensus ermittelt. Der Hochrechnungsfaktor für die Behinderung (EF957) wird nach dem gleichen Verfahren wie im Originalfile berechnet (siehe Abschnitt 1.2.6).

2.3 Anonymisierungsmaßnahmen, Regionalangaben

Weiterhin unterscheidet sich das SUF vom Originalfile des Mikrozensus dadurch, dass bestimmte Variablen im SUF, bedingt durch die Anonymisierung, in klassierter und vergrößerter Form verfügbar sind. So sind im SUF die Regionalangaben Bundesland und eine ver-

größte Gemeindegrößenklasse, die nicht für die kleinen Bundesländer Bremen und Saarland ausgewiesen ist, enthalten. Es darf keine einzelne Gemeinde mit weniger als 500.000 Einwohner/-innen identifizierbar sein. Bei mehreren Gemeinden in einer Klasse müssen diese insgesamt in jedem Bundesland mindestens 400.000 Einwohner/-innen umfassen. Im Originalfile ist die Gemeindegrößenklasse zum Stand 31.12.2019 umgesetzt. Im SUF wurde die Gemeindegrößenklasse mit Stand 31.12.2018 umgesetzt, um die Kompatibilität zu anderen Regionalklassifikationen zu wahren und die Zusammenfassung von Gemeinden vorzunehmen. Neben den aus Datenschutzgründen notwendigen Vergrößerungen können daher Verteilungsunterschiede zwischen Ergebnissen des SUF und der Originaldaten methodisch mit dieser Anpassung zusammenhängen.

Bei Auswertungen nach Landesteilen wird von den Statistischen Ämtern i. d. R. nicht mehr zwischen dem früheren Bundesgebiet inkl. West-Berlin und den neuen Bundesländern inkl. Ost-Berlin unterschieden, sondern nach früherem Bundesgebiet ohne Berlin vs. neuen Ländern einschließlich Berlin. Dies hängt auch damit zusammen, dass nach einer Gebietsreform die ehemalige Ost-West-Aufteilung Berlins nicht mehr trennscharf abzubilden ist. Die Auswahlbezirke des Mikrozensus können nach wie vor den früheren Gemeindeteilen West- und Ost-Berlin korrekt zugeordnet werden (Bömermann/Gruber 2010: 23). Im SUF kann mithilfe eines speziellen Codes der Variable Gemeindegrößenklasse weiterhin zwischen West- und Ost-Berlin (EF563 = 9 Berlin-Ost) unterschieden werden. Für die Anpassung des Mikrozensus an die laufende Bevölkerungsfortschreibung liegen nur Daten zu den neuen Gebietsabgrenzungen (12 statt 23 Bezirke) vor. Deshalb ist bei der Hochrechnung (Gewichtungsvariablen EF951 ff.) mit regionalen Unschärfen zu rechnen.

Auch bei den weiteren Variablen des SUF werden je nach Bedarf Merkmalsvergrößerungen vorgenommen, sodass jede Ausprägung in den univariaten Verteilungen mindestens 5.000 Fälle in der Grundgesamtheit aufweist. Die Ausprägungen der Variable Staatsangehörigkeit werden so aggregiert, dass jede Gruppe von Staatsangehörigkeiten in der Grundgesamtheit wenigstens 50.000 Einwohner/-innen umfasst. Im Fall von Vergrößerungen wird im SUF i. d. R. die am stärksten besetzte Kategorie ausgewiesen.

Zusätzlich zu den im SUF enthaltenen Variablen können im Rahmen eines Aufenthalts als Gastwissenschaftler/-in in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder weitere Variablen genutzt werden. Die Auflistung dieser Variablen finden Sie am Ende der [thematischen Übersicht](#).

2.4 Die wichtigsten Veränderungen gegenüber 2018 in Stichworten

Im Folgenden werden nur die bedeutendsten Änderungen gegenüber 2018 genannt. Auf eine detaillierte Darstellung wird aus Platzgründen verzichtet. Weitere Informationen zu den Änderungen des Mikrozensus 2019 im Vergleich zum Mikrozensus 2018 (bzw. 2015 für die Zusatzprogramme zur Krankenversicherung und den Zusatzangaben zur gegenwärtigen Haupttätigkeit) enthalten sowohl das [Datenhandbuch](#), als auch die Datenbeschreibungen im Mikrodaten-Informationssystem [MISSY](#). Die [Variablen-Zeitpunkte-Matrix](#) in MISSY gibt zusätzliche Hinweise zur zeitlichen Vergleichbarkeit von Variablen des Mikrozensus ab 1973.

- Aufgrund von Änderungen im Fragebogen wurde in den Variablen EF350-EF355, EF358-EF363 und EF2007-EF2008 die Antwortkategorie „Weiß nicht“ neu eingefügt.
- Bei der Frage zur im Haushalt vorwiegend gesprochene Sprache (EF377) wurde Rumänisch als eigenständige Antwortkategorie neu aufgenommen.
- Im Vergleich zum Berichtsjahr 2017 hat sich der Inhalt der Variablen des Ad-hoc-Moduls geändert (EF2600-EF2615).

- Aufgrund eines Aufbereitungsfehlers liegt die Information, ob das Angebot an Betreuungseinrichtungen für Kinder, pflegebedürftige Personen oder Menschen mit Behinderung Einfluss darauf hat, teilzeitbeschäftigt zu sein oder keine bezahlte Tätigkeit zu suchen, nur für den Aspekt der Teilzeit vor. Nichterwerbstätige werden dem Missingwert -5 zugeordnet (EF179u1- EF179u5).
- Die Erhebung 2019 umfasst folgende Zusatzprogramme:
 - Angaben zur Krankenversicherung (EF456-EF473),
 - Zusatzangaben zu weiteren Eigenschaften der Haupttätigkeit für Erwerbstätige (EF172, EF175).

2.5 Möglichkeit der Längsschnittverknüpfung

Im Abschnitt 1.2.3 wird darauf hingewiesen, dass der Mikrozensus eine Wiederholungsbefragung mit teilweiser Überlappung der Erhebungseinheiten ist. Nach der Umstellung der Substichprobenziehung ab dem SUF 2012 und mit der Bereitstellung längsschnittkonsistenter Ordnungsnummern besteht die Möglichkeit, die Querschnittsdaten bis einschließlich des SUF 2015 zusammenzuführen. Eine Zusammenführung mit den Querschnittsdaten ab dem SUF 2016 ist aufgrund der Erneuerung der kompletten Stichprobe im Jahr 2016 nicht möglich. Die Querschnittsdaten des SUF 2016 können mit den Querschnittsdaten der SUF bis einschließlich 2019 zusammengeführt werden. Für leichtere Auswertungen im Längsschnitt sind seit 2015 folgende Identifikatoren enthalten: idpers (Längsschnittorientierter Personenidentifikator), idpersx (Querschnittsorientierter Personenidentifikator), idhh (Längsschnittorientierter Haushaltsidentifikator) und idhhx (Querschnittsorientierter Haushaltsidentifikator). Wie ein Kurzzeitpanel erstellt werden kann und was hierbei zu beachten ist, entnehmen Sie bitte dem Skript [„Möglichkeiten der Verknüpfung von Mikrozensus-Querschnittserhebungen ab 2012 zu Panels“](#).

2.6 Zur Abgrenzung der Bevölkerungsbegriffe und zur Hochrechnung

Zum Vergleich mit den vom Statistischen Bundesamt veröffentlichten Zahlen sind die jeweiligen Abgrenzungen (siehe Anhang A2) sowie [Definitionen](#) und Hochrechnungsfaktoren zu berücksichtigen.

Die Hochrechnungsfaktoren EF951 „Standardhochrechnungsfaktor Quartal (in 1000)“, EF952 „Standardhochrechnungsfaktor Jahr (in 1000)“ und EF957 „Hochrechnungsfaktor Behinderung: Jahr (in 1000)“ des Mikrozensus SUF stammen aus dem vollen Mikrozensus. Sie rechnen bereits auf 1.000 Personen bzw. Wohnungen, Haushalte, Familien usw. der Population basierend auf der laufenden Bevölkerungsfortschreibung nach dem Zensus 2011 hoch (siehe Abschnitt 2.2).

Für deskriptive Querschnittsauswertungen zum Erhebungsjahr werden Sätze nicht benötigt, die in EF30 sowie weiteren Variablen mit -7 „Entfällt (Auswahlbezirk ohne befragten Haushalt)“ und -8 „Entfällt (Gewerblich genutzte Wohnung, leerstehende Wohnung, ausgefallene Privathaushalte und Gemeinschaftsunterkünfte)“ codiert sind. Bei der Berechnung des Stichprobenfehlers sind jedoch auch Auswahlbezirke ohne erhobene Personen, sogenannte Nullbezirke, zu berücksichtigen (siehe [Statistisches Bundesamt 2020](#): Anhang B).

Infolge der kontinuierlichen Erhebung des Mikrozensus treten sogenannte Jahresüberhänge auf. Dies sind Haushalte, die in $t-1$ (mit t = aktuelles Erhebungsjahr) hätten befragt werden sollen, in der Regel aufgrund schwerer Erreichbarkeit nicht befragt werden konnten und deren Befragung erst in t erfolgte. Je nach Befragungszeitpunkt durch eine/-n Intervie-

wer/-in in t werden Haushalte der Jahresüberhänge mit dem Erhebungsprogramm des Vorjahres oder mit dem Erhebungsprogramm des aktuellen Jahres befragt. Ebenso zu den Jahresüberhängen zählen Selbstausfüller/-innen, die den schriftlichen Fragebogen aus $t-1$ erst in t nach einem festgelegten Abschlusstermin für $t-1$ an das zuständige Statistische Landesamt zurückschicken. Inhaltlich bedeutet dies, dass für Haushalte aus dem Jahresüberhang, die mit dem Erhebungsprogramm des Vorjahres befragt wurden, erhebungsbedingt keine Angaben zu dem aktuellen vierjährigen Zusatzprogramm vorliegen und ebenso keine Angaben zu den Fragen, die neu in das Erhebungsprogramm aufgenommen wurden. Im SUF werden bei Fragen des vierjährigen Zusatzprogramms bzw. bei neu in das Erhebungsprogramm aufgenommenen Fragen die zugehörigen Variablen mit einem eigenen Missing-Code (-6) gekennzeichnet. Im Originalfile werden diese Fälle der Kategorie „Ohne Angabe“ zugeordnet.

Im Fall von Jahresüberhängen (EF5b=1, 2) ist EF4 mit Haushalten des Jahresbestandes bzw. ohne Jahresüberhang (EF5b=0) identisch. Bei eigenen Abgrenzungen auf der Ebene von Haushalten, Lebensformen oder Familien muss für eine eindeutige Identifikation der Einheiten die folgende Sortierfolge verwendet werden: EF1, EF3, EF12, EF5b und EF4 (sowie nach Bedarf Nr. der Lebensform im Haushalt (EF27) oder Nr. der Familie im Haushalt (EF25)).

2.7 Zugang zum Mikrozensus Scientific Use File und Beratung

Die Mikrozensus SUF können von unabhängigen inländischen Forschungseinrichtungen gegen eine Bereitstellungsgebühr von 250 Euro bei den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder beantragt werden.¹⁴

Interessenten/-innen und Nutzer/-innen des Mikrozensus können sich auch an das GML bei GESIS wenden und Beratungsleistungen in Anspruch nehmen.

¹⁴ Unter bestimmten Bedingungen werden für Promovierende und Studierende Ermäßigungen gewährt.

3 Literaturhinweise

3.1 Allgemein

- Afentakis, A.; Bihler, W. (2005): Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005. [Wirtschaft und Statistik \(10\): 1039-1048](#).
- Bömermann, H.; Gruber, G. (2010): Gebietsgliederungen in Berlin. [Zeitschrift für amtliche Statistik 5+6/2010: 22-25](#).
- Eurostat (2020): [EU Labour Force Survey ad hoc module 2019 on work organisation and working time arrangements. Quality Assessment Report](#). Luxembourg.
- Herter-Eschweiler, R; Schimpl-Neimanns, B. (2018): [Möglichkeiten der Verknüpfung von Mikrozensus-Querschnittserhebungen ab 2012 zu Panels](#). Bonn/Mannheim.
- Statistisches Bundesamt (2020): [Qualitätsbericht Mikrozensus 2019](#). Wiesbaden.

[Veröffentlichungen](#) auf der Datengrundlage von SUF finden sich auf der Webseite des GML

3.2 Ergebnisse des Mikrozensus 2019

- Statistisches Bundesamt (2020): [Fachserie 1 Reihe 3, Bevölkerung und Erwerbstätigkeit, Haushalte und Familien](#), Ergebnisse des Mikrozensus 2019. Wiesbaden.
- Statistisches Bundesamt (2020): [Fachserie 1 Reihe 2.2, Bevölkerung und Erwerbstätigkeit, Bevölkerung mit Migrationshintergrund](#), Ergebnisse des Mikrozensus 2019. Wiesbaden.
- Statistisches Bundesamt (2020): [Fachserie 1 Reihe 4.1, Bevölkerung und Erwerbstätigkeit, Erwerbsbeteiligung der Bevölkerung, Ergebnisse des Mikrozensus zum Arbeitsmarkt](#), Mikrozensus 2019. Wiesbaden.
- Statistisches Bundesamt (2020): [Bildungsstand der Bevölkerung – Ergebnisse des Mikrozensus 2019](#). Wiesbaden.
- Statistisches Bundesamt (2020): [Fachserie 13 Reihe 1.1, Sozialleistungen, Angaben zur Krankenversicherung](#), Mikrozensus 2019. Wiesbaden.

Unter [Themen](#) kann auf weitere kostenlose Publikationen aus dem Bereich Mikrozensus zugegriffen werden. Des Weiteren können aus der Datenbank Genesis-Online kostenlos [Tabellen](#) abgerufen werden.

3.3 Mikrozensusgesetz, EU-Verordnungen

3.3.1 Mikrozensusgesetz 2017

- Mikrozensusgesetz vom 7. Dezember 2016 (BGBl. I S. 2826), das durch Artikel 178 der Verordnung vom 19. Juni 2020 (BGBl. I S. 1328) geändert worden ist. [[Download](#)]

3.3.2 EU-Verordnungen zur Arbeitskräfteerhebung

- Verordnung (EG) Nr. 577/98 des Rates vom 9. März 1998 zur Durchführung einer Stichprobenerhebung über Arbeitskräfte in der Gemeinschaft (ABl. L 77 vom 14.

März 1998, S. 3), zuletzt geändert durch Verordnung (EG) Nr. 545/2014 des Europäischen Parlaments und des Rates vom 15. Mai 2014 (ABl. L 163 vom 29. Mai 2014, S. 10). [\[Download\]](#).

- Durchführungsverordnung (EU) 2017/2384 der Kommission vom 19. Dezember 2017 zur Festlegung der technischen Merkmale des Ad-hoc-Moduls 2019 über Arbeitsorganisation und Arbeitszeitgestaltung. [\[Download\]](#)

A Anhang

A1 Informationen zur Nutzung der Setups

Dieser Abschnitt dokumentiert, wie die Systemfiles auf Basis der Rohdaten erstellt wurden bzw. mittels der Setups des German Microdata Lab (GML) der GESIS selbst erstellt werden können.

Die vom GML bereitgestellten Setups für das Mikrozensus SUF 2019 dienen zum Einlesen des Rohdatenmaterials und zum Erstellen von Systemfiles für die Programmpakete SPSS, SAS und Stata. Sie beinhalten Programmanweisungen zur Definition von fehlenden Werten sowie zum Versehen der Variablen und ihrer Ausprägungen mit entsprechenden Labels. Die Setups werden im Dateiformat PC, Dos/Windows angeboten. Eine Umsetzung auf das Dateiformat Unix kann zum Beispiel mit Notepad++ oder Textpad++ durchgeführt werden.

Spezifika der Statistikprogramme:

- Am Anfang der Setups sind in der **Configuration Section** die vollständigen lokalen Dateinamen (einschließlich Laufwerkskennzeichen und Verzeichnis) zu nennen. Ansonsten sind im Setup keine weiteren Änderungen vorzunehmen.

IBM SPSS Statistics 24.0.0.1	Stata/MP 17.1 for Windows	SAS 9.4 for Windows
	<code>local log_file "LOG_FILENAME.log" ;</code>	
<code>file handle raw_data / name = "RAW_DATA_FILE- NAME" / lrec1 = 1216.</code>	<code>local raw_data_file "RAW_DATA_FILENAME" ;</code>	<code>%let raw_data_filename = 'RAW_DATA_FILENAME' ;</code>
		<code>%let sas_file_dirname = 'SAS_FILE_DIREC- TORY_PATH' ;</code>
<code>file handle spss_refer- ence / name = "SPSS_FILENAME.sav"</code>	<code>local stata_file "STATA_FILENAME.dta" ;</code>	<code>%let sas_file_basename = SAS_FILENAME ;</code>

- **Stata:** Das Setup steht mit der Zeichencodierung Unicode (UTF-8) bereit und ist mit Stata ab Version 14 ablauffähig. Für die Verwendung mit älteren Versionen kann es z. B. mithilfe von MS-Edit in die Zeichencodierung ANSI bzw. Windows-1252 umgesetzt werden.
- **SPSS:** Das Setup mit der Zeichencodierung Windows-1252 ist sowohl mit Version 24 als auch mit älteren Versionen ablauffähig, wenn entsprechend unter „Bearbeiten“ | „Optionen“ | „Sprache“ | „Zeichencodierung ...“ | „[x] Schriftsystem der Ländereinstellung ...“ eingestellt ist.
- **Missing Values:** In SPSS können vorliegende Werte als benutzerdefinierte Missings deklariert werden. Sie werden bei Auszählungen mit entsprechenden Werten und Labels ausgewiesen, zählen aber i. d. R. bei statistischen Modellen nicht als gültige Werte. In SAS und Stata können dagegen vorliegende Werte nur nach Recodierung als fehlende

Werte definiert werden. Im SPSS-Setup werden benutzerdefinierte Missings spezifiziert. Diese Definitionen sind zwar auch in den SAS- und Stata-Setups enthalten, im Unterschied zum SPSS-Setup jedoch auskommentiert, da sonst die Originalwerte der Rohdaten (-1, ..., -9) im Systemfile durch benutzerdefinierte Missing-Zeichen (.a, ..., .h) ersetzt würden. Bei Bedarf können diese im Setup unter dem Kommentar „Definition of usermissing values“ stehenden Programmanweisungen durch Entfernen der Kommentarzeichen (/*, */) am Beginn und Ende des Anweisungsblocks aktiviert werden.

- **Voreinstellungen:** Um eine reibungsfreie Aufbereitung der Daten zu gewährleisten, empfiehlt es sich, die in den Setups vorgesehenen Voreinstellungen nicht zu verändern.

IBM SPSS Statistics 24.0.0.1	Stata/MP 17.1 for Windows	SAS 9.4 for Windows
Set decimal dot. * Ggf. ändern in "... yes." oder "... no." und Kommentarzeichen * löschen. * set unicode no. set olang = GERMAN.	set more off set dp comma #delimiter ;	options compress = binary ;

Ist das entsprechende Systemfile erstellt, kann die einfache Fallzahl $n = 596.858$ (ohne Gewichtung, ohne Selektion) zur Kontrolle, ob der Rohdatensatz fehlerfrei eingelesen wurde, mit der des erstellten Datensatzes verglichen werden. Unterscheiden sich die Fallzahlen, weist dies meist auf einen Fehler beim Einlesen hin. Zusätzlich können die folgenden Verteilungen zur Prüfung des fehlerfreien Ablaufs des Setups herangezogen werden. Weitere Randverteilungen ausgewählter Variablen des Mikrozensus SUF 2019 sind im [GESIS Web](#) zu finden.

Die Datei enthält insgesamt 67.886 Sätze, die in EF30 sowie weiteren Variablen mit -7 „Entfällt (Auswahlbezirk ohne befragten Haushalt)“ und -8 „Entfällt (Gewerblich genutzte Wohnung, leerstehende Wohnung, ausgefallene Privathaushalte und Gemeinschaftsunterkünfte)“ codiert sind. Diese Sätze werden für Querschnittsauswertungen zum Erhebungsjahr nicht benötigt. Hierfür stehen $n = 528.972$ ($= 596.858 - 67.886$) Sätze zur Verfügung.

A2 Verteilungen zur Prüfung der Systemfiles bzw. des Ablaufs der Setups

Merkmal	Selektion	Hochrechnung	in 1.000
Bevölkerungskonzepte			
Bevölkerung am Haupt- und Nebenwohnsitz (ungewichtete Fallzahl (EF30>0): n = 529.610)	EF30>=1 & EF30<=3	EF952	82.794
Bevölkerung am Hauptwohnsitz	EF30>=1 & EF30<=2	EF952	81.848
Bevölkerung am Hauptwohnsitz, Erwerbstätige	EF30>=1 & EF30<=2 & EF29=1	EF952	42.378
Bevölkerung in Privathaushalten	EF31=1	EF952	82.794
Bevölkerung in Familien/Lebensformen am Hauptwohnsitz (Lebensformenkonzept)	EF761>=1 & EF761<=2	EF952	81.933
Lebensformen am Hauptwohnsitz (Lebensformenkonzept)	EF761>=1 & EF761<=2 & EF762=1	EF952	42.085
Bevölkerung am Familienwohnsitz (trad. Familienkonzept ¹⁵)	EF863=1	EF952	81.966
Familien (trad. Familienkonzept)	EF863=1 & EF864=1	EF952	45.321
Mikrozensus-Grundprogramm			
Stellung im Beruf (EF117) x Erwerbstyp (EF29)¹⁶	EF30>=1 & EF30<=2	EF952	
Erwerbstätige	EF29=1		
Selbstständige			3.950
Mithelfende Familienangehörige			114
Beamte (inkl. Zeit-/ Berufssoldaten (inkl. BGS und Bereitschaftspolizei), Beamtenanwärter, freiwillig Wehrdienstleistende)			2.042
Angestellte (inkl. geringfügig beschäftigte Personen, Personen im Freiwilligendienst)			26.440
Arbeiter			8.310
Auszubildende (inkl. Volontäre, Praktikanten)			1.522
Erwerbstätige insgesamt			42.378
Erwerbslose	EF29=2		1.371
Erwerbspersonen	EF29>=1 & EF29<=2		43.749
Nichterwerbspersonen	EF29>2		38.098

¹⁵ Für die bis einschließlich Mikrozensus 2004 verwendeten traditionellen Familienkonzepte werden in den aktuellen Fachserien keine Verteilungen veröffentlicht.

¹⁶ Siehe Statistisches Bundesamt (2020), Fachserie 1 Reihe 4.1, Tab. 1.2.

Merkmal	Selektion	Hochrechnung	in 1.000
Privathaushalte: Haushaltsgröße (EF20)¹⁷		EF952	
1 Person	EF31=1 & EF662>=1 & EF662<=4		17.596
2 Personen			13.785
3 Personen			4.959
4 Personen			3.789
5 Personen und mehr			1.419
Haushalte insgesamt			41.550
Mikrozensus-Zusatzprogramm Krankenversicherung			
Krankenversicherung (EF456)¹⁸		EF952	
	EF30>=1 & EF30<=2		
In einer gesetzlichen Krankenversicherung selbst pflichtversichert			50.044
In einer gesetzlichen Krankenversicherung selbst freiwillig versichert			3.623
In einer gesetzlichen Krankenversicherung als Familienangehöriger versichert			16.220
In einer privaten Krankenversicherung selbst versichert			6.628
In einer privaten Krankenversicherung als Familienangehöriger versichert			2.240
Nicht bekannt, wie krankenversichert			68
Nein, nicht krankenversichert			200
Keine Angabe			226
EU-Labour Force Survey – Arbeitsorganisation und Arbeitszeitgestaltung			
Ort der Hauptarbeitsstätte (EF2612) 0,1 %-Un- terstichprobe, freiwillige Beantwortung	EF6=2 & EF32=1	EF956	
In den Räumlichkeiten des Arbeitgebers bzw. in eigenen Büro- oder Geschäftsräumen			29.585
Zu Hause			1.401
In Räumlichkeiten oder auf Grundstücken von Kunden			3.341
Im Auto oder anderen Verkehrsmitteln (z.B. Taxi, LKW, Bahn, Flugzeug)			1.343
An einem anderen Ort			1.703
Keine Angabe			4.167

¹⁷ Siehe Statistisches Bundesamt (2020), Fachserie 1 Reihe 3, Tab. 1.2.1.

¹⁸ Siehe Statistisches Bundesamt (2020), Fachserie 13 Reihe 1.1, Tab. 2.

Als Referenz zur Plausibilitätsprüfung der Mikrozensus SUF dienen die in den Fachserien des Statistischen Bundesamtes veröffentlichten Ergebnisse des jeweiligen Mikrozensus. Für den Mikrozensus 2019 sind dies insbesondere die Fachserie 1 Bevölkerung und Erwerbstätigkeit: Reihe 3 (Haushalte und Familien), Reihe 4.1 (Erwerbsbeteiligung der Bevölkerung) und Reihe 2.2 (Bevölkerung mit Migrationshintergrund). Bei diesen Vergleichen sind die von den Statistischen Ämtern verwendeten Bevölkerungs- und Erwerbskonzepte sowie die Hochrechnung zu beachten.

Bedingt durch die Ziehung der 70 %-Substichprobe ergeben sich Abweichungen zwischen den Häufigkeiten der Variablen des Mikrozensus SUF und den in den Fachserien des Statistischen Bundesamtes veröffentlichten Zahlen bzw. der Original-Mikrozensusdaten; siehe Tabelle [„Überprüfung der Randverteilung“](#). Die meisten Variablen des SUF weichen nur in sehr geringem Maße von den veröffentlichten Daten ab. Größere relative Abweichungen sind nur bei Merkmalen festzustellen, die mit sehr geringen Fallzahlen besetzt sind.

Bei der Aufbereitung des Mikrozensus SUF 2019 wurden die Variablenlabels nach der erstmals für das Mikrozensus SUF 2003 verwendeten Systematik des Mikrodaten-Informationssystems ([MISSY](#)) erstellt. Die Variablenlabels enthalten die Nummer der entsprechenden Frage im Selbstausfüllerbogen, den grundsätzlichen inhaltlichen Bezug der Variable und eventuelle Spezifizierungen; außerdem sind Freiwilligkeit und Zugehörigkeit zur Unterstichprobe gekennzeichnet. Ein im Vergleich mit früheren Files abweichendes Label bedeutet also i. d. R. nicht, dass sich am Inhalt der entsprechenden Variable etwas geändert hat.

Zur besseren Lesbarkeit und aus Platzgründen wird in den Setups zum Mikrozensus ab 2006 - im Unterschied zu den Setups früherer Erhebungen - bei den Value Labels anstatt der gleichzeitigen Verwendung männlicher und weiblicher Sprachformen das generische Maskulinum verwendet.