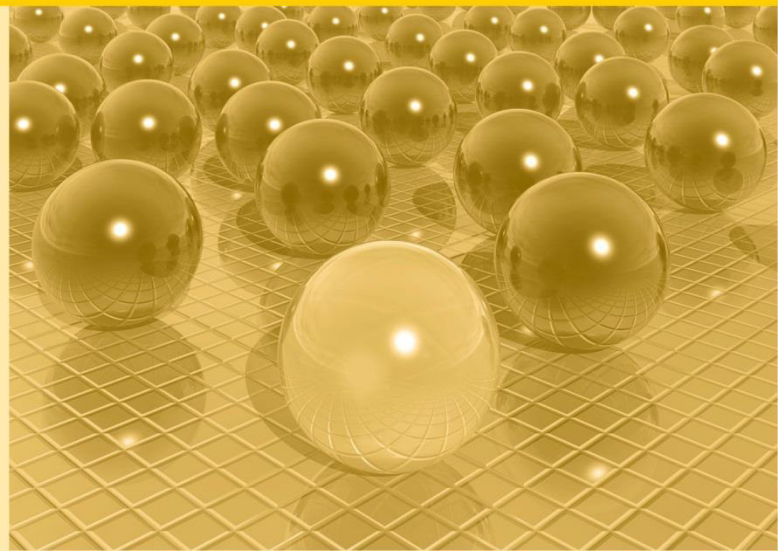


# Metadatenreport



Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus 2021 am  
Gastwissenschaftsarbetsplatz sowie per kontrollierter Datenfernverarbeitung  
(sog. On-Site-Zugangswege/-Nutzungen)

DOI: 10.21242/12211.2021.00.00.1.1.0

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](http://www.forschungsdatenzentrum.de)  
E-Mail: [forschungsdatenzentrum@it.nrw.de](mailto:forschungsdatenzentrum@it.nrw.de)

### Fachliche Informationen

zu dieser Veröffentlichung:

Forschungsdatenzentrum der  
Statistischen Ämter der Länder  
– Düsseldorf –  
Tel.: 0211 9449-2871  
Fax: 0211 9449-8087  
[forschungsdatenzentrum@it.nrw.de](mailto:forschungsdatenzentrum@it.nrw.de)

### Informationen zum Datenangebot:

Statistisches Bundesamt  
Forschungsdatenzentrum  
Tel.: 0611 75-2420  
Fax: 0611 75-3915  
[forschungsdatenzentrum@destatis.de](mailto:forschungsdatenzentrum@destatis.de)

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

Erscheinungsfolge: unregelmäßig  
Erschienen im März 2024

Diese Publikation wird kostenlos als PDF-Datei zum Download unter [www.forschungsdatenzentrum.de](http://www.forschungsdatenzentrum.de) angeboten.

© Information und Technik Nordrhein-Westfalen, Düsseldorf, 2024  
(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 2021 am Gastwissenschaftsarbetsplatz sowie per kontrollierter Datenfernverarbeitung (sog. On-Site-Zugangswege/-Nutzungen) (EVAS-Nummer: 12211, 12231, 12241, 12251). Version 1. DOI: 10.21242/12211.2021.00.00.1.1.0. Düsseldorf 2024.

# Metadatenreport

Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus 2021 am Gastwissenschaftsarbeitsplatz sowie per kontrollierter Datenfernverarbeitung (sog. On-Site-Zugangswege/-Nutzungen)

DOI: 10.21242/12211.2021.00.00.1.1.0

Version 1



## Inhalt

|  |           |
|--|-----------|
| <b>Einleitung</b> .....  | <b>3</b>  |
| <b>1. Datenaufbereitung in den FDZ</b> .....                   | <b>4</b>  |
| <b>1.1 Datenaufbereitung</b> .....                             | <b>4</b>  |
| 1.1.1 Missingkodierung.....                                    | 4         |
| 1.1.2 Hochrechnungs-/Gewichtungsvariablen.....                 | 6         |
| 1.1.3 SILC Indizes .....                                       | 7         |
| <b>1.2 Anonymisierungsmaßnahmen</b> .....                      | <b>7</b>  |
| <b>1.3 Methodik der Verknüpfung</b> .....                      | <b>8</b>  |
| <b>2. Produkt</b> .....  | <b>9</b>  |
| <b>2.1 Hinweise zur Qualität des Mikrozensus 2021</b> .....    | <b>9</b>  |
| <b>2.2 Merkmale und Merkmalsbeschreibung</b> .....             | <b>10</b> |
| 2.2.1 Merkmalsdefinitionen.....                                | 10        |
| 2.2.2 Datensatzbeschreibung / Schlüsselverzeichnis .....       | 11        |
| <b>2.3 Vergleichbarkeit der Merkmale über die Zeit</b> .....   | <b>12</b> |
| 2.3.1 Variablennamen .....                                     | 12        |
| 2.3.2 Missingkodierung .....                                   | 12        |
| 2.3.3 Haushalte und Lebensformen.....                          | 13        |
| 2.3.4 Migrationstypisierungen.....                             | 13        |
| 2.3.5 Einkommen und Lebensbedingungen (SILC) .....             | 13        |
| 2.3.6 Informations- und Kommunikationstechnologien (IKT) ..... | 13        |
| 2.3.7 Zusatzprogramm 2021 .....                                | 14        |
| 2.3.8 Adhoc-Modul 2021 .....                                   | 15        |
| <b>2.4 Eckwerte relevanter Merkmale</b> .....                  | <b>16</b> |
| 2.4.1 Verteilung der Bevölkerung am Hauptwohnsitz .....        | 17        |
| 2.4.2 Privathaushalte am Hauptwohnsitz.....                    | 18        |
| <b>2.5 Auswertbare regionale Ebenen</b> .....                  | <b>19</b> |
| <b>2.6 Produktversionen</b> .....                              | <b>20</b> |
| <b>3. Praktische Hinweise</b> .....                            | <b>21</b> |

|   |           |
|---|-----------|
| <b>3.1 Hinweise zur Geheimhaltung.....</b>                        | <b>21</b> |
| 3.1.1 Gesetzliche Grundlagen der statistischen Geheimhaltung..... | 21        |
| 3.1.2 Geheimhaltung von Ergebnissen .....                         | 21        |
| 3.1.3 Praktische Tipps zur Vermeidung von Geheimhaltungsfällen .  | 22        |
| <b>3.2 FAQ .....</b>  | <b>22</b> |
| <b>3.3 Verfügbare Tools .....</b>                                 | <b>29</b> |
| <b>Anhang.....</b>  | <b>31</b> |

## Einleitung

Dieser Metadatenreport soll Forschenden dabei helfen, die Daten des Mikrozensus On-Site-Materials 2021<sup>1</sup> sachgerecht auszuwerten. Bei den On-Site-Daten handelt es sich um die Daten, die zur Nutzung am Gastwissenschaftsarbetsplatz (GWAP) oder über eine kontrollierte Datenfernverarbeitung (KDFV) vorgesehen sind. Er gibt einen Überblick über Datenaufbereitung, Dokumentation, Qualität, bereitgestellte Merkmale, Eckwerte relevanter Merkmale und praktische Hinweise.

Weitere Informationen zum Datenangebot und zum Datenzugang sind zudem auf den Webseiten der [Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder](#) abrufbar. Hier findet sich u.a. der [Metadatenreport Teil I Statistik](#), der diesen Metadatenreport mit allgemeinen und methodischen Informationen zum Mikrozensus ab dem Berichtsjahr 2020 ergänzt.

Bei weiteren Fragen können sich interessierte Personen und Nutzende des Mikrozensus an das [Forschungsdatenzentrum der Statistischen Ämter der Länder – Standort NRW](#) (insbesondere bei Fragen zur Datenaufbereitung) und an das [Forschungsdatenzentrum des Statistischen Bundesamtes](#) wenden.

Neben der Nutzung der formal anonymen On-Site-Daten an einem Gastwissenschaftsarbetsplatz oder über eine kontrollierte Datenfernverarbeitung, gibt es auch die Möglichkeit, Scientific-Use-Files am eigenen Arbeitsplatz in der wissenschaftlichen Einrichtung innerhalb Deutschlands auszuwerten. Produktspezifische Informationen zum Scientific-Use-File (MZ-SUF) werden im [Metadatenreport Teil II zum MZ-SUF](#) bereitgestellt. Darüber hinaus stehen auf dem [Mikrodaten-Informationssystem \(MISSY\) der GESIS](#) ausführliche Informationen und Auswertungshilfen zum MZ-SUF zur Verfügung.

Weitere allgemeine Informationen zum Mikrozensus sind auf den Seiten des Statistischen Bundesamtes ([Was ist der Mikrozensus?](#)) und der [Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder](#) abrufbar.

---

<sup>1</sup> Die Aufbereitung und Dokumentation des formal anonymisierten Einzelmaterials erfolgt durch das Forschungsdatenzentrum der Statistischen Ämter der Länder – Standort NRW.

# **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. Die Datenaufbereitung findet für den Mikrozensus 2021 letztmalig mit dem Programm SPSS statt. Bei der Konvertierung in andere Datenformate kann es aufgrund softwarespezifischer Besonderheiten zu Veränderungen von Variablenformaten kommen.

Im Erhebungsteil SILC und dem erstmals im Mikrozensus enthaltenen Erhebungsteil IKT wurden un plausible Angaben aus den Variablen und Hochrechnungsfaktoren entfernt. Außerdem sind in bezogen auf die einzelnen Erhebungsteile mit Ausnahme der IKT nur vollständige Haushalte in dem jeweiligen Erhebungsteil miteingeflossen. Das bedeutet, dass bei einer fehlenden Angabe einer Person der ganze Haushalt aus dem entsprechenden Befragungsteil entfernt wurde.

### **1.1.1 Missingkodierung**

Durch das Forschungsdatenzentrum wird eine Ausfallkennung zur Kenntlichmachung der Ausfallursache umgesetzt. Weiterhin wird geprüft, ob sich die Filterführung des Fragebogens im Datenmaterial widerspiegelt. Für Personengruppen, denen gemäß Ausfalltypisierung bestimmte Fragen nicht gestellt wurden, werden für die betroffenen Merkmale die in Tabelle 1 beschriebenen Werte codiert. Liegt ein Fehlwert in den Daten vor, wird der dazu passende Missingcode hinterlegt. Treten mehrere Ausfallgründe bei derselben befragten Person auf, wird der Missingcode eingetragen, der in Tabelle 1 als erstes referenziert ist. Wird beispielsweise eine Person unter 15 Jahren aus einer Gemeinschaftsunterkunft (GU) erhoben, wird die Ausfallkennung -1 zugewiesen. Bei einer unter 15-jährigen Person, die nicht erwerbstätig ist, wird der Missingcode -3 hinterlegt. Die Abfolge der Ausfalltypisierung wird vorgenommen, um die Zuordnung der Ausfalltypisierung zu vereinfachen, da andernfalls verschiedene Kombinationen eines Ausfalls geprüft werden müssten.



| <b>Missingcode</b> | <b>Beschreibung</b>   |
|--------------------|---|
| -7 <sup>2</sup>    | Auswahlbezirke (AWB) ohne Befragung   |
| -1                 | Gemeinschaftsunterkunft (GU)  |
| -4                 | Nicht in der jeweiligen Unterstichprobe oder kein geeigneter Hochrechnungsfaktor für die jeweilige Unterstichprobe  |
| -3                 | <p>Person unter 15 (bzw. 16) Jahren<sup>3</sup></p> <p>Im Erhebungsteil IKT auch: Person unter 16 Jahren oder älter als 74 Jahre</p> <p>Filtervariablen:</p> <ul style="list-style-type: none"> <li>• EA0100P</li> <li>• DG0400P</li> <li>• ER1001P</li> <li>• ES1403P</li> </ul>   |
| -2                 | <p>Nichterwerbstätige</p> <p>Filtervariablen:</p> <ul style="list-style-type: none"> <li>• EA0502P</li> <li>• EA0800P</li> <li>• EA0700P</li> <li>• EA0802P</li> <li>• EJ0501P</li> </ul>   |
| -6                 | <p>Nebenwohnsitze für SILC: Personen am Nebenwohnsitz entfallen in der SILC Unterstichprobe, bei Haushaltsvariablen entfallen im SILC-Teil Haushalte, in denen keine Person ab 16 Jahren den Hauptwohnsitz hat</p> <p>Filtervariablen:</p> <ul style="list-style-type: none"> <li>• AA0200P</li> <li>• AA0500H</li> </ul> |

<sup>2</sup> -7 wird im Schlüsselverzeichnis nicht mit aufgeführt, wird aber mit Ausnahme einzelner Identifier und Hochrechnungsfaktoren in die Daten codiert und gelabelt.

<sup>3</sup> Der Code -3 wird auch dort verwendet, wo die Altersgrenze nicht 15 Jahre, sondern 16 Jahre ist (z. B. im Erhebungsteil SILC).

|    |   |
|----|---|
|    | <ul style="list-style-type: none"> <li>• BA0400H</li> <li>• CA0400H</li> <li>• DC0700P</li> <li>• EB0800P</li> <li>• ED1801P</li> <li>• EA1200P</li> <li>• DG0302H</li> <li>• ER1000P</li> <li>• ES1402P</li> </ul> |
| -5 | Spezifischer Filterausfall, Ausfallgrund wird variablenspezifisch benannt <sup>4</sup>  |
| -9 | Filterfehler oder technisch zugelassene Fehlwerte bei Fragen mit Auskunftspflicht, keine Angabe bei Fragen ohne Auskunftspflicht  |

*Tabelle 1: Ausfallkennungen*

Im Mikrozensus On-Site-Material 2018 und 2019 wurde zusätzlich der Missingcode -8 für Leerstand und ausgefallene Privathaushalte vergeben. Im Mikrozensus 2021 wird dieser nicht programmiert, da im Datenmaterial 2021 keine ausgefallenen Haushalte (kein Leerstand oder unbewohnte Wohnungen) enthalten sind. Anders als 2020 sind ab 2021 aber wieder Auswahlbezirke im Datenmaterial, die zur Bruttostichprobe gehören, aber keine Befragungsergebnisse aufweisen (z. B. da keine Wohnbevölkerung vorhanden war oder der ganze AWB eine Unit-Non-Response hatte). Dies erhöht die Transparenz, wenn eine Verknüpfung zu Panels gewünscht ist (vgl. Abschnitt 1.3). Für solche Auswahlbezirke wird bei Erhebungsvariablen und daraus abgeleiteten Variablen der Ausfallcode -7 verwendet.

Item-Nonresponse bei freiwilligen Fragen wird ab 2021 nicht mehr mit 9, 99, 999 etc., sondern ebenfalls mit -9 codiert. Die Ausprägungen 9, 99, 999 etc. für „keine Angabe“ werden nur noch verwendet, wenn diese Kategorie explizit im Fragebogen vorgesehen ist.

### **1.1.2 Hochrechnungs-/Gewichtungsvariablen**

Fälle ohne einen Hochrechnungsfaktor werden aus dem Datenmaterial entfernt. Anders als im Erhebungsjahr 2020, sind aber Auswahlbezirke der Bruttostichprobe als eigener Fall im Datenmaterial enthalten, wenn keine Einzeldatensätze für diesen Auswahlbezirk vorliegen.

<sup>4</sup> Die technische Beschreibung der Filterführung im Wertelabel bezieht sich auf positive Werte der filternden Variablen. Durch nachträgliche Korrekturen kann es im Einzelfall dazu kommen, dass sich nicht alle Filter in den On-Site-Daten reproduzieren lassen.

Die Hochrechnungsfaktoren werden so angeglichen, dass sie ohne zusätzliche Multiplikation (mit 1000) auf die Gesamtbevölkerung hochgerechnet werden können.

### **1.1.3 SILC Indizes**

Ab dem Berichtsjahr 2020 enthält der Mikrozensus den Erhebungsteil zu Einkommen und Lebensbedingungen (SILC). Aus Anonymitätsgründen können die meisten SILC-Merkmale nicht für die Off-Site-Nutzung im Scientific-Use-File (MZ-SUF) bereitgestellt werden. Daher werden für das MZ-SUF Variablen mehrere Fragen zu jeweils einem Index zusammengefasst. Diese Indizes werden auch On-Site neben den ursprünglichen Erhebungsmerkmalen bereitgestellt.

Folgende SILC-Indizes stehen 2021 zur Verfügung:

- Index\_Finanz\_Situation
- Index\_Sozialkapital
- Index\_Lebenssituation
- Index\_Zahlungsrueckstaende
- Index\_Oeffentliche\_Zahlungen
- Index\_Arztkosten
- Index\_Situation\_Kinder
- Index\_Arztkosten\_Kind

Wo möglich liegen den Indizes SILC-Zielvariablen, die auch für Lieferungen an Eurostat verwendet werden, zugrunde. In Fällen, in denen keine entsprechenden verwendbaren Zielvariablen vorliegen, werden die Erhebungsmerkmale verwendet. Alle für die Berechnung der einzelnen Indizes verwendeten Erhebungs- oder Zielvariablen sind dem [Schlüsselverzeichnis](#) zu entnehmen. Liegen den Indizes freiwillige Fragen zugrunde, ist mit erhöhtem Item-Nonresponse und Ausfällen bei den Indizes zu rechnen.

Die Missingcodes werden bei den SILC-Indizes entsprechend den Missingcodes der zugrunde liegenden Variablen vergeben. Sofern mindestens eine der zugrunde liegenden Variablen den Fehlwert -9 aufgrund von Nichtbeantwortung aufweist, wird der gesamte Index mit -9 ausgewiesen.

## **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 sind systemfrei sortiert, sodass 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, sodass 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, auf Individualebene die Wellen ab 2020 miteinander zu einem Paneldatensatz zu verknüpfen. Hierfür stehen ab 2020 verkettete Identifikatoren für Auswahlbezirke (idawb), Haushalte (idhh) und Personen (idpers) zur Verfügung. Unterjährige Wiederholungsbefragungen im Rahmen der Arbeitskräfteerhebung (LFS) lassen sich über die vorgesehene Bogenart (AWBAUSWAHLTEIL=4) identifizieren. Sofern aus methodischen Gründen keine Dubletten innerhalb eines Erhebungsjahres gewünscht sind, ist eine Entfernung der Fälle erforderlich. In diesem Fall ist zu beachten, dass die Hochrechnungsfaktoren HR100QQ und HR100JQ sowie die Hochrechnungsfaktoren für die Kernvariablen nicht für eine Hochrechnung auf die Gesamtbevölkerung geeignet sind, da relevante Fälle entfernt werden. Auswirkungen auf die Hochrechnung mit HR100JJ gibt es nicht, da der Hochrechnungsfaktor für LFS-Strukturvariablen bei Wiederholungsbefragungen unbelegt ist. In den ebenfalls enthaltenen querschnittsorientierten Identifikatoren (idawbx, idhbx, idpersx) sind die Wiederholungsbefragungen bereits mit einer abweichenden ID versehen. Nähere Informationen zur Panelverknüpfung sind [Brocker und Mühlenfeld 2020](#) sowie [Herter-Eschweiler und Schimpl-Neimanns 2018](#) und [Backes und Pforr 2023](#) zu entnehmen. Die Arbeitspapiere beziehen sich auf die Erhebungsjahre 2012 bis 2015 bzw. 2016 bis 2019. Die Erkenntnisse sind aber in Teilen auch auf die Jahre ab 2020 zu übertragen. Für panelverknüpfte Datensätze aus den Mikrozensusjahren 2020 und 2021 steht für die Unterstichprobe SILC der Hochrechnungsfaktor HR200L2 zur Verfügung. Für die übrigen Erhebungsteile und Unterstichproben sind keine Längsschnitt-Hochrechnungsfaktoren verfügbar.

## 2. Produkt

### 2.1 Hinweise zur Qualität des Mikrozensus 2021

Das Erhebungsjahr 2020 des Mikrozensus stand im Zeichen von zwei besonderen Herausforderungen. Zum einen wurde für den MZ 2020 ein komplett neues IT-System aufgebaut, dessen Einführung von technischen Problemen begleitet war. Zum anderen erschwerte die Corona-Pandemie die Vorbereitung und Umsetzung der Datenerhebung in den Haushalten. Beide Herausforderungen wirkten sich auch auf die Erhebung 2021 aus – wenn auch in deutlich geringerem Maße im Vergleich zur Erhebung 2020.

Die mit 35% hohe Ausfallquote des Jahres 2020 bei den Endergebnissen im Bundesschnitt konnte deutlich reduziert werden. Sie liegt im Jahr 2021 bei ca. 14%. Die Ausfallquoten für die freiwilligen Fragen, z. B. zu den Lebensbedingungen aus der seit 2020 integrierten SILC-Unterstichprobe und der 2021 erstmalig integrierten IKT-Unterstichprobe, sind in der Regel deutlich höher. Die Antwortausfälle können nicht als zufällig angenommen werden. Teilweise gibt es auch 2021 größere regionale Unterschiede. Informationen zu länderspezifischen Besonderheiten für das Jahr 2021 finden sich auf den Webseiten der Statistischen Ämter der Länder zum Mikrozensus. Für die IKT-Unterstichprobe (maximal 3,5% der MZ-Stichprobe) sind die [Ergebnisse 2021 auf Bundesländer-Ebene nur eingeschränkt interpretierbar](#), da die Stichprobenauswahl und die Hochrechnung in erster Linie darauf ausgerichtet und optimiert sind, präzise Ergebnisse für den Bund bereitstellen zu können. Für die Hochrechnung wurden daher verschiedene länderübergreifende Zusammenlegungen vorgenommen.

Die Erhebungssituation 2021 hat zur Folge, dass Auswertungen – insbesondere in Kombination von fachlicher und regionaler Tiefe – vereinzelt nicht die vom Mikrozensus sonst gewohnte Belastbarkeit aufweisen. Die Einschränkungen sind allerdings deutlich geringer als im Erhebungsjahr 2020.

Grundsätzlich gilt, dass aufgrund methodischer Veränderungen (verschiedene Unterstichproben und damit einhergehende veränderte Hochrechnungsverfahren, unterjährige Rotation einzelner Teile, verändertes Fragenprogramm) und dem Einsatz von Online-Fragebögen (CAWI) zur Datenerhebung sowie der genannten Besonderheiten in der Erhebung die Ergebnisse ab 2020 nur eingeschränkt mit den Ergebnissen der Vorjahre vergleichbar sind. Detaillierte Hinweise hierzu finden sich auf den Seiten des Statistischen Bundesamts zur [Neuregelung des Mikrozensus ab 2020](#) und in den offiziellen [Qualitätsberichten zum Mikrozensus](#), der [Unterstichprobe SILC](#) sowie der [Unterstichprobe IKT](#).

Vor dem Hintergrund der genannten Herausforderungen in den Erhebungen 2020 und 2021 sollten keine Veränderungsanalysen zu Vorjahren durchgeführt werden und keine kausalen Interpretationen bezüglich möglicher Effekte der Corona-Pandemie abgeleitet werden.

## 2.2 Merkmale und Merkmalsbeschreibung

### 2.2.1 Merkmalsdefinitionen

Folgende Klassifikationen finden in dem beschriebenen Produkt Anwendung:

- Klassifikation der Berufe, Ausgabe 2010, überarbeitete Fassung von 2020 (KldB 2010 ü. F.) (bis 5-Steller):  
<https://statistik.arbeitsagentur.de/DE/Navigation/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-Fassung2020/Systematik-Verzeichnisse/Systematik-Verzeichnisse-Nav.html>
- Datenbank der Dokumentationskennzahlen (DKZ, Fassung 2020) (8-Steller basierend auf dem 5-Steller der KldB):  
Bundesagentur für Arbeit (auf Anfrage)
- Internationale Standardklassifikation der Berufe, Ausgabe 2008 (ISCO-08) (4-Steller):  
<http://www.ilo.org/public/english/bureau/stat/isco/isco08/> (englisch)
- Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008) (bis 5-Steller):  
<https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/Downloads/klassifikation-wz-2008-3100100089004.pdf?blob=publicationFile&v=5>
- 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> (englisch)
- Bildungsfelder ISCED Fields of Education and Training (ISCED-F 2013):  
[https://www.statistik.at/KDBWeb/kdb\\_DownloadsAnzeigen.do?KDBtoken=ignore&&UFRUF=klass&&NAV=DE&&KLASSID=10527&&KLASSNAME=ISCED-F](https://www.statistik.at/KDBWeb/kdb_DownloadsAnzeigen.do?KDBtoken=ignore&&UFRUF=klass&&NAV=DE&&KLASSID=10527&&KLASSNAME=ISCED-F)  
(Übersetzung der Klassifikation von Statistik Austria)
- Hauptfachrichtungen (7-Steller basierend auf dem 4-Steller der ISCED-Bildungsfelder): Statistisches Bundesamt  
Auch 2021 stehen wieder zu jeder Hauptfachrichtungsvariablen eine auf die bis 2017 verwendeten Zweisteller umgeschlüsselte Variable (DC0603PA, EP1103PA, ER0703PA, ER0713PA) zur Verfügung. Aufgrund methodischer Änderungen bei der Zuordnung ist eine Vergleichbarkeit mit Vorjahresergebnissen nicht möglich.
- Staatsangehörigkeits- und Gebietssystematik (außereuropäische Staaten zum Teil zusammengefasst siehe Schlüsselverzeichnisse):  
<https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/staatsangehoerigkeit-gebietsschluessel.html>

- Amtlicher Gemeindeschlüssel (AGS) / Regionalschlüssel (RS)<sup>5</sup>:  
<https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/inhalt.html>

### 2.2.2 Datensatzbeschreibung / Schlüsselverzeichnis

Der Mikrozensus ist eine Haushalts- und Personenstatistik. Die vollständige Liste der im Datensatz vorhandenen Variablen ist dem [Schlüsselverzeichnis](#) zu entnehmen. Sofern Variablen das Ergebnis einer Frage in einem oder mehreren Fragebögen wiedergeben, sind die zugehörigen Fragebögen mitsamt der Fragennummer in der Spalte „Fragennummer“ enthalten.

Aufgrund der 2020 eingeführten Neukonzeption des Mikrozensus, welche die Integration weiterer Haushaltserhebungen<sup>6</sup> sowie Neuerungen bei der europäischen Arbeitskräfteerhebung (LFS)<sup>7</sup> umfasst, besteht die Mikrozensus-Befragung ab 2020 aus einem Kernfragenprogramm, das alle befragten privaten Haushalte beantworten, und weiteren Erhebungsteilen, welche jeweils lediglich eine Unterstichprobe der Haushalte erhält. Das Kernprogramm und die verschiedenen Erhebungsteile werden nicht modular hintereinander erhoben, sondern das resultierende Frageprogramm verzahnt die Inhalte thematisch. Um für Mikrozensusnutzende einen kompakten Gesamtüberblick über alle enthaltenen Fragen und die Struktur der verschiedenen Fragebögen zu geben, wurde ein [Masterfragebogendokument](#) erstellt, welches die Informationen aus allen sieben Fragebögen enthält und angibt, welche der Fragen in welchen Fragebögen enthalten sind.

Zu den Erhebungsteilen SILC und IKT gibt es neben den erhobenen Merkmalen Zielvariablen, die für amtliche Veröffentlichungen verwendet werden und für alle inhaltlichen Auswertungen empfohlen werden.

Weitere Informationen zu den Inhalten und der Methodologie von SILC ist der englischsprachigen Publikation [EU statistics on income and living conditions \(EU-SILC\) methodology](#) und von IKT der ebenfalls englischsprachigen Publikation [ICT usage in households and by individuals](#) zu entnehmen.

---

<sup>5</sup> Im Mikrozensus 2021 sind 6.585 von ca. 11.000 Gemeinden in Deutschland mit mindestens einem Auswahlbezirk vertreten. Bedingt durch die Zufallsstichprobe (1% der Auswahlbezirke) werden Befragungen jedes Jahr nicht in allen (kleineren) Gemeinden durchgeführt. Die Stichprobe und damit die belegten Gemeinden rotieren jedes Jahr. Ein Teil der Auswahlbezirke wird jedes Jahr ersetzt. Mehr Informationen zum Stichprobendesign sind dem [Metadatenreport Teil 1](#) zu entnehmen. Andere regionale Fragen (z. B. Arbeitsstätten) sind gar nicht bis auf Gemeindeebene regionalisierbar.

<sup>6</sup> europäische Statistik der Einkommens- und die Lebensbedingungen (SILC) und ab 2021 europäische Haushaltsbefragung zum privaten Einsatz von Informations- und Kommunikationstechnologien (IKT-Haushalte)

<sup>7</sup> unterjährige Wiederholungsbefragung, Konzeption als Unterstichprobe

[Schlüsselverzeichnis](#)

[Erhebungsunterlagen](#)

## **2.3 Vergleichbarkeit der Merkmale über die Zeit**

Vor dem Hintergrund der in Abschnitt 2.1 skizzierten methodischen Neuerungen des Mikrozensus 2020 und der Einschränkungen in der Qualität des Mikrozensus ab 2020 wird von Veränderungsanalysen zu Vorjahren generell abgeraten, auch wenn die Ausfallquote im Jahr 2021 mit ca. 14% niedriger als im Jahr 2020 mit ca. 35% ausfällt. Alle Neuregelungen zum Mikrozensus ab 2020 können [hier](#) nachgelesen werden.

### **2.3.1 Variablennamen**

Ab dem Erhebungsjahr 2020 wurde die Bezeichnung der Variablen geändert. Variablennamen beginnen nicht mehr mit EF (Eingabefeld) und sind nicht mehr durchgängig nummeriert. Namen von Erhebungsmerkmalen bestehen aus zwei Buchstaben und vier Ziffern. Danach folgt die Kennung der Erhebungsebene (P: Person, H: Haushalt, L: Lebensform). Gegebenenfalls folgen weitere Untergliederungen mit dem Buchstaben U. Die klassischen typisierten Merkmale beginnen mit T. Die SILC-Zielvariablen beginnen mit D (Haushaltsregister) R (Personenregister), H (Haushaltsdaten) oder P (Personendaten). Danach folgen ein weiterer Buchstabe und drei Ziffern. Die Benennung der IKT-Zielvariablen folgt keinem einheitliches Schema. Außerdem gibt es einen anderen Umgang mit Haupt- und Nebenwohnsitzhaushalten als bisher, insbesondere im SILC-Befragungsteil: Die Befragungen für die in 2020 neu integrierten Unterstichprobe SILC und die in 2021 neu integrierte Unterstichprobe IKT erfolgen nur in Haushalten, in denen mindestens eine Person ab 16 Jahren ihren Hauptwohnsitz hat. Weitere Informationen hierzu finden sich in [Hochgürtel und Weinmann 2020; 93](#).

Im Schlüsselverzeichnis sind bei Variablen, bei denen es eine Kontinuität zu Erhebungsjahren vor 2020 gibt, die früheren Variablennamen aufgeführt. Im Detail können sich Frageformulierungen oder Antwortkategorien unterscheiden. Für die Variablen, die auch im Scientific-Use-File enthalten sind, gibt die [Variablen-Zeitpunkte-Matrix](#) in MISSY Hinweise zur zeitlichen Vergleichbarkeit von Variablen des Mikrozensus ab 1973.

### **2.3.2 Missingkodierung**

Im Mikrozensus 2021 gibt es im Vergleich zu 2020 eine Erweiterung der Missingcodes:

Der Code -7 wird für die Auswahlbezirke (AWB) ohne eine Befragung vergeben. Die -7 wird 2021 im Schlüsselverzeichnis nicht explizit ausgewiesen. Weitere Informationen zur Missingcodierung finden sich im Abschnitt 1.1.1.



### **2.3.3 Haushalte und Lebensformen**

Vor dem Hintergrund der genannten Qualitätseinschränkungen (Abschnitt 2.1) gibt es bei der Analyse von (kleinen) Teilpopulationen auch für das Erhebungsjahr 2021 Schwierigkeiten bezüglich der Vergleichbarkeit zu den Vorjahren. Hierfür sind ggf. auch Veränderungen in der Frageformulierung und den Antwortkategorien verantwortlich.

### **2.3.4 Migrationstypisierungen**

Die Zahlen zur Bevölkerung nach Migrationshintergrund ab dem Mikrozensus 2020 sind nur eingeschränkt mit den Vorjahren vergleichbar. Dies gilt insbesondere für einige Teilpopulationen (z. B. als Deutsche Geborene, Eingebürgerte), die umso stärker schwanken, je kleiner diese sind (z. B. Differenzierung nach Geburtsland). Neben den in Abschnitt 2.1 erläuterten generellen Umstellungen und methodisch-technischen Einschränkungen sind Umstellungen in der Erhebung/Frageformulierung sowie in der Methodik der Typisierung des Migrationshintergrundes<sup>8</sup> hierfür ursächlich. Durch die methodische Weiterentwicklung der Typisierung des Migrationshintergrundes werden Personen, die die deutsche Staatsangehörigkeit von Geburt an besitzen besser abgebildet.

Ab 2021 stehen mit TM0020P und TM0021P Variablen zur Verfügung, die ein alternatives Konzept von Einwanderungsstatus und Einwanderungsgeschichte abbilden (vgl. [StBA 2023](#)).

### **2.3.5 Einkommen und Lebensbedingungen (SILC)**

Die früher eigenständige Statistik zu Einkommen und Lebensbedingungen („Leben in Europa“, EU-SILC) ist ab 2020 ein Erhebungsteil des Mikrozensus (SILC). Bis zu 12% der ausgewählten Haushalte werden neben dem Kernprogramm zu Einkommen, Wohnen, Gesundheit und Lebensbedingungen befragt. Durch den Wechsel der Erhebung von einer freiwilligen zu einer in Teilen auskunftspflichtigen Befragung ist ein inhaltlicher Vergleich der Daten der Erhebungsjahre ab 2020 mit Vorjahren nicht möglich. Grundsätzlich besteht im Rahmen des Mikrozensus Auskunftspflicht. Allerdings wird aufgrund der fortgesetzten Freiwilligkeit von Fragen zu Lebensbedingungen ein hoher Anteil an fehlenden Werten (Missings) erzeugt.

### **2.3.6 Informations- und Kommunikationstechnologien (IKT)**

Die früher eigenständige Statistik zu Informations- und Kommunikationstechnologien in privaten Haushalten ist ab dem Erhebungsjahr 2021 als freiwillige Unterstichprobe im Mikrozensus enthalten. Durch die Umstrukturierung der IKT-Erhebung ist ein Vergleich der

---

<sup>8</sup> Personen, die angegeben haben, die deutsche Staatsangehörigkeit als (Spät-)Aussiedler/in erlangt zu haben, aber gleichzeitig vor 1950 nach Deutschland zugewandert sind, werden als Vertriebene und somit als Person ohne Migrationshintergrund umgesetzt. Gleiches gilt auch für die Personen, die zu ihren externen Eltern angegeben haben, dass diese als (Spät-)Aussiedler/in vor 1950 nach Deutschland zugewandert sind. Für alle anderen Personen wird diese Abgrenzung auf Basis des Zuzugsjahres nicht vorgenommen.

Ergebnisse des Jahres 2021 mit den Vorjahren nur eingeschränkt möglich. Weitere Informationen hierzu finden sich [hier](#).

### 2.3.7 Zusatzprogramm 2021

Zusätzlich zum Kernprogramm enthält der Mikrozensus im Rhythmus von jeweils vier Jahren wiederkehrende Zusatzprogramme. Mit der Neukonzeption der LFS als Unterstichprobe, wurden die Zusatzprogramme mit Ausnahme des Zusatzprogramms Wohnen in diese Unterstichprobe ausgelagert. Dies hat aufgrund des damit verbundenen geringeren Auswahlsatzes negative Konsequenzen für die fachliche und regionale Differenzierbarkeit der Zusatzprogramme. Im Erhebungsjahr 2021 wurde das Zusatzprogramm zur Schichtarbeit und zum Gesundheitszustand erhoben.

Gesundheit:

- DF0100P Waren Sie in den letzten 4 Wochen krank?
- DF0200P Wie lange dauert/-e Ihre Krankheit an?
- DF0300P Waren Sie in den letzten 4 Wochen in ärztlicher Behandlung oder in einem Krankenhaus?
- DF0400P Waren Sie in den letzten 4 Wochen unfallverletzt?
- DF0500P Welcher Art war Ihr Unfall?
- DF0600P Wie lange dauert/-e Ihre Unfallverletzung an?
- DF0700P Waren Sie in den letzten 4 Wochen wegen Ihrer Unfallverletzung in ärztlicher Behandlung oder in einem Krankenhaus?
- DF0800P Wie groß sind Sie?
- DF0900P Wie viel wiegen Sie?
- ET0100P Rauchen Sie gegenwärtig?
- ET0200P Haben Sie früher einmal geraucht?
- ET0301P In welchem Alter haben Sie angefangen zu rauchen?
- ET0400P Was rauchen bzw. rauchten Sie überwiegend?
- ET0500P Wie viele Zigaretten rauchen bzw. rauchten Sie täglich?

Schichtarbeit:

- ED3900P Wie viele Stunden haben Sie durchschnittlich zwischen 23 und 6 Uhr gearbeitet?
- ED4100P Haben Sie in der Berichtswoche und den 3 Wochen davor in der Frühschicht gearbeitet?
- ED4200P Haben Sie in der Berichtswoche und den 3 Wochen davor in der Spätschicht gearbeitet?

- ED4300P Haben Sie in der Berichtswoche und den 3 Wochen davor in der Nachtschicht gearbeitet?
- ED4400P Haben Sie in der Berichtswoche und den 3 Wochen davor in der Tagschicht gearbeitet?
- ED4500P Haben Sie in der Berichtswoche und den 3 Wochen davor in einer sonstigen Schicht gearbeitet?

### **2.3.8 Adhoc-Modul 2021**

Im Rahmen der in den Mikrozensus integrierten Arbeitserhebung (LFS) werden in der Regel jährlich Ad-hoc-Module durchgeführt, die dazu dienen, detaillierte Informationen zu politisch relevanten Themen der EU zu sammeln, welche nicht Teil des LFS-Standardfrageprogramms sind. Im Jahr 2021 wurde das Ad-hoc-Modul „Gründe für Migration, Arbeitszeitgestaltung und Arbeitsmarktsituation von Zuwanderern und ihren direkten Nachkommen“ erhoben. Die Fragen des Ad-hoc-Moduls erhält nur der Teil der Haushalte der LFS-Unterstichprobe, die gemäß Rotationsschema erstmals befragt werden. Ihre Beantwortung ist freiwillig.

- EB0210P Alles in allem, wie zufrieden sind Sie mit Ihrer Haupttätigkeit?
- EB0214P Sind Sie in Ihrer derzeitigen Haupttätigkeit schon einmal diskriminiert worden?
- EB0215P Was war der Hauptgrund der Diskriminierung?
- EB0211P Welche Aussage trifft auf Sie zu?
- EB0212P Haben Sie gegen Bezahlung gearbeitet, bevor Sie nach Deutschland gekommen sind?
- EB0400P Diese Frage bezieht sich auf Ihre Fähigkeiten. Unter Fähigkeiten verstehen wir alle Kenntnisse, Fertigkeiten, Kompetenzen und Erfahrungen, die Sie in Ihrer Aus- und Weiterbildung sowie im Arbeitsleben erworben haben. Bitte vergleichen Sie die Anforderungen Ihrer letzten Arbeit vor Ihrer Einreise nach Deutschland mit den Anforderungen Ihrer derzeitigen Arbeitsstelle.
- EV0600P Welche Aussage trifft auf Sie zu?
- EV0601P Denken Sie an die Zeit, bevor Sie nach Deutschland gekommen sind. Wie gut konnten Sie Deutsch sprechen, bevor Sie nach Deutschland gekommen sind?
- EV0602P Und wie gut sprechen Sie die deutsche Sprache heute?
- EV0622P Haben Sie hier in Deutschland einen Deutschkurs besucht?
- EV0623P War dies ein allgemeiner oder berufsbezogener Deutschkurs?
- EV0624P Aus welchem Grund haben Sie keinen Deutschkurs besucht?
- EV0603P Welche Aussage trifft auf Sie zu?

- EV0604P Haben Sie schon einmal in Deutschland gearbeitet?
- EV0605P Haben Sie jemals in Deutschland nach einer Arbeit gesucht?
- EV0606P Hatten Sie Probleme, einen passenden Job zu finden? Der Ihrem höchsten beruflichen Abschluss entspricht?
- EV0607P Aus welchem Grund hatten Sie Probleme, einen passenden Job zu finden?
- EB0700P Wie lange hat es nach der Ankunft in Deutschland gedauert, bis Sie Ihre erste bezahlte Arbeitsstelle gefunden haben?
- EP1301P Welche Aussage trifft auf Ihren höchsten Bildungsabschluss zu?
- EP1313P/EP1313T In welchem (heutigen) Land haben Sie Ihren höchsten Bildungsabschluss erworben?
- EP1320P Haben Sie die Anerkennung dieses Abschlusses hier in Deutschland beantragt?
- EP1321P Wurde dieser Abschluss anerkannt?
- EP1322P Aus welchem Grund haben Sie keine Anerkennung des Abschlusses beantragt?
- EP1500P Hat bzw. hatte Ihr Vater einen allgemeinen Schulabschluss?
- EP1600P Welchen höchsten Abschluss hat bzw. hatte Ihr Vater?
- EP1700P Hat oder hatte Ihr Vater einen beruflichen Ausbildungsabschluss oder einen Hochschulabschluss?
- EP1800P Welchen höchsten Abschluss hat bzw. hatte Ihr Vater?
- EP2000P Hat bzw. hatte Ihre Mutter einen allgemeinen Schulabschluss?
- EP2100P Welchen höchsten Abschluss hat bzw. hatte Ihre Mutter?
- EP2200P Hat bzw. hatte Ihre Mutter einen beruflichen Ausbildungsabschluss oder einen Hochschulabschluss?
- EP2300P Welchen höchsten Abschluss hat bzw. hatte Ihre Mutter?

## 2.4 Eckwerte relevanter Merkmale

Die Standardauswertungen der amtlichen Statistik beziehen sich anders als vor 2020 für alle Erhebungsteile ausschließlich auf Hauptwohnsitzhaushalte<sup>9</sup>. Diese können mit der Variable TH0201H identifiziert werden. Auf Personenebene steht TPBevlkrg zur Verfügung. Diese Variable wird im Rahmen des aktuellen Konzeptes nicht mehr von der amtlichen Statistik verwendet. Die Tabellen im Abschnitt 2.4.1 sind auf Personenebene ausgegeben, die Tabellen im Abschnitt 2.4.2 hingegen auf Haushaltsebene (Haupt- und Nebenwohnsitze).

---

<sup>9</sup> Haushalte, in denen mindestens eine Person ab 16 Jahren am Hauptwohnsitz wohnt

## 2.4.1 Verteilung der Bevölkerung in Hauptwohnsitzhaushalten

|        |            | TPGeschlecht Geschlecht |         |                  |                        |
|--------|------------|-------------------------|---------|------------------|------------------------|
|        |            | Häufigkeit              | Prozent | Gültige Prozente | Kumulierte<br>Prozente |
| Gültig | 1 Männlich | 40.727.217              | 49,5    | 49,5             | 49,5                   |
|        | 2 Weiblich | 41.619.727              | 50,5    | 50,5             | 100,0                  |
| Gesamt |            | 82.346.944              | 100,0   | 100,0            |                        |

|        |   | AB0500P Familienstand |         |                     |                        |
|--------|---|-----------------------|---------|---------------------|------------------------|
|        |   | Häufigkeit            | Prozent | Gültige<br>Prozente | Kumulierte<br>Prozente |
| Gültig | 1 Ledig   | 34.758.123            | 42,2    | 42,2                | 42,2                   |
|        | 2 Verheiratet   | 36.657.567            | 44,5    | 44,5                | 86,7                   |
|        | 3 Verwitwet   | 5.222.579             | 6,3     | 6,3                 | 93,1                   |
|        | 4 Geschieden  | 5.565.960             | 6,8     | 6,8                 | 99,8                   |
|        | 5 Eingetragene<br>Lebenspartnerschaft   | 116.601               | ,1      | ,1                  | 100,0                  |
|        | 6 Eingetragene<br>Lebenspartnerin/<br>eingetragener<br>Lebenspartner verstorben | 9.821                 | ,0      | ,0                  | 100,0                  |
|        | 7 Eingetragene<br>Lebenspartnerschaft<br>aufgehoben                             | 16.293                | ,0      | ,0                  | 100,0                  |
| Gesamt |   | 82.346.944            | 100,0   | 100,0               |                        |

Die hier abgebildeten Tabellen sind in ähnlicher Form dem [Statistischen Bericht zu den Endergebnissen des Mikrozensus](#) zu entnehmen.

## 2.4.2 Privathaushalte am Haupt- und Nebenwohnsitz

|        |                                       | Haushaltsmitglieder2 AnzahlHaushaltsmitglieder |              |                     |                        |
|--------|---------------------------------------|--|--------------|---------------------|------------------------|
|        |                                       | Häufigkeit                                     | Prozent      | Gültige<br>Prozente | Kumulierte<br>Prozente |
| Gültig | 1 Einpersonenhaushalt                 | 17.610.960                                     | 42,2         | 42,2                | 42,2                   |
|        | 2 Zweipersonenhaushalt                | 13.842.812                                     | 33,2         | 33,2                | 75,4                   |
|        | 3 Dreipersonenhaushalt                | 4.879.766                                      | 11,7         | 11,7                | 87,1                   |
|        | 4 Vierpersonenhaushalt                | 3.915.031                                      | 9,4          | 9,4                 | 96,5                   |
|        | 5 Fünf- oder mehr<br>Personenhaushalt | 1.450.064                                      | 3,5          | 3,5                 | 100,0                  |
|        | <b>Gesamt</b>                         | <b>41.698.633</b>                              | <b>100,0</b> | <b>100,0</b>        |                        |

|        |                               | LAND Bundesland   |              |                     |                        |
|--------|-------------------------------|-------------------|--------------|---------------------|------------------------|
|        |                               | Häufigkeit        | Prozent      | Gültige<br>Prozente | Kumulierte<br>Prozente |
| Gültig | 1 Schleswig-Holstein          | 1.496.686         | 3,6          | 3,6                 | 3,6                    |
|        | 2 Hamburg                     | 997.838           | 2,4          | 2,4                 | 6,0                    |
|        | 3 Niedersachsen               | 3.998.550         | 9,6          | 9,6                 | 15,6                   |
|        | 4 Bremen                      | 359.726           | ,9           | ,9                  | 16,4                   |
|        | 5 Nordrhein-Westfalen         | 8.775.824         | 21,0         | 21,0                | 37,5                   |
|        | 6 Hessen                      | 3.144.192         | 7,5          | 7,5                 | 45,0                   |
|        | 7 Rheinland-Pfalz             | 1.974.918         | 4,7          | 4,7                 | 49,8                   |
|        | 8 Baden-Württemberg           | 5.382.206         | 12,9         | 12,9                | 62,7                   |
|        | 9 Bayern                      | 6.559.683         | 15,7         | 15,7                | 78,4                   |
|        | 10 Saarland                   | 474.337           | 1,1          | 1,1                 | 79,5                   |
|        | 11 Berlin                     | 2.045.349         | 4,9          | 4,9                 | 84,4                   |
|        | 12 Brandenburg                | 1.282.240         | 3,1          | 3,1                 | 87,5                   |
|        | 13 Mecklenburg-<br>Vorpommern | 843.159           | 2,0          | 2,0                 | 89,5                   |
|        | 14 Sachsen                    | 2.142.795         | 5,1          | 5,1                 | 94,7                   |
|        | 15 Sachsen-Anhalt             | 1.121.199         | 2,7          | 2,7                 | 97,4                   |
|        | 16 Thüringen                  | 1.099.930         | 2,6          | 2,6                 | 100,0                  |
|        | <b>Gesamt</b>                 | <b>41.698.633</b> | <b>100,0</b> | <b>100,0</b>        |                        |

Im Anhang befindet sich die SPSS-Syntax für das Programm SPSS für die oben aufgeführten Tabellen auf Personen- und Haushaltsebene.

## 2.5 Auswertbare regionale Ebenen

Ein Hochrechnungsfaktor für tiefregionalisierte Ergebnisse ist im Mikrozensus 2021 nicht verfügbar. Bei regionalisierten Auswertungen sind die auch 2021 noch vorhandenen qualitativen Einschränkungen die in Abschnitt 2.1 beschrieben sind zu beachten.

- Bundesebene (NUTS 0)
- Land: Landesebene (NUTS 1)
- Regierungsbezirke/Statistische Bezirke (NUTS 2)
- Nielseengebiete
- Kreis: Kreisebene (nur zu methodischen Zwecken, nicht zur Ergebnisdarstellung empfohlen) (NUTS 3)
- Gemeindeebene (nur zu methodischen Zwecken, nicht zur Ergebnisdarstellung empfohlen<sup>10</sup>) (LAU)
- Anpassungsschichten (tiefste räumliche Ebene, die von der amtlichen Statistik zur Ergebnisdarstellung empfohlen wird) (zudem: regionale Schicht und regionale Untergruppe)<sup>11</sup>
- Raumordnungsregionen
- Planungsregionen
- Arbeitsmarktregionen
- Arbeitsagenturbezirke
- Verflechtungsbereich Zentralität
- Regionstyp
- Verdichtungsräume

Stichtag des Gebietstandes der Regionalkennungen: 31.12.2020

Vereinzelte Auswahlbezirke auf Ebene der Anpassungs- und Regionalschicht falsch zugeordnet. Eine Korrektur erfolgt nicht, da sonst eine Hochrechnung auf die Gesamtbevölkerung der regionalen Ebene nicht mehr möglich wäre.

Label Verflechtungsbereich Zentralität (neu):

---

<sup>10</sup> Es steht kein geeigneter Hochrechnungsfaktor zur Verfügung. Viele kleine Gemeinden entfallen vollständig, da kein Auswahlbezirk in die Stichprobe gelangt ist. Der Klumpeneffekt der Stichprobe hat besonders starke Auswirkungen bei kleineren Gebietseinheiten.

<sup>11</sup> 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. Die regionalen Untergruppen entsprechen häufig den Kreisen.

|       |   |
|-------|---|
| 00010 | Oberzentrum   |
| 00011 | Teil eines Oberzentrums                                       |
| 00020 | Mittelzentrum mit Teilfunktion eines Oberzentrums             |
| 00021 | Teil eines Mittelzentrums mit Teilfunktion eines Oberzentrums |
| 00030 | Mittelzentrum   |
| 00031 | Teil eines Mittelzentrums                                     |
| 00040 | Untzentrum mit Teilfunktion eines Mittelzentrums              |
| 00041 | Teil eines Untzentrums mit Teilfunktion eines Mittelzentrums  |
| 00050 | Untzentrum  |
| 00051 | Teil eines Untzentrums  |
| 00060 | Kleinzentrum mit Teilfunktion eines Untzentrums               |
| 00061 | Teil eines Kleinzentrums mit Teilfunktion eines Untzentrums   |
| 00070 | Kleinzentrum  |
| 00071 | Teil eines Kleinzentrums                                      |
| 00090 | Gemeinde ohne zentralörtliche Funktion                        |

Die Gemeindegrößenklassen werden zum Stichtag 31.12.2021 ausgewiesen und sind daher nicht kompatibel mit den nichtadministrativen Gebietseinheiten, die zum 31.12.2020 ausgewiesen werden.

## **2.6 Produktversionen**

Die Versionsnummer des Datensatzes ist der Variablen DOI zu entnehmen.

10.21242/12211.2021.00.00.1.1.0

- Erstveröffentlichung



## **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 keiner befragten Person 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 Gastwissenschafts Arbeitsplatz oder per Kontrollierter Datenfernverarbeitung erzeugt werden, vor ihrer Freigabe an die Nutzenden 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. Auch die Fachabteilungen der Statistischen Ämter stellen sicher, dass veröffentlichte Ergebnisse absolut anonym sind. Ab

dem Erhebungsjahr 2020 werden von amtlicher Seite keine Ergebnisse aus dem Mikrozensus veröffentlicht, die auf weniger als 71 Befragten basieren<sup>12</sup>. Die amtliche Statistik veröffentlicht weiterhin nur hochgerechnete auf die Tausenderstelle gerundete Stichprobenergebnisse. Auch dies leistet einen Beitrag zur Anonymität der Ergebnisse.

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 der Broschüre [„Übersicht Geheimhaltungsregeln“](#) enthält Informationen darüber, welche Geheimhaltungsregeln auf welche Statistiken anzuwenden sind.

### **3.1.3 Praktische Tipps zur Vermeidung von Geheimhaltungsfällen**

Treten in den erstellten Analysen (Outputs) 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 erhalte ich Zugang zu den MZ-Daten?**

Die Mikrozensus-Daten können gegen Zahlung bei den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder beantragt werden. Die Datennutzung ist

---

<sup>12</sup> Vor 2020 richtete sich die Veröffentlichungsgrenze nach hochgerechneten Zahlen und lag bei 5.000 Fällen.

entgeltpflichtig. Auf den Seiten der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder finden Sie nähere Informationen zum [Datenangebot](#), [Datenzugang](#), [Beantragung](#) und [Bedingungen](#).

### **Wo erhalte ich Auskunft, wenn ich Fragen zum Datenangebot habe?**

Informationen zum Datenangebot und zum Datenzugang sind auf den Seiten der [Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder](#) abrufbar. Weitere Informationen mit einem Fokus auf den MZ-SUF, die aber auch für die Analyse der On-Site-Daten nützlich sind, stehen auf dem [Mikrodaten-Informationssystem \(MISSY\) der GESIS](#) zur Verfügung. Hierzu gehören u.a. Masterfragebogen, Datenhandbuch mit Randauszählungen, Tools zur Umsetzung sozialwissenschaftlicher Konzepte, Variablen-Zeitpunkte-Matrix, und Hinweise und Syntax zur Verknüpfung von MZ-Querschnitterhebungen zu Paneldatensätzen.

Bei weiteren Fragen können sich interessierte Personen und Nutzende des Mikrozensus an das Forschungsdatenzentrum der Statistischen Ämter der Länder – Düsseldorf (insbesondere bei Fragen zum Datenzugang und Datenaufbereitung), das Forschungsdatenzentrum des Statistischen Bundesamtes und an das German Microdata Lab (GML) bei GESIS (insbesondere bei inhaltlichen Fragen und Fragen zum Angebot in MISSY) wenden.

### **Wie werden in den Mikrozensusdaten Haushalte selektiert?**

Für den Mikrozensus 2021 sind bereits IDs zur Personen- (idpers, idpersx), Haushalts- (idhh, idhhx) und Auswahlbezirkskennung (idawb, idawbx) im Datensatz enthalten. In den IDs der längsschnitorientierten Daten, IDs ohne x am Ende, kann es auf Grund von Wiederholungsbefragungen des LFS-Teils zu Dopplungen der IDs kommen. Dopplungsfrei, d.h. eindeutig sind die IDs mit x am Ende, welche für die querschnittsorientierten Daten genutzt werden können. Die Wiederholungsbefragungen sind bei der Erstellung von Jahresergebnissen einzubeziehen. Dies gilt nicht für LFS-Strukturvariablen, die nur in den Fragebögen 2 und 3 erhoben werden. Bei der Verknüpfung über mehrere Erhebungsjahre kann es gewünscht sein, jeden Haushalt nur einmal je Erhebungsjahr zu berücksichtigen. Hierfür wird eine Entfernung der Wiederholungsbefragungen AWBAUSWAHLTEIL = 4 empfohlen.

Die Bildung dieser IDs erfolgt für die jeweiligen Einheiten durch die Aneinanderreihung folgender Variablen:

idpers: LAND AWBNummerFremd HHNummerFremd PERNr

idpersx: LAND AWBNummerFremd TPBerichtsquartal HHNummerFremd PERNr

idhh: LAND AWBNummerFremd HHNummerFremd

idhxx: LAND AWBNummerFremd TPBerichtsquartal HHNummerFremd

idawb: LAND AWBNummerFremd

idawbx: LAND AWBNummerFremd TPBerichtsquartal

Leerzeichen werden bei der Bildung der Identifikatoren durch Nullen ersetzt.

### **Wie können einer befragten Person Informationen zur im Haushalt lebenden Mutter und zum im Haushalt lebenden Vater zugeordnet werden?**

Unter Verwendung der Haushaltsidentifikationsnummer (idhxx) kann über die Merkmale TL0702P und TL0802P die Personennummer der Mutter bzw. des Vaters ermittelt werden. Der zugehörige Einzeldatensatz der Mutter/des Vaters lässt sich über die Personennummer PERNr finden. Die Informationen können dann satzübergreifend der jeweiligen Referenzperson zugespielt werden.

### **Wie ist das Rotationsschema der Haushalte aus der LFS-Substichprobe aufgebaut?**

Um die unterjährigen Veränderungen auf dem Arbeitsmarkt besser analysieren zu können, wurde das Rotationsschema für die MZ-Erhebung ab 2020 angepasst. Im Gegensatz zum Kern- und SILC-Programm, welche einmal jährlich abgefragt werden, rotiert der LFS-Teil in kürzeren Abständen. Haushalte, die für den LFS-Teil ausgewählt wurden, werden im 2-(2)-2-Schema befragt. Das bedeutet, dass die Haushalte in zwei aufeinanderfolgenden Quartalen befragt werden, anschließend zwei Quartale pausieren und dann wieder zwei Quartale in Folge befragt werden. Verwendet werden dabei die Fragebögen in der Reihenfolge 3-4-2-4. Eine genauere Beschreibung dazu ist in den [Neuregelungen des Mikrozensus](#) zu finden.

### **Wann wird welcher Hochrechnungsfaktor verwendet?**

Aufgrund der verschiedenen Substichproben, die seit dem Berichtsjahr 2020 in den Daten vorliegen, stehen auch für die Erhebung 2021 erneut mehr Hochrechnungsfaktoren zur Verfügung als zuvor. Aus der Übersicht in Tabelle 3 lässt sich erkennen, für die Analyse welcher Substichproben welche Hochrechnungsfaktoren vorgesehen sind. Eine detaillierte Beschreibung der Hochrechnungsfaktoren ist im [Metadatenreport Teil I Statistik](#) in Kapitel 2.6 zu finden. Werden Analysen mit Variablen aus verschiedenen Substichproben durchgeführt, sollte der Hochrechnungsfaktor verwendet werden, der für die Analyse der kleinsten Substichprobe ausgelegt ist. Werden bspw. Analysen mit Variablen des Kernprogramms und des LFS-Teils durchgeführt, so sollte der LFS-Hochrechnungsfaktor verwendet werden.

- Für die Merkmale aus dem Kernprogramm wird für Hochrechnungen auf das Jahr der Standardhochrechnungsfaktor HR000JJ verwendet. Quartalsergebnisse werden mit dem Hochrechnungsfaktor HR000QQ berechnet. Dafür können über die Variable TPBerichtsquartal einzelne Berichtsquartale selektiert werden. Darüber hinaus besteht mit HR000JQ die Möglichkeit, die Daten auf den Durchschnitt der vier Quartale hochzurechnen.
- Strukturmerkmale aus der LFS-Substichprobe (Fragebögen 2 und 3) werden für Jahresergebnisse mit dem Merkmal HR100JJ hochgerechnet. Entsprechend dem Kernprogramm stehen auch für die LFS-Merkmale Hochrechnungsfaktoren für die Quartale (HR100QQ) und den Quartalsdurchschnitt (HR100JQ) zur Verfügung. Hierbei ist vor allem zu beachten, dass Haushalte im LFS-Programm mehrfach im Jahr befragt werden (siehe oben). Bei der Auswertung von Merkmalen, die unterjährig wiederholt erhoben werden (Fragebögen 2, 3 UND 4) empfiehlt sich, mit HR100JQ hochzurechnen, sodass nicht nur das Ergebnis der Erstbefragung in die Auswertung eingeht. Weitere Hochrechnungsfaktoren für LFS-Variablen sind HR100MO für Merkmale des LFS-Ad-hoc-Moduls und HR100BH für Auswertungen, die Menschen mit Behinderung betreffen.
- Für die SILC-Substichprobe wird der Hochrechnungsfaktor HR200JJ verwendet. Werden Auswertungen für Kinder unter 13 Jahren erstellt, so kann der neue Hochrechnungsfaktor HR200KI verwendet werden. Ebenfalls in 2021 neu hinzugekommen ist der Hochrechnungsfaktor HR200L2. Dieser kann für Längsschnittanalysen genutzt werden.
- Die IKT Merkmale können mit den Hochrechnungsfaktoren HR300PN und HR300HH hochgerechnet werden. Dabei ist HR300PN für Auswertungen auf Personenebene und HR300HH für Auswertungen auf Haushaltsebene zu verwenden.

| Hochrechnungs-faktor | Stich-probe | Summe      | Anmerkung  |
|----------------------|-------------|------------|--|
| HR000JJ              | Kern        | 84.603.477 | Zur Verwendung bei Berechnungen mit Jahresergebnissen. Abweichungen zu den anderen Summen folgen aus dem 2-Stufen-Vorgehen zur Bestimmung der Hochrechnungsfaktoren. |
| HR000JQ              | Kern        | 82.238.398 | Zur Verwendung bei Berechnungen mit Quartalsdurchschnitt zur Jahresauswertung.   |

|         |      |                           |   |
|---------|------|---------------------------|---|
| HR000QQ | Kern | 328.953.591 <sup>13</sup> | Zur Verwendung bei Berechnungen mit Quartalsergebnissen. Werden nicht wie hier die Jahresergebnisse verwendet, sondern nur die Daten eines Quartals summieren sich die Werte auf den Quartalsdurchschnitt von 82.238.398. |
| HR100JJ | LFS  | 82.227.687                | Zur Verwendung bei Berechnung mit Jahresergebnissen.  |
| HR100JQ | LFS  | 82.227.687                | Zur Verwendung bei Berechnungen mit Quartalsdurchschnitt zur Jahresauswertung.  |
| HR100QQ | LFS  | 328.910.748 <sup>13</sup> | Zur Verwendung bei Berechnungen mit Quartalsergebnissen. Werden nicht wie hier die Jahresergebnisse verwendet, sondern nur die Daten eines Quartals summieren sich die Werte auf 82.227.687.                              |
| HR100MO | LFS  | 62.060.457                | Zur Verwendung bei Berechnungen von Variablen des LFS-Ad-Hoc-Moduls.  |
| HR100BH | LFS  | 10.325.782                | Sonderhochrechnungsfaktor für Berechnungen, die Menschen mit Behinderungen betreffen  |
| HR200JJ | SILC | 82.346.944                | Zur Verwendung bei Berechnungen des SILC-Querschnitts. Es empfiehlt sich, SILC-Hochrechnungsfaktoren auf die Zielvariablen anzuwenden.  |
| HR200KI | SILC | 10.043.718                | Zur Verwendung bei Berechnungen die Kinder unter 13 betreffen. Es empfiehlt sich, SILC-Hochrechnungsfaktoren auf die Zielvariablen anzuwenden   |
| HR200L2 | SILC | 82.260.703                | Zur Verwendung bei Längsschnittanalysen. Es empfiehlt sich, SILC-Hochrechnungsfaktoren auf die Zielvariablen anzuwenden   |
| HR300PN | IKT  | 60.667.535                | Zur Verwendung bei Auswertungen, die IKT auf Personenebene betreffen.   |
| HR300HH | IKT  | 56.793.209                | Zur Verwendung bei Auswertungen, die IKT auf Haushaltsebene betreffen. Wird der Hochrechnungsfaktor auf Haushaltsebene angewendet, dann ergibt sich eine Summe von 42.685.780.  |

*Tabelle 1: Hochrechnungsfaktoren mit Gesamtbevölkerungszahlen der hochgerechneten Jahresergebnisse. Angaben sind auf ganze Zahlen gerundet.*

<sup>13</sup> Da jedes Quartal einzeln berechnet wird, entspricht der über das ganze Jahr summierte Wert dem Vierfachen der Bevölkerungszahl.

## **Wieso unterscheiden sich die hochgerechneten Gesamtfallzahlen zwischen verschiedenen Variablen bzw. zwischen verschiedenen Hochrechnungsfaktoren?**

Die Unterschiede in den hochgerechneten Gesamtfallzahlen kommen aus verschiedenen Gründen zustande. Differenzen zwischen Jahres- und Quartalshochrechnungsfaktoren können sich daraus ergeben, dass in die Jahreshochrechnungsfaktoren tendenziell mehr für die Bevölkerung bekannte Eckwerte (z. B. Altersgruppen, Geschlecht, Staatsangehörigkeit sowie regionale Verteilungen) eingehen, als in die Quartalshochrechnungsfaktoren. Genauere Informationen finden sich in [Metadatenreport Teil I Statistik](#) in Abschnitt 2.6 sowie in [Schmidt und Stein 2021](#).

Abweichungen der LFS-Stichprobe kommen durch europäische Konsistenzanforderungen zustande. Diese sehen vor, dass Quartals- und Jahresergebnisse der LFS-Unterstichprobe miteinander konsistent sein müssen. Diese Anforderungen wurden gegenüber der nationalen Anforderung (Ergebniskonsistenz zwischen den einzelnen Unterstichproben) präferiert umgesetzt. Um die europäischen Anforderungen zu erfüllen, werden für amtliche Veröffentlichungen die LFS-Strukturmerkmale (sprich LFS-Jahresergebnisse) am MZ-Kern-Quartalsdurchschnitt hochgerechnet. Das MZ-Kern-Jahresergebnis entspricht nicht dem Quartalsdurchschnitt, sondern wurde eigenständig hochgerechnet.

## **Warum sind einige Variablen weniger stark besetzt als andere?**

Mit Umsetzung der Neuerungen ab dem Mikrozensus 2020 wurden Unterstichproben eingeführt, die zur Folge haben, dass nicht alle Fragen allen zu befragenden Haushalten gestellt werden. Für den Mikrozensus werden 1% der deutschen Bevölkerung befragt. Allen Befragten in Privathaushalten werden unabhängig von der Zugehörigkeit zu einer Unterstichprobe die Fragen des Mikrozensus-Kernprogramms gestellt. Gemäß Mikrozensusgesetz werden in bis zu 45% der Auswahlbezirke Befragungen zur Arbeitsmarktbeteiligung (LFS) durchgeführt. Der Anteil zu Einkommen und Lebensbedingungen (SILC) macht bis zu 12% aus und der Anteil zu Informations- und Kommunikationstechnologien (IKT) macht bis zu 3,5% die kleinste Unterstichprobe aus. Die realisierten Auswahlsätze können von diesen Angaben abweichen. Die genannten Unterstichproben sind überschneidungsfrei. LFS-, SILC- und IKT-Variablen sind entsprechend den oben genannten Auswahlsätze weniger häufig belegt als die Kern-Variablen. Ebenso ist zu beachten, dass einige Fragen freiwillig sind und die Anzahl der Beobachtungen deshalb geringer ausfallen kann. Die Einschränkungen in der Qualität des Mikrozensus in den

Erhebungsjahren 2020 und 2021, welche in Kapitel 2.1 näher erläutert werden, tragen ebenfalls zu Fehlwerten bei der Erhebung 2021 bei.

Der Artikel [„Die Neuerungen des Mikrozensus ab 2020“](#) enthält weitere Informationen zu den Erhebungsteilen. Informationen zu den Unterstichproben, sowie zur Anzahl an freiwilligen und auskunftspflichtigen Fragen je Befragungsprogramm sind dem [Qualitätsbericht zum Mikrozensus 2021<sup>14</sup>](#) zu entnehmen.

### **Warum beträgt die Anzahl der Befragten mehr als 1% (etwa 832 000) der deutschen Bevölkerung?**

Die tatsächliche Menge der Befragungen ist größer als die Stichprobengröße von 1%, da 7/9 des LFS-Anteils aufgrund der unterjährigen Wiederholungsbefragung zwei Mal im Erhebungsjahr befragt wird.

Mehr Informationen zur Stichprobenziehung auf Ebene der Auswahlbezirke und der Anzahl an Befragungen können dem [Qualitätsbericht des Mikrozensus 2021<sup>14</sup>](#) unter Kapitel 3 entnommen werden.

### **Was muss beim Abgleich der Daten mit den Tabellen der Fachserien und anderen amtlichen Veröffentlichungen beachtet werden?**

Um die Haushaltstabellen und Tabellen der Lebensformen nachzubilden, müssen die Daten nach Hauptwohnsitzhaushalten gefiltert werden. In der [Fachserie 1 Reihe 2.2](#) und im [Statistischen Bericht zu den Endergebnissen des Mikrozensus](#) werden nur Hauptwohnsitzhaushalte betrachtet. Diese können mittels der Variable TH0201H selektiert werden. In den Erhebungsteilen SILC, LFS und IKT werden nur Hauptwohnsitzhaushalte befragt.

Für Tabellen, in denen das Alter dargestellt wird, wird die Altersvariable TPALTER\_1 verwendet. Für Veröffentlichungen aus SILC wird RB081 oder RB082 verwendet.

Für das Geschlecht wird TPGeschlecht verwendet. Bei SILC wird RB090 verwendet. Personen mit diversem oder ohne Eintrag des Geschlechts in das Personenstandsregister sind in diesen Variablen mit einer Wahrscheinlichkeit von je 50% männlich oder weiblich zugeordnet. Im Erhebungsmerkmal AB0200P ist das Geschlecht mit vier möglichen Merkmalsausprägungen ausgewiesen. Die Ausprägungen 3 und 4 werden jedoch deutlich überschätzt und weisen weit

---

<sup>14</sup> Die Fallzahlen im Qualitätsbericht beruhen nicht auf den Endergebnissen und weichen deshalb von diesen ab.



höhere Anteile als in den Personenstandsregistern auf. Das Erhebungsmerkmal selbst wird nicht für amtliche Veröffentlichungen verwendet.

Die Erhebungsmerkmale aus SILC sind unplausibilisiert und werden nicht unmittelbar für Veröffentlichungen verwendet. Für die Lieferung an die EU und amtliche Veröffentlichungen werden die SILC-Zielvariablen DB100, RB031 bis RG\_11, HB100 bis HY170N, PB120 bis PY143G, RCH010 bis HD240 und HK010 bis PK030 verwendet. Diese werden auch für die Nutzung in den Forschungsdatenzentren empfohlen. Um veröffentlichte Tabellen nachbilden zu können, müssen die SILC-Zielvariablen verwendet werden.

### **Auf welchen regionalen Ebenen lassen sich die Daten sinnvoll auswerten?**

Die Variablen aus dem Kern- und LFS-Material können von NUTS-0 bis zur regionalen Anpassungsschicht ausgewertet werden. Die Unterstichproben SILC und IKT hingegen sind so konzipiert, dass empfohlen wird, diese nur auf NUTS-0-Ebene (Deutschland) auszuwerten. Werden Variablen dieser Unterstichproben auf feineren regionalen Ebenen ausgewertet, kann es zu Verzerrungen kommen.

### **Warum gibt es inkonsistente Personeninformationen zwischen verschiedenen Erhebungsteilen?**

Die verschiedenen Erhebungsteile haben unterschiedliche Abzugsdaten. Der Datenabzug aus SILC beispielsweise findet aufgrund von EU-Lieferverpflichtungen deutlich früher statt, als der Abzug des Kernmaterials. Daher kann es sein, dass im Kernmaterial Korrekturen von Personenmerkmalen, wie TPALTER\_1 stattfinden, die in den entsprechenden SILC-Zielvariablen (z.B. RB081) nicht mehr nachträglich korrigiert werden. So kann es vereinzelt auch dazu kommen, dass eine Person in SILC einem anderen Haushalt (HB030) zugeordnet wird, als im Kernmaterial (idhh).

### **3.3 Verfügbare Tools**

Syntaxen zur Umsetzung der sozialwissenschaftlichen Konzepte ESeG - European Socioeconomic Groups, ESeC - European Socioeconomic Classification, ISEI - Internationaler Sozioökonomischer Index des beruflichen Status und CASMIN-Bildungsklassifikation stehen für die Programmpakete SPSS und Stata auf dem Mikrodaten-Informationssystem (MISSY) der GESIS zur Verfügung. Für das Erhebungsjahr 2020 werden keine Syntaxen zur Umsetzung von ESeG, ESeC und ISEI angeboten, da die amtliche Statistik aufgrund von Qualitätsbedenken von einer Veröffentlichung der ISCO-Variablen im MZ-SUF 2020 absieht,

die als Grundlage für die Operationalisierung der o.a. Tools dienen. Im Datenmaterial zum Erhebungsjahr 2021 sind die ISCO-Variablen ISCO4D und ISCOPR3D wieder enthalten.

# Anhang

## SPSS-Syntax zu Abschnitt 2.4

\*Filtern nach Hauptwohnsitzhaushalten\*.

if (TH0201H=1) Hauptwohnsitz=1.

filter by Hauptwohnsitz.

EXECUTE.

WEIGHT HR000JJ.

\*\*\*\*\*

\*Häufigkeitstabelle für Geschlecht\*.

VARIABLE LABELS TPGeschlecht 'Geschlecht'.

VALUE LABELS TPGeschlecht

1 'Männlich'

2 'Weiblich'.

FREQUENCIES TPGeschlecht.

\*\*\*\*\*

\*Häufigkeitstabelle für Familienstand\*.

VARIABLE LABELS AB0500P 'Familienstand'.

FREQUENCIES AB0500P.

\*\*\*\*\*

\*

filter off.

use all.

EXECUTE.

COMPUTE Haushaltsmitglieder2 = 0.

if (NpersHH =1) Haushaltsmitglieder2=1.

if (NpersHH =2) Haushaltsmitglieder2=2.

if (NpersHH =3) Haushaltsmitglieder2=3.

if (NpersHH =4) Haushaltsmitglieder2=4.

if (NpersHH >4) Haushaltsmitglieder2=5.

FREQUENCIES Haushaltsmitglieder2.

EXECUTE.

ALTER TYPE Haushaltmitglieder2 (F1.0).

VARIABLE LABELS Haushaltmitglieder2 'AnzahlHaushaltsmitglieder'.

VALUE LABELS

Haushaltsmitglieder2

1 'Einpersonenhaushalt'

2 'Zweipersonenhaushalt'

3 'Dreipersonenhaushalt'

4 'Vierpersonenhaushalt'

5 'Fünf- oder mehr Personenhaushalt'.

EXECUTE.

\*Datensatz auf Haushaltsebene und Privathaushalte reduzieren.

if (TH0602P=1) and (TPPrivatHH=1) Haushalt=1.

filter by Haushalt.

EXECUTE.

\*Haushaltsgröße.

FREQUENCIES Haushaltmitglieder2.

EXECUTE.

\*Häufigkeitstabelle für Bundesland\*.

VARIABLE LABELS Land 'Bundesland'.

FREQUENCIES Land.

EXECUTE.

Statistische Ämter des Bundes und der Länder,  
Metadatenreport – Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus 2021 am  
Gastwissenschaftsarbeitsplatz sowie per kontrollierter Datenfernverarbeitung (sog. On-Site-Zugangswege/-  
Nutzungen)

Fotorechte Umschlag: ©artSILENCEcom – Fotolia.com