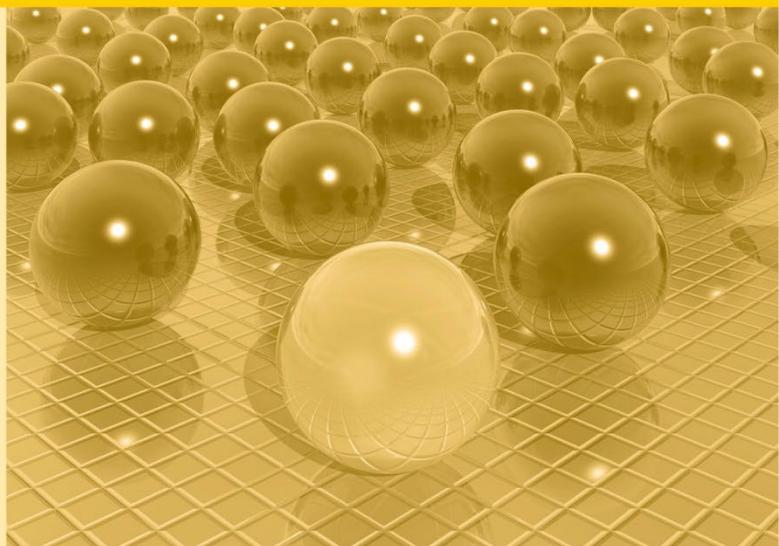


Metadatenreport



Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus 2020 am Gastwissenschaftsarbetsplatz sowie per kontrollierter Datenfernverarbeitung (sog. On-Site-Zugangswege/-Nutzungen)

DOI: 10.21242/12211.2020.00.00.1.1.0

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
E-Mail: 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

Informationen zum Datenangebot:

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

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

Erscheinungsfolge: unregelmäßig
Erschienen im Mai 2023

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

© Information und Technik Nordrhein-Westfalen, Düsseldorf, 2023
(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 2020 am Gastwissenschaftsarbetsplatz sowie per kontrollierter Datenfernverarbeitung (sog. On-Site-Zugangswege/-Nutzungen) (EVAS-Nummer: 12211, 12231, 12241, 12251). Version 1. DOI: 10.21242/12211.2020.00.00.1.1.0. Düsseldorf 2023

Metadatenreport

Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus 2020 am Gastwissenschaftsarbeitsplatz sowie per kontrollierter Datenfernverarbeitung (sog. On-Site-Zugangswege/-Nutzungen)

DOI: 10.21242/12211.2020.00.00.1.1.0

Version 1

Inhalt

Einleitung	3
1. Datenaufbereitung in den FDZ	4
1.1 Datenaufbereitung	4
1.1.1 Missingkodierung.....	4
1.1.2 Hochrechnungs-/Gewichtungsvariablen.....	5
1.1.3 SILC Indizes	5
1.2 Anonymisierungsmaßnahmen	6
1.3 Methodik der Verknüpfung	6
2. Produkt	7
2.1 Hinweise zur Qualität des Mikrozensus 2020	7
2.2 Merkmale und Merkmalsbeschreibung	8
2.2.1 Merkmalsdefinitionen.....	8
2.2.2 Datensatzbeschreibung / Schlüsselverzeichnis	9
2.3 Vergleichbarkeit der Merkmale über die Zeit	10
2.3.1 Variablennamen	10
2.3.2 Missingkodierung	10
2.3.3 Haushalte und Lebensformen.....	11
2.3.4 Migrationstypisierungen.....	11
2.3.5 Variablen zur Arbeitssituation	12
2.3.6 Einkommen und Lebensbedingungen (SILC)	12
2.3.7 Zusatzprogramm 2020	12
2.3.8 Adhoc-Modul 2020	13
2.4 Eckwerte relevanter Merkmale und Merkmalskombinationen	14
2.5 Auswertbare regionale Ebenen	16
2.6 Produktversionen	17
3. Praktische Hinweise	18
3.1 Hinweise zur Geheimhaltung	18

3.1.1 Gesetzliche Grundlagen der statistischen Geheimhaltung.....	18
3.1.2 Geheimhaltung von Ergebnissen	18
3.1.3 Praktische Tipps zur Vermeidung von Geheimhaltungsfällen .	19
3.2 FAQ	19
3.3 Verfügbare Tools.....	25
Anhang.....	26

Einleitung

Dieser Metadatenreport soll Forschenden dabei helfen, die Daten des Mikrozensus On-Site-Materials 2020¹ sachgerecht auszuwerten. Bei den On-Site-Daten handelt es sich um die Daten, die zur Nutzung am Gastwissenschaftsarbeitsplatz (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 Seiten 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 Gastwissenschaftsarbeitsplatz oder über eine kontrollierte Datenfernverarbeitung, gibt es auch die Möglichkeit, Scientific-Use-Files am eigenen Arbeitsplatz in der wissenschaftlichen Einrichtung 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 ([Häufige Fragen zum Mikrozensus](#)).

¹ 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.

Im Erhebungsteil SILC wurden unplausible Angaben aus den SILC-Variablen und SILC-Hochrechnungsfaktoren entfernt. Außerdem sind in den gesamten Mikrozensus nur vollständige Haushalte 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 beim selben Befragten auf, wird der Missingcode eingetragen, der in Tabelle 1 als erstes referenziert ist. Wird beispielsweise eine Person unter 15 Jahren aus einer Gemeinschaftsunterkunft (GU) befragt, 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.

Missingcode	Beschreibung
-1	Gemeinschaftsunterkunft (GU)
-4	Nicht in der Substichprobe des Ad-hoc Moduls bzw. <Benennung der jeweiligen Unterstichprobe bzw. des Zusatzprogramms>
-3	Person unter 15 (bzw. 16) Jahre ²

² Der Code -3 wird auch dort verwendet, wo die Altersgrenze nicht 15 Jahre, sondern 16 Jahre ist (z.B. im Erhebungsteil SILC).

-2	Nichterwerbstätige
-6	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
-5	Spezifischer Filterausfall, Ausfallgrund wird variablenspezifisch benannt
-9	Filterfehler oder technisch zugelassene Fehlwerte bei Fragen mit Auskunftspflicht, Personen ohne substichprobenspezifischen Hochrechnungsfaktor bei SILC- oder LFS-Unterstichprobe

Tabelle 1: Ausfallkennungen

Im Mikrozensus On-Site-Material 2018 und 2019 wurden zusätzlich der Missingcode -8 für Leerstand und ausgefallene Privathaushalte sowie der Missingcode -7 für Auswahlbezirke (AWBs) ohne potenziell zu befragende Haushalte vergeben. Im Mikrozensus 2020 werden diese beiden Ausfälle nicht programmiert, da im Datenmaterial 2020 nur freigegebene Haushalte (kein Leerstand oder unbewohnte Wohnungen und nur AWBs mit zu befragenden Haushalten) vorhanden sind.

Item Nonresponse bei freiwilligen Fragen wurden mit 9, 99, 999 oder Ähnlichem codiert.

1.1.2 Hochrechnungs-/Gewichtungsvariablen

Fälle ohne einen Kernhochrechnungsfaktor werden aus dem Datenmaterial entfernt. 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 zur Verfügung:

- Index_Wohnsituation

- Index_Wohnumgebung
- Index_Finanz_Situation
- Index_Sozialkapital
- Index_Lebenssituation
- Index_Kinderbetreuung_Std
- Index_Gesundheitszustand

Die zugrundeliegenden Erhebungsmerkmale für die beiden Indizes Index_Kinderbetreuung_Std und Index_Gesundheitszustand weisen im aktuellen Berichtsjahr 2020 allerdings hohe Ausfallquoten auf.

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, so dass anhand der Position des einzelnen Falls im Datenmaterial kein Rückschluss auf einzelne Personen möglich ist.

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

1.3 Methodik der Verknüpfung

Der Mikrozensus stellt kein bereits verknüpftes Produkt dar, allerdings besteht die Möglichkeit, 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, idhxx, idpersx) sind die Wiederholungsbefragungen bereits mit einer abweichenden ID versehen. Nähere Informationen zur Panelverknüpfung sind [Brockner und Mühlenfeld 2020](#) zu entnehmen. Das FDZ-Arbeitspapier bezieht sich auf die Erhebungsjahre 2012 bis 2015. Die Erkenntnisse sind aber auch auf die Jahre 2016 bis 2019 und in Teilen auch auf die Jahre ab 2020 zu übertragen.

2. Produkt

2.1 Hinweise zur Qualität des Mikrozensus 2020

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. Zentrale Einschränkungen, die sich auf die Datenqualität des Mikrozensus auswirken, werden im Folgenden aufgeführt.

Im Jahr 2020 war es nicht möglich, Erhebungsbeauftragte im gewohnten Maße für die Befragungen einzusetzen. Ausgangs- und Kontaktbeschränkungen infolge der Corona-Pandemie verhinderten überwiegend die gewohnte persönliche Vor-Ort-Befragung der Haushalte (CAPI) und erschwerten zusätzlich die Durchführung der (Vor-)Begehungen von Gebäuden, die zur Stichprobenkonkretisierung notwendig sind. Zudem wurde das Mahnwesen in vielen Ländern nicht oder nur teilweise umgesetzt.

Hieraus resultierte eine vom MZ sonst unbekannt hohe Ausfallquote. Im Bundesschnitt gibt es einen Ausfall von rund 35% bei den Endergebnissen. Die Ausfallquoten für die freiwilligen Fragen, z.B. zu den Lebensbedingungen aus der neu integrierten SILC-Unterstichprobe, sind in der Regel noch deutlich höher. Die Antwortausfälle können nicht als zufällig angenommen werden. Zudem sind sie regional, u.a. zwischen den Bundesländern, und zeitlich sehr unterschiedlich verteilt. Informationen zu länderspezifischen Besonderheiten für das Jahr 2020 finden sich auf den Web-Seiten der Statistischen Ämter der Länder zum Mikrozensus.

Die Erhebungssituation 2020 hat zur Folge, dass Auswertungen – insbesondere in Kombination von fachlicher und regionaler Tiefe – nicht durchgängig die vom Mikrozensus sonst gewohnte Qualität aufweisen. Die Ergebnisse ab 2020 sind nur eingeschränkt mit den Ergebnissen der Vorjahre vergleichbar. Darüber hinaus gibt es 2020 eine Reihe von Auffälligkeiten in verschiedenen Themenbereichen zu beachten, z.B. zur Abbildung gleichgeschlechtlicher Paare, des Migrationshintergrundes oder von atypisch Beschäftigten. Hinweise hierzu finden sich auf den Seiten des Statistischen Bundesamts zur [Neuregelung des Mikrozensus ab 2020](#) und in den offiziellen [Qualitätsberichten zum Mikrozensus](#) und der [Unterstichprobe SILC](#).

Die amtliche Statistik hat sich vor dem Hintergrund der genannten Schwierigkeiten darauf verständigt, nur stark eingeschränkt Ergebnisse unterhalb der Landesebene zu veröffentlichen. Vor dem Hintergrund der genannten Besonderheiten des Erhebungsjahres sollten keine Veränderungsanalysen zu Vorjahren durchgeführt werden und auch keine kausalen Interpretationen bezüglich möglicher Effekte der Corona-Pandemie abgeleitet

werden. Für wissenschaftliche Projekte werden die Daten in der bisherigen regionalen und inhaltlichen Tiefe bereitgestellt. Dies dient insbesondere dazu, eine Verknüpfung mit externen Merkmalen zu ermöglichen.

Es ist zu beachten:

- Vor dem Hintergrund der skizzierten Einschränkungen werden Veränderungsanalysen zu Vorjahren sowie Analysen unterhalb der Landesebene (NUTS-1) für das Jahr 2020 nicht empfohlen.
- Auch ein Vergleich zwischen den Bundesländern ist aufgrund der teils großen Unterschiede in der Erhebungspraxis für dieses Berichtsjahr nicht sinnvoll.
- Die neue Struktur des Mikrozensus (verschiedene Unterstichproben und damit einhergehende veränderte Hochrechnungsverfahren, unterjährige Rotation einzelner Teile, verändertes Fragenprogramm) und der erstmalige Einsatz von Online-Fragebögen (CAWI) zur Datenerhebung sind bei der Interpretation von Ergebnissen ebenfalls zu berücksichtigen.

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](https://www.destatis.de/DE/Navigation/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-Fassung2020/Systematik-Verzeichnisse/Systematik-Verzeichnisse-Nav.html)
- 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&&A

[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 2020 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 Gebietsystematik (außereuropäische Staaten zum Teil zusammengefasst siehe Schlüsselverzeichnisse):
<https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/staatsangehoerigkeit-gebietsschluessel.html>
- Amtlicher Gemeindeschlüssel (AGS) / Regionalschlüssel (RS)³:
<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 des Fragebogens wiedergeben, ist die zugehörige Fragennummer in der Spalte „Fragebogennummer“ enthalten.

Aufgrund der 2020 eingeführten Neukonzeption des Mikrozensus, welche die Integration weiterer Haushaltserhebungen umfasst (die europäische Arbeitskräfteerhebung (LFS), die europäische Statistik der Einkommens- und die Lebensbedingungen (SILC) und ab 2021 die europäische Haushaltsbefragung zum privaten Einsatz von Informations- und Kommunikationstechnologien (IKT-Haushalte)), besteht die Mikrozensus-Befragung ab 2020 aus einem verkürzten 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 den Mikrozensusnutzenden einen kompakten Gesamtüberblick über alle enthaltenen Fragen und die Struktur der verschiedenen Fragebögen zu geben, wurde ein [Masterfragebogendokument](#) erstellt, welches die Informationen aus allen fünf Fragebögen enthält und angibt, welche der Fragen in welchen Fragebögen enthalten sind.

³ Im Mikrozensus sind jährlich Auswahlbezirke aus etwa 7.200 von 11.000 Gemeinden in Deutschland vertreten. Andere regionale Fragen (z. B. Arbeitsstätten) sind nicht bis auf Gemeindeebene regionalisierbar.

Zum Erhebungsteil SILC gibt es neben den unplausibilisierten Erhebungsmerkmalen Zielvariablen, die für amtliche Veröffentlichungen verwendet werden und für alle inhaltlichen Auswertungen herangezogen werden sollen.

Weitere Informationen zu den Inhalten und der Methodologie von SILC ist der englischsprachigen Publikation [EU statistics on income and living conditions \(EU-SILC\) methodology](#) zu entnehmen.

[Schlüsselverzeichnis](#)

[Erhebungsunterlagen](#)

2.3 Vergleichbarkeit der Merkmale über die Zeit

Vor dem Hintergrund der in Abschnitt 2.1 skizzierten methodischen Neuerungen und der Einschränkungen in der Qualität des Mikrozensus 2020 wird von Veränderungsanalysen zu Vorjahren generell abgeraten. Alle Neuregelungen zum Mikrozensus 2020 können [hier](#) nachgelesen werden.

2.3.1 Variablennamen

Eine weitere Neuerung im Erhebungsjahr 2020 ist die veränderte Bezeichnung der Variablen. Variablennamen beginnen nicht mehr mit EF (Eingabefeld) und sind nicht mehr durchnummeriert. 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. Typisierte Merkmale beginnen mit T. Außerdem gibt es einen anderen Umgang mit Haupt- und Nebenwohnsitzhaushalten als bisher, insbesondere im SILC-Befragungsteil: Die Befragungen für die neu integrierten Unterstichproben SILC und ab 2021 IKT erfolgen nur in Haushalten, in denen mindestens eine Person über 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 Vorjahren 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

Die Struktur der Missings hat sich im Vergleich zu 2019 ebenfalls verändert:

Die Missingcodes -7 und -8 finden im neuen Mikrozensus 2020 keine Anwendung mehr, da nur noch realisierte Interviews in den Datensatz miteinbezogen wurden.

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

Da es keine Jahresüberhänge mehr gibt, wird der Code -6 ab 2020 für Filter aufgrund von Nebenwohnsitzen verwendet. Personenfragen des SILC-Erhebungsteils (im Fragebogen 5) werden nur Personen am Hauptwohnsitz gestellt. Haushalten, in denen nicht mindestens eine Person ab 16 Jahren den Hauptwohnsitz hat beantworten die Fragen des SILC-Erhebungsteils einschließlich der Haushaltsfragen dementsprechend nicht..

Missingcodes werden hierarchisch vergeben. Weitere Informationen zur Missingkodierung 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 Schwierigkeiten in der Vergleichbarkeit zu den Vorjahren. So hat sich zum Beispiel die Zahl der gleichgeschlechtlichen Paare im Mikrozensus 2020 im Vergleich mit 2019 verdoppelt, die von verheirateten gleichgeschlechtlichen Paaren gar verdreifacht. Hierfür sind ggf. auch Veränderungen in der Frageformulierung und den Antwortkategorien verantwortlich. Auch ein Zusammenhang mit einer weniger validen Erfassung des Geschlechts wird angenommen. Eine inhaltlich Begründete Veränderung dieses Ausmaßes erscheint unplausibel.

2.3.4 Migrationstypisierungen

Die Zahlen zur Bevölkerung nach Migrationshintergrund aus 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 Migrationshintergrundes⁴ hierfür ursächlich. Durch die methodische Weiterentwicklung der Typisierung des Migrationshintergrundes werden die mit deutscher Staatsangehörigkeit Geborenen besser abgebildet.

⁴ 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 Zuzugjahres nicht vorgenommen.

2.3.5 Variablen zur Arbeitssituation

Bei Analysen zum Thema Arbeitsmarkt besteht eine zusätzliche Unsicherheit bei der Bewertung der Ergebnisse, da sich pandemiebedingt die Situation auf dem deutschen Arbeitsmarkt in vielen Bereichen deutlich verändert hat. So können bei den Ergebnissen nur bedingt Aussagen getroffen werden, ob diese auf reale Entwicklungen oder auf die oben beschriebenen Einschränkungen zurückzuführen sind. Mit zunehmender Gliederungstiefe nehmen diese Unsicherheiten zu (z. B. bei Erwerbslosenquoten in tiefer regionaler oder demographischer Gliederung). Ein weiteres Problem gibt es bei der Erfassung von Zeitarbeit insbesondere in CAWI. Aus diesem Grund ist die Vergleichbarkeit im Hinblick auf den Mikrozensus 2021 eingeschränkt.

2.3.6 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 des Erhebungsjahres 2020 mit den 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.

2.3.7 Zusatzprogramm 2020

Zusätzlich zum Kernprogramm enthält der Mikrozensus vierjährige Zusatzprogramme. Im Erhebungsjahr 2020 wurde das Zusatzprogramm zu den Pendlereigenschaften von Schüler*innen, Studierenden sowie Erwerbstätigen erhoben.

Pendlerverhalten Erwerbstätige

- EC2100P Gehen bzw. fahren Sie üblicherweise von dieser Wohnung zu Ihrer Arbeitsstätte?
- EC2200P Wie weit ist der Hinweg zu Ihrer Arbeitsstätte, z.B. zum Betriebsgelände, Dienstgebäude?
- EC2300P Wie lange brauchen Sie normalerweise für den Hinweg zu Ihrer Arbeitsstätte?
- EC2400P Welches Verkehrsmittel benutzen Sie normalerweise auf dem Hinweg zu Ihrer Arbeitsstätte?
- EC2700P Nutzen Sie ein weiteres Verkehrsmittel, mit dem Sie eine wesentliche Strecke für den Hinweg zu Ihrer Arbeitsstätte zurücklegen?
- EC2701P Welches weitere Verkehrsmittel nutzen Sie?

Pendlerverhalten Schüler*innen und Studierende

- DC1100P Liegt die (zuletzt) besuchte Schule/Hochschule in der Gemeinde, in der Sie wohnen?
- DC1201P Liegt Ihre Schule/Hochschule in Deutschland?
- DC1300P Gehen oder fahren Sie üblicherweise von dieser Wohnung zu Ihrer Schule/Hochschule?
- DC1400P Wie weit ist der Hinweg zu Ihrer Schule/Hochschule?
- DC1500P Wie lange brauchen Sie normalerweise für den Hinweg zu Ihrer Schule/Hochschule?
- DC1600P Welches Verkehrsmittel benutzen Sie normalerweise auf dem Hinweg zu Ihrer Schule/Hochschule?
- DC1700P Nutzen Sie ein weiteres Verkehrsmittel, mit dem Sie eine wesentliche Strecke auf dem Hinweg zu Ihrer Schule/Hochschule zurücklegen?
- DC1800P Welches weitere Verkehrsmittel nutzen Sie?

2.3.8 Adhoc-Modul 2020

Im Rahmen der in den Mikrozensus integrierten Arbeitskräfteerhebung (LFS) werden normalerweise 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 2020 wurde das Ad-hoc-Modul „Arbeitsunfälle und andere arbeitsbedingte Gesundheitsprobleme“ durchgeführt. Die Fragen des Ad-hoc-Moduls erhält nur ein Teil der Haushalte der LFS-Unterstichprobe. Ihre Beantwortung ist freiwillig.

Gesundheitliche Belastungen bei der Arbeit

- EK0200P Sind Sie gegenwärtig erwerbstätig?
- EE0100P Sind Sie bei Ihrer Arbeit körperlichen Belastungen ausgesetzt, die Ihre Gesundheit schädigen könnten?
- EE0200P Sind Sie bei Ihrer Arbeit seelischen Belastungen ausgesetzt, die Ihr Wohlbefinden beeinträchtigen?

Fragen zu Arbeitsunfällen

- EK0100P Was trifft auf Ihre gegenwärtige Situation zu?
- EK0300P Hatten Sie in den letzten 12 Monaten einen Arbeitsunfall, bei dem Sie sich verletzt haben?
- EK0400P Handelte es sich bei Ihrem letzten Arbeitsunfall um einen Unfall im Straßenverkehr?
- EK0500P Bei welcher Tätigkeit hat sich der letzte Arbeitsunfall ereignet?

- EK0600P Mussten Sie Ihre Erwerbstätigkeit wegen des letzten Arbeitsunfalls zeitweise unterbrechen?
- EK0700P Konnten Sie Ihre Arbeit nach dem letzten Arbeitsunfall mittlerweile wieder aufnehmen?
- EK0800P Wie lange konnten Sie wegen Ihres Arbeitsunfalls keiner Erwerbstätigkeit nachgehen?

Fragen zu arbeitsbedingten Gesundheitsproblemen (keine Arbeitsunfälle)

- EK0900P Hatten Sie in den letzten 12 Monaten Gesundheitsprobleme, die durch Ihre Arbeit verursacht oder verschlimmert wurden?
- EK1000P Welcher der folgenden arbeitsbedingten Beschwerden beeinträchtigt bzw. beeinträchtigte Sie am meisten?
- EK1100P Bei welcher Tätigkeit wurde das Gesundheitsproblem, das Ihre Gesundheit am meisten beeinträchtigt bzw. beeinträchtigte, verursacht oder verschlimmert?
- EK1200P Sind Sie durch das arbeitsbedingte Gesundheitsproblem, das Ihre Gesundheit am meisten beeinträchtigt bzw. beeinträchtigte, bei der Arbeit oder im Privatleben eingeschränkt?
- EK1300P Mussten Sie wegen des arbeitsbedingten Gesundheitsproblems, das Ihre Gesundheit am meisten beeinträchtigt bzw. beeinträchtigte, Ihre Erwerbstätigkeit zeitweise unterbrechen?
- EK1400P Konnten Sie Ihre Arbeit mittlerweile wiederaufnehmen?
- EK1500P Wie lange konnten Sie wegen Ihres arbeitsbedingten Gesundheitsproblems, das Ihre Gesundheit am meisten beeinträchtigt bzw. beeinträchtigte, nicht arbeiten?

2.4 Eckwerte relevanter Merkmale und Merkmalskombinationen

Die Standardauswertungen der amtlichen Statistik beziehen sich anders als in früheren Berichtsjahren für alle Erhebungsteile ausschließlich auf Haushalte am Hauptwohnsitz.

Diese können mit der Variablen TH0201H identifiziert. Auf Personenebene steht TPBevlkrg zur Verfügung.

Verteilung der Bevölkerung am Hauptwohnsitz

		TPGeschlecht Geschlecht			
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 Männlich	40.646.156	49,5	49,5	49,5
	2 Weiblich	41.528.356	50,5	50,5	100,0

Gesamt	82.174.512	100,0	100,0
--------	------------	-------	-------

Familienstand

		AB0500P Familienstand			Kumulierte
		Häufigkeit	Prozent	Gültige Prozente	Prozente
Gültig	1 ledig	34.350.271	41,8	41,8	41,8
	2 verheiratet	37.103.740	45,2	45,2	87,0
	3 verwitwet	5.203.139	6,3	6,3	93,3
	4 geschieden	5.380.250	6,5	6,5	99,8
	5 eingetragene Lebenspartnerschaft	113.698	,1	,1	100,0
	6 eingetragene Lebenspartnerin/ eingetragener Lebenspartner verstorben	7.315	,0	,0	100,0
	7 eingetragene Lebenspartnerschaft aufgehoben	16.098	,0	,0	100,0
	Gesamt	82.174.512	100,0	100,0	

Privathaushalte am Hauptwohnsitz

		Haushaltsmitglieder2 AnzahlHaushaltsmitglieder			Kumulierte
		Häufigkeit	Prozent	Gültige Prozente	Prozente
Gültig	1,00 Einpersonenhaushalt	16.471.945	20,0	20,0	20,0
	2,00 Zweipersonenhaushalt	27.552.890	33,5	33,5	53,6
	3,00 Dreipersonenhaushalt	14.760.642	18,0	18,0	71,5
	4,00 Vierpersonenhaushalte	15.856.722	19,3	19,3	90,8
	5,00 Fünf- und mehr Personenhaushalte	7.532.313	9,2	9,2	100,0
	Gesamt	82.174.512	100,0	100,0	

LAND Bundesland					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 Schleswig-Holstein	2.864.716	3,5	3,5	3,5
	2 Hamburg	1.833.182	2,2	2,2	5,7
	3 Niedersachsen	7.896.827	9,6	9,6	15,3
	4 Bremen	675.526	,8	,8	16,1
	5 Nordrhein-Westfalen	17.707.854	21,5	21,5	37,7
	6 Hessen	6.226.500	7,6	7,6	45,3
	7 Rheinland-Pfalz	4.034.410	4,9	4,9	50,2
	8 Baden-Württemberg	11.041.188	13,4	13,4	63,6
	9 Bayern	12.978.902	15,8	15,8	79,4
	10 Saarland	970.640	1,2	1,2	80,6
	11 Berlin	3.628.756	4,4	4,4	85,0
	12 Brandenburg	2.479.949	3,0	3,0	88,0
	13 Mecklenburg- Vorpommern	1.588.240	1,9	1,9	90,0
	14 Sachsen	4.004.483	4,9	4,9	94,8
	15 Sachsen-Anhalt	2.146.754	2,6	2,6	97,4
	16 Thüringen	2.096.586	2,6	2,6	100,0
	Gesamt	82.174.512	100,0	100,0	

Im Anhang befindet sich die entsprechende Syntax für das Programm SPSS für die oben aufgeführten Tabellen. Die hier abgebildeten Tabellen sind in ähnlicher Form der [Fachserie 1 Reihe 2.2](#) zu entnehmen.

2.5 Auswertbare regionale Ebenen

Für das Erhebungsjahr 2020 wird aus den in Unterkapitel 2.1 genannten Gründen von einer Auswertung der Daten unterhalb der Landesebene (NUTS 1) abgeraten. Ein geeigneter Hochrechnungsfaktor für die Kreis- oder Bezirksebene steht 2020 nicht zur Verfügung.

- Bundesebene (NUTS 0)

- Land: Landesebene (NUTS 1)
- Reg: Regierungsbezirke/Statistische Bezirke (NUTS 2)
- Nielsegebiete
- Kreis: Kreisebene (NUTS 3)
- Gemeindeebene (nur zu methodischen Zwecken, nicht zur Ergebnisdarstellung empfohlen⁵) (LAU)
- Anpassungsschichten (regionale Schicht und regionale Untergruppe)⁶
- Raumordnungsregionen
- Planungsregionen
- Arbeitsmarktregionen
- Arbeitsagenturbezirke

Stichtag des Gebietstandes der Regionalkennungen: 31.12.2019

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

2.6 Produktversionen

Die Versionsnummer des Datensatzes ist der Variablen DOI zu entnehmen.

1.0.0 (DOI: 10.21242/12211.2020.00.00.1.1.0)

- Erstveröffentlichung

⁵ 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.

⁶ 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.

3. Praktische Hinweise

3.1 Hinweise zur Geheimhaltung

3.1.1 Gesetzliche Grundlagen der statistischen Geheimhaltung

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

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

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

3.1.2 Geheimhaltung von Ergebnissen

Um die gesetzlich vorgeschriebene Geheimhaltung von Einzelfällen in den Daten sicherzustellen, müssen alle Ergebnisse, die am 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 ist dies gewährleistet, da von amtlicher Seite keine Ergebnisse aus dem Mikrozensus veröffentlicht werden, die auf weniger als 71 Befragten basieren.

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

Seit dem Berichtsjahr 2020 hat sich die Veröffentlichungspraxis der amtlichen Statistik geändert. Es werden von amtlicher Seite keine Ergebnisse veröffentlicht, die auf weniger als 71 Personen basieren.

3.1.3 Praktische Tipps zur Vermeidung von Geheimhaltungsfällen

Treten in den erstellten Analysen Geheimhaltungsfälle auf, werden diese Werte von den FDZ zur Sicherstellung der Geheimhaltung durch ein Sperrmuster ersetzt. Gerade in Kreuztabellen entstehen so durch die notwendige Sekundärsperre schnell viele Lücken in den Auswertungen. Da eine einmal zur Sekundärsperre 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 einer Bereitstellungsgebühr bei den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder bestellt werden. Diese beträgt 250 Euro pro Statistik und Erhebungsjahr (unter bestimmten Bedingungen werden für Promovierende und Studierende Ermäßigungen gewährt). 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 Panels.

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 2020 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 bezüglich LFS-Strukturvariablen, die nicht mit dem verkürzten Fragebogen 4 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

idhhx: LAND AWBNummerFremd TPBerichtsquartal HHNummerFremd

idawb: LAND AWBNummerFremd

idawbx: LAND AWBNummerFremd TPBerichtsquartal

Leerzeichen werden bei der Bildung der Identifikatoren durch Nullen ersetzt.

Wie können einem 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 Personnummer der Mutter bzw. des Vaters ermittelt werden. Der zugehörige Einzeldatensatz der Mutter/des Vaters lässt sich über die Personnummer 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 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 mehr Hochrechnungsfaktoren als bisher zur Verfügung. 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. Mit dem Hochrechnungsfaktor HR000QQ können Quartalsergebnisse berechnet werden. Mit TPBerichtsquartal können dafür einzelne Berichts quartale selektiert werden. Darüber hinaus steht mit HR000JQ der Durchschnitt der vier Quartale zur Verfügung.
- Strukturmerkmale aus der LFS-Substichprobe (Fragebögen 2 und 3) werden auf das Jahr mit dem Merkmal HR100JJ hochgerechnet. Entsprechend des Kernprogramms 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.
- Für die SILC-Substichprobe wird der Hochrechnungsfaktor HR200JJ verwendet.
- Weitere Hochrechnungsfaktoren sind HR100MO für Merkmale des Ad-hoc-Moduls und HR100BH für Menschen mit Behinderung.

Hochrechnungsfaktor	Stichprobe	Summe	Anmerkung
HR000JJ	Kern	84.358.363	Zur Verwendung bei Berechnungen mit Jahresergebnissen. Abweichungen zu den anderen Summen folgen aus dem 2-Stufen-Vorgehen zur Bestimmung der Hochrechnungsfaktoren und sind den Komplikationen im Erhebungsablauf geschuldet .
HR000JQ	Kern	82.216.296	Zur Verwendung bei Berechnungen mit Quartalsdurchschnitt zur Jahresauswertung.
HR000QQ	Kern	328.865.187 ⁷	Zur Verwendung bei Berechnungen mit Quartalsergebnissen. Werden nicht wie hier die Jahresergebnisse verwendet, sondern nur die Daten eines Quartals summieren sich die hochgerechneten Werte auf den Quartalsdurchschnitt, der bei Verwendung des Hochrechnungsfaktors HR000JQ resultiert.
HR100JJ	LFS	82.206.996	Zur Verwendung bei Berechnung mit Jahresergebnissen.
HR100JQ	LFS	82.206.996	Zur Verwendung bei Berechnungen mit Quartalsdurchschnitt zur Jahresauswertung.

⁷ Da jedes Quartal einzeln berechnet wird, entspricht der über das ganze Jahr summierte Wert dem Vierfachen der Bevölkerungszahl.

HR100QQ	LFS	328.827.981 ⁸	Zur Verwendung bei Berechnungen mit Quartalsergebnissen. Werden nicht wie hier die Jahresergebnisse verwendet, sondern nur die Daten eines Quartals summieren sich die Werte auf 82.206.996.
HR100MO	LFS	62.038.285	Zur Verwendung bei Berechnungen von Variablen des LFS-Ad-Hoc-Moduls.
HR100BH	LFS	10.482.824	Zur Verwendung bei Berechnungen mit Jahresergebnissen für Menschen mit Behinderungen.
HR200JJ	SILC	82.174.512	Zur Verwendung bei Berechnungen des SILC-Querschnitts. Die SILC-Erhebung findet in einem begrenzten Zeitraum des Jahres statt, deshalb gibt es keine weiteren Hochrechnungsfaktoren.

Tabelle 1: Hochrechnungsfaktoren mit Gesamtbevölkerungszahlen der hochgerechneten Jahresergebnisse

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. Aufgrund der Einschränkungen bei der Erhebung 2020 wird in amtlichen Veröffentlichungen für die MZ-Kern-Jahresergebnisse nicht auf den Quartalsdurchschnitt zur Hochrechnung zurückgegriffen.

⁸ Da jedes Quartal einzeln berechnet wird, entspricht der über das ganze Jahr summierte Wert dem Vierfachen der Bevölkerungszahl.

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 sollen in 45% der Auswahlbezirke Befragungen zur Arbeitsmarktbeteiligung (LFS) durchgeführt werden. Der Anteil zu Einkommen und Lebensbedingungen (SILC) macht etwa 12% aus. Die realisierte Auswahlsätze können davon abweichen. Die beiden genannten Substichproben sind überschneidungsfrei. Entsprechend geringer als bei Kern-Variablen sind LFS- und SILC-Variablen belegt. Ebenso ist zu beachten, dass einige Fragen freiwillig sind und die Anzahl der Beobachtungen deshalb geringer ausfallen kann. Die Problematik des Mikrozensus 2020, welche in Kapitel 2.1 näher erläutert wird, trägt ebenfalls zu vermehrten Fehlwerten 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 [Qualitätsbericht zum Mikrozensus 2020](#) zu entnehmen.

Warum beträgt die Anzahl der Befragten weniger als 1% (etwa 831 600) der deutschen Bevölkerung?

Hauptgrund für die geringere Zahl der Befragten Personen und Haushalte im Erhebungsjahr 2020 ist die hohe Ausfallquote. Diese ist auf technische Schwierigkeiten und die pandemische Lage zurückzuführen.

Erläutert wird die Problematik und die Folgen für die Auswertbarkeit des Datenmaterials in Abschnitt 2.1. des hier vorliegenden Reports sowie in [„Die Neuerungen des Mikrozensus ab 2020“](#) und dem [Qualitätsbericht zum Mikrozensus 2020](#).

Was muss beim Abgleich der Daten mit den Tabellen der Fachserie beachtet werden?

Um die Haushaltstabellen und Tabellen der Lebensformen der Fachserien nachzubilden, müssen die Daten nach Hauptwohnsitzhaushalten gefiltert werden. In der Fachserie werden nur Haushalte am Hauptwohnsitz betrachtet. Diese können mittels der Variable TH0201H selektiert werden. Gleiches gilt für Ergebnisse von Eurostat zu den Erhebungsteilen SILC und

LFS. Bei Tabellen zum Migrationsstatus von Personen wird nicht durchgängig nach Hauptwohnsitzen gefiltert.

Für Tabellen in denen das Alter verwendet wird, wird die Altersvariable TPALTER_1 verwendet.

Für das Geschlecht wird TPGeschlecht verwendet. Personen mit diversem oder ohne Eintrag des Geschlechts in das Personenstandsregister sind in dieser 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 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 DB010 bis DB135, RB010 bis RL070, HB010 bis HY170N und PB010 bis PY200G verwendet. Diese werden auch für die Nutzung in den Forschungsdatenzentren empfohlen.

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 einmalig 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.

Anhang

SPSS-Syntax zu Abschnitt 2.4

Filtern nach Bevölkerung am Hauptwohnsitz

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.

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.

Häufigkeitstabelle für Personenhaushalte

VARIABLE LABELS Haushaltsmitglieder2 'AnzahlHaushaltsmitglieder'.

VALUE LABELS

Haushaltsmitglieder2

1 'Einpersonenhaushalt'

2 'Zweipersonenhaushalt'

3 'Dreipersonenhaushalt'

4 'Vierpersonenhaushalte'

5 'Fünf- und mehr Personenhaushalte'.

EXECUTE.

FREQUENCIES Haushaltsmitglieder2.

EXECUTE.

Häufigkeitstabelle für Bundesland

VARIABLE LABELS Land 'Bundesland'.

FREQUENCIES Land.

Statistische Ämter des Bundes und der Länder,
Metadatenreport – Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus 2020 am
Gastwissenschaftsarbeitsplatz sowie per kontrollierter Datenfernverarbeitung (sog. On-Site-Zugangswege/-
Nutzungen)

Fotorechte Umschlag: ©artSILENCEcom – Fotolia.com