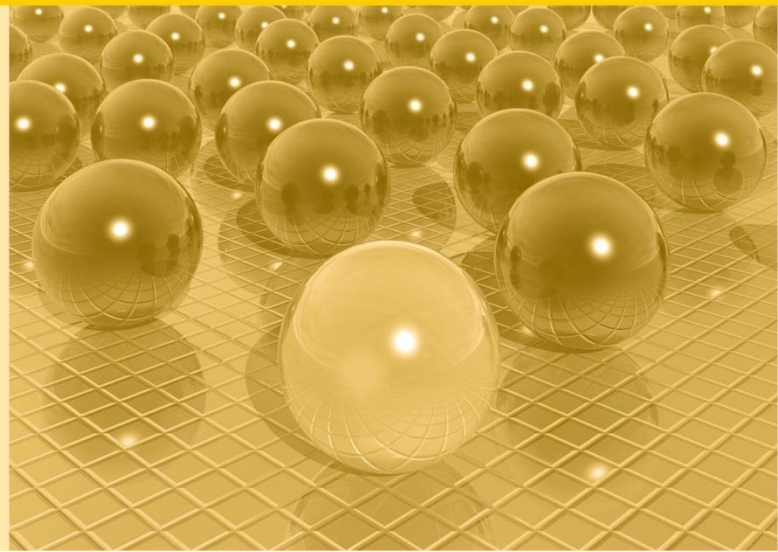


Metadatenreport



Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus
Scientific Use Files 2021

DOI: 10.21242/12211.2021.00.00.3.1.0

Version 1

Impressum

Herausgeber: Statistische Ämter des Bundes und der Länder
GESIS - Leibniz-Institut für Sozialwissenschaften, German Microdata Lab
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 März 2024

Diese Publikation wird kostenlos als PDF-Datei zum Download unter 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, Statistisches Bundesamt, GESIS - Leibniz-Institut für Sozialwissenschaften: Metadatenreport. Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus Scientific Use Files 2021 (EVAS-Nummer: 12211, 12231, 12241, 12251). Version 1. DOI: 10.21242/12211.2021.00.00.3.1.0. Düsseldorf 2024.

Metadatenreport

Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus
Scientific Use Files 2021

DOI: 10.21242/12211.2021.00.00.3.1.0

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 im MZ-SUF	6
1.2 Anonymisierungsmaßnahmen	6
1.2.1 Substichprobenziehung	6
1.2.2 Identifikatoren.....	9
1.2.3 Gesperrte Merkmale.....	9
1.2.4 Vergrößerungen von Merkmalsausprägungen.....	12
1.3 Methodik der Verknüpfung	13
2. Produkt	15
2.1 Hinweise zur Qualität des Mikrozensus 2021	15
2.2 Merkmale und Merkmalsbeschreibung	16
2.2.1 Merkmalsdefinitionen.....	16
2.2.2 Datensatzbeschreibung	16
2.3 Vergleichbarkeit der Merkmale über die Zeit	17
2.3.1 Variablennamen	17
2.3.2 Missingkodierung.....	18
2.3.3 Haushalte und Lebensformen	18
2.3.4 Migrationstypisierungen	18
2.3.5 Einkommen und Lebensbedingungen (SILC)	19
2.3.6 Zusatzprogramm 2021	19
2.3.7 Adhoc-Modul 2021	20
2.4. Eckwerte relevanter Merkmale	22
2.4.1 Verteilung der Bevölkerung in Hauptwohnsitzhaushalten	22
2.4.2 Privathaushalte am Haupt- und Nebenwohnsitz	23
2.5 Auswertbare regionale Ebenen	24

2.6 Produktversionen	24
3. Praktische Hinweise.....	25
3.1 Hinweise zur Geheimhaltung	25
3.2 FAQ	25
3.3 Verfügbare Tools	32
Anhang	33

Einleitung

Dieser Metadatenreport soll Forschenden dabei helfen, die Daten des Mikrozensus Scientific Use Files (MZ-SUF) 2021¹ sachgerecht auszuwerten. 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. Ausführliche Informationen und Auswertungshilfen zum MZ-SUF, u. a. Masterfragebogen, Datenhandbuch mit Randauszählungen, Tools zur Umsetzung sozialwissenschaftlicher Konzepte, Variablen-Zeitpunkte-Matrix, Verknüpfung von MZ-Querschnitterhebungen zu Panels, stehen auf dem [Mikrodaten-Informationssystem \(MISSY\) der GESIS](#) zur Verfügung.

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

Neben der Nutzung der Scientific-Use-Files am eigenen Arbeitsplatz in der wissenschaftlichen Einrichtung innerhalb Deutschlands gibt es auch die Möglichkeit, weniger stark anonymisierte Datensätze an einem Gastwissenschaftsarbeitsplatz oder über eine kontrollierte Datenfernverarbeitung auszuwerten. Produktspezifische Informationen zur On-Site-Nutzung werden im [Metadatenreport Teil II zur On-Site-Nutzung](#) bereitgestellt.

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.

¹ Die Aufbereitung und Dokumentation des faktisch anonymisierten Einzelmaterials erfolgt durch das Forschungsdatenzentrum der Statistischen Ämter der Länder – Standort NRW und das Statistische Bundesamt, Referat F 36 – Mikrozensus - Auswertung und Analyse in Kooperation mit dem German Microdata Lab der GESIS.

1. Datenaufbereitung in den FDZ

1.1 Datenaufbereitung

Das Datenmaterial wird einer Vollzählkeits- 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.

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

Im MZ-SUF 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.

Missingcode	Beschreibung
-7	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) Jahren²</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"> • EA0100P, DG0400P, ER1001P, ES1403P
-2	<p>Nichterwerbstätige</p> <p>Filtervariablen:</p> <ul style="list-style-type: none"> • EA0502P, EA0800P, EA0700P, EA0802P, EJ0501P
-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"> • AA0200P, AA0500H, BA0400H, CA0400H, DC0700P, EB0800P, ED1801P, EA1200P, DG0302H, ER1000P, ES1402P
-5	<p>Spezifischer Filterausfall, Ausfallgrund wird variablenspezifisch benannt³</p>
-9	<p>Filterfehler oder technisch zugelassene Fehlwerte bei Fragen mit Auskunftspflicht, keine Angabe bei Fragen ohne Auskunftspflicht</p>

Tabelle 1: Ausfallkennungen

Im MZ-SUF vor 2020 wurde zusätzlich der Missingcode -8 für Leerstand und ausgefallene Privathaushalte vergeben. Im MZ-SUF 2021 wird dieser nicht programmiert, da im Datenmaterial 2021 keine ausgefallenen Haushalte (kein Leerstand oder unbewohnte Wohnungen oder der ganze AWB eine Unit-Non-Response hatte) enthalten sind. Anders als 2020 sind ab 2021 aber wieder Auswahlbezirke im Datenmaterial, die zur Bruttostichprobe

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

³ 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.

gehören, aber keine Befragungsergebnisse aufweisen (z. B. da keine Wohnbevölkerung vorhanden war). Dies erhöht die Robustheit der Substichprobenziehung bezüglich Ausfällen (vgl. Abschnitt 1.2.1) und 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 im MZ-SUF

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. Die Hochrechnungsfaktoren werden so angeglichen, dass sie ohne zusätzliche Multiplikation (mit 1000) auf die Gesamtbevölkerung hochgerechnet werden können. Zudem werden die Hochrechnungsfaktoren im MZ-SUF skaliert, indem alle Hochrechnungsfaktoren im gesamten MZ durch 0,7 geteilt werden. Durch diese Skalierung kommt es zu minimalen Abweichungen zwischen Ergebnissen des MZ-SUF und den veröffentlichten Ergebnissen. Auf eine Poststratifizierung bzw. Kalibrierung der Hochrechnungsfaktoren des MZ-SUF wird seit dem Erhebungsjahr 2020 verzichtet.

1.2 Anonymisierungsmaßnahmen

Die angewendeten Anonymisierungsverfahren im MZ-SUF bestehen aus Verfahren zur Informationsreduktion für einzelne Merkmalstragende und für einzelne Merkmale.

1.2.1 Substichprobenziehung

Der MZ-SUF ist eine faktisch anonymisierte 70% Substichprobe der AWBs des Mikrozensus. Seit 2012 werden als Auswahlseinheiten für die Substichprobe die AWBs innerhalb eines Rotationsviertels herangezogen.

Um zu gewährleisten, dass die Substichprobe in bestimmten Merkmalen nur geringe zufallsbedingte Abweichungen aufweist, wird das gesamte Mikrozensusmaterial zunächst nach folgenden Merkmalen sortiert, um eine geschichtete Substichprobenziehung zu erreichen:

- Kennung Rotationsviertel (AWBRotationsviertel)
- Bundesland (Land)

- Summe der Befragungen im AWB (Dezile)
- Kennung über Grundauswahl und Aktualisierung der AWBs (AWBANSGROESSENKLASSE)
- Kennung über die Unterstichprobe (ABBZUSATZPROGRAMM)
- Regierungsbezirk (GBTRegierungsbezirk)
- Regionale Anpassungsschicht (AWBLFDNRANPASSUNGSSCHICHTSTICH)
- Regionale Schicht (AWBLFDNRREGSCHICHTSTICHPROBE)
- Regionale Untergruppe (AWBLFDNRREGREGUNTERGRUPPESTICH)
- Gemeindegrößenklasse (GBTGEMEINDEGROESSENKLASSEAKTUE)
- Nr. des AWB (AWBNummerFremd)

Zwar wurde bei der Grundauswahl bzw. wird bei der Aktualisierung der AWBs eine einheitliche Größe der jeweiligen Anschriftengrößenklassen angestrebt, sie unterscheiden sich jedoch fluktuationsbedingt erheblich. Um eine daraus resultierende Erhöhung der Fehlervarianz zu begrenzen, erfolgt die zusätzliche Aufnahme einer Sortierung nach der Anzahl der Befragungen in einem AWB. Die Größenklassen der AWB (Dezile) werden berechnet, indem man die Anzahl der Befragungen pro AWB ermittelt und diese dann auf AWB-Ebene aggregiert.

Jeweils zehn in der Reihenfolge der Sortierung aufeinanderfolgende AWBs bilden eine Schicht. Bei einem Wechsel in den Merkmalen AWBRotationsviertel, Land oder Dezile gibt es einen Sortierwechsel (d.h. die Durchnummerierung startet wieder bei 1), sodass auch unvollständige Schichten zugelassen werden. Im Hinblick auf die Analysemöglichkeit nach Bundesländern sowie die anzustrebende Homogenität der AWBs innerhalb einer Schicht, die varianzreduzierend wirkt, werden bei Sortierwechseln von Rotationsviertel, Bundesland und Größenklassen (Dezile) der AWBs unvollständige Schichten mit weniger als zehn AWBs zugelassen. Die letzte Schicht kann ebenfalls weniger als zehn AWBs umfassen.

Zur Ziehung der Substichprobe werden für jede Schicht drei reproduzierbare, ganzzahlige und gleichverteilte Zufallszahlen in einer Spanne von 1-10 gezogen. AWBs, deren Schichtnummer nun einer der gezogenen Zufallszahlen entsprechen, werden gekennzeichnet und im MZ-SUF nicht mehr berücksichtigt. Wurde also für eine Schicht die Zufallszahl 6 gezogen, wird der AWB mit der Schichtnummer 6 markiert und nicht mehr mit in den MZ-SUF aufgenommen (usw.). Der Stichprobenumfang jeder vollständigen Schicht beträgt somit 7. Die Anzahl gezogener Auswahlbezirke in unvollständigen Schichten ist zufällig, die Ziehungswahrscheinlichkeit von 7/10 bleibt aber erhalten. Mit der so gezogenen einfachen Stichprobe ohne Zurücklegen werden gleiche Auswahlwahrscheinlichkeiten für alle AWBs des MZ-SUF erreicht.

Umstellung der Substichprobenziehung ab dem Berichtsjahr 2021

Der Bruttodatensatz umfasst alle zu befragenden Auswahlbezirke. Der Nettodatensatz entspricht einer Teilmenge des Bruttodatensatz und ist vermindert um die Anzahl der Ausfälle.

Im Berichtsjahr 2020 wurde eine Substichprobe des Nettodatensatzes gezogen. Ab dem Berichtsjahr 2021 wird das System umgestellt auf eine Substichprobenziehung des Bruttodatensatzes, um robuster bezüglich Ausfällen zu sein.

Durch die Umstellung der Substichprobenziehung werden leere Auswahlbezirke (d.h. Auswahlbezirke die aufgrund von Ausfällen keine befragte Haushalte enthalten) im MZ-SUF abgebildet. Diese Fälle haben nur eine Beobachtung mit leeren Einträgen in allen Merkmalen, außer in den AWB-IDs und den Merkmalen LAND, AWBNummerFremd, AWBAUSWAHLTEIL und Substichprobe.

Berücksichtigung des Längsschnitts bei der Substichprobenziehung in 2021

Ab dem Berichtsjahr 2020 sind Verknüpfungen der Querschnitte mit bis zu 3 Folgejahren (max. 4 Jahre) möglich, da Haushalte bis zu 4 Mal befragt werden. Diese Rotation und die Möglichkeit zur Panelbildung soll in der Substichprobenziehung berücksichtigt werden.

Der MZ-SUF 2021 umfasst die folgenden Teilmengen:

1. alle Auswahlbezirke im MZ-SUF 2020, die auch im Berichtsjahr 2021 gemäß Auswahlplan befragt wurden und
2. eine Substichprobe von 70% aller neu hineinrotierten Auswahlbezirke im Berichtsjahr 2021 gemäß Auswahlplan und
3. eine Substichprobe von 70% aller Auswahlbezirke im Bruttodatensatz des Berichtsjahr 2020, die im Berichtsjahr 2021 weiterbefragt wurden und im Nettodatensatz 2020 nicht vorhanden waren (d.h. keine Ziehungswahrscheinlichkeit in den MZ-SUF 2020 hatten) sowie
4. die SILC Nachverfolgung.

Alle Auswahlbezirke im MZ-SUF 2020, die im Berichtsjahr 2021 weiterbefragt wurden, sind automatisch Bestandteil des MZ-SUF 2021. Eine Substichprobenziehung von 70% der Auswahlbezirke wird durchgeführt für die neu hineinrotierten Auswahlbezirke im Bruttodatensatz des Berichtsjahr 2021 gemäß Auswahlplan. Eine weitere Substichprobenziehung von 70% wird durchgeführt für alle Auswahlbezirke im Bruttodatensatz des Berichtsjahr 2020, die im Berichtsjahr 2021 weiterbefragt wurden und im Nettodatensatz 2020 nicht vorhanden waren (d.h. keine Ziehungswahrscheinlichkeit in den MZ-SUF 2020 hatten).

Berücksichtigung des Längsschnitts bei der Substichprobenziehung ab 2022

Ab 2022 umfasst der MZ-SUF des aktuellen Berichtsjahres (t)

1. alle Auswahlbezirke im MZ-SUF aus dem Vorjahr (t-1), die auch im aktuellen Berichtsjahr (t) gemäß Auswahlplan zu befragen sind und
2. eine Substichprobe von 70% aller neu hineinrotierten Auswahlbezirke im aktuellen Berichtsjahr (t) gemäß Auswahlplan sowie
3. die SILC Nachverfolgung.

Alle Auswahlbezirke im MZ-SUF des Vorjahres (t-1), die im aktuellen Berichtsjahr weiterbefragt werden, sind automatisch Bestandteil des MZ-SUF des aktuellen Berichtsjahres (t). Die Substichprobenziehung von 70% der Auswahlbezirke erfolgt ausschließlich für die neu hineinrotierten Auswahlbezirke im Bruttodatensatz des aktuellen Berichtsjahres gemäß Auswahlplan.

1.2.2 Identifikatoren

Direkte Identifikatoren und Hilfsmerkmale werden aus dem Datenmaterial entfernt. AWB-Nummer, Haushaltsnummer, Gebäudenummer und Wohnungsnummer werden zudem verfremdet. Die Ordnungsnummern des AWBs, des Haushalts im AWB und der Person im Haushalt sind systemfrei sortiert, sodass anhand der Position des einzelnen Falls im Datenmaterial kein Rückschluss auf einzelne Personen möglich ist.

1.2.3 Gesperrte Merkmale

Es gibt eine Reihe von Merkmalen, die Forschenden am Gastwissenschaftsarbetsplatz (GWAP) einsehen und dort sowie per Kontrollierter Datenfernverarbeitung (KDFV) auswerten können, die aber nicht mit in den MZ-SUF aufgenommen werden. Solche Merkmale werden als „gesperrte Merkmale“ bezeichnet.

Hierzu gehören einige Stichprobenkennzeichen, um persönliche oder regionale Rückschlüsse zu vermeiden. Zudem werden Gebietskennzeichen für nicht-administrative Raumeinheiten nicht für den MZ-SUF freigegeben.

Merkmalsbezeichnung	Variable
Berichtsmonat	TPBerichtsmonat, ABBReferenzmonat
Berichtswoche	TPBerichtswoche, ABBERferenzwocheBefragung

Befragungswoche	PBFBefragungswoche
Regionalkennungen	GBTKreis, GBTRegierungsbezirk, GBTOrtsteil, GBTGemeinde, AWBEZKENNZEICHENMEHREREGEMEIN, AWBLFDNRANPASSUNGSSCHICHTAKTUE, AWBLFDNRANPASSUNGSSCHICHTSTICH, AWBLFDNRREGSCHICHTAKTUELL, AWBLFDNRREGSCHICHTSTICHPROBE, AWBLFDNRREGGREGUNTERGRUPPEAKTUE, AWBLFDNRREGGREGUNTERGRUPPESTICH, GBTGEMEINDEGROESSENKLASSEZENSU, AWBAnsGroessenklasse, AWBANZZERLEGUNGSTEILGEB
Gemeinde der Arbeitsstätte	EC0303P
Stichprobenkennzeichen	AWBStichprobenNR, AWBRotationsviertel, AWBKennzDispropAuswahlsaeetze, AWBWOCHENKENNZEICHEN13, AWBWOCHENKENNZEICHEN14, AWBAKTUALISIERUNGSJAHR, AWBTeilschichtkennzeichen, AWBUnterschichtkennzeichen, AWBNRGebaeudeteil, AWBRVVQUARTALSKENNZEICHEN, AWBAUSWAHLTEIL, AWBANZWOHNKUMANSCHRIFTEN, AWBWOHNBERECHBEVOELKERUNG, AWBZONENNUMMER, ABBNULLBEZIRK
Raumeinheit (nicht-administrativ)	Stadt_Land_Glied_EU, BIK_Stadtregion_neu, Regionstyp_grund, Regionstyp_diff, Kreistyp, Arbeitsmarktregion, Raumordnungsregion, Planungsregion, Arbeitsagenturbezirk, Bundestagswahlkreis_von, Bundestagswahlkreis_bis, Gemeindetyp, Verdichtungsraum, EUFoerdergebiet, LUZ, Verflechtungsbereich_Zentr
Nettoeinkommen (im letzten Monat): Betrag	DG0101P
Haushaltsnettoeinkommen (im letzten Monat): Betrag, nur CAPI/CAWI	DG0300P
Merkmale aus Erhebungsteil SILC	Es werden nur ausgewählte SILC-Merkmale in den MZ-SUF integriert

Merkmale aus Erhebungsteil IKT	IKT-Merkmale werden erst ab 2022 in den MZ-SUF integriert
--------------------------------	---

Tabelle 2: Gesperrte Merkmale

Informationen aus dem Erhebungsteil zu den Informations- und Kommunikationstechnologien (IKT)) werden erstmalig ab dem Erhebungsjahr 2022 in den MZ-SUF integriert.

Aus dem Erhebungsteil zu den Einkommen und Lebensbedingungen (Substichprobe EU-SILC) werden nur ausgewählte Informationen in den MZ-SUF aufgenommen, welche es erlauben, die Konzeption des MZ-SUF hinsichtlich der allgemein vorgenommenen Anonymisierungsmaßnahmen konsistent zu den Berichtsjahren vor 2020 zu gestalten.

Ab dem Berichtsjahr 2020 werden erstmalig SILC-Merkmale der Beobachtungseinheit zu Indizes zusammengefasst und in den MZ-SUF integriert. Als Index werden im MZ-SUF Variablen bezeichnet, die sich durch die Berechnung aus den Werten mehrerer Merkmale für ein neu generiertes Konstrukt (z. B. „Wohnumgebung“) ergeben. Das Ziel der Indexbildung besteht darin, die verschiedenen Indikatoren eines Konstruktes zu einer Messgröße zusammenzufassen. Die Ausprägungen der im MZ-SUF integrierten Indizes ergeben sich aus der Summe der Werte der Merkmale.

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. Wenn 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. Weitere Informationen zur Bildung der Indizes können zudem dem [Datenhandbuch](#) entnommen werden. 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.

Ab dem Berichtsjahr 2021 werden zudem einige Merkmale zu subjektiven Einschätzungen und Verhaltensweisen aufgenommen, von denen aufgrund der Subjektivität in den Angaben ein sehr niedriges Deanonymisierungsrisiko ausgeht. Dies gilt insbesondere auch deshalb, da alle ausgewählten Merkmale auf freiwillig zu beantwortenden Fragen basieren. Es handelt sich dabei 2021 um folgende Variablen: HS120, HS150, HI010, HI040, PW010, PW191, RCH010, RCH020, PK020.

Des Weiteren beinhaltet der MZ-SUF Merkmale, die zur SILC-Unterstichprobe gehören, die aber bis einschließlich Berichtsjahr 2019 Teil des Kern-MZ waren.

1.2.4 Vergrößerungen von Merkmalsausprägungen

Das MZ-SUF unterscheidet sich u.a. vom Originalfile des MZ dadurch, dass bestimmte Merkmale im MZ-SUF, bedingt durch die Anonymisierung, in vergrößerter Form verfügbar sind. Vergrößerungen von Merkmalsausprägungen werden wie folgt angewendet:

1. Keine einzelne Gemeinde mit weniger als 500 000 Einwohnern darf identifizierbar sein.
2. Werden mehrere Gemeinden zu Größenklassen zusammengefasst, muss die Gemeindegrößenklasse in jedem Bundesland mindestens 400 000 Einwohner umfassen.
3. Angaben über Staatsangehörigkeit oder Geburtsland werden so weit aggregiert, dass eine Nationalität oder Gruppe von Nationalitäten mindestens 50 000 Einwohner in der Grundgesamtheit umfassen. Zusammenfassungen erfolgen unter Beachtung der räumlichen Lage bzw. wirtschaftlichen Verflechtungen. Zusammenfassungen innerhalb der EU richten sich nach den Beitrittsjahren der EU und nicht zwingend nach geografischer Nähe. Die Zuordnung erfolgt nach den Kriterien: EU, EFTA, EU-Beitrittskandidaten, sonstige Staaten.
4. Alle weiteren Variablen werden so weit aggregiert, dass eine ausgewiesene Merkmalsausprägung mindestens 5 000 Einwohner in der Grundgesamtheit umfasst. Die Merkmalsvergrößerung bei metrisch skalierten Angaben wird über eine Zusammenfassung direkt benachbarter Merkmalswerte vorgenommen. Im Fall von Vergrößerungen wird im MZ-SUF in der Regel die am stärksten besetzte Kategorie ausgewiesen.
5. Für die Variablen zur Klassifikation der Berufe (KldB 2010) wird ein besonderes Verfahren angewendet. Die KldB strukturiert die Berufe anhand zweier Dimensionen: der Berufsfachlichkeit und dem Anforderungsniveau, mit der die Komplexität der ausgeübten Tätigkeit erfasst wird. Eine sachgerechte Merkmalsvergrößerung der Berufsgattungen, um noch sinnvolle Analysen zu ermöglichen ist unter Beachtung der Anonymisierungsregeln nicht möglich. Um dennoch die beiden Dimensionen der Berufsklassifikation

(Berufsfachlichkeit und Anforderungsniveau) im MZ-SUF abbilden zu können, wird ein zweistufiges Verfahren eingesetzt, wobei die vierte Position in der Signierziffer der KldB 2010 unberücksichtigt bleibt. Somit enthalten die Variablen zur KldB die 1.-3. Stelle und die 5. Stelle des ursprünglichen Klassifizierungscodes. Zuerst werden die Variablen auf Basis des 3-Stellers vergrößert, so dass jede Ausprägung mindestens 5 000 Einwohner in der Grundgesamtheit umfasst. Anschließend wird die ursprüngliche 5. Stelle (das Anforderungsniveau) wieder an den nun vergrößerten 3-Steller angespielt. Bei Ausprägungen die bei diesen (künstlichen) 4-Stellern weniger als 5 000 Einwohner in der Grundgesamtheit umfassen, wird die 4. Stelle anschließend unter Beibehaltung der konditionalen Verteilung zufallsbasiert neu vergeben (unter Berücksichtigung der Variablen TPBvlkrg, TPGeschlecht und EP0800P).

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.⁴ 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 [Brocker und Mühlenfeld 2020](#) sowie [Herter-Eschweiler und Schimpl-Neimanns 2018](#) und [Backes und Pforr 2023](#) zu entnehmen. Die

⁴ Wie im Metadatenreport Teil I näher beschrieben, stellt die Stichprobe des Mikrozensus ein rotierendes Panel dar. Seit dem Erhebungsjahr 2012 werden als Auswahlseinheiten für die Substichprobe die Auswahlbezirke innerhalb eines Rotationsviertels herangezogen. Damit wird seit 2012 die Möglichkeit geschaffen, selbstständig mit den MZ-SUFs Paneldatensätze zu erzeugen. Infolge der Erneuerung der gesamten Mikrozensusstichprobe im Jahr 2016 sind Verknüpfungen der Querschnittsdaten ab dem Erhebungszeitpunkt 2012 nur bis einschließlich der Erhebung 2015 möglich. Zudem sind aufgrund der Weiterentwicklung des Mikrozensus ab 2020 Verknüpfungen der Querschnittsdaten ab dem Erhebungszeitpunkt 2016 nur bis einschließlich der Erhebung 2019 möglich.

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.

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 als auf die 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.

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.) (3-Steller + Anforderungsniveau):
<https://statistik.arbeitsagentur.de/DE/Navigation/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-Fassung2020/Systematik-Verzeichnisse/Systematik-Verzeichnisse-Nav.html>
- Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008) (3-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&&AUFUF=klass&&NAV=DE&&KLASSID=10527&&KLASSNAME=ISCED-F
(Übersetzung der Klassifikation von Statistik Austria)
Auch 2021 stehen zu den Hauptfachrichtungsvariablen EP1103PUG4, ER0703PUG4 und ER0713PUG4 auf die bis 2017 verwendeten Zweisteller umgeschlüsselte Variablen (EP1103PA, ER0703PA, ER0713PA) zur Verfügung.
- Staatsangehörigkeits- und Gebietssystematik (außereuropäische Staaten zum Teil zusammengefasst siehe Schlüsselverzeichnisse):
<https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/staatsangehoerigkeit-gebietsschluesel.html>

2.2.2 Datensatzbeschreibung

Der Mikrozensus ist eine Haushalts- und Personenstatistik. Die vollständige Liste der im MZ-SUF vorhandenen Variablen, ist dem [Datenhandbuch](#) zu entnehmen.⁵ Sofern Variablen das Ergebnis einer einzelnen Frage in einem oder mehreren Fragebögen wiedergeben, sind die zugehörigen Fragennummern und der Fragetext enthalten. Zudem sind im Datenhandbuch

⁵ Neben Informationen zu den Personen eines Haushalts bzw. einer Gemeinschaftsunterkunft liegen im Mikrozensus auch Informationen über die sekundären Erhebungseinheiten Haushalt, Lebensform und Familie vor. Zur Analyse dieser unterschiedlichen Ebenen werden einige Informationen, sogenannte Bandsatzergänzungen, nachträglich vom Statistischen Bundesamt auf Basis der direkt erhobenen Fragen generiert. Zur Vereinfachung der Datenaufbereitung sind im Scientific Use File (SUF) des MZ ab 2015 solche Bandsatzergänzungen nicht mehr in den bereitgestellten Daten enthalten. Diese Typisierungen können für viele Fragestellungen relevant sein. Sie enthalten jedoch überwiegend Angaben, die aus anderen in den Daten enthaltenen Variablen eigenständig erstellt werden können. Der Bericht „Einführung in die eigenständige Erstellung von Typisierungen am Beispiel des Mikrozensus Scientific Use Files 2014“ ([Börlin 2020](#)) zeigt am Beispiel des MZ SUF 2014, wie diese Typisierungen mithilfe der in den Daten vorliegenden Angaben selbst erstellt werden können.

Randverteilungen (ungewichtete Fallzahlen) der Variablen, erläuternde Kommentare und die vollständigen Label (das Statistikprogramm SPSS übernimmt nur 120 Zeichen) enthalten.

Aufgrund der 2020 eingeführten Neukonzeption des Mikrozensus, welche die Integration weiterer Haushaltserhebungen⁶ sowie Neuerungen bei der europäischen Arbeitskräfteerhebung (LFS)⁷ 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. Zudem enthält das Dokument Informationen zu den korrespondierenden Variablennamen, welche im MZ-SUF enthalten sind.

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 typisierten Merkmale der Mikrozensus-Hauptbefragung beginnen mit T. Die SILC-Zielvariablen beginnen mit D (Haushaltsregister), R (Personenregister), H (Haushaltsdaten) oder P (Personendaten).

⁶ Es handelt sich hierbei um die europäische Statistik der Einkommens- und die Lebensbedingungen (SILC) und ab 2021 die europäische Haushaltsbefragung zum privaten Einsatz von Informations- und Kommunikationstechnologien (IKT-Haushalte). Aus dem SILC-Erhebungsteil werden nur ausgewählte Informationen in den MZ-SUF aufgenommen. Informationen aus dem IKT-Erhebungsteil werden erst ab dem Erhebungsjahr 2022 in den MZ-SUF integriert.

⁷ unterjährige Wiederholungsbefragung, Konzeption als Unterstichprobe

Danach folgen ein weiterer Buchstabe und drei Ziffern. Die Benennung der IKT-Zielvariablen folgt keinem einheitlichen Schema. Außerdem gibt es einen neuen Umgang mit Haupt- und Nebenwohnsitzhaushalten, insbesondere im SILC-Befragungsteil: Die Befragungen für die in 2020 neu integrierte Unterstichproben 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](#).

Die [Variablen-Zeitpunkte-Matrix](#) in MISSY gibt 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 Missingkodierung um den Missingcode -7. Der Code -7 wird für die Auswahlbezirke (AWB) ohne Befragungen vergeben. Weitere Informationen zur Missingcodierung finden sich im Abschnitt 1.1.1.

2.3.3 Haushalte und Lebensformen

Vor dem Hintergrund der genannten Qualitätseinschränkungen (Unterkapitel 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 Unterkapitel 2.1 erläuterten generellen Umstellungen und methodisch-technischen Einschränkungen sind Umstellungen in der Erhebung/Frageformulierung sowie in der Methodik der Typisierung des Migrationshintergrundes⁸ 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).

⁸ 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.

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 (MZ-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.

2.3.6 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 Auswahlrates 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?

2.3.7 Adhoc-Modul 2021

Im Rahmen der in den Mikrozensus integrierten Arbeitskräfteerhebung (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?
- 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 Hauptwohnsitzhaushalte⁹. 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). Im Anhang befindet sich die SPSS-Syntax für das Programm SPSS für die im Folgenden aufgeführten Tabellen auf Personen- und Haushaltsebene.

2.4.1 Verteilung der Bevölkerung in Hauptwohnsitzhaushalten

Geschlecht

		TPGeschlecht Geschlecht			
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 Männlich	40.712.564	49,4	49,4	49,4
	2 Weiblich	41.665.496	50,6	50,6	100,0
Gesamt		82.378.060	100,0	100,0	

Familienstand

		AB0500P Familienstand			
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 Ledig	34.847.448	42,3	42,3	42,3
	2 Verheiratet	36.601.104	44,4	44,4	86,7
	3 Verwitwet	5.226.630	6,3	6,3	93,1
	4 Geschieden	5.563.605	6,8	6,8	99,8
	5 Eingetragene Lebenspartnerschaft	113.802	,1	,1	100,0
6 Eingetragene Lebenspartnerin/ eingetragener Lebenspartner verstorben		8.800	,0	,0	100,0
	7 Eingetragene Lebenspartnerschaft aufgehoben	16.671	,0	,0	100,0
Gesamt		82.378.060	100,0	100,0	

⁹ Haushalte, in denen mindestens eine Person ab 16 Jahren am Hauptwohnsitz wohnt.

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

Anzahl Haushaltsmitglieder

		Haushaltsmitglieder2 AnzahlHaushaltsmitglieder		Gültige	Kumulierte
		Häufigkeit	Prozent	Prozente	Prozente
Gültig	1 Einpersonenhaushalt	17.648.515	42,3	42,3	42,3
	2 Zweipersonenhaushalt	13.835.829	33,2	33,2	75,5
	3 Dreipersonenhaushalt	4.873.110	11,7	11,7	87,1
	4 Vierpersonenhaushalt	3.906.128	9,4	9,4	96,5
	5 Fünf- oder mehr Personenhaushalt	1.458.918	3,5	3,5	100,0
	Gesamt	41.722.500	100,0	100,0	

Bundesland

		LAND Bundesland		Gültige	Kumulierte
		Häufigkeit	Prozent	Prozente	Prozente
Gültig	1 Schleswig-Holstein	1.509.906	3,6	3,6	3,6
	2 Hamburg	991.418	2,4	2,4	6,0
	3 Niedersachsen	4.007.981	9,6	9,6	15,6
	4 Bremen	355.785	,9	,9	16,5
	5 Nordrhein-Westfalen	8.792.400	21,1	21,1	37,5
	6 Hessen	3.132.341	7,5	7,5	45,0
	7 Rheinland-Pfalz	1.971.528	4,7	4,7	49,8
	8 Baden-Württemberg	5.378.724	12,9	12,9	62,7
	9 Bayern	6.547.062	15,7	15,7	78,3
	10 Saarland	489.362	1,2	1,2	79,5
	11 Berlin	2.045.558	4,9	4,9	84,4
	12 Brandenburg	1.293.902	3,1	3,1	87,5
	13 Mecklenburg- Vorpommern	838.278	2,0	2,0	89,5
	14 Sachsen	2.143.226	5,1	5,1	94,7
	15 Sachsen-Anhalt	1.128.306	2,7	2,7	97,4
	16 Thüringen	1.096.723	2,6	2,6	100,0
	Gesamt	41.722.500	100,0	100,0	

2.5 Auswertbare regionale Ebenen

- Bundesebene (NUTS 0)
- Land: Landesebene (NUTS 1)
- Land * Gemeindegrößenklasse

2.6 Produktversionen

Die Versionsnummer des Datensatzes ist der Variablen DOI zu entnehmen.

1.0.0 (DOI: 10.21242/12211.2021.00.00.3.1.0)

- Erstveröffentlichung

3. Praktische Hinweise

3.1 Hinweise zur Geheimhaltung

Bei den MZ-SUF handelt es sich um faktisch anonyme Mikrodaten. Mikrodaten werden als faktisch anonym bezeichnet, wenn eine Deanonymisierung zwar nicht gänzlich ausgeschlossen werden kann, die Angaben jedoch „nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft“ der jeweiligen Merkmalsträgerin beziehungsweise dem jeweiligen Merkmalsträger zugeordnet werden können (§16 Abs. 6 BStatG). Die MZ-SUF bieten im Vergleich zu den On-Site-Zugangswegen ein geringeres Analysepotenzial, sind jedoch so konzipiert, dass sie sich für einen großen Teil der wissenschaftlichen Forschungsvorhaben eignen. Durch die faktische Anonymisierung der Mikrodaten dürfen sie außerhalb der geschützten Infrastruktur der amtlichen Statistik verwendet werden. Voraussetzung hierfür ist, dass die beantragende Institution ihren Sitz in Deutschland hat und dass die bereitgestellten Daten nur in den Räumen der beantragenden wissenschaftlichen Einrichtungen innerhalb Deutschlands genutzt werden. Zudem müssen alle Datennutzenden zur statistischen Geheimhaltung nach § 16 Abs. 7 BStatG verpflichtet werden (<https://www.forschungsdatenzentrum.de/de/zugang>). Insbesondere sind Handlungen zu unterlassen, die darauf abzielen oder geeignet sind, anonymisierte statistische Einzelangaben zu deanonymisieren.

3.2 FAQ

Wie erhalte ich Zugang zu den MZ-SUF?

Die Mikrozensus-Daten können entgeltpflichtig bei den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder beantragt werden. 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 MZ-SUF habe?

Ausführliche Informationen und Auswertungshilfen zum MZ-SUF, u. a. Masterfragebogen, Datenhandbuch mit Randauszählungen, Tools zur Umsetzung sozialwissenschaftlicher Konzepte, Variablen-Zeitpunkte-Matrix, Verknüpfung von MZ-Querschnitterhebungen zu Panels, stehen auf dem [Mikrodaten-Informationssystem \(MISSY\) der GESIS](#) zur Verfügung. Weitere Informationen zum Datenangebot und zum Datenzugang sind zudem auf den Seiten der [Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder](#) abrufbar.

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 bezüglich 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

idhhx: 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 (idhhx) 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 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](#)¹⁰ 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, welche Hochrechnungsfaktoren für die Analyse welcher Substichproben 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

¹⁰ https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2019/06/neuregelung-mikrozensus-062019.pdf?__blob=publicationFile

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.

Hochrechnungs-faktor	Stich-probe	Summe ¹¹	Anmerkung
HR000JJ	Kern	84.600.880	Zur Verwendung bei Berechnungen mit Jahresergebnissen. Abweichungen zu den anderen Summen folgen aus dem 2-Stufen-Vorgehen zur Bestimmung der Hochrechnungsfaktoren.
HR000JQ	Kern	82.293.327	Zur Verwendung bei Berechnungen mit Quartalsdurchschnitt zur Jahresauswertung.
HR000QQ	Kern	329.173.306 ¹²	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 82.293.327.
HR100JJ	LFS	82.034.927	Zur Verwendung bei Berechnung mit Jahresergebnissen.
HR100JQ	LFS	82.021.078	Zur Verwendung bei Berechnungen mit Quartalsdurchschnitt zur Jahresauswertung.
HR100QQ	LFS	328.084.314 ¹³	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	61.743.049	Zur Verwendung bei Berechnungen von Variablen des LFS-Ad-Hoc-Moduls.

¹¹ Bedingt durch die Ziehung der 70% Substichprobe ergeben sich kleinere Abweichungen zwischen den hier berichteten hochgerechneten Summen des MZ-SUF und den in den Fachserien des Statistischen Bundesamtes veröffentlichten Zahlen bzw. der Original-Mikrozensusdaten.

¹² Da jedes Quartal einzeln berechnet wird, entspricht der über das ganze Jahr summierte Wert dem Vierfachen der Bevölkerungszahl.

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

HR100BH	LFS	10.241.691	Sonderhochrechnungsfaktor für Berechnungen, die Menschen mit Behinderungen betreffen.
HR200JJ	SILC	82.469.256	Zur Verwendung bei Berechnungen des SILC-Querschnitts. Es empfiehlt sich, SILC-Hochrechnungsfaktoren auf die Zielvariablen anzuwenden.

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 Kapitel 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. Die realisierten Auswahlsätze können davon abweichen. Die beiden genannten Substichproben sind überschneidungsfrei.

Entsprechend weniger häufig 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 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 Neuregelung des Mikrozensus ab 2020“](#) enthält weitere Informationen zu den Erhebungsteilen. Informationen zu den Unterstichproben, sowie zur Anzahl der freiwilligen und auskunftspflichtigen Fragen je Befragungsprogramm sind dem [Qualitätsbericht zum Mikrozensus 2021¹⁴](#) zu entnehmen.

Warum beträgt die Anzahl der Befragten mehr als 0,7% (etwa 582 000) der deutschen Bevölkerung?

Die tatsächliche Menge der Befragungen ist größer als die Stichprobengröße von 0,7%, da 7/9 des LFS-Anteils aufgrund der unterjährigen Wiederholungsbefragung zwei Mal im Erhebungsjahr befragt wird.

Mehr Informationen zur Stichprobenziehung auf Ebene der Auswahlbezirke können dem [Qualitätsbericht des Mikrozensus 2021](#) 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 das Geschlecht wird TPGeschlecht verwendet. Personen mit diversem oder ohne Eintrag des Geschlechts in das Personenstandsregister sind in dieser Variable mit einer Wahrscheinlichkeit von je 50% männlich oder weiblich zugeordnet.

Bedingt durch die Ziehung der 70% Substichprobe ergeben sich Abweichungen zwischen den Häufigkeiten der Variablen des MZ-SUF und den in den Fachserien des Statistischen

¹⁴ Die Fallzahlen im Qualitätsbericht beruhen nicht auf den Endergebnissen und weichen deshalb von diesen ab. Zudem bezieht sich der Qualitätsbericht auf die volle 1% Stichprobe.

Bundesamtes veröffentlichten Zahlen bzw. der Original-Mikrozensusdaten. Die meisten Variablen des SUF sollten nur in geringem Maße von den veröffentlichten Daten abweichen. Größere relative Abweichungen können sich bei Merkmalen ergeben, die mit sehr geringen Fallzahlen besetzt sind.

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).

Wie werden die Systemfiles für die Programmpakete SPSS, SAS und Stata erstellt?

Die vom GML der GESIS bereitgestellten Setups für das Mikrozensus SUF 2021 dienen zum Einlesen des Rohdatenmaterials und zum Erstellen von Systemfiles für die Programmpakete SPSS, SAS und Stata. Sie beinhalten Programmanweisungen zur Definition von fehlenden Werten sowie zum Versehen der Variablen und ihrer Ausprägungen mit entsprechenden Labels.

Spezifika der Statistikprogramme:

- Am Anfang der Setups sind in der **Configuration Section** die vollständigen lokalen Dateinamen (einschließlich Laufwerkskennzeichen und Verzeichnis) zu nennen. Ansonsten sind im Setup keine weiteren Änderungen vorzunehmen.
- Stata: Das Setup steht mit der Zeichencodierung Unicode (UTF-8) bereit und ist mit Stata ab Version 14 ablauffähig. Für die Verwendung mit älteren Versionen kann es z. B. mithilfe von MS-Edit oder Notepad++ in die Zeichencodierung ANSI bzw. Windows-1252 umgesetzt werden.
- SAS: Das Setup steht mit der Zeichencodierung ANSI bereit und ist mit SAS ab Version 9 ablauffähig. Für die Verwendung mit älteren Versionen kann es notwendig sein, z. B. mithilfe von MS-Edit oder Notepad++ die Zeichencodierung zu ändern.
- SPSS: Das Setup mit der Zeichencodierung ANSI ist sowohl mit Version 24 als auch mit älteren Versionen ablauffähig, wenn entsprechend unter „Bearbeiten“ | „Optionen“ |

„Sprache“ | „Zeichencodierung ...“ | „[x] Schriftsystem der Ländereinstellung ...“ eingestellt ist.

- **Missing Values:** In SPSS können vorliegende Werte als benutzerdefinierte Missings deklariert werden. Sie werden bei Auszählungen mit entsprechenden Werten und Labels ausgewiesen, zählen aber i. d. R. bei statistischen Modellen nicht als gültige Werte. In SAS und Stata können dagegen vorliegende Werte nur nach Rekodierung als fehlende Werte definiert werden. Im SPSS-Setup werden benutzerdefinierte Missings spezifiziert. Diese Definitionen sind zwar auch in den SAS- und Stata-Setups enthalten, im Unterschied zum SPSS-Setup jedoch auskommentiert, da sonst die Originalwerte der Rohdaten (-1, ..., -9) im Systemfile durch benutzerdefinierte Missing-Zeichen (.a, ..., .h) ersetzt würden. Bei Bedarf können diese im Setup unter dem Kommentar „Definition of user-missing values“ stehenden Programmanweisungen durch Entfernen der Kommentarzeichen (/*, */) am Beginn und Ende des Anweisungsblocks aktiviert werden.
- **Voreinstellungen:** Um eine reibungsfreie Aufbereitung der Daten zu gewährleisten, empfiehlt es sich, die in den Setups vorgesehenen Voreinstellungen nicht zu verändern.

Ist das entsprechende Systemfile erstellt, kann die einfache Fallzahl $n = 638.332$ (ohne Gewichtung, ohne Selektion) zur Kontrolle, ob der Rohdatensatz fehlerfrei eingelesen wurde, mit der des erstellten Datensatzes verglichen werden. Zusätzlich können die Eckwerte aus Abschnitt 2.4 zur Prüfung des fehlerfreien Ablaufs des Setups herangezogen werden. Hierbei ist zu beachten, dass sich die Eckwerte in Abschnitt 2.4.1 auf die hochgerechnete Bevölkerung (HR000JJ) in Hauptwohnsitzhaushalten (TH0201H=1) beziehen und in Abschnitt 2.4.2 auf die hochgerechneten Privathaushalte (TH0602P=1) and (TPPrivatHH=1).

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. Im Datenmaterial zum Erhebungsjahr 2021 sind die ISCO-Variable 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 Haushaltmitglieder2.

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 Scientific Use Files 2021

Fotorechte Umschlag: ©artSILENCEcom – Fotolia.com