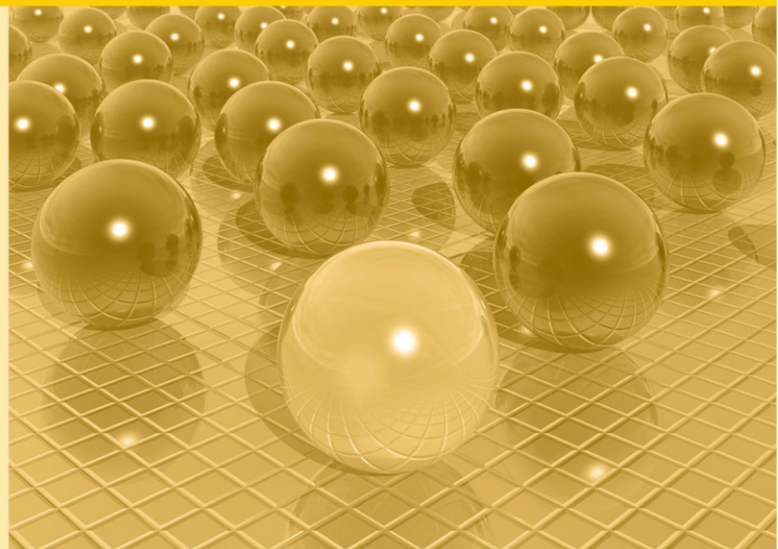


Metadatenreport



Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus
Scientific Use Files 2024

DOI: 10.21242/12211.2024.00.00.3.1.0

Version 1



GESIS Leibniz-Institut
für Sozialwissenschaften



STATISTISCHE ÄMTER
DES BUNDES UND DER LÄNDER
FORSCHUNGSDATENZENTREN

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 9449-2939
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 Oktober 2026

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

© Information und Technik Nordrhein-Westfalen, Düsseldorf, 2026
(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 2024 (EVAS-Nummer: 12211, 12231, 12241, 12251). Version 1. DOI: 10.21242/12211.2024.00.00.3.1.0. Düsseldorf 2026

Metadatenreport

Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus
Scientific Use Files 2024

DOI: 10.21242/12211.2024.00.00.3.1.0

Version 1

Inhalt

Abkürzungsverzeichnis	3
Einleitung.....	4
1 Datenaufbereitung in den FDZ	5
1.1 Datenaufbereitung	5
1.1.1 Missingkodierung	5
1.1.2 Hochrechnungs-/Gewichtungsvariablen.....	8
1.1.3 SILC-Indizes	8
1.2 Anonymisierungsmaßnahmen.....	9
1.2.1 Substichprobenziehung	9
1.2.2 Identifikatoren	11
1.2.3 Gesperrte Merkmale	12
1.2.4 Vergrößerungen von Merkmalsausprägungen	13
1.3 Methodik der Verknüpfung mehrerer MZ Jahrgänge	14
2 Produkt	16
2.1 Hinweise zur Qualität des Mikrozensus 2024	16
2.2 Merkmale und Merkmalsbeschreibung	16
2.2.1 Merkmalsdefinitionen	16
2.2.2 Datensatzbeschreibung	17
2.3 Vergleichbarkeit der Merkmale über die Zeit	18
2.3.1 Variablennamen.....	18
2.3.2 Migrationstypisierungen	19
2.3.3 Einkommen und Lebensbedingungen (SILC).....	19
2.3.4 Informations- und Kommunikationstechnologien (IKT)	20
2.3.5 Zusatzprogramm.....	20
2.3.6 Ad-hoc-Modul	21
2.3.7 SILC-Zusatzprogramme.....	22
2.3.8 IKT-Zusatzprogramme	23
2.4 Eckwerte relevanter Merkmale.....	24
2.4.1 Verteilung der Bevölkerung in Hauptwohnsitzhaushalten....	25

2.4.2	Privathaushalte am Hauptwohnsitz	25
2.5	Auswertbare regionale Ebenen.....	26
2.6	Produktversionen	26
3	Praktische Hinweise	27
3.1	Hinweise zur Geheimhaltung	27
3.2	FAQ	28
3.2.1	Wann wird welcher Hochrechnungsfaktor verwendet?	28
3.2.2	Wieso unterscheiden sich die hochgerechneten Gesamtfallzahlen zwischen verschiedenen Variablen bzw. zwischen verschiedenen Hochrechnungsfaktoren?	31
3.2.3	Warum beträgt die Anzahl der Befragungen (etwa 700.00) mehr als 0,7% der deutschen Bevölkerung?	31
3.2.4	Das Statistische Bundesamt hat die Berechnung der Armutsgefährdungsquote umgestellt. Kann ich mit den FDZ-Daten die alte Berechnung nachbilden?	31
3.2.5	Die IKT-Unterstichprobe ist freiwillig. Warum gibt es im MZ 2024 keine -9er-Fehlwerte?	32
3.3	Verfügbare Tools	32

Abkürzungsverzeichnis

AWB	Auswahlbezirk
BStatG	Bundesstatistikgesetz
CAWI	Computergestütztes Webinterview
ESeC	European Socioeconomic Classification
ESeG	European Socioeconomic Groups
FDZ	Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder
GML	German Microdata Lab
GU	Gemeinschaftsunterkunft
GWAP	Gastwissenschaftsarbetsplatz
IKT	Europäische Statistik zu Informations- und Kommunikationstechnologien
ISEI	Internationaler Sozioökonomischer Index des beruflichen Status
KDFV	Kontrollierte Datenfernverarbeitung
LFS	Europäische Arbeitskräfteerhebung (Labor Force Survey)
MDR	Metadatenreport
MISSY	Mikrodaten-Informationssystem der GESIS
MZ	Mikrozensus
SILC	Europäische Erhebung über Einkommen und Lebensbedingungen (Statistics on Income and Living Conditions)
SUF	Scientific-Use-Files

Einleitung

Dieser Metadatenreport soll Forschenden dabei helfen, die Daten des Mikrozensus Scientific Use Files (MZ-SUF) 2024¹ 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, über eine kontrollierte Datenfernverarbeitung oder per Remote Access auszuwerten. Produktspezifische Informationen zur [On-Site-Nutzung](#) werden im Metadatenreport Teil II zur On-Site-Nutzung bereitgestellt. Für den [Remote Access SUF](#) existiert ein modifiziertes Schlüsselverzeichnis, aber kein eigenständiger MDR.

Weitere allgemeine Informationen zum MZ sind auf den Seiten des Statistischen Bundesamtes ([Was ist der Mikrozensus?](#)) und der [Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder](#) (FDZ) 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 F36 – 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 Vollzähligkeits- und Vollständigkeitskontrolle unterzogen, bei der geprüft wird, ob alle Erhebungsmerkmale und die dazugehörigen typisierten Merkmale, die für das Mikrozensusjahr vorgesehen sind, belegt sind. Nicht belegte Merkmale werden aus dem Datenmaterial entfernt. Gleichzeitig wird darauf geachtet, dass die einzelnen Variablen die korrekten Formate aufweisen. Die Datenaufbereitung findet seit dem MZ 2022 überwiegend mit dem Statistikprogramm R statt. Für die Arbeit mit R und dem R-Datensatz ist zwingend das Paket „[haven](#)“ zu laden, da sonst nicht mit den gelabelten Daten gearbeitet werden kann. Datensätze werden neben dem R-Datensatz auch in spezifischen Formaten für die Programme SPSS und Stata angeboten.²

Mit Ausnahme der IKT-Unterstichprobe fließen nur vollständige Haushalte in die einzelnen Erhebungsteile ein. Das bedeutet, dass bei fehlenden Angaben mindestens einer Person aus einem Haushalt die Angaben des gesamten Haushalts aus dem entsprechenden Befragungsteil entfernt werden.

1.1.1 Missingkodierung

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

² Aufgrund fehlender Nachfrage wird ab dem MZ 2024 kein SAS-Datensatz mehr bereitgestellt. Bei Bedarf können Nutzende den Stata-Datensatz eigenständig in SAS importieren. Bei Fragen zum Import oder bei Problemen können sich die Nutzenden an das [Forschungsdatenzentrum der Statistischen Ämter der Länder – Standort NRW](#) wenden.

Missingcode	Beschreibung
-7	Auswahlbezirke (AWB) ohne Befragung
-1	Nicht erfragt, da wohnhaft in einer Gemeinschaftsunterkunft (GU)
-4	Nicht in der jeweiligen Unterstichprobe oder Hochrechnungsfaktor in der jeweiligen Unterstichprobe nicht belegt auch: nicht im Kernprogramm, wenn ausschließlich Ausprägungen in IKT oder SILC ³
-3	Keine Angabe aufgrund einer Altersgrenze: Person unter 15 Jahren (Kern und LFS), unter 16 Jahren (SILC), unter 16 Jahren oder älter als 74 Jahre (IKT), bzw. unter 15 Jahren oder über 34 Jahre (LFS lang) Filtervariablen (nicht im MZ-SUF enthalten): ➤ <i>dc0202p</i> ➤ <i>ea0100p</i> ➤ <i>ed1803p</i> ➤ <i>ea1201p</i> ➤ <i>dg0400p</i> ➤ <i>ep1205p</i> ➤ <i>er1001p</i> ➤ <i>es1403p</i>
-2	Keine Angabe aufgrund von Nichterwerbstätigkeit Filtervariablen: ➤ <i>ea0502p</i> ➤ <i>ea0800p</i> ➤ <i>ea0700p</i> ➤ <i>ea0802p</i>
-6	Nebenwohnsitze (für SILC und IKT)

³ Eine Erklärung dazu findet sich in der entsprechenden FAQ unten auf der Homepage der FDZ [„Wieso gibt es in den Daten ab 2022 Personen, bei denen nur die SILC- oder IKT-Variablen belegt sind, aber nicht die Kernvariablen?“](#)

	Personen am Nebenwohnsitz entfallen in der SILC- und in der IKT-Unterstichprobe. Bei Haushaltsvariablen entfallen im SILC-Teil Haushalte, in denen keine Person ab 16 Jahren den Hauptwohnsitz hat. Filtervariablen (abgesehen von aa0200p nicht im MZ-SUF enthalten): ➤ aa0200p ➤ aa0500h ➤ ba0400h ➤ ca0400h ➤ ca0702h ➤ dc0201p ➤ eb0800p ➤ ed1801p ➤ ea1200p ➤ dg0302h ➤ er1000p ➤ es1402p
-5	spezifischer Filterausfall, Ausfallgrund wird variablenspezifisch benannt ⁴
-9	Filterfehler oder technisch zugelassene Fehlwerte bei Fragen mit Auskunftspflicht, keine Angabe bei Fragen ohne Auskunftspflicht

Tabelle 1: Ausfallkennungen

Im MZ-SUF vor 2020 wurde zusätzlich der Missingcode -8 für Leerstand und ausgefallene Privathaushalte vergeben. Im Mikrozensus 2024 wird dieser nicht programmiert, da im Datenmaterial 2024 keine Informationen über ausgefallene Haushalte (Leerstand bzw. unbewohnte Wohnungen) vorliegen. Anders als 2020 sind ab 2021 aber wieder AWB im Datenmaterial, die zur Bruttostichprobe gehören, aber keine Befragungsergebnisse aufweisen (z.B. da keine Wohnbevölkerung vorhanden war oder alle Personen im AWB Unit-Non-Response aufwiesen). Hieraus ergeben sich Vorteile für die Erstellung eines Panels bestehend aus mehreren Jahrgängen des MZ ([vgl. Abschnitt 1.3](#)) und bei der

⁴ 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 Daten reproduzieren lassen.

Substichprobenziehung (vgl. [Abschnitt 1.2.1](#)). Für solche AWB wird bei Erhebungsvariablen und daraus abgeleiteten Variablen der Ausfallcode -7 verwendet.

Item-Non-Response bei freiwilligen Fragen wird seit 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

Der neue Standardhochrechnungsfaktor lautet ab 2022 *hr000tr*. Nähere Informationen zur geänderten Methodik sind in den FAQ zur Hochrechnung [im Abschnitt 3.1](#) und [auf der Homepage der FDZ](#) zu finden. Die Hochrechnungsfaktoren werden so angeglichen, dass sie ohne zusätzliche Multiplikation (mit 1000) auf die Gesamtbevölkerung hochrechnen. 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.

Ab dem MZ 2023 sind alle verfügbaren Kern- und LFS-Hochrechnungsfaktoren auf den fortgeschriebenen Bevölkerungszahlen basierend auf dem Zensus 2022 berechnet⁵.

Fälle ohne einen Hochrechnungsfaktor werden aus dem Datenmaterial entfernt. Dabei bleiben jedoch (wie oben beschrieben) AWB der Bruttostichprobe, in denen keine Einzeldatensätze vorliegen, als eigene Fälle im Datenmaterial enthalten.

1.1.3 SILC-Indizes

Ab dem Berichtsjahr 2020 enthält der MZ den Erhebungsteil zu SILC. Zur Sicherstellung der faktischen Anonymität können die meisten SILC-Merkmale nicht für die Off-Site-Nutzung im SUF bereitgestellt werden. Daher werden für das SUF mehrere Merkmale zu jeweils einem Index zusammengefasst.

Folgende SILC-Indizes stehen 2024 zur Verfügung:

- *index_finanz_situation*
- *index_lebenssituation*
- *index_kinderbetreuung_std*
- *index_zahlungsrueckstaende*

⁵ Für das Berichtsjahr 2021 und 2022 stehen sowohl Hochrechnungsfaktoren basierend auf dem Zensus 2011, als auch Hochrechnungsfaktoren basierend auf dem Zensus 2022 zur Verfügung.

- *index_oeffentliche_zahlungen*
- *index_arztkosten*
- *index_situation_kinder*
- *index_arztkosten_kind*

Den Indizes liegen grundsätzlich SILC-Zielvariablen⁶ zugrunde. In Fällen, in denen keine entsprechenden Zielvariablen vorliegen, werden die Erhebungsmerkmale verwendet. Alle für die Berechnung der einzelnen Indizes verwendeten Erhebungs- oder Zielvariablen sind dem [Schlüsselverzeichnis des MZ-On-Site](#) zu entnehmen. Weitere Informationen zur Bildung der Indizes finden sich zudem im [Datenhandbuch des MZ-SUF](#). Liegen den Indizes freiwillige Fragen zugrunde, ist mit erhöhtem Item-Non-Response und entsprechenden Ausfällen bei den Indizes zu rechnen.

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

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)

⁶ Zielvariablen werden gemäß den Vorgaben von Eurostat aus den Erhebungsvariablen abgeleitet und zusätzlich aufbereitet. Sie werden für alle inhaltlichen Auswertungen empfohlen und auch für amtliche Veröffentlichungen verwendet.

- Kennung über Grundauswahl und Aktualisierung der AWBs (awbangroessenklasse)
- Kennung über die Unterstichprobe (awbauswahlteil)
- Regierungsbezirk (gbtregierungsbezirk)
- Regionale Anpassungsschicht (awblfdnranpassungsschichtstich)
- Regionale Schicht (awblfdnrregschichtstichprobe)
- Regionale Untergruppe (awblfdnrreguntergruppestich)
- 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). 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. Mit dem Berichtsjahr 2021 wurde das System auf eine Substichprobenziehung des Bruttodatensatzes umgestellt, um robuster gegenüber Ausfällen zu sein.

Durch die Umstellung der Substichprobenziehung sind leere Auswahlbezirke (d.h. Auswahlbezirke, die aufgrund von Ausfällen keine befragten Haushalte enthalten) im MZ-SUF enthalten. Diese AWBs werden durch eine Beobachtung dargestellt, die lediglich Angaben zu den Merkmalen `id`, `idawb`, `land`, `awbnummerfremd`, `gemeindegroessenklasse`, `awbangroessenklasse` und `substichprobe` enthält. In den übrigen Merkmalen weisen sie die Ausprägung „-7 Entfällt (AWB ohne befragte Haushalte)“ auf.

Berücksichtigung des Längsschnitts bei der Substichprobenziehung ab 2022

Ab dem Berichtsjahr 2020 sind längsschnittliche Verknüpfungen der Querschnitte von Auswahlbezirken, Haushalten und Personen mit bis zu 3 Folgejahren möglich, da Haushalte in bis zu 4 aufeinander folgenden Jahren befragt werden.⁷ Diese Rotation und die Möglichkeit zur Panelbildung wird in der Substichprobenziehung berücksichtigt.

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

⁷ Eine Zusammenführung mit den Querschnittsdaten vor dem SUF 2020 ist aufgrund der Vergabe neuer Identifikationsnummern im Mikrozensus 2020 nicht möglich.

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 Forschende 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.⁸

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
Regionalkennungen	gbtkreis, gbtregierungsbezirk, awblfdnranpassungsschichtaktue, awblfdnrregschichtaktuell, awblfdnrreguntergruppeaktue, awbangroessenklasse
Gemeinde der Arbeitsstätte	ec0303p
Stichprobenkennzeichen	awbstichprobennr, awbrotationsviertel, awbrvquartalskennzeichen,
Raumeinheit (nicht-administrativ)	stadt_land_glied_eu, regionstyp_grund, regionstyp_diff, kreistyp, arbeitsmarktregion, raumordnungsregion, planungsregion, arbeitsagenturbezirk, gemeindetyp, verdichtungsraum, verflechtungsbereich_zentr
Merkmale aus Erhebungsteil SILC	Es werden nur ausgewählte SILC-Merkmale in den MZ-SUF integriert
Merkmale aus Erhebungsteil IKT	Es werden nur Zielvariablen in den MZ-SUF integriert

Tabelle 2: Gesperrte Merkmale

⁸ Im Remote-Access-SUF sind einige dieser Merkmale enthalten, weitere Infos hierzu finden Sie im entsprechenden [Schlüsselverzeichnis](#).

Seit dem Erhebungsjahr 2022 werden Zielvariablen⁹ aus dem Erhebungsteil zu den Informations- und Kommunikationstechnologien (IKT) 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. Der MZ-SUF beinhaltet Merkmale, die zur SILC-Unterstichprobe gehören, die zuvor bis einschließlich des Berichtsjahres 2019 Teil des Kern-MZ waren. Seit dem Berichtsjahr 2020 werden SILC-Merkmale der Beobachtungseinheit zu Indizes zusammengefasst und in den MZ-SUF integriert (vgl. [Abschnitt 1.1.3](#)). Seit dem Berichtsjahr 2021 werden zudem einige SILC-Zielvariablen aufgenommen, von denen ein sehr niedriges Deanonymisierungsrisiko ausgeht. Zudem werden ab dem Berichtsjahr 2023 Variablen aufgenommen, die im aktuellen Jahr lediglich in der SILC-Unterstichprobe erhoben werden, in vorangegangenen Jahren jedoch zusätzlich auch im Kernfragebogen enthalten waren.

1.2.4 Vergrößerungen von Merkmalsausprägungen

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

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

⁹ Zu den Erhebungsteilen SILC und IKT gibt es neben den Erhebungsmerkmalen auch Zielvariablen. Zielvariablen werden gemäß den Vorgaben von Eurostat aus den Erhebungsvariablen abgeleitet und zusätzlich aufbereitet. Sie werden für alle inhaltlichen Auswertungen empfohlen und auch für amtliche Veröffentlichungen verwendet.

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 *tpbevltkrg*, *tpgeschlecht* und *ep0800p*).

1.3 Methodik der Verknüpfung mehrerer MZ Jahrgänge

Der MZ bietet die Möglichkeit, auf Individualebene die Wellen ab 2020 zu einem Paneldatensatz zu verknüpfen. Hierfür stehen verkettete Identifikatoren für AWB (*idawb*), Haushalte (*idhh*) und Personen (*idpers*) zur Verfügung.

Unterjährige Wiederholungsbefragungen im Rahmen der Arbeitskräfteerhebung (LFS) lassen sich über die vorgesehene Fragebogenart (*awbauswahlteil* = 4) identifizieren. Sofern aus methodischen Gründen keine Dubletten innerhalb eines Erhebungsjahres gewünscht sind, ist eine Entfernung dieser Fälle erforderlich. Dann ist jedoch zu beachten, dass die Hochrechnungsfaktoren *hr100qq* und *hr100jq* sowie die Hochrechnungsfaktoren für die Kernvariablen nicht für eine Hochrechnung auf die Gesamtbevölkerung genutzt werden können, da relevante Fälle entfernt wurden. Der Hochrechnungsfaktor für LFS-Strukturvariablen *hr100jj* ist dagegen bei Wiederholungsbefragungen unbelegt und kann zur Hochrechnung verwendet werden. In den ebenfalls enthaltenen querschnittsorientierten Identifikatoren (*idawbx*, *idhhx*, *idpersx*) sind die Wiederholungsbefragungen bereits mit abweichenden IDs versehen.

Nähere Informationen zur Panelverknüpfung sind [Brocker und Mühlenfeld 2020](#) sowie [Herter-Eschweiler und Schimpl-Neimanns 2018](#) und [Backes und Pforr 2025](#) zu entnehmen. Die Arbeitspapiere beziehen sich auf die Erhebungsjahre 2012 bis 2015 bzw. 2016 bis 2019, die darin enthaltenen Erkenntnisse sind aber in Teilen auch auf die Jahre ab 2020 übertragbar.

2 Produkt

2.1 Hinweise zur Qualität des Mikrozensus 2024

„Das neue Mikrozensus-System hat sich seit der Neugestaltung im Jahr 2020 mittlerweile etabliert. Das Erhebungsjahr 2024 lief wie bereits 2023 ohne Einschränkungen bei der Erhebungsdurchführung. Die Ausfallquote kann auf zwei unterschiedliche Weisen gemessen werden: Zum einen kann die Bevölkerungsschätzung auf Basis der Zensusergebnisse 2022 als Bruttowert herangezogen werden. Eine Veröffentlichung der Ausfallquote auf dieser Basis findet seit 2020 statt. Alternativ kann auch eine [sic!] systeminternes Brutto gebildet werden - also die Anzahl der in der Feldarbeit vor Ort registrierten Haushalte. Diese Referenz [sic!] wurden in den Veröffentlichungen der Ausfallquote bis 2020 herangezogen. Bei den Endergebnissen aus dem Mikrozensus 2024 auf Bundesebene liegt die Ausfallquote unter Heranziehung der Bevölkerungsvorausschätzung bei ca. 7,1 % und bei einer Ausrichtung auf das systeminterne Brutto bei ca. 5,7% und bleibt damit auf einem ähnlichen Niveau wie 2023.“ ([Qualitätsbericht - Mikrozensus 2024 \(Kapitel 4.1 – Absätze 5 und 6\)](#)).

Die Ausfallquoten für die freiwilligen Fragen sind in der Regel deutlich höher.

Grundsätzlich gilt, dass die Ergebnisse des MZ ab dem Jahr 2020 aufgrund methodischer Veränderungen (verschiedene Unterstichproben und damit einhergehende veränderte Hochrechnungsverfahren, unterjährige Rotation einzelner Teile, verändertes Fragenprogramm) und aufgrund des Einsatzes von Online-Fragebögen (CAWI) zur Datenerhebung nur eingeschränkt mit den Ergebnissen der Jahre vor 2020 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](#), sowie (in englischer Sprache) auf den Seiten von Eurostat zur [Unterstichprobe SILC](#) und zur [Unterstichprobe IKT](#).

2.2 Merkmale und Merkmalsbeschreibung

2.2.1 Merkmalsdefinitionen

Folgende Klassifikationen finden im MZ-SUF 2024 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>

- Internationale Standardklassifikation der Berufe, Ausgabe 2008 (ISCO-08) (4-Steller):
<https://www.ilo.org/publications/international-standard-classification-occupations-2008-isco-08-structure> (englisch)
- Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008) (3-Steller):
<https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/klassifikation-wz-2008.html>
- Bildungsklassifikation International Standard Classification of Education, Ausgabe 2011 (ISCED-2011):
<https://unesdoc.unesco.org/ark:/48223/pf0000219109.locale=en> (englisch)
- Bildungsfelder (Hauptfachrichtungen) ISCED Fields of Education and Training (ISCED-F 2013):
https://www.statistik.at/KDBWeb/kdb_DownloadsAnzeigen.do?KDBtoken=ignore&&AUFRUF=klass&&NAV=DE&&KLASSID=10527&&KLASSNAME=ISCED-F
(Übersetzung der Klassifikation von Statistik Austria)
- Staatsangehörigkeits- und Gebietssystematik:
<https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/staatsangehoerigkeit-gebietsschluesel.html>

2.2.2 Datensatzbeschreibung

Der MZ ist eine Haushalts- und Personenstatistik. Die vollständige Liste der im MZ-SUF vorhandenen Variablen, ist dem [Datenhandbuch](#) zu entnehmen.¹⁰ 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 Randverteilungen (ungewichtete Fallzahlen) der Variablen, erläuternde Kommentare und die vollständigen Label (das Statistikprogramm SPSS übernimmt bei den Wertelabels nur 120 Zeichen; Stata übernimmt bei den Variablenlabels nur 80 Zeichen¹¹) enthalten.

Aufgrund der 2020 eingeführten Neukonzeption des MZ, welche die Integration weiterer Haushaltserhebungen¹² sowie Neuerungen¹³ bei der europäischen Arbeitskräfteerhebung

¹⁰ 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 Typisierungen mithilfe der in den Daten vorliegenden Angaben selbst erstellt werden können.

¹¹ Ab dem MZ-SUF 2022 werden in Stata die vollständigen Variablenlabels in den notes hinterlegt.

¹² Ab 2020: Europäische Statistik der Einkommens- und die Lebensbedingungen (SILC), ab 2021: Europäische Haushaltsbefragung zum privaten Einsatz von Informations- und Kommunikationstechnologien (IKT-Haushalte)

¹³ Unterjährige Wiederholungsbefragungen sowie Konzeption als Unterstichprobe des MZ

(LFS) umfasst, besteht die Mikrozensus-Befragung seit 2020 aus einem Kernfragenprogramm, das alle befragten privaten Haushalte beantworten, und weiteren Erhebungsteilen, welche jeweils lediglich von einer Unterstichprobe der Haushalte beantwortet werden. Das Kernprogramm und die verschiedenen Erhebungsteile werden nicht nacheinander erhoben, sondern das resultierende Frageprogramm verzahnt die Inhalte thematisch. Für einen kompakten Gesamtüberblick über alle enthaltenen Fragen und die Struktur der verschiedenen Fragebögen wurde ein [Masterfragebogendokument](#) erstellt, welches die Informationen aus allen [sieben Fragebögen](#) für Privathaushalte enthält und angibt, welche Frage in welchem Fragebogen enthalten ist. Zudem enthält das Dokument Informationen zu den korrespondierenden Variablennamen, welche im MZ-SUF enthalten sind (Variablen, die auf mehreren Fragen beruhen werden im Masterfragebogen nicht angegeben).

Aus dem SILC-Erhebungsteil werden nur ausgewählte Informationen in den MZ-SUF aufgenommen. Informationen aus dem IKT-Erhebungsteil werden seit dem Erhebungsjahr 2022 in den MZ-SUF integriert. Nähere Informationen zu den Inhalten und der Methodologie von SILC sind der englischsprachigen Publikation [EU statistics on income and living conditions \(EU-SILC\) methodology](#) zu entnehmen. Nähere Informationen zu Inhalt und Methodologie der IKT können Sie dem Methodological Manual unter [ICT usage in households and by individuals](#) entnehmen.

2.3 Vergleichbarkeit der Merkmale über die Zeit

Vor dem Hintergrund der methodischen Neuerungen des MZ 2020 wird von Veränderungsanalysen zu den Jahren vor 2020 generell abgeraten. Alle Neuregelungen zum MZ ab 2020 können auf [der Website des StBA](#) nachgelesen werden. Zudem müssen bei Veränderungsanalysen zwischen 2020/2021 und nachfolgenden Jahren die im [Metadatenreport 2021](#) beschriebenen Einschränkungen der Datenqualität der MZ 2020 und 2021 berücksichtigt werden.

2.3.1 Variablennamen

Im Vergleich zu den Erhebungsjahren vor 2022 werden Variablennamen konsequent kleingeschrieben.

Variablennamen von Erhebungsmerkmalen aus dem Kern- und LFS-Fragebogen 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*.

Die SILC-Zielvariablen beginnen mit *d* (Haushaltsregister), *r* (Personenregister), *h* (Haushaltsdaten) oder *p* (Personendaten). Danach folgen ein weiterer Buchstabe und drei Ziffern. Die Benennung der IKT-Zielvariablen folgt keinem einheitlichen Schema.

Außerdem gibt es im Vergleich zu den Erhebungsjahren vor 2020 einen anderen Umgang mit Haupt- und Nebenwohnsitzhaushalten, insbesondere im SILC-Befragungsteil: Die Befragungen in den SILC- und IKT-Unterstichproben 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 Migrationstypisierungen

Die Zahlen zur Bevölkerung mit Migrationshintergrund ab dem MZ 2020 sind nur eingeschränkt mit den Jahren vor 2020 vergleichbar. Je spezifischer und kleiner die betrachtete Teilpopulation ist (z. B. als Deutsche Geborene, Eingebürgerte, Differenzierung nach Geburtsland), desto größer sind die Unterschiede zu den Zahlen vor 2020. Hierfür sind insbesondere Umstellungen in der Erhebung/Frageformulierung sowie in der Methodik der Typisierung des Migrationshintergrundes¹⁴ ursächlich. Durch die methodische Weiterentwicklung der Typisierung des Migrationshintergrundes können Personen, die die deutsche Staatsangehörigkeit von Geburt an besitzen, besser abgebildet werden.

Ab 2021 stehen mit *tm0020p* bis *tm0026h* Variablen zur Verfügung, die mit der Einwanderungsgeschichte ein neues Konzept abbilden. Dieses basiert auf einer Empfehlung der Fachkommission der Bundesregierung zu den Rahmenbedingungen der Integrationsfähigkeit (vgl. [StBA 2023](#)). Anders als beim Migrationshintergrund, der sich aus der Staatsangehörigkeit zum Geburtszeitpunkt ableitet, ist bei der Einwanderungsgeschichte die tatsächliche Wanderungserfahrung das ausschlaggebende Kriterium.

2.3.3 Einkommen und Lebensbedingungen (SILC)

Die früher eigenständige Statistik zu Einkommen und Lebensbedingungen („Leben in Europa“, EU-SILC) ist ab 2020 ein Erhebungsteil des MZ. Bis zu 12% der ausgewählten Haushalte im

¹⁴ 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 nicht im selben Haushalt wohnenden 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.

MZ werden neben dem Kernprogramm zusätzlich auch zu Einkommen, Wohnen, Gesundheit und Lebensbedingungen befragt. Durch den Wechsel der Erhebung von einer freiwilligen zu einer in Teilen auskunftspflichtigen Befragung ist ein inhaltlicher Vergleich der Daten der Erhebungsjahre ab 2020 mit Vorjahren nicht sinnvoll.

Grundsätzlich besteht im Rahmen des MZ eine Auskunftspflicht. Allerdings wird aufgrund der fortgesetzten Freiwilligkeit einiger Fragen ein hoher Anteil an fehlenden Werten (Missings) in diesen erzeugt. Im Rahmen der Generierung von Zielvariablen erfolgen daher zahlreiche Imputationen. Ob eine Imputation durchgeführt wurde, ist dem Wert der korrespondierenden Flagvariable (gekennzeichnet durch Variablennamenendung „_f“) zu entnehmen.

Die bereitgestellten SILC-Zielvariablen werden aus Gründen der Konsistenz mit den EU-Veröffentlichungen und der schon vorher stattgefundenen Plausibilisierung nur minimal bearbeitet. Daher kommt es teilweise zu fehlenden -5er-Filtern, hohen -9er Werten und anderen Labels als in den zugehörigen Erhebungsvariablen.

2.3.4 Informations- und Kommunikationstechnologien (IKT)

Die früher eigenständige IKT-Befragung in privaten Haushalten ist ab dem Erhebungsjahr 2021 als freiwillige Unterstichprobe im MZ enthalten. Durch die Umstrukturierung der IKT-Erhebung ist ein Vergleich der Ergebnisse ab dem Erhebungsjahr 2021 mit den Vorjahren nur eingeschränkt möglich. Weitere Informationen hierzu finden sich auf [der Website des StBA zur IKT-Erhebung](#).

Aus Gründen der Konsistenz mit EU-Veröffentlichungen werden bei den IKT-Zielvariablen die Mehrfachantwortfragen anders kodiert als im restlichen Material. Insbesondere sind Ausschlusskategorien (bspw. „keine Angabe“ oder „keines davon“) als eigenständige Variablen im Material enthalten und werden nicht in die Variablen der anderen Antwortmöglichkeiten integriert.

Seit dem Erhebungsjahr 2023 werden alle Antwortausfälle auf Merkmalsebene imputiert. Ob eine Antwort imputiert wurde oder nicht, ist den Flag-Variablen (*Variablenname_f*) zu entnehmen. Weitere Informationen dazu sind dem [Qualitätsbericht](#) der IKT 2024 zu entnehmen.

2.3.5 Zusatzprogramm

Zusätzlich zum Kernprogramm enthält der MZ im Rhythmus von jeweils vier Jahren wiederkehrende Zusatzprogramme. Mit der Neukonzeption der LFS als Unterstichprobe, wurden die Zusatzprogramme mit Ausnahme des Zusatzprogramms Wohnen in diese Unterstichprobe ausgelagert. Dies hat aufgrund des damit verbundenen geringeren

Auswahlsatzes negative Konsequenzen für die fachliche und regionale Differenzierbarkeit der Zusatzprogramme. Das Zusatzprogramm 2024 enthält „Fragen zu Pendlereigenschaften von Schülern und Schülerinnen, Studenten und Studentinnen sowie Erwerbstätigen“.

- dc1100p Schule/Hochschule: Lage in der Wohnsitzgemeinde
- tp_dc1100p Schule/Hochschule: Lage im selben Bundesland wie Wohnsitz
- dc1201p Schule/Hochschule: Bundesland
- dc1300p Schule/Hochschule: Hinweg ab Wohnsitz
- dc1400p Schule/Hochschule: Entfernung
- dc1500p Schule/Hochschule: Zeitaufwand Hinweg
- dc1600p Schule/Hochschule: Verkehrsmittel Hinweg
- dc1700p Schule/Hochschule: weiteres Verkehrsmittel: Ja/Nein
- dc1800p Schule/Hochschule: weiteres Verkehrsmittel
- ec2100p Arbeitsstätte: Hinweg ab Wohnsitz
- ec2200p Arbeitsstätte: Distanz Hinweg ab Wohnsitz
- ec2300p Arbeitsstätte: Dauer Hinweg ab Wohnsitz
- ec2400p Arbeitsstätte: Verkehrsmittel Hinweg
- ec2700p Arbeitsstätte: Weiteres Verkehrsmittel Hinweg: Ja/Nein
- ec2701p Arbeitsstätte: Weiteres Verkehrsmittel Hinweg

2.3.6 Ad-hoc-Modul

Im Rahmen des in den MZ integrierten 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 2024 erhebt das Ad-hoc-Modul Informationen zu „Ausbildungs- und Arbeitserfahrungen junger Menschen“. Die Fragen des Ad-hoc-Moduls erhalten nur Haushalte der LFS-Unterstichprobe, die gemäß Rotationsschema erstmals befragt werden (Fragebogen 3). Ihre Beantwortung ist freiwillig. Ab dem Berichtsjahr 2022 bekommen alle Personen, die gemäß Stichprobenplan Fragebogen 3 beantworten sollten, den entsprechenden Hochrechnungsfaktor *hr100mo*, nicht nur die Personen, welche die Fragen auch tatsächlich beantwortet haben.

- ep1205p Alter gruppiert
- ep1206p Schulische oder berufliche Ausbildung begonnen, ohne Abschluss beendet
- ep1207p Art der Ausbildung
- ep1208p Grund für Beendigung der Ausbildung ohne Abschluss
- ep1209p Abgeschlossene Ausbildung
- ep1210p Gegenwärtige berufliche Situation

- *ep1211p* Höchster Bildungsabschluss stimmt mit Anforderungen in aktueller Haupttätigkeit überein
- *ep1212p* Fachrichtung des höchsten Bildungsabschlusses stimmt mit aktueller Haupttätigkeit überein
- *ep1213p* Fähigkeiten stimmen mit Anforderungen in aktueller Haupttätigkeit überein
- *ep1214p* Höchster Bildungsabschluss stimmte mit Anforderungen in letzter Haupttätigkeit überein
- *ep1215p* Fachrichtung des höchsten Bildungsabschlusses stimmte mit der letzten Haupttätigkeit überein
- *ep1216p* Fähigkeiten stimmten mit Anforderungen in letzter Haupttätigkeit überein

2.3.7 SILC-Zusatzprogramme

Zusätzlich zu den jährlichen Variablen wird der SILC-Erhebungsteil durch 3- sowie 6-jährliche Zusatzprogramme und Ad-hoc-Fragen ergänzt. Eine Übersicht, welche Fragen in jedem Jahr ergänzt werden, gibt es auf der [Seite von Eurostat](#). Im Jahr 2024 wurden hier Fragen zu den Themen „Kinder“ (3-jährlich) und „Zugang zu Dienstleistungen“ (6-jährlich) gestellt. Im MZ SUF 2024 sind die folgenden Variablen aus den Zusatzprogrammen enthalten:

Kinder:

- *rc380* Ungedeckter Bedarf an Kinderbetreuung in einer Einrichtung
- *rc390* Hauptgrund warum keine Kinderbetreuung in einer Einrichtung genutzt wird
- *hc040* Zurechtkommen des Haushalts mit Kosten für Kinderbetreuung
- *rch010* Gesundheitszustand (Kind)
- *rch020* Einschränkung der Aktivitäten aufgrund von Gesundheitsproblemen (Kind)
- *index_situation_kinder* Index Situation Kinder
- *index_arztkosten_kind* Index Arztkosten Kind (letzte 12 Monate)

Zugang zu Dienstleistungen:

- *pc330* Sich bei Kontakt mit Ämtern oder öffentlichen Diensten diskriminiert fühlen (i.d.l. 12 Monaten)
- *pc340* Sich bei der Wohnungssuche diskriminiert fühlen (i.d.l. 5 Jahren)
- *pc350* Sich in Bildungseinrichtung diskriminiert fühlen (i.d.l. 12 Monaten)

- *pc360* Sich im Geschäft / im öffentlichen Raum
diskriminiert fühlen (i.d.l. 12 Monaten)

2.3.8 IKT-Zusatzprogramme

Zusätzlich zu den jährlichen Variablen wird der IKT-Erhebungsteil durch unregelmäßige Fragen erweitert. Eine Übersicht, welche Fragen in jedem Jahr ergänzt werden, gibt es auf der [Seite von Eurostat](#). Im Jahr 2024 wurden hier Fragen zu den Themen „Internet of Things (IoT)“ und „Green IT“ gestellt.

Internet of Things (IoT) (i.d.l. 3 Monaten):

- *iot_dem2* Nutzung smarter Geräte: Geräte zum Energiemanagement
- *iot_dsec2* Nutzung smarter Geräte: Sicherheitsgeräte
- *iot_dha2* Nutzung smarter Geräte: Haushalts- oder Gartengeräte
- *iot_dva2* Nutzung smarter Geräte: virtuelle Assistenten
- *iot_dx2* Nutzung smarter Geräte: kein Gerät
- *iot_bdk2* Grund Nichtnutzung smarter Geräte: Existenz nicht bekannt
- *iot_bnn2* Grund Nichtnutzung smarter Geräte: kein Bedarf
- *iot_bcst2* Grund Nichtnutzung smarter Geräte: Kosten zu hoch
- *iot_blc2* Grund Nichtnutzung smarter Geräte: Kompatibilitätsprobleme mit
anderen Geräten
- *iot_blsk2* Grund Nichtnutzung smarter Geräte: fehlende Kenntnis smarte Geräte
bedienen zu können
- *iot_bcpp2* Grund Nichtnutzung smarter Geräte: Bedenken bzgl. Privatsphäre
- *iot_bcsc2* Grund Nichtnutzung smarter Geräte: Bedenken bzgl. IT-Sicherheit
- *iot_bcsh2* Grund Nichtnutzung smarter Geräte: Bedenken bzgl.
gesundheitlicher/persönlicher Sicherheit
- *iot_both2* Grund Nichtnutzung smarter Geräte: sonstige Gründe
- *iot_iutv2* Nutzung smarter Geräte zu Hause: Smart-TV
- *iot_iugc2* Nutzung smarter Geräte zu Hause: Spielekonsole
- *iot_iuha2* Nutzung smarter Geräte zu Hause: Heim-Audio-System/Smart-Speaker
- *iot_psec2* Probleme mit smarten Geräten: Sicherheit/Datenschutz
- *iot_pshe2* Probleme mit smarten Geräten: Sicherheit/Gesundheit
- *iot_pdu2* Probleme mit smarten Geräten: Schwierigkeiten bei Verwendung
- *iot_poth2* Probleme mit smarten Geräten: andere Schwierigkeiten
- *iot_px2* Probleme mit smarten Geräten: keine Probleme

- *iot_dcs2* Nutzung smarter Geräte: smarte Accessoires/Kleidung/Schuhe
- *iot_dhe2* Nutzung smarter Geräte: Gesundheitsgeräte/Pflegeroboter
- *iot_dtoy2* Nutzung smarter Geräte: smartes Spielzeug
- *iot_dcar2* Nutzung smarter Geräte: PKW mit eingebauter Internetverbindung
- *iot_use2* Nutzung smarter Geräte

Green IT:

- *eco_dmob* Verwertung nach Nutzung: Smartphone
- *eco_dlt* Verwertung nach Nutzung: Laptop/Tablet
- *eco_dpc* Verwertung nach Nutzung: Desktop-PC
- *eco_pp* Wichtige Kaufkriterien (Smartphone, Tablet, Laptop, PC): Preis
- *eco_phd* Wichtige Kaufkriterien (Smartphone, Tablet, Laptop, PC):
Hardware-Spezifikationen
- *eco_pecd* Wichtige Kaufkriterien (Smartphone, Tablet, Laptop, PC): Ökodesign
- *eco_peg* Wichtige Kaufkriterien (Smartphone, Tablet, Laptop, PC):
Möglichkeit von zusätzlicher Garantie
- *eco_pee* Wichtige Kaufkriterien (Smartphone, Tablet, Laptop, PC):
Energieeffizienz
- *eco_ptbs* Wichtige Kaufkriterien (Smartphone, Tablet, Laptop, PC):
Ankauf/kostenlose Rücknahme von Alt-Gerät
- *eco_px* Wichtige Kaufkriterien (Smartphone, Tablet, Laptop, PC): kein Kriterium
- *eco_pbx* Wichtige Kaufkriterien (Smartphone, Tablet, Laptop, PC):
kein Gerät gekauft
- *eco_bds* Wichtige Kaufkriterien (Smartphone, Tablet, Laptop, PC):
Marke/Design/Größe

2.4 Eckwerte relevanter Merkmale

Für Standardveröffentlichungen aus dem MZ wird die Abgrenzung nach "Bevölkerung in Hauptwohnsitzhaushalten"¹⁵ mithilfe von *th0201h == 1* verwendet.

Die Variable *tpbevltkrg* kann verwendet werden, um die alte Abgrenzung "Bevölkerung am Ort der Hauptwohnung" zu bilden, die in den Jahren vor 2020 verwendet wurde.

¹⁵ Haushalte, in denen mindestens eine Person ab 16 Jahren am Hauptwohnsitz wohnt.

Die Tabellen im [Abschnitt 2.4.1](#) beziehen sich auf die Personenebene, die Tabellen im [Abschnitt 2.4.2](#) hingegen auf die Haushaltsebene (Hauptwohnsitzhaushalte).

Die hier abgebildeten Tabellen sind in ähnlicher Form der [GENESIS-Online-Datenbank](#) zu entnehmen.

2.4.1 Verteilung der Bevölkerung in Hauptwohnsitzhaushalten

Dichotomes Geschlecht (tpgeschlecht)

value	label	n	percent
1	Männlich	40,860,383	49.42
2	Weiblich	41,817,012	50.58
	Gesamt	82,677,395	100.00

Familienstand (ab0500p)

value	label	n	percent
1	Ledig	35,576,192	43.03
2	Verheiratet	36,130,473	43.70
3	Verwitwet	5,242,904	6.34
4	Geschieden	5,727,825	6.93
	Gesamt	82,677,394	100.00

2.4.2 Privathaushalte am Hauptwohnsitz

Anzahl Haushaltsmitglieder (npershh)¹⁶

value	label	n	percent
1	Einpersonenhaushalt	17,471,722	42.41
2	Zweipersonenhaushalt	13,521,157	32.82
3	Dreipersonenhaushalt	4,803,996	11.66
4	Vierpersonenhaushalt	3,844,730	9.33

¹⁶ In den offiziellen Veröffentlichungen der amtlichen Statistik wird die Variable aa0400h genutzt. Diese ist im SUF nicht verfügbar. Es kann daher zu leichten Abweichungen kommen.

value	label	n	percent
5	Fünf- oder mehr Personenhaushalt	1,557,412	3.78
	Gesamt	41,199,017	100.00

Im Anhang befindet sich das R-Script, mit dem die oben aufgeführten Tabellen auf Personen- und Haushaltsebene erstellt wurden.

2.5 Auswertbare regionale Ebenen

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

2.6 Produktversionen

Die Versionsnummer des Datensatzes ist der Variable *doi* zu entnehmen.

10.21242/12211.2024.00.00.3.1.0

- Erstveröffentlichung

3 Praktische Hinweise

3.1 Hinweise zur Geheimhaltung

Im Rahmen der Geheimhaltung soll sichergestellt werden, dass keine Rückschlüsse auf einen Einzelfall (z.B. Person, Haushalt, Betrieb, Einrichtung) gezogen werden können. Statistische Geheimhaltung wird überall dort angewendet, wo statistische Ergebnisse oder Einzeldaten die geschützten Räume der amtlichen Statistik verlassen.

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

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

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

Ab dem Berichtsjahr 2024 enthält der Metadatenreport Teil II ausschließlich jahresspezifische FAQ. Jahresübergreifende FAQ sind auf den Seiten der [Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder](#) abrufbar.

3.2.1 Wann wird welcher Hochrechnungsfaktor verwendet?

Aufgrund der verschiedenen Unterstichproben stehen verschiedene Hochrechnungsfaktoren zu Verfügung. [Tabelle 3](#) stellt dar, 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. Alle Hochrechnungsfaktoren basieren auf der Bevölkerungsfortschreibung und dem Zensus 2022. 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 die Gesamtbevölkerung der Hochrechnungsfaktor *hr000tr* verwendet. Dieser tiefregionalisierte Hochrechnungsfaktor löst seit dem MZ 2022 den Standardhochrechnungsfaktor *hr000jj* ab. Die Berechnung folgt der in [Schmidt und Stein 2021](#) beschriebenen bisherigen Methodik, wobei der Anpassungsrahmen nun aber durch die regionalen Untergruppen (gekreuzt mit Staatsangehörigkeit, Geschlecht und regionaler Anpassungsschicht) ergänzt wird. 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. Analog zum Kernprogramm stehen auch für die LFS-Merkmale Hochrechnungsfaktoren für die Quartale (*hr100qq*) und den Quartalsdurchschnitt (*hr100jq*) zur Verfügung. Hierbei ist

vor allem zu beachten, dass Haushalte im LFS-Programm mehrfach im Jahr befragt werden ([siehe oben](#)). Bei der Auswertung von Merkmalen, die unterjährig wiederholt erhoben werden (Fragebögen 2, 3 und 4) empfiehlt es sich, mit *hr100jq* hochzurechnen, sodass nicht nur das Ergebnis der Erstbefragung in die Auswertung eingeht. Mit *hr100mo* steht ein weiterer Hochrechnungsfaktor explizit für das LFS-Ad-Hoc-Modul zur Verfügung. Für Auswertungen zu Menschen mit Behinderung steht der Hochrechnungsfaktor *hr100bh* ab dem Erhebungsjahr 2022 nicht mehr zur Verfügung, da für eine Weiterführung der Methodik verlässliche Angaben zum Anteil der in Einrichtungen lebenden behinderten Menschen fehlen. Stattdessen empfiehlt es sich, für Auswertungen zu Menschen mit Behinderung den Hochrechnungsfaktor *hr100jj* zu verwenden.

- Für die SILC-Substichprobe wird der Hochrechnungsfaktor *hr200jj* verwendet. Werden Auswertungen für Kinder unter 13 Jahren erstellt, so kann der neue Hochrechnungsfaktor *hr200ki* verwendet werden.
- Die IKT Merkmale können mit den Hochrechnungsfaktoren *hr300pn* und *hr300hh* hochgerechnet werden. Dabei ist *hr300pn* für Auswertungen auf Personenebene und *hr300hh* für Auswertungen auf Haushaltsebene zu verwenden.

Hochrechnungs- faktor	Stich- probe	Summe Personen	Anmerkungen
hr000tr	Kern	84.631.044	Zur Verwendung bei Berechnungen mit Jahresergebnissen. Abweichungen zu den anderen Summen folgen aus dem 2-Stufen-Vorgehen zur Bestimmung der Hochrechnungsfaktoren.
hr000jq	Kern	82.785.280	Zur Verwendung bei Berechnungen mit Quartalsdurchschnitt zur Jahresauswertung.
hr000qq	Kern	331.141.121	Zur Verwendung bei Berechnungen mit Quartalsergebnissen. Die Summe der Quartalsergebnisse entspricht dem vierfachen des Quartalsdurchschnitts.
hr100jj	LFS	82.602.959	Zur Verwendung bei Berechnung mit Jahresergebnissen.

hr100jq	LFS	82.817.251	Zur Verwendung bei Berechnungen mit Quartalsdurchschnitt zur Jahresauswertung.
hr100qq	LFS	331.269.004	Zur Verwendung bei Berechnungen mit Quartalsergebnissen. Die Summe der Quartalsergebnisse entspricht dem vierfachen des Quartalsdurchschnitts.
hr100mo	LFS	18.252.172	Zur Verwendung bei Berechnungen von Variablen des LFS-Ad-Hoc-Moduls „Ausbildungs- und Arbeitserfahrungen junger Menschen“. Befragt werden Personen von 15 bis 34 Jahren. Es wird auf den entsprechenden Anteil der Bevölkerung hochgerechnet.
hr200jj	SILC	82.047.342	Zur Verwendung bei Berechnungen des SILC-Querschnitts.
hr200ki	SILC	10.115.840	Zur Verwendung bei Berechnungen zu Kindern unter 13 Jahren. Es empfiehlt sich, SILC-Hochrechnungsfaktoren auf die Zielvariablen anzuwenden.
hr300pn	IKT	62.221.675	Zur Verwendung bei Auswertungen zur IKT auf Personenebene. Befragt werden Personen von 16 bis 74 Jahren. Es wird auf den entsprechenden Anteil der Bevölkerung hochgerechnet
hr300hh	IKT	54.898.690	Zur Verwendung bei Auswertungen zur IKT auf Haushaltsebene. Wird der Hochrechnungsfaktor auf Haushaltsebene angewendet, dann ergibt sich eine Summe von 34.125.696.

Tabelle 3: Hochrechnungsfaktoren mit Gesamtbevölkerungszahlen der hochgerechneten Jahresergebnisse. Angaben sind auf ganze Zahlen gerundet.

3.2.2 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 im [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 wird eigenständig hochgerechnet.

3.2.3 Warum beträgt die Anzahl der Befragungen (etwa 700.00) mehr als 0,7% 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 2024](#) unter Kapitel 3 entnommen werden.

3.2.4 Das Statistische Bundesamt hat die Berechnung der Armutsgefährdungsquote umgestellt. Kann ich mit den FDZ-Daten die alte Berechnung nachbilden?

Mit den FDZ-On-Site-Daten können Armutsgefährdungsquoten sowohl auf Basis des MZ-Kernprogramms als auch auf Basis der EU-SILC berechnet werden. Die amtliche Statistik veröffentlicht seit 2025 nur noch die Quote auf Basis der EU-SILC-Daten, um für die breite Öffentlichkeit einheitliche Werte zu kommunizieren. Mit der Berechnung auf Basis von EU-

SILC wird ein europaweit einheitlicher Messwert erzeugt, was einen konsistenten EU-Vergleich ermöglicht.

Aufgrund der detaillierten Abfrage der einzelnen Einkommensbestandteile in EU-SILC wird das Einkommen von den Befragten vollständiger angegeben (zum Beispiel inklusive staatlicher Leistungen wie Kindergeld, Kinderzuschlag, BAföG, Pflegegeld oder Wohngeld) und auch nicht monatlich anfallende Einkommen oder kleinere Beträge werden erfasst. Dies ist ein entscheidender qualitativer Vorteil von den Einkommensangaben in EU-SILC gegenüber dem MZ-Kernprogramm. Hinzukommt, dass die Einkommensangaben in den EU-SILC-Daten detaillierter erfasst sind als die Einkommensangaben im Kernmaterial. Während die Angaben im Kernmaterial in 24 Klassen angegeben sind, sind die EU-SILC-Daten auf den Euro gerundet erfasst.

Zur Berechnung der Armutsgefährdungsquote wird das Haushaltseinkommen des Vorjahres bestimmt und mittels eines Gewichtungsschlüssels auf die Personen des Haushaltes verteilt. Dabei erhält die Hauptbezugsperson das Gewicht 1. Weitere Erwachsene und Kinder ab 14 Jahren erhalten das Gewicht 0,5, Kinder unter 14 Jahren das Gewicht 0,3. Das verfügbare Haushaltseinkommen wird nun durch die Summe der Gewichte dividiert. Der ermittelte Quotient wird als bedarfsgewichtetes Äquivalenzeinkommen bezeichnet. Anschließend wird das mittlere Haushaltseinkommen mittels des Medians bestimmt. 60 % dieses Medianwertes stellen den Schwellenwert für Armutsgefährdung dar.

Detailliertere Informationen dazu sind auf der [Homepage des Statistischen Bundesamts](#) zu finden.

3.2.5 Die IKT-Unterstichprobe ist freiwillig. Warum gibt es im MZ 2024 keine -9er-Fehlwerte?

Alle Antwortausfälle auf Merkmalsebene wurden in der IKT 2024 vom Bundesamt imputiert. Ob es sich bei einer Beobachtung um eine geschätzte Angabe handelt oder nicht, ist den IKT Flag-Variablen (*Variablenname_f*) zu entnehmen. Siehe dazu auch Kapitel 2.3.4.

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 R, SPSS und Stata auf dem [Mikrodaten-Informationssystem der GESIS \(MISSY\)](#) zur Verfügung.

Anhang

R-Skript zu Abschnitt 2.4

```
# load packages

library(haven)

library(dplyr)

library(tidyr)

library(labelled)

library(flextable)

library(officer)


freq <- function(data, var, weight) {

  # create frequency table

  table <- data %>%

    count(!sym(var), wt = !!sym(weight)) %>%

    mutate(

      percent = round((n / sum(n) * 100),2),

      label = haven::as_factor(!sym(var)),

      value = as.numeric(!sym(var)),

      n = round(n,0)

    ) %>%

    select(value, label, n, percent)


  # add row "Gesamt"

  table <- table %>%

    add_row(value = 0, label = "Gesamt", n=sum(table$n), percent=sum(table$percent))


  #create flextable from table

  ft <- flextable(table) %>%

    set_caption(get_variable_labels(data[[var]])) %>%

    theme_vanilla() %>%

    autofit()


  return(ft)

}
```

```

load('path/to/file/MZ 2024 (SUF).Rdata')

mz <- SUF

#filter: Personen in Hauptwohnsitzhaushalten
mzPers <- mz[which(mz$th0201h == 1),]

#recode household member [5 or more]
mz$HHMitglieder <- mz$npershh
mz$HHMitglieder[which(mz$HHMitglieder >= 5)] <- 5
mz$HHMitglieder <- set_value_labels(mz$HHMitglieder,c("Einpersonenhaushalt" = 1,
               "Zweipersonenhaushalt" = 2,
               "Dreipersonenhaushalt" = 3,
               "Vierpersonenhaushalt" = 4,
               "Fünf- oder mehr Personenhaushalt" = 5))

#recode ab0500p [eingetragene Lebenspartnerschaft into Ehe]
mzPers[which(mzPers$ab0500p == 5), "ab0500p"] <- 2
mzPers[which(mzPers$ab0500p == 6), "ab0500p"] <- 3
mzPers[which(mzPers$ab0500p == 7), "ab0500p"] <- 4

#filter: Hauptwohnsitzhaushalte
mzHH <- mz[which(mz$th0201h == 1 & mz$th0602p == 1 & mz$tpprivathh == 1),]

ftgender <- freq(mzPers, "tpgeschlecht", "hr000tr")
ftfamily <- freq(mzPers, "ab0500p", "hr000tr")
ftpersons <- freq(mzHH, "HHMitglieder", "hr000tr")

#Write the tables into a word document...
doc <- read_docx() %>%
  body_add_flextable(ftgender) %>%
  body_add_flextable(ftfamily) %>%
  body_add_flextable(ftpersons)

#... and save it.
print(doc, target = "my_table.docx")

```

