

## Mikrozensus Scientific Use File 2007: Dokumentation und Datenaufbereitung

*Florian Thirolf, Delia Jäger, Malte Langeheine,  
Yvonne Lechert, Freya Lohel, Tony Siegel,  
Bernhard Schimpl-Neimanns*



GESIS-Technical Reports 2010|05

## Mikrozensus Scientific Use File 2007: Dokumentation und Datenaufbereitung

*Florian Thirolf, Delia Jäger, Malte Langeheine,  
Yvonne Lechert, Freya Lohel, Tony Siegel,  
Bernhard Schimpl-Neimanns*

## GESIS-Technical Reports

GESIS – Leibniz-Institut für Sozialwissenschaften  
Postfach 12 21 55  
68072 Mannheim  
Telefon: (0621) 1246 - 263  
Telefax: (0621) 1246 - 100  
E-Mail: [bernhard.schimpl-neimanns@gesis.org](mailto:bernhard.schimpl-neimanns@gesis.org)

ISSN: 1868-9043 (Print)  
ISSN: 1868-9051 (Online)  
Herausgeber,  
Druck und Vertrieb: GESIS - Leibniz-Institut für Sozialwissenschaften  
Lennéstraße 30, 53113 Bonn

# Mikrozensus Scientific Use File 2007: Dokumentation und Datenaufbereitung

Florian Thirolf, Delia Jäger, Malte Langeheine, Yvonne Lechert, Freya Lohel, Tony Siegel,  
Bernhard Schimpl-Neimanns

## Zusammenfassung

Die Aufbereitung und Dokumentation des faktisch anonymisierten Einzelmaterials des Mikrozensus als Scientific Use File erfolgt durch das Statistische Bundesamt, Gruppe VIII C - Mikrozensus, in Kooperation mit dem German Microdata Lab der GESIS. In dem vorliegenden Methodenbericht werden das Erhebungsprogramm und der Stichprobenplan des Mikrozensus im Allgemeinen sowie die Besonderheiten des Mikrozensus 2007 skizziert, der die Zusatzprogramme Angaben zur Krankenversicherung sowie Zusatzangaben zur Erwerbstätigkeit und das Ad-hoc-Modul Arbeitsunfälle, Arbeit und Gesundheit der Arbeitskräfteerhebung der EU umfasst. Außerdem werden die Datenaufbereitung und Substichprobenziehung sowie die Qualitätsprüfung des Scientific Use Files beschrieben.

*Schlüsselwörter: Mikrozensus, Scientific Use File.*

**Title: German Microcensus Scientific Use File 2007: Documentation and Data Preparation**

## Summary

The processing of the factually anonymized microdata of the Microcensus into a scientific use file and its documentation is carried out by the German Federal Statistical Office, Group VIII C - Microcensus in cooperation with the German Microdata Lab of GESIS. The following report gives a general outline of the Microcensus' survey program and sampling plan. Furthermore, the unique features of the Microcensus 2007 will be discussed. The survey 2007 includes the supplementary programs for information on health insurance, additional data on employment and the ad hoc module on accidents at work and work-related health problems made available by the EU Labour Force Survey. In addition, a description of the processing of data, sub-sampling and the quality testing of the scientific use file is provided within this report.

*Keywords: German Microcensus, Scientific Use File.*



## 1 Einleitung

---

Seit dem Bundesstatistikgesetz von 1987 (§ 16, Abs. 6, BStatG) können anonymisierte Individualdatenfiles der amtlichen Statistik durch die statistischen Ämter an unabhängige inländische Forschungseinrichtungen weitergegeben werden. Das Scientific Use File (SUF) ist eine faktisch anonymisierte 70 %-Substichprobe der Haushalte des Mikrozensus, wobei alle Personen eines ausgewählten Haushalts enthalten sind. Statt einer Haushaltssubstichprobe wird in den Erhebungsjahren, in denen das Zusatzprogramm zur Wohnsituation (z. B. 2002 und 2006) durchgeführt wird, eine Wohnungssubstichprobe gezogen. Das SUF ist eine systematische Zufallsauswahl aus den Originaldaten des Statistischen Bundesamtes. Das Ziehungsverfahren orientiert sich eng am Stichprobendesign des Mikrozensus. Das Konzept der faktischen Anonymität bezieht sich auf das sogenannte Grundfile, das neben der Ziehung einer Substichprobe die eingeschränkte Weitergabe von Regionalinformationen vorsieht. Im SUF werden nur die Länder und vergrößerte Angaben zur Gemeindegrößenklasse ausgewiesen. Des Weiteren werden Merkmalsausprägungen vergrößert, weil Staatsangehörigkeiten oder Gruppen von Nationalitäten in der Grundgesamtheit mindestens 50.000 Personen umfassen müssen. Für andere Merkmale liegt dieses Anonymisierungskriterium bei 5.000 Fällen.

Die Aufbereitung und Dokumentation des SUF erfolgt durch das Statistische Bundesamt, Gruppe VIII C – Mikrozensus, in Kooperation mit dem German Microdata Lab (GML), dem Servicezentrum für amtliche Mikrodaten der GESIS. Durch die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder werden Rohdaten im ASCII-Format sowie Daten im dokumentierten Format (Systemfiles) der am häufigsten verwendeten Statistikprogramme PASW (früher: SPSS), SAS und STATA bereitgestellt.

Um die Datendokumentation und -auswertungsmöglichkeiten für die Nutzer des SUF weiter zu verbessern, sind ab der Erhebung 2007 zwischen der Mikrozensusgruppe des Statistischen Bundesamtes und GML die Schritte zur Datenaufbereitung und der -dokumentation neu gestaltet worden. Ein Ergebnis ist die Bereitstellung von Systemfiles, außerdem wird das bisherige Schlüsselverzeichnis durch ein neues Datenhandbuch mit umfassenden Informationen zu den Variablen ersetzt. Diese und weitere Informationen stehen auf den Web-Seiten von GESIS zur Verfügung.<sup>1</sup>

In diesem Bericht werden im Folgenden nach einem Überblick zum Mikrozensus und zum SUF 2007 die Datenaufbereitung und Schritte zur Prüfung der Daten beschrieben, die wichtigsten Veränderungen im Vergleich zum SUF 2006 benannt und die Verwendung der amtlichen Bevölkerungskonzepte sowie die Hochrechnung beispielhaft erläutert.

---

<sup>1</sup> Weitere Informationen sind auf den Web-Seiten des Statistischen Bundesamtes (siehe u. a. [Statistik von A bis Z](#)) und der Forschungsdatenzentren der Statistischen Ämter (siehe [Datenangebot, Mikrozensus](#)) des Bundes und der Länder abrufbar.

## 2 Kurzbeschreibung des Mikrozensus

---

Der Mikrozensus ist die amtliche Repräsentativitätsstatistik über die Bevölkerung und den Arbeitsmarkt in Deutschland mit einem Auswahlsatz von einem Prozent der Haushalte und der Bevölkerung. Sie wird seit 1957 in Westdeutschland und seit 1991 auch in den neuen Bundesländern durchgeführt. Der Mikrozensus dient der Bereitstellung statistischer Informationen über die wirtschaftliche und soziale Lage der Bevölkerung sowie über die Erwerbstätigkeit, den Arbeitsmarkt und die Ausbildung. Er schreibt die Ergebnisse der Volkszählung fort. Darüber hinaus dient er der Rationalisierung anderer amtlicher Statistiken. Die Stichprobenerhebung über Arbeitskräfte in der Europäischen Union (EU Labour Force Survey), einschließlich der in diesem Rahmen durchgeführten Ad-hoc-Module, ist in den Mikrozensus integriert. Im Folgenden werden nur die Haupteigenschaften skizziert (für Details siehe den [Qualitätsbericht zum Mikrozensus 2007](#) des Statistischen Bundesamtes (2008a).

**Grundgesamtheit** Zur Grundgesamtheit zählt die gesamte wohnberechtigte Bevölkerung in Privathaushalten und Gemeinschaftsunterkünften am Haupt- und Nebenwohnsitz. Nicht zur Erhebungsgesamtheit gehören Angehörige ausländischer Streitkräfte sowie ausländischer diplomatischer Vertretungen mit ihren Familienangehörigen. Personen ohne Wohnung (Obdachlose) haben im Mikrozensus aufgrund der Anlage als Flächenstichprobe keine Erfassungschance.

**Berichtszeitraum** Seit 2005 wird das Befragungsvolumen gleichmäßig über alle Kalenderwochen des Jahres verteilt (kontinuierliche Erhebung). Der Berichtszeitraum ist die letzte Kalenderwoche vor der Befragung (gleitende Berichtswoche).

**Fragenprogramm** Das Fragenprogramm des Mikrozensus besteht aus einem gleich bleibenden Grundprogramm, das sozioökonomische Grundinformationen liefert und die laufende Beobachtung des Arbeitsmarktes ermöglicht.

Neben den jährlich zu erhebenden Merkmalen des Grundprogramms werden in vierjährigen Zusatzprogrammen weitere Fragen gestellt. Das Zusatzprogramm umfasst 2007 Angaben zur Krankenversicherung sowie Zusatzangaben zur Erwerbstätigkeit.

Darüber hinaus wurde im Rahmen des EU Labour Force Survey 2007 das Ad-hoc-Modul mit Fragen zum Thema „Arbeitsunfälle, Arbeit und Gesundheit“ durchgeführt. Der Auswahlsatz des Ad-Hoc-Moduls beträgt 0,1 %.

**Stichprobenplan** Der Mikrozensus ist eine einstufig geschichtete Flächen- oder Klumpenstichprobe, das heißt, aus dem Bundesgebiet werden Auswahlbezirke ausgewählt, in denen alle Haushalte und Personen befragt werden. Sowohl die Auswahlbezirke als Primäreinheiten als auch die Sekundäreinheiten (Haushalte und Personen) haben die gleiche Auswahlwahrscheinlichkeit von einem Prozent. Die Auswahlbezirke der Grundausswahl wurden aus Ergebnissen der Volkszählung 1987 gebildet; für die neuen Bundesländer wurde auf der Basis des Bevölkerungsregisters Statistik eine vergleichbare Auswahlgrundlage erstellt.

Die Stichprobe der Grundausswahl wird durch Berücksichtigung der Neubautätigkeit über die Meldungen zur Bautätigkeitsstatistik aktualisiert (Neubauauswahl). Jährlich werden ein Viertel aller in der Stichprobe enthaltenen Auswahlbezirke ausgetauscht.



Die aus einem Auswahlbezirk wegziehenden Haushalte und Personen werden nicht weiter befragt, sondern durch zuziehende Haushalte und Personen ersetzt. Der Mikrozensus ist somit eine Wiederholungsbefragung mit teilweiser Überlappung der Erhebungseinheiten (partielle Rotation).

**Erhebungsverfahren** Die Interviews werden i. d. R. als mündliche Befragung mit Hilfe eines Laptops (CAPI) durchgeführt. Die Haushalte können auch selbst einen schriftlichen Fragebogen ausfüllen. Zulässig sind Proxy-Interviews, d. h., ein erwachsenes Haushaltsmitglied darf stellvertretend für andere Haushaltsmitglieder antworten (Fremdauskünfte für Personen ab 15 Jahren: ca. 27 %). Für die ausgewählten Haushalte und die darin lebenden Personen besteht gemäß der gesetzlichen Regelungen Auskunftspflicht. Einige Fragen sind freiwillig zu beantworten.

**Hochrechnung** Bei der Hochrechnung wird ein zweistufiges Verfahren verwendet. Im ersten Schritt der Ausfallkorrektur wird von der Nettostichprobe der erfolgreich befragten Haushalte auf die Bruttostichprobe aller zu befragenden Haushalte hochgerechnet. Der Unit-Nonresponse beträgt im Mikrozensus seit 2005 ca. fünf Prozent der Haushalte. Im zweiten Schritt wird diese Bruttostichprobe an Eckzahlen aus der laufenden Bevölkerungsfortschreibung angepasst. Ab 2005 werden dabei auch Ergebnisse des Ausländerzentralregisters verwendet. Die Anpassung erfolgt quartalsweise auf unterschiedlichen regionalen Ebenen. Mit Ausnahme des Ad-hoc-Moduls liegen für alle Merkmalsbereiche gleiche Hochrechnungsfaktoren vor.

### 3 Das Scientific Use File Mikrozensus 2007

---

Wie bereits einleitend erwähnt, ist das Scientific Use File (SUF) eine faktisch anonymisierte 70%-Substichprobe; i. d. R. eine Haushaltssubstichprobe.<sup>2</sup> Das Ziehungsverfahren orientiert sich eng am Stichprobendesign des Mikrozensus.

Die Haushaltssubstichprobe ist eine systematische Zufallsauswahl aus den Originaldaten. Die Daten werden dazu nach Berichtsquartal, Bundesland, Regierungsbezirk, Gemeindegrößenklasse, Anzahl der Personen im Haushalt, Gebäudegrößenklasse, Auswahlbezirksnummer und Haushaltsnummer sortiert. (Das Merkmal Gebäudegrößenklasse wurde im SUF 2006 erstmals verwendet.) Durch diese Anordnung ist gewährleistet, dass die Substichprobe hinsichtlich dieser Merkmale nur geringe zufallsbedingte Abweichungen zum Originalmaterial aufweist.

Bei Gemeinschaftsunterkünften erhält jede Person eine eigene fortlaufende „Haushalts“-Nummer. Zur Stichprobenziehung dient die letzte (Einer-) Stelle der Haushaltsnummer. Alle Haushalte, deren Einerstelle nicht mit drei zufällig gezogenen einstelligen Zahlen übereinstimmen, gelangen nach diesem sogenannten Schlussziffernverfahren in die Stichprobe (70 % der Haushalte). Nicht enthalten sind sogenannte Nullbezirke, das sind Auswahlbezirke ohne Befragte (z. B. leer stehende Häuser). Diese werden vor der Substichprobenziehung gelöscht.

Weiterhin unterscheidet sich das SUF von den Originaldaten des Mikrozensus dadurch, dass bestimmte Variablen im SUF, bedingt durch die Anonymisierung, in klassierter und vergrößerter Form verfügbar sind.<sup>3</sup> So sind im SUF die Regionalangaben Bundesland und eine vergrößerte Gemeindegrößenklasse, die nicht für die kleinen Länder Bremen und Saarland ausgewiesen ist, enthalten. Mithilfe eines speziellen Codes der Variablen Gemeindegrößenklasse (EF563=9) kann zwischen West- und Ost-Berlin unterschieden werden. Auch bei den weiteren Variablen des SUF werden je nach Bedarf Merkmalsvergrößerungen vorgenommen, sodass jede Ausprägung in den univariaten Verteilungen mindestens 5.000 Fälle in der Grundgesamtheit aufweist. Die Ausprägungen der Variable Nationalität werden so aggregiert, dass jede Nationalitätengruppe in der Grundgesamtheit wenigstens 50.000 Einwohner umfasst. Im Fall von Vergrößerungen wird im SUF i. d. R. die am stärksten besetzte Kategorie ausgewiesen.

Für Schätzungen des Stichprobenfehlers enthält das SUF Informationen zum Stichprobendesign (Schimpl-Neimanns 2010): Mit den Variablen Bundesland (EF1) und Gebäudegrößenklasse (EF570) sowie den Angaben zum Berichtsquartal (EF12) stehen Schichtungsmerkmale zur Verfügung. Die Information zur Klumpung liegt in Form der Auswahlbezirksnummer (EF3) vor. Des Weiteren können die Hochrechnungsfaktoren für Quartals- und Jahresdurchschnittsergebnisse (EF951 und EF952) sowie für die Merkmale der 0,1 % (Ad-hoc-) Substichprobe (EF955) genutzt werden. Diese Hochrechnungsfaktoren sind bereits mit dem Kehrwert der Auswahlwahr-

---

<sup>2</sup> In Erhebungsjahren des Zusatzprogramms mit Fragen zur Wohnsituation wird im Unterschied zu den anderen Erhebungen nicht eine Haushaltssubstichprobe, sondern eine 70%-Substichprobe der Wohnungen gezogen.

<sup>3</sup> Zusätzlich zu den im SUF enthaltenen Variablen können im Rahmen eines Aufenthalts als Gastwissenschaftler in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder weitere Variablen genutzt werden. Die Auflistung dieser Variablen findet sich am Ende der thematischen Übersicht.

scheinlichkeit der Substichprobe (100/70) multipliziert und rechnen auf 1.000 der Population hoch.

### 3.1 Die wichtigsten Veränderungen gegenüber 2006

Um Nutzern, die bereits mit dem SUF früherer Erhebungszeitpunkte gearbeitet haben, die Auswertung zu erleichtern, werden im Folgenden die wichtigsten Veränderungen gegenüber dem MZ 2006 skizziert.

Der MZ 2007, und damit auch das SUF, enthalten im Unterschied zur Erhebung 2006 keine sogenannten Jahresüberhänge; das waren Haushalte bzw. Personen, die im Erhebungszeitraum (2005) i. d. R. aufgrund schwerer Erreichbarkeit nicht, sondern erst im Folgejahr (2006) befragt werden konnten (Statistisches Bundesamt 2009b: 4). In diesem Zusammenhang ist auch darauf hinzuweisen, dass die damit einhergehende ungleiche Verteilung des Befragungsvolumens auf die Zeiträume teilweise zu erheblichen Verzerrungen bei Haushaltsergebnissen führt.

Um insbesondere die Erfassung des ILO-Erwerbsstatus zu verbessern, sind seit der Erhebung 2005 einige Fragen zur Arbeitsmarktbeteiligung (siehe insbes. EF77-EF86) und die Feldarbeit verändert worden. Verteilungsunterschiede können deshalb sowohl mit Veränderungen des Arbeitsmarktes als auch methodisch mit geänderten Fragen zusammenhängen (siehe Köhne-Finster und Lingnau 2009; Körner und Puch 2009). Auch im Vergleich zwischen 2007 und 2006 weisen einige Fragen einen anderen Wortlaut auf bzw. enthalten andere Antwortkategorien.

Im Rahmen der 0,1%-Substichprobe (siehe EF6) des Ad-hoc-Moduls „Arbeitsunfälle, Arbeit und Gesundheit“ (EF1040-EF1051) sind 2007 weitere ergänzende Fragen ohne Auskunftspflicht zur Erwerbstätigkeit gestellt worden. Dies betrifft die Variablen:

- EF118 Leitungstätigkeit (F127)
- EF177 Arbeitsvertrag: Zeitarbeitsvermittlungsfirma (F126)
- EF178 Arbeitsvermittlung: Agentur für Arbeit (F128)
- EF179 Nichtarbeitsuche oder Teilzeitbeschäftigung: Hauptgrund (F129)

Um die Analysemöglichkeiten des SUF weiter zu verbessern, wurden 2007 bei einigen Merkmalen die Kriterien der Zusammenfassung modifiziert. Dies ist bei zeitlich vergleichenden Auswertungen zu berücksichtigen.

In Tabelle 1 sind die von Änderungen betroffenen Variablen aufgeführt. Die Änderungen sind in eckigen Klammern und kursiver Schrift kurz kommentiert. Weitergehende Hinweise können auch dem Datenhandbuch entnommen werden, das in der detaillierten Variablenbeschreibung ebenfalls erläuternde Kommentare mit Hinweisen zur zeitlichen Vergleichbarkeit sowie den Fragetext des Selbstausfüllerbogens bzw. des CAPI-Interviews enthält.

Tabelle 1: Variablen im Mikrozensus 2007 und 2006 mit Änderungen

	Variable	Fragennummer	
		2007	2006
EF1	Bundesland [Code im MZ 2007 (wie bis 2004) gemäß amtlichem Gemeindeschlüssel; im MZ 2005 und 2006 nach alphabetischer Sortierung.]		
EF10	Baujahr der Wohnung [Im MZ 2006 (Zusatzprogramm Wohnsituation: EF494) differenzierter erhoben.]	19	122
EF49	Familienstand [Neue Merkmalskategorien: „eingetragener Lebenspartner verstorben“; „eingetragene Lebenspartnerschaft aufgehoben“] Siehe auch EF735, EF757, EF765, EF823, EF847	8	8
EF77	Bezahlte Tätigkeit (Berichtswoche)	21	20
EF78	Elternzeit, (Sonder-) Urlaub o.Ä. (Berichtswoche)	21b	21
EF79	Mithelfender Familienangehöriger oder Gelegenheitstätigkeit (Berichtswoche)	22	22
EF80	Sonstige bezahlte Tätigkeit	23	23
EF81	Geringfügige Beschäftigung (Berichtswoche)	24	24
EF82	Geringfügige Beschäftigung (Berichtsw.): einzige oder hauptsächl. Erwerbstätigk.	25	24a
EF83	Frühere Erwerbstätigkeit	27	26
EF85	Erwerbs- oder Berufstätigkeit (Berichtsw.)	21a	
EF86	Geringfügige Beschäftigung (Berichtsw.): Elternzeit, (Sonder-) Urlaub o.Ä. [Neue Frage seit 2007]	24a	
EF172	Überw. ausgeübte Tätigkeit [Zusatzprogramm 2007]	55	58
EF173	Betrieb: Gliederung in Abteilungen [Zusatzprogramm 2007]	56	
EF174	Abteilung des Arbeitsplatzes [Zusatzprogramm 2007]	56a	
EF175	Stellung im Betrieb [Zusatzprogramm 2007]	56b	
EF176	Überw. ausgeübte Tätigkeit: Computernutzung [Zusatzprogramm 2007]	55a	
EF190	Arbeitsstätte: Staat (teilw. Zusammenfassung von Unterregionen) [Neue Zusammenfassungen im SUF nach den Kriterien: 1. Aufnahmedatum in die EU, 2. Räumliche Lage.] Siehe auch EF455	57c	54c
EF203	2. Erwerbstätigk.: Beruf (KldB92) [Neue Zusammenfassungen im SUF zur Vermeidung von Zusammenfassungen über Berufsgruppen hinweg. Sofern eine Berufsgruppe zusammengefasst wurde, wird der Code der Berufsgruppe (Zweisteller) angegeben.]	61	58
EF214	2. Erwerbstätigk.: Wirtschaftszweig Betrieb (WZ03) [Neue Zusammenfassungen im SUF zur Vermeidung von Zusammenfassungen über WZ-Abteilungen hinweg. Sofern eine WZ-Abteilung zusammengefasst wurde, wird der Code der Abteilung (Zweisteller) angegeben.]	62	59
EF244	Arbeitslosengeld I oder II [Neuer Fragentext: Hartz IV]	76	73

Variable	Fragennummer	
	2007	2006
EF312 Höchster berufl. Ausbildungs- oder Hochschul-/ Fachhochschulabschluss [Kategorie 11 „Vorbereitungsdienst für den mittleren Dienst in der öffentlichen Verwaltung“ 2007 extra; bis 2006 in Kategorie 3 „Abschluss einer Lehrausbildung ...“ enthalten.] Siehe auch EF316, EF744, EF832, EF856	91a	88a
EF313 Höchster berufl. Abschluss: Hauptfachrichtung (HFR03) [Neue Zusammenfassungen im SUF nach den Kriterien: 1. ISCED-Ausbildungsfelder, 2. ISCED-Bildungsfelder]	91b	88b
EF328 Allg./Berufl. Weiterbildung (letzte): Fachrichtung [Neue Zusammenfassungen im SUF nach den Kriterien: 1. ISCED-Ausbildungsfelder, 2. ISCED-Bildungsfelder, 3. ISCED breitere Felder.]	97	94
EF330 Allg./Berufl. Weiterbildung (im letzten Jahr): Stunden [Differenziertere Darstellung im SUF 2007]	95a	92a
EF369 1. Staatsangehörigkeit [Neue Zusammenfassungen im SUF nach den Kriterien: 1. Aufnahmedatum in die EU, 2. Räumliche Lage] Siehe auch EF374, EF733, EF755, EF821, EF845	101a	98a
EF372 Erwerb deutsche Staatsangehörigk.: Geburt/Einbürgerung/Spätaussiedler [Änderung der Frage und Merkmalsausprägungen]	103	100
EF373 Einbürgerung/Zuzug als Spätaussiedler: Jahr [Änderung der Frage]	103a	100a
EF401 Überwiegender Lebensunterhalt [Neue Antwortkategorien] Siehe auch EF741, EF829, EF853	105	104
EF421 Öffentl. Zahlung: Wohngeld [Neue Antwortkategorien] Siehe auch EF422-EF426, EF439	109	108
EF521 Letzte Erwerbstätigk.: Beruf nach ISCO-88 COM [Für die Umschlüsselung von der Klassifikation der Berufe zu ISCO fehlen für den früheren Beruf Angaben zur Betriebsgröße. Siehe zum Schätzverfahren die Ausführungen in Klassifikation der Berufe (KldB-92) in der Signierung für den Mikrozensus - Statistisches Bundesamt, Ausgabe 1992 - in Verbindung mit dem Umsteigerschlüssel für die Internationale Standardklassifikation der Berufe zur Anwendung in der Europäischen Gemeinschaft (ISCO-88 COM) [ <a href="#">pdf</a> ].]		

## 4 Aufbereitung der Rohdaten zum Mikrozensus Scientific Use File 2007 und Überprüfung der Plausibilität des erstellten Systemfiles

---

Als Ergebnis der Datenaufbereitung liegen Rohdaten im numerischen ASCII-Format sowie nun auch sogenannte Systemfiles der Statistikprogramme PASW (früher SPSS), SAS und STATA im dokumentierten Format vor. Das frühere Schlüsselverzeichnis ist durch ein Datenhandbuch mit zusätzlichen Informationen (Fragetext, Kommentare zu Änderungen (s. o.) und ungewichtete Häufigkeitsauszählungen) ersetzt worden. Somit können Einzelforscher nach Erhalt der Daten umgehend mit den eigentlichen Analysen beginnen. Bei der Erstellung der maschinenlesbaren Datendokumentation wurden die Variablen- und Wertelabels<sup>4</sup> sowie die thematische Übersicht nach der Systematik des Mikrodaten-Informationssystems (MISSEY) erstellt. Die Variablenlabels enthalten die Nummer der entsprechenden Frage im Selbstausfüllerbogen, den grundsätzlichen inhaltlichen Bezug der Variable und eventuelle Spezifizierungen zu dieser Variablen; außerdem sind Freiwilligkeit und Zugehörigkeit zur Unterstichprobe gekennzeichnet. Diese Standardisierung und Anpassung an internationale Dokumentationsstandards für sozialwissenschaftliche Daten unterstützt auch die Replizierbarkeit von Analysen. Des Weiteren werden im GESIS-Web wie bisher Fragebogen und Klassifikationen sowie Mikrodaten-Tools zur Umsetzung internationaler Klassifikationen oder Skalen bereitgestellt.

Schließlich erfolgen im Rahmen der Datenaufbereitung Plausibilitätskontrollen, die nicht nur Einzelforschern zur Kontrolle dienen können, ob die erhaltenen Daten fehlerfrei verarbeitbar sind. Ergänzende Beispiele im GESIS-Web zur Abgrenzung der amtlichen Bevölkerungsbegriffe und den für die Hochrechnung verwendeten Gewichtungsvariablen zeigen zugleich, wie mit Auswertungen des SUF zentrale Eckwerte aus den amtlichen Veröffentlichungen des Mikrozensus replizieren werden können.

Als Referenz zur Plausibilitätsprüfung der Mikrozensus SUF dienen die in den Fachserien des Statistischen Bundesamtes veröffentlichten Ergebnisse des jeweiligen Mikrozensus. Für den Mikrozensus 2007 sind dies insbesondere die Fachserie 1 Bevölkerung und Erwerbstätigkeit: Reihe 3 (Haushalte und Familien), Reihe 4.1.1 (Stand und Entwicklung der Erwerbstätigkeit), Reihe 4.1.2 (Beruf, Ausbildung und Arbeitsbedingungen der Erwerbstätigen) und Reihe 2.2 (Bevölkerung mit Migrationshintergrund) sowie Fachserie 13 Sozialleistungen, Angaben zur Krankenversicherung. Bei diesen Vergleichen sind die von den Statistischen Ämtern verwendeten Bevölkerungs- und Erwerbskonzepte sowie die Hochrechnung zu beachten (vgl. die Seiten Abgrenzung und Hochrechnung der Bevölkerungsbegriffe im Mikrozensus 2007 sowie die Definitionenkataloge für die Merkmale des Grundprogramms und das Zusatzprogramm Krankenversicherung im GESIS-Web).

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

---

<sup>4</sup> Geschlechtsspezifische Berufsbezeichnungen sind in den Setups zum Mikrozensus ab 2006 - im Unterschied zu den Setups früherer Erhebungen - ausschließlich aus Platzgründen und zur besseren Lesbarkeit in der männlichen Form genannt.

Tabelle 2: Bevölkerungsbegriffe im SUF 2007 (Jahresdurchschnitt in 1.000, hochgerechnet mit EF952)

Bevölkerung am Haupt- und Nebenwohnsitz <sup>a</sup> (EF30>=1 & EF30<=3)  83.213		
Bevölkerung am Hauptwohnsitz <sup>b</sup> (EF30>=1 & EF30<=2)  82.250	Bev. am Nebenwohnsitz <sup>c</sup> (EF30=3) 963	
Bevölkerung in Privathaushalten <sup>d</sup> (EF31=1)  82.373		Bev. in Gemeinschaftsun- terk. <sup>e</sup> (EF31=2) 840
Bevölkerung in Lebensformen am Hauptwohnsitz <sup>f</sup> [Bezugsperson (BP) am Hauptwohnsitz] (EF761>=1 & EF761<=2)  81.544	- [BP am Nebenwohns.] (EF761>2) 829	

- a Zur Bevölkerung am Haupt- und Nebenwohnsitz (früher: „wohnberechtigte Bevölkerung“), zählen alle in der Gemeinde wohnhaften Personen. (Mehrfachzählungen sind möglich.)
- b Zur Bevölkerung am Hauptwohnsitz („Bevölkerung“) zählen alle Personen mit nur einer Wohnung (EF30=1) sowie Personen mit mehreren Wohnungen am Ort ihrer Hauptwohnung (EF30=2).
- c Zur Bevölkerung am Nebenwohnsitz zählen Personen mit mehreren Wohnungen, die am Nebenwohnsitz befragt wurden.
- d Zur Bevölkerung in Privathaushalten zählen alle Personen, die am Haupt- oder Nebenwohnsitz allein oder zusammen mit anderen Personen eine wirtschaftliche Einheit (Privathaushalt) bilden. (Mehrfachzählungen sind möglich.)
- e Zur Bevölkerung in Gemeinschaftsunterkünften gehören alle Personen, die dort wohnen, bzw. gemeldet sind, und nicht für sich wirtschaften, d. h. keinen eigenen Haushalt führen. (Ohne Personen, die sich nur vorübergehend (weniger als drei Monate) dort aufhalten.)
- f Die Bevölkerung in Lebensformen am Hauptwohnsitz umfasst keine Personen in Gemeinschaftsunterkünften. Die Lebensform wird als zusammengehörige Einheit betrachtet, unabhängig davon, ob einzelne Personen eventuell einen weiteren Wohnsitz haben. (Für die Typisierung ist der (Haupt-) Wohnsitz der Bezugsperson relevant. Potenzielle Mehrfachzählungen werden dadurch vermieden.)

Tabelle 3: Verteilungen zur Prüfung der Systemfiles bzw. des Ablaufs der Setups

	Bevölkerung in Privathaus- halten	Bevölkerung am Ort der Hauptwoh- nung	Bevölkerung in Lebensfor- men am Hauptwohnsitz	Bevölkerung a. O. d. Haupt- wohnung, Erwerbstätige	Fallzahlen
Hochrechnung, Gewichtungsvariable	EF952	EF952	EF952	EF952	keine
Insgesamt (in 1.000; außer Fallzahlen)	82.373	82.250	81.544	38.151	483.595
EF117: Stellung im Beruf x EF29: Erwerbstyp; Selektion: Bevölkerung am Ort der Hauptwohnung (ef30>=1 & ef30<=2); (siehe Fachserie 1, Reihe 4.1.1, Band 2, Tabelle 1.1, S. 3)					
Selbstständige		4.147			24.408
Mithelf. Familienangeh.		397			2.319
Beamte (inkl. Soldaten)		2.236			12.758
Angestellte (inkl. kfm./ techn. Auszubilden- de u. Zivildienstleistende)		19.896			116.154
Arbeiter (inkl. gew. Aus- zubildende)		11.475			64.893
Erwerbstätige insges.		38.151			220.532
Erwerbslose		3.624			19.915
Erwerbspersonen		41.775			240.447
Nichterwerbspersonen		40.475			243.148
Insgesamt		82.250			483.595
EF663: Personen im Haushalt: Anzahl (Haushaltstypisierung); Selektion: Bevölkerung in Privathaushalten (EF31=1), Haushaltsbezugsperson [Haupteinkommensbezieher] (EF662>=1 & EF662<=4); (siehe Fachserie 1, Reihe 3, Tabelle 1.1, S. 26)					
1 Person	15.384				88.522
2 Personen	13.519				80.376
3 Personen	5.318				30.360
4 Personen	4.075				23.510
5 Personen und mehr	1.443				8.258
Haushalte insgesamt	39.738				231.026
EF118: Leitungstätigkeit (FB) (0,1 %); Selektion: Unterstichprobe (EF6=2), erwerbstätige Bevölkerung am Hauptwohnsitz (EF29=1 & EF30>=1 & EF30<=2); GewichtungsvARIABLE: ef955 Hochrechnungsfaktor Ad-hoc- Substichprobe: Jahr					
Ja				4.141	2.438
Trifft nicht zu, da selbst- ständig				4.137	2.460
Nein				24.131	14.139
Keine Angabe				5.689	3.276
Insgesamt				38.097	22.313



## Literatur

---

- Afentakis, Anja, und Wolf Bihler, 2005: Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005. *Wirtschaft und Statistik* (10): 1039-1048. URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Querschnittsveroeffentlichungen/WirtschaftStatistik/Mikrozensus/Hochrechnungunterjaehrig.property=file.pdf>.
- Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder: Datenangebot, Mikrozensus. URL: <http://www.forschungsdatenzentrum.de/bestand/mikrozensus/index.asp>.
- Köhne-Finster, Sabine; Lingnau, Andreas, 2009: Untersuchung der Datenqualität erwerbsstatistischer Angaben im Mikrozensus. Ergebnisse des Projekts "Nachbefragung im Mikrozensus/ LFS". *Wirtschaft und Statistik* 12/2008: 1067-1088. URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Querschnittsveroeffentlichungen/WirtschaftStatistik/Mikrozensus/UntersuchungMikrozensus.property=file.pdf>.
- Körner, Thomas; Puch, Katharina, 2009: Der Mikrozensus im Kontext anderer Arbeitsmarktstatistiken. Ergebnisunterschiede und ihre Hintergründe. *Wirtschaft und Statistik* (6): 528-552. URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Querschnittsveroeffentlichungen/WirtschaftStatistik/Arbeitsmarkt/MikrozensusArbeitsmarktstatistiken.property=file.pdf>.
- Schimpf-Neimanns, Bernhard, 2010: Varianzschätzung für Mikrozensus Scientific Use Files ab 2005. *GESIS-Technical Reports 2010/03*. Mannheim. URL: [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_methodenberichte/2010/TechnicalReport\\_10-03.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2010/TechnicalReport_10-03.pdf).
- Statistisches Bundesamt, 2008a: Qualitätsbericht Mikrozensus 2007. Wiesbaden. URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Qualitaetsberichte/Mikrozensus/Mikrozensus2007.property=file.pdf>.
- Statistisches Bundesamt, 2008b: Fachserie 1 Bevölkerung und Erwerbstätigkeit, Reihe 3 Haushalte und Familien. Ergebnisse des Mikrozensus 2007. Wiesbaden. [[Download](#)] URL: <https://www-ec.destatis.de/csp/shop/sfg/bpm.html.cms.cBroker.cls?cmspath=struktur,vollanzeige.csp&ID=1022463>.
- Statistisches Bundesamt, 2008c: Fachserie 1 Bevölkerung und Erwerbstätigkeit, Reihe 4.1.1 Stand und Entwicklung der Erwerbstätigkeit. Mikrozensus 2007. [Verschiedene Bände; [Band 1 Download](#), [Band 2 Download](#)] URL: <https://www-ec.destatis.de/csp/shop/sfg/bpm.html.cms.cBroker.cls?cmspath=struktur,vollanzeige.csp&ID=1022695>; <https://www-ec.destatis.de/csp/shop/sfg/bpm.html.cms.cBroker.cls?cmspath=struktur,vollanzeige.csp&ID=1022696>.
- Statistisches Bundesamt, 2008d: Fachserie 1 Bevölkerung und Erwerbstätigkeit, Reihe 4.1.2 Beruf, Ausbildung und Arbeitsbedingungen der Erwerbstätigen. Band 1: Allgemeine und methodische Erläuterungen. Mikrozensus 2007. Wiesbaden. [Verschiedene Bände; [Band 1 Download](#), [Band 2 Download](#)] URL: <https://www-ec.destatis.de/csp/shop/sfg/bpm.html.cms.cBroker.cls?cmspath=struktur,vollanzeige.csp&ID=1022697>; <https://www-ec.destatis.de/csp/shop/sfg/bpm.html.cms.cBroker.cls?cmspath=struktur,vollanzeige.csp&ID=1022698>.
- Statistisches Bundesamt, 2008e: Fachserie 13 Sozialleistungen, Reihe 1.1 Angaben zur Krankenversicherung. Ergebnisse des Mikrozensus 2007. Wiesbaden. [[Download](#)] URL: <https://www-ec.destatis.de/csp/shop/sfg/bpm.html.cms.cBroker.cls?cmspath=struktur,vollanzeige.csp&ID=1023230>.
- Statistisches Bundesamt, 2009a: Fachserie 1 Bevölkerung und Erwerbstätigkeit, Reihe 2.2 Bevölkerung mit Migrationshintergrund. Ergebnisse des Mikrozensus 2007. Wiesbaden. [[Download](#)] URL: <https://www-ec.destatis.de/csp/shop/sfg/bpm.html.cms.cBroker.cls?cmspath=struktur,vollanzeige.csp&ID=1023127>.
- Statistisches Bundesamt, 2009b: Mikrozensus: Haushaltszahlen ab 2005 [[Methodenpapier](#)]. Wiesbaden. URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/>

Wissenschaftsforum/MethodenVerfahren/Mikrozensus/MikrozensusHaushaltszahlen,property=file.pdf.

Statistisches Bundesamt: Statistik von A bis Z: Mikrozensus. URL: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Presse/abisz/Mikrozensus,templateId=renderPrint.psml>.

Statistisches Bundesamt, GESIS (Hrsg.), 2010a: Thematische Übersicht zum Mikrozensus Scientific Use File 2007. URL: [http://www.gesis.org/dienstleistungen/daten/amtliche-mikrodaten/mikrozensus/grundfile/mz2007/Thematische\\_Uebersicht\\_MZ2007.pdf](http://www.gesis.org/dienstleistungen/daten/amtliche-mikrodaten/mikrozensus/grundfile/mz2007/Thematische_Uebersicht_MZ2007.pdf).

Statistisches Bundesamt, 2010b: Klassifikation der Berufe (KldB-92) in der Signierung für den Mikrozensus - Statistisches Bundesamt, Ausgabe 1992 - in Verbindung mit dem Umsteigerschlüssel für die Internationale Standardklassifikation der Berufe zur Anwendung in der Europäischen Gemeinschaft (ISCO-88 COM). URL: [http://www.gesis.org/fileadmin/upload/dienstleistung/daten/amtl\\_mikrodaten/mz/Grundfile/allgemein/kldb92\\_isco88com.pdf](http://www.gesis.org/fileadmin/upload/dienstleistung/daten/amtl_mikrodaten/mz/Grundfile/allgemein/kldb92_isco88com.pdf).

Letzter Zugriff bei allen Web-Dokumenten: 02. 02. 2010.

## Anhang

---

```
* SPSS-Syntax zu Tab. 3: Verteilungen zur Prüfung der Systemfiles bzw. des
  Ablaufs der Setups .

* PASW Statistics 17.0, Release 17.0.2 (früher SPSS).

CD "<Arbeitsverzeichnis>".

get file='<mz2007.sav>'.

missing values all().

* Bevoelkerung am Hauptwohnsitz .
compute hws = (EF30>=1 & EF30<=2).

* Bevoelkerung in Privathaushalten.
compute bphh = (EF31=1).

* Bevoelkerung in Lebensformen am Hauptwohnsitz .
compute blf = (EF761>=1 & EF761<=2).

* Erwerbspersonen.
compute ep = (EF29>=1 & EF29<=2).

* Erwerbstaetige.
compute et = (EF29=1).

* Bevoelkerung am Hauptwohnsitz, Erwerbstaetige. .
compute hwsep = (EF30>=1 & EF30<=2 & EF29=1).

* Stellung im Beruf (Erwerbstaetige).
if (EF29=1) stib=EF117.
recode stib (2 1=1) (3=2) (4 9 10=3) (5 7 11=4) (6 8=5).
value label stib
  1 'Selbststaendige'
  2 'Mithelfende Familienangehoerige'
  3 'Beamte (inkl. Zeit-/Berufssoldat (einschl. BGS und
    Bereitschaftspolizei), Grundwehrdienstleistender)'
  4 'Angestellte (inkl. kaufm./techn. Auszubildende, Zivildienstleistender)'
  5 'Arbeiter (inkl. gewerbl. Auszubildende)'.

* Erwerbslose.
compute elo = (EF29=2).

* Nichterwerbspersonen.
compute nep = (EF29>=3 & EF29<=4).

* Hochrechnung: Jahresdurchschnitt.
compute w952 = EF952.

* Fallzaehlung.
compute fz = 1.

* =====.
temporary.
* INSGESAMT .
weight off.
means tables w952 fz by bphh hws blf hwsep /cells sum.
```

```

temporary.
* EF117: STELLUNG IM BERUF x EF29: ERWERBSTYP .
weight off.
* w952: Bev. aO d HW - gewichtet.
means tables w952 fz by stib elo ep nep by hws /cells sum.

temporary.
* EF117: Stellung im Beruf x EF29: Erwerbstyp .
weight off.
* fz: Fallzahlen - ungewichtet.
means tables fz by stib et elo ep nep /cells sum.

* HAUSHALTSGROESSE: EF663.
recode EF663 (5 thru 9=5) (else=copy) into v663.
* Mehrpersonenhaushalte.
recode EF663 (1=1) (2 thru 9=2) into v663m.
temporary.
* Haushalte:= Haushaltsbezugsperson = Haupteinkommensbezieher.
select if (EF31=1 & EF662>=1 & EF662<=4).
weight off.
means tables w952 fz by v663m v663 /cells sum.

* SUBSTICHPROBENMERKMALE.
* Filter: EF6 Unterstichprobe, Fragenprogramm .
* Gewichtung: EF955 .
* --- 0,1%-Substichprobenmerkmale: --- .
* EF118 Leitungstaetigkeit (F127)
* EF177 Arbeitsvertrag: Zeitarbeitsvermittlungsfirma (F126)
* EF178 Arbeitsvermittlung: Agentur fuer Arbeit (F128)
* EF179 Nichtarbeitsuche oder Teilzeitbeschaeftigung: Hauptgrund (F129)
* EF1040-EF1051 Ad-hoc-Modul "Arbeitsunfaelle, Arbeit und Gesundheit" .
* ----- .
* EF118 LEITUNGSTAETIGKEIT (F127) (FB) (0,1%) .
temporary.
compute w955 = EF955.
* Substichprobe & Erwerbstaetige & Bevoelkerung am Hauptwohnsitz.
select if (EF6=2 & EF29=1 & EF30>=1 & EF30<=2).
means tables w955 fz by EF118 /cells sum.

```