

# Metadatenreport



Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus 2016 am  
Gastwissenschaftsarbetsplatz sowie per kontrollierter Datenfernverarbeitung  
(sog. On-Site-Zugangswege/-Nutzungen)

DOI: 10.21242/12211.2016.00.00.1.1.1

Version 1

## Impressum

Herausgeber: Statistische Ämter des Bundes und der Länder  
Herstellung: Information und Technik Nordrhein-Westfalen  
Telefon 0211 9449-01 • Telefax 0211 9449-8000  
Internet: [www.forschungsdatenzentrum.de](http://www.forschungsdatenzentrum.de)  
E-Mail: [forschungsdatenzentrum@it.nrw.de](mailto:forschungsdatenzentrum@it.nrw.de)

### Fachliche Informationen

zu dieser Veröffentlichung:

Forschungsdatenzentrum der  
Statistischen Ämter der Länder  
– Düsseldorf –  
Tel.: 0211 9449-2877  
Fax: 0211 9449-8087  
[forschungsdatenzentrum@it.nrw.de](mailto:forschungsdatenzentrum@it.nrw.de)

### Informationen zum Datenangebot:

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

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

Erscheinungsfolge: unregelmäßig  
Erschienen im Dezember 2020

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

© Information und Technik Nordrhein-Westfalen, Düsseldorf, 2020  
(im Auftrag der Herausbergemeinschaft)

Vervielfältigung und Verbreitung, nur auszugsweise, mit Quellenangabe gestattet. Alle übrigen Rechte bleiben vorbehalten.

Fotorechte Umschlag: ©artSILENCE –Fotolia.com

### Empfohlene Zitierung:

Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder: Metadatenreport. Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus 2016 am Gastwissenschafts Arbeitsplatz sowie per kontrollierter Datenfernverarbeitung (EVAS-Nummer: 12211, 12212, 12213). Version 1. DOI: 10.21242/12211.2016.00.00.1.1.1 Düsseldorf 2020.

# Metadatenreport

Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus 2016 am Gastwissenschaftsarbetsplatz sowie per kontrollierter Datenfernverarbeitung (sog. On-Site-Zugangswege/-Nutzungen)

DOI: 10.21242/12211.2016.00.00.1.1.1

Version 1



## Inhalt

<b>1. Datenaufbereitung in den FDZ.....</b>	<b>2</b>
<b>1.1 Datenaufbereitung .....</b>	<b>2</b>
<b>1.2 Anonymisierungsmaßnahmen.....</b>	<b>3</b>
<b>1.3 Methodik der Verknüpfung.....</b>	<b>3</b>
<b>2 Produkt.....</b>	<b>4</b>
<b>2.1 Merkmale und Merkmalsbeschreibung.....</b>	<b>4</b>
2.1.1 Merkmalsdefinitionen .....	4
2.1.2 Datensatzbeschreibung / Schlüsselverzeichnis.....	4
<b>2.2 Vergleichbarkeit der Merkmale über die Zeit .....</b>	<b>5</b>
<b>2.3 Eckwerte relevanter Merkmale und Merkmalskombinationen....</b>	<b>7</b>
<b>2.4 Auswertbare regionale Ebene.....</b>	<b>9</b>
<b>2.5 Produktversionen.....</b>	<b>9</b>
<b>3 Praktische Hinweise.....</b>	<b>11</b>
<b>3.1 Hinweise zur Geheimhaltung .....</b>	<b>11</b>
3.1.1 Gesetzliche Grundlagen der statistischen Geheimhaltung.....	11
3.1.2 Geheimhaltung von Ergebnissen .....	11
3.1.3 Praktische Tipps zur Vermeidung von Geheimhaltungsfällen.	12
<b>3.2 FAQ.....</b>	<b>12</b>
<b>3.3 Verfügbare Tools.....</b>	<b>14</b>
<b>Anhang.....</b>	<b>15</b>

# 1. Datenaufbereitung in den FDZ

## 1.1 Datenaufbereitung

Das Datenmaterial wird einer Vollzähligkeits- und Vollständigkeitskontrolle unterzogen, bei der geprüft wird, ob alle Erhebungsmerkmale und die dazugehörigen, typisierten Merkmale, die für das Mikrozensusjahr vorgesehen sind, belegt sind. Nicht belegte Merkmale werden aus dem Datenmaterial entfernt. Gleichzeitig wird darauf geachtet, dass die einzelnen Variablen die korrekten Formate aufweisen.

Im Zuge des Abgleichs mit den zentralen Veröffentlichungen<sup>1</sup> des Statistischen Bundesamts (siehe Metadatenreport Teil I, Kapitel 4 Zentrale Veröffentlichungen) werden die Hochrechnungsfaktoren so angeglichen, dass sie ohne zusätzliche Multiplikation auf die Gesamtbevölkerung hochrechnen<sup>2</sup>. Fälle, die keinen positiven Hochrechnungsfaktor besitzen, werden aus dem Datenmaterial entfernt. Dies betrifft ca. 9% der Fälle aus dem Originaldatenmaterial. Für die betroffenen Haushalte liegt kein Interview vor, weil es sich beispielsweise um leerstehende Wohnungen oder weggefallene Haushalte handelt.

Die einzelnen Merkmale werden auf inhaltliche Konsistenz geprüft. Inkonsistenzen werden im Schlüsselverzeichnis (siehe auch Kapitel 2.1.2) dokumentiert und, sofern diese plausibel auflösbar sind, bereinigt. Dabei wird gewährleistet, dass die inhaltliche Aussage der Angaben nicht verändert wird.

Weiterhin wird geprüft, ob sich die Filterführung laut Fragebogen im Datenmaterial widerspiegelt. Für Personengruppen, denen gemäß Filterführung bestimmte Fragen nicht gestellt wurden, werden für die betroffenen Merkmale systemfehlende Werte vergeben (Ausprägung „leer“). Treten Inkonsistenzen bei der Filterführung auf, werden diese im Schlüsselverzeichnis (siehe auch Kapitel 2.1.2) dokumentiert und gegebenenfalls bereinigt, ohne die inhaltliche Aussage der Angaben zu verändern.

Zudem wird überprüft, ob unterschiedliche fachliche Ebenen, wie Haushalte, Familien und Lebensformen korrekt abgegrenzt und ausgewertet werden können.

Außerdem wird der in den Daten vorhandene Gebietsstand ermittelt. Die einzelnen Regionalkennungen (siehe Abschnitt 2.4 Auswertbare regionale Ebene) werden in diesem Zuge auf inhaltliche

---

<sup>1</sup> Exemplarisch finden Sie im Anhang dieses Reports Häufigkeitstabellen sowie die dazugehörige SPSS-Syntax mit Verweis auf die jeweilige Publikation des Statistischen Bundesamtes.

<sup>2</sup> Nähere Informationen zu den in diesem Produkt vorliegenden Hochrechnungsfaktoren finden Sie im Abschnitt 3.2 FAQ „Wann wird welcher Hochrechnungsfaktor verwendet?“.

Konsistenz geprüft. Treten Inkonsistenzen auf, werden diese im Schlüsselverzeichnis (siehe auch Kapitel 2.1.2) dokumentiert und gegebenenfalls bereinigt.

## **1.2 Anonymisierungsmaßnahmen**

Direkte Identifikatoren und Hilfsmerkmale werden aus dem Datenmaterial entfernt.

Die Ordnungsnummern des Auswahlbezirks, des Haushalts im Auswahlbezirk und der Person im Haushalt werden systemfrei sortiert, so dass anhand der Position des einzelnen Falls im Datenmaterial kein Rückschluss auf einzelne Personen möglich ist.

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

## **1.3 Methodik der Verknüpfung**

Der Mikrozensus stellt kein bereits verknüpftes Produkt dar, allerdings besteht die Möglichkeit, mit Hilfe der Merkmale EF 1- EF4 und EF 63, auf Individualebene die Wellen 2016-2019 miteinander zu verknüpfen.

Da sich 2016 die Auswahlgrundlage der Stichprobe verändert hat, ist keine Verknüpfung des Erhebungsjahres 2016 mit den Vorjahren möglich.

Genauere Informationen zur neuen Stichprobe finden Sie unter: <https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2016/06/die-neue-mikrozensusstichprobe-062016.pdf>

## 2 Produkt

### 2.1 Merkmale und Merkmalsbeschreibung

#### 2.1.1 Merkmalsdefinitionen

Folgende Klassifikationen finden in dem beschriebenen Produkt Anwendung:

- Klassifikation der Berufe, Ausgabe 2010 (KIdB 2010)  
<https://statistik.arbeitsagentur.de/DE/Navigation/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KIdB2010/Systematik-Verzeichnisse/Systematik-Verzeichnisse-Nav.html>
- Internationale Standardklassifikation der Berufe, Ausgabe 2008 (ISCO-08)  
<http://www.ilo.org/public/english/bureau/stat/isco/isco08/>
- Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008)  
<https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/Downloads/klassifikation-wz-2008-3100100089004.pdf>
- Bildungsklassifikation International Standard Classification of Education, Ausgabe 2011 (ISCED-2011)  
<http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf>
- Staats- und Gebietssystematik  
[https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/Staatsangehoerigkeitsgebietsschluessel\\_pdf.pdf](https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/Staatsangehoerigkeitsgebietsschluessel_pdf.pdf)
- Amtlicher Gemeindeschlüssel (AGS)/ Regionalschlüssel (RS)  
<https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/inhalt.html>

#### 2.1.2 Datensatzbeschreibung / Schlüsselverzeichnis

Der Mikrozensus ist eine Haushalts- und Personenstatistik. Die vollständige Liste der im Datensatz vorhandenen Variablen finden Sie im Schlüsselverzeichnis. Sofern Variablen das Ergebnis einer Frage des Fragebogens wiedergeben, finden Sie die zugehörige Fragennummer in der Spalte „Fragebogennummer“ des Schlüsselverzeichnisses. Das Zusatzprogramm beschäftigt sich mit „Ausbildungs- und Berufspendlern“. Den vollständigen Fragebogen mitsamt des Ad-hoc-Moduls „Junge Menschen auf dem Arbeitsmarkt“ finden Sie bei den Erhebungsunterlagen.

Schlüsselverzeichnis

[https://www.forschungsdatenzentrum.de/sites/default/files/mz\\_2016\\_svz.pdf](https://www.forschungsdatenzentrum.de/sites/default/files/mz_2016_svz.pdf)

Erhebungsunterlagen

[https://www.forschungsdatenzentrum.de/sites/default/files/mz\\_2016\\_eu.pdf](https://www.forschungsdatenzentrum.de/sites/default/files/mz_2016_eu.pdf)

2016 wurde das Ad-Hoc-Modul „Junge Menschen auf dem Arbeitsmarkt“ abgefragt.

## 2.2 Vergleichbarkeit der Merkmale über die Zeit

Folgende Merkmale weisen im Vergleich zum Vorjahr Veränderungen auf:

Haushalts-/Personenidentifikation

- EF10 (Baualter der Wohnung): Die Ausprägungen „1 – vor 1987“ wurde in „1- vor 2011“ und „2 - 1987 bis 1990“ wurde in „2 – 2011 oder später“ umbenannt. Die Ausprägung „3 – 1991 oder später“ entfällt. Die Ausprägungen haben sich geändert, da diese mit der Auswahl gekoppelt sind. Eine Vergleichbarkeit ist nicht gegeben.

Gegenwärtige (Haupt-)Erwerbstätigkeit

- EF126 (Grund der Befristung der Tätigkeit): Die Ausprägung „1 – Ausbildung“ wurde entfernt. Die Ausprägung „5 – Aus anderen Gründen“ wurde in „5 – Tätigkeit als Auszubildende/r mit Ausbildungsvergütung“ umgeändert. Die Ausprägungen „6 – Andere Art der Ausbildung“ und „7 – Anderer Hauptgrund“ wurden neu hinzugefügt. Eine Vergleichbarkeit ist eingeschränkt gegeben

Weg zur Schule/Hochschule/Lage der Schule/Hochschule

- EF299 (Lage der Schule/Hochschule in demselben Bundesland): Die Ausprägung 2 wurde in Ausprägung 8 geändert.

Schul-/Ausbildungsabschluss

- EF312 (Höchster beruflicher Ausbildungs- oder Hochschul-/Fachhochschulabschluss): Die Ausprägung „08- Meister-/in“ wurde neu hinzugefügt. Die Ausprägung „18- Techniker/-in oder gleichwertiger Fachabschluss“ wurde neu hinzugefügt. Eine Vergleichbarkeit ist eingeschränkt gegeben.

Unterhalt, Einkommen

- EF400 (Bezug, öffentliche Zahlungen/ Unterstützungen, Betreuungsgeld): Die Variable ist neu hinzugekommen.

#### Angaben zum Haupteinkommensbezieher im Haushalt

- EF744 (Höchster beruflicher Ausbildungsabschluss des Haupteinkommensbeziehers mit Grad): Die Ausprägungen „1 – Anlernausbildung oder berufliches Praktikum“ und „2- Berufsvorbereitungsjahr“ wurden entfernt. Die Ausprägung „18- Techniker/-in oder gleichwertiger Fachabschluss“ wurde neu hinzugefügt. Die Ausprägung „88 – Kein beruflicher Abschluss“ wurde in „88 – Kein Abschluss, aber in schulischer oder beruflicher Ausbildung“ umbenannt. Die Ausprägung „89 – Kein Abschluss und nicht in schulischer oder beruflicher Ausbildung“ wurde neu hinzugefügt. Eine Vergleichbarkeit ist eingeschränkt gegeben.

#### Angaben zu den Familien und Lebensformen

- EF832 (Höchster beruflicher Ausbildungsabschluss der Bezugsperson in der Familie/Lebensform mit Grad): Die Ausprägungen „1 – Anlernausbildung oder berufliches Praktikum“ und „2- Berufsvorbereitungsjahr“ wurden entfernt. Die Ausprägung „18- Techniker/-in oder gleichwertiger Fachabschluss“ wurde neu hinzugefügt. Die Ausprägung „88 – Kein beruflicher Abschluss“ wurde in „88 – Kein Abschluss, aber in schulischer oder beruflicher Ausbildung“ umbenannt. Die Ausprägung „89 – Kein Abschluss und nicht in schulischer oder beruflicher Ausbildung“ wurde neu hinzugefügt. Eine Vergleichbarkeit ist eingeschränkt gegeben.
- EF856 (Höchster beruflicher Ausbildungsabschluss des Ehe- oder Lebenspartners): Die Ausprägungen „1 – Anlernausbildung oder berufliches Praktikum“ und „2- Berufsvorbereitungsjahr“ wurden entfernt. Die Ausprägung „18- Techniker/-in oder gleichwertiger Fachabschluss“ wurde neu hinzugefügt. Die Ausprägung „88 – Kein beruflicher Abschluss“ wurde in „88 – Kein Abschluss, aber in schulischer oder beruflicher Ausbildung“ umbenannt. Die Ausprägung „89 – Kein Abschluss und nicht in schulischer oder beruflicher Ausbildung“ wurde neu hinzugefügt. Eine Vergleichbarkeit ist eingeschränkt gegeben.

#### Ad-hoc-Modul

- Das Ad-hoc-Modul ändert sich jährlich. Das Ad hoc Modul 2016 „Junge Menschen auf dem Arbeitsmarkt“ umfasst die EFs ab EF1080.

Eine weitere Besonderheit im Jahr 2016 ist die hohe Zahl der in Deutschland Schutzsuchenden, die in Aufnahmeeinrichtungen lebten. Bei der überwiegenden Zahl dieser Aufnahmeeinrichtungen handelt es sich nicht um Gebäude mit Wohnraum, so dass die dort lebenden Personen keine Chance hatten im Mikrozensus erfasst zu werden. Zu beachten ist dies bei den Zahlen zur nichtdeutschen Bevölkerung. Nach dem Vorliegen der Ergebnisse des Zensus 2011 wurde eine neue Grundauswahl

gezogen und das Verfahren der Aktualisierung beibehalten. Auf Basis von Baugenehmigungen werden Flächen oder Anschriften identifiziert, auf denen neue Wohnungen entstehen. Diese werden der Grundausswahl hinzugefügt (vgl. Bihler und Zimmermann 2016: 27).

In den Jahren 2016 und 2017 gab es Probleme bei der Bereitstellung von Bevölkerungszahlen seitens der Bevölkerungsstatistik. Neben anderen Statistiken war der Mikrozensus in besonderer Weise von den damit einhergehenden Verzögerungen betroffen. Für die Hochrechnung, die quartalsweise erfolgt, wird mit vorausgeschätzten Bevölkerungszahlen gearbeitet, da die Ergebnisse der Bevölkerungsfortschreibung zum Zeitpunkt der Hochrechnung noch nicht vorliegen. Aufgrund der Verzögerungen bei der Bevölkerungsfortschreibung musste der Vorausschätzungszeitraum teilweise auf bis zu 15 Monate verlängert werden. Da ab Ende 2015 eine außergewöhnliche Bevölkerungsentwicklung zu beobachten ist, die die Qualität der Vorausschätzungen beeinflusst, wurden für das Jahr 2016 die Eckwerte Anfang 2017 neu geschätzt und die Hochrechnung wiederholt. Die Hochrechnungseckwerte sind dennoch 2016 und 2017 mit mehr Unsicherheiten belastet, als dies üblicherweise der Fall ist. Insbesondere auf regionaler Ebene kann sich dieser Effekt deutlich niederschlagen. Da es weitere interagierende Effekte (zum Beispiel die neue Stichprobe 2016) gibt, sind die einzelnen Einflüsse kaum zu quantifizieren.

Genauere Informationen finden Sie unter:

<https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2016/06/die-neue-mikrozensusstichprobe-062016.pdf>

## 2.3 Eckwerte relevanter Merkmale und Merkmalskombinationen

### Verteilung der Bevölkerung am Hauptwohnsitz

**Tab. 1: EF46NEU Geschlecht**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	männlich	40.682.621	49,4	49,4	49,4
	weiblich	41.742.221	50,6	50,6	100,0
Gesamt		82.424.842	100,0	100,0	

**Tab. 2: EF49NEU Familienstand**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	ledig	33.932.263	41,2	41,2	41,2
	verheiratet	37.031.977	44,9	44,9	86,1
	verwitwet	11.460.602	13,9	13,19	100,0
	Gesamt	82.424.842	100,0	100,0	

**Privathaushalte am Haupt- und Nebenwohnsitz****Tab. 3: EF20NEU Privathaushalte kategorisiert**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	Einpersonenhaushalte	16.832.558	41,1	41,1	41,1
	Zweipersonenhaushalte	13.914.918	34,0	34,0	75,1
	Dreipersonenhaushalte	5.028.269	12,3	12,3	87,3
	Vierpersonenhaushalte	3.806.701	9,3	9,3	96,6
	Fünf- und mehr Personenhaus- halte	1.377.237	3,4	3,4	100,0
	Gesamt	40.959.684	100,0	100,0	

**Tab. 4: EF1 Bundesland**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	Schleswig-Holstein	1.437.184	3,5	3,5	3,5
	Hamburg	973.810	2,4	2,4	5,9
	Niedersachsen	3.941.242	9,6	9,6	15,5
	Bremen	359.446	,9	,9	16,4
	Nordrhein-Westfalen	8.707.279	21,3	21,3	37,6
	Hessen	3.026.605	7,4	7,4	45,0
	Rheinland-Pfalz	1.947.099	4,8	4,8	49,8
	Baden-Württemberg	5.201.012	12,7	12,7	62,5
	Bayern	6.358.399	15,5	15,5	78,0
	Saarland	495.137	1,2	1,2	79,2
	Berlin	1.964.855	4,8	4,8	84,0
	Brandenburg	1.241.333	3,0	3,0	87,0

Mecklenburg-Vorpommern	838.350	2,0	2,0	89,1
Sachsen	2.173.699	5,3	5,3	94,4
Sachsen-Anhalt	1.174.508	2,9	2,9	97,3
Thüringen	1.119.726	2,7	2,7	100,0
Gesamt	40.959.684	100,0	100,0	

Im Anhang finden Sie diese und weitere relevante Tabellen mit der Syntax für das Programm SPSS sowie allen Fundstellen in den Fachserien des Statistischen Bundesamtes.

## 2.4 Auswertbare regionale Ebene

- Bundesebene (NUTS 0)
- Landesebene (NUTS 1)
- Regierungsbezirksebene (NUTS 2)
- Kreisebene (ggf. eingeschränkte Aussagekraft durch geringe Zellbesetzungen) (NUTS 3)
- Gemeindeebene (ggf. eingeschränkte Aussagekraft durch geringe Zellbesetzungen) (LAU)
- Anpassungsschichten (regionale Schicht und regionale Untergruppe)<sup>3</sup>
- Raumordnungsregionen
- Arbeitsmarktregionen
- Nielseengebiete
- Arbeitsagenturbezirke

## 2.5 Produktversionen

### 1.0.0

- Erstveröffentlichung

#### 1.1.1 (DOI 10-21242-12211-2016-00-00-1-1-0)

- Der Amtliche Gemeindeschlüssel (AGS) des Mikrozensus wurde an das Gemeindeverzeichnis angepasst. Betroffen davon sind folgende Städte und Gemeinden:
  - Berlin

<sup>3</sup> In Berlin und Hamburg ist eine Auswertung auf Bezirksebene möglich. In München kann eine Aufteilung nach innerstädtischen Bezirken und nach Bezirken in der Randlage der Stadt München vorgenommen werden.

- Hamburg
- Aachen
- Die Variablen EF313, EF313U1 und EF313U2 wurden in den Datensatz aufgenommen
- Es wurde eine eindeutige Benennung der Anpassungsschicht (EF564), der Regionalen Schicht (EF566) und der Regionalen Untergruppe (EF568) vorgenommen.

#### 1.2.0 (DOI 10-21242-12211-2016-00-00-1-1-1)

- Die Nielsegebiete wurden an den Stand ab 2008 angepasst.
- Aufgrund einer inkonsistenten Schichtung zweier Kreise in Rheinland-Pfalz wurde eine eindeutige Zuordnung über die Regionale Anpassungsschicht (EF564) vorgenommen.
- Die Variablen EF3, EF4 und EF63 wurden zur Bildung einer jahresübergreifenden Personen-ID aufgenommen.
- Die Arbeitsmarktregion Berlin wurde hinzugefügt.
- Der fehlerhafte Hochrechnungsfaktor für tiefregionalisierte Ergebnisse in Rheinland-Pfalz wurde entfernt.

## **3 Praktische Hinweise**

### **3.1 Hinweise zur Geheimhaltung**

#### **3.1.1 Gesetzliche Grundlagen der statistischen Geheimhaltung**

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

Die Geheimhaltung in der amtlichen Statistik ist in § 16 Bundesstatistikgesetz (BStatG) geregelt und beinhaltet, dass Einzelangaben über persönliche und sachliche Verhältnisse, die für eine Bundesstatistik angegeben werden, von den jeweils durchführenden statistischen Stellen geheim zu halten sind, soweit es keine anderslautenden Bestimmungen gibt. Dies wird auch als Statistikgeheimnis bezeichnet. Das Statistikgeheimnis verpflichtet die amtliche Statistik, die erhaltenen Informationen zu schützen, d. h. 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. Die amtliche Statistik muss deshalb sicherstellen, dass die erhobenen Daten keinem Befragten zugeordnet werden können.

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

#### **3.1.2 Geheimhaltung von Ergebnissen**

Um die gesetzlich vorgeschriebene Geheimhaltung von Einzelfällen in den Daten sicherzustellen, müssen alle Ergebnisse, die am Gastwissenschaftsarbetsplatz oder per Kontrollierter Datenfernverarbeitung erzeugt werden, vor ihrer Freigabe an den Nutzer von den FDZ einer Geheimhaltungsprüfung unterzogen werden. Dabei stellen die FDZ sicher, dass die Ergebnisse absolut anonym sind und eine Reidentifikation einzelner Befragter nach menschlichem Ermessen ausgeschlossen werden kann. Entsprechend handeln auch die Fachabteilungen der Statistischen Ämter vor der Veröffentlichung von Ergebnissen.

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

Die Broschüre finden Sie hier:

[https://www.forschungsdatenzentrum.de/sites/default/files/fdz\\_broschuere\\_regelungen\\_2019.pdf](https://www.forschungsdatenzentrum.de/sites/default/files/fdz_broschuere_regelungen_2019.pdf)

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

Treten in den erstellten Analysen Geheimhaltungsfälle auf, werden diese Werte von den FDZ zur Sicherstellung der Geheimhaltung durch ein Sperrmuster ersetzt. Gerade in Kreuztabellen entstehen so durch die notwendige Sekundärspernung schnell viele Lücken in den Auswertungen. Da eine einmal zur Sekundärspernung herangezogene Tabellenzelle auch in allen folgenden Analysen gesperrt werden muss (tabellenübergreifende Geheimhaltung) – auch, wenn es in der neu erstellten Tabelle nicht nötig wäre – ist es sinnvoll, bei jeder Ergebniserstellung darauf zu achten, dass möglichst keine Geheimhaltungsfälle erzeugt werden. Treten in einem Output Geheimhaltungsfälle auf, steht es dem betreuenden FDZ frei, die Prüfung und Freigabe des Outputs abzulehnen.

Um Geheimhaltungsfälle in den Analysen zu vermeiden, sollte immer darauf geachtet werden, dass die erstellten Analysen auf ausreichend großen Fallzahlen beruhen. Bei geringen Fallzahlen empfiehlt es sich, Variablenausprägungen zusammenzufassen und damit größere Fallzahlen zu erzielen.

## **3.2 FAQ**

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

In den Veröffentlichungen der amtlichen Statistik wird eine Selektion nach allen Privathaushalten (EF31=1) und der Bezugsperson im Haushalt (EF37=1) vorgenommen. Alternativ dazu ist es möglich eine eindeutige Haushalts-ID über die Aneinanderreihung der Eingabefelder Bundesland (EF1), Auswahlbezirk (EF3), Haushaltsnummer (EF4) und Kennzeichnung von Überhanginterviews (EF5U2) zu generieren und das Datenmaterial anhand dieser ID über alle Personen des Haushalts zu aggregieren. Sofern es sich nicht um ein Überhanginterview handelt (EF5U2=missing), empfiehlt es sich der Haushalts-ID eine 0 anzuhängen, um Dubletten aufgrund der Verschiebung von Stellen zu vermeiden.

## **Wie können Mütter und Väter ausgewertet werden?**

Zunächst muss eine Haushaltsidentifikationsnummer erstellt werden. Hierfür werden folgende Variablen verkettet:

- EF1 Bundesland
- EF3 Auswahlbezirk
- EF4 Haushaltsnummer
- EF5U2 Kennzeichnung der Jahresüberhänge

Sofern EF5U2 missing ist, handelt es sich nicht um ein Überhang Interview. Um Dubletten durch die Verschiebung von Stellen zu vermeiden, ist in diesen Fällen eine 0 anzuhängen. Über die Eingabefelder EF53 und EF55 besteht die Möglichkeit die Personnummer der Mutter bzw. des Vaters zu ermitteln. Bei Überhanginterviews ist diesen Nummern EF5U2 anzuhängen. Über EF5 können diese Nummern Personen nun im Haushalt identifiziert werden. EF5 ist die EF5U2 bereits angehängt.

## **Wieso sind in den Daten doppelte Personennummern enthalten?**

Die Dopplung entsteht durch die in den Daten enthaltenen Jahresüberhänge aus dem Vorjahr. Über das Merkmal EF5U2 können diese von den tatsächlich in diesem Jahr befragten Fällen abgegrenzt werden.

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

Für Ergebnisse auf Ebene der Anpassungsschichten und höher (zum Beispiel Bundesländer) wird der Standardhochrechnungsfaktor (EF952) verwendet. Für Ergebnisse unterhalb der Anpassungsschichten (zum Beispiel Kreisebene) bis hinunter zur regionalen Untergruppe wird der Hochrechnungsfaktor für tiefregionalisierte Ergebnisse (EF953) empfohlen. Bei der quartalsweisen statistischen Analyse (die Ergebnisse basieren also nur auf der Anzahl der Personen, die in dem jeweiligen Quartal befragt wurden) findet der Standardhochrechnungsfaktor für die Quartale (EF951) Anwendung. Eine ausführliche Erläuterung der Modelle für die einzelnen Faktoren ist dem Metadatenreport Teil I in Abschnitt 2.6 zu entnehmen.

Beinhaltet eine Analyse auf Bundesebene nur Personen der zehnpromzentigen EU-Unterstichprobe des Ad-Hoc-Modules, ist der Hochrechnungsfaktor EF956 zu verwenden.

Im Rahmen des Zusatzprogramms wurde auf freiwilliger Basis das Thema „Junge Menschen (15-34 Jahre) auf dem Arbeitsmarkt erfragt. Der Hochrechnungsfaktor für das Ad-hoc-Modul wird nur für

Personen berechnet, da Auswertungen hier im Haushalts- und Familienzusammenhang von untergeordneter Bedeutung sind. Als Eingangsfaktoren für die gebundene Hochrechnung, die nach gleichen Methodik wie der Standardhochrechnungsfaktor berechnet wird, wird neben den dort herangezogenen Kompensationsfaktoren (siehe Afentakis/Bihler 2005: 1044) des Weiteren noch die Erhebungsart als erklärende Variable herangezogen. Als Anpassungsmodell wird für

- Deutschland der Erwerbstyp differenziert nach Geschlecht und Altersklassen, für die
- Bundesländer der Erwerbstyp differenziert nach dem Geschlecht bzw. nur die Altersklasse und für die
- Gebiete Ost und West differenziert nach der Staatsangehörigkeit (deutsch, nicht deutsch) und Geschlecht

gewählt.

### **3.3 Verfügbare Tools**

Für dieses Produkt werden seitens der FDZ keine weiterführenden Tools angeboten.

# Anhang

## Ausgewählte Häufigkeitstabellen und SPSS-Syntax (Anhang zu Abschnitt 2.3):

### Bevölkerung (Personen am Hauptwohnsitz)<sup>4</sup>:

\*\*\*\* Bevölkerung am Hauptwohnsitz auswählen.

\*\*\*\*\*

IF (EF30=1 OR EF30=2) Hauptwohnsitz=1.

FILTER BY Hauptwohnsitz.

EXECUTE.

\*\*\* Hochrechnung: Hochrechnungsfaktor (Personenfaktor) in der regionalen.

\*\*\*\* Anpassungsschicht (EF952).

\*\*\*\*\*

WEIGHT BY EF952.

EXECUTE.

VARIABLE LABELS

EF46 'EF46NEU Geschlecht'.

EXECUTE.

ADD VALUE LABELS

EF46

1 'männlich'

2 'weiblich'.

EXECUTE.

\*\*\*\* Häufigkeiten für Geschlecht (EF46).

FREQUENCIES

VARIABLES= EF46

/ORDER= ANALYSIS.

#### EF46NEU Geschlecht

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	männlich	40.682.621	49,4	49,4	49,4
	weiblich	41.742.221	50,6	50,6	100,0
Gesamt		82.424.842	100,0	100,0	

\*\*\*\* Familienstand entsprechend der Publikationen zusammenfassen.

\*\*\*\*\*

RECODE EF49 (4=3) (5=2) (6=3) (7=3) (ELSE=COPY) INTO EF49NEU.

EXECUTE.

FORMATS EF49NEU (F1.0).

EXECUTE.

<sup>4</sup> Vgl. Zahlen in DESTATIS Fachserie 1, Reihe 4.1.1, 2016 [https://www.destatis.de/GPStatistik/servlets/MCRFileNodeServlet/DEHeft\\_derivate\\_00051965/2010410167004\\_korr29102019.pdf](https://www.destatis.de/GPStatistik/servlets/MCRFileNodeServlet/DEHeft_derivate_00051965/2010410167004_korr29102019.pdf) Tabellen 1.2, 4.1 und 4.3

VARIABLE LABELS  
 EF49NEU 'EF49NEU Familienstand'.  
 EXECUTE.

ADD VALUE LABELS  
 EF49NEU  
     1 'ledig'  
     2 'verheiratet'  
     3 'verwitwet'.

EXECUTE.

\*\*\*\* Häufigkeiten für Familienstand (EF49)  
 FREQUENCIES  
 VARIABLES= EF49NEU  
 /ORDER= ANALYSIS.

**EF49NEU Familienstand**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	ledig	33.932.263	41,2	41,2	41,2
	verheiratet	37.031.977	44,9	44,9	86,1
	verwitwet	11.460.602	13,9	13,19	100,0
	Gesamt	82.424.842	100,0	100,0	

\*\*\*\* Überwiegender Lebensunterhalt EF401. Wird entsprechend der Publikation umkodiert.

\*\*\*\*\*

RECODE EF401 (7=2) (5 THRU 9=3) (ELSE=COPY) INTO EF401NEU.  
 EXECUTE.

FORMATS EF401NEU (F1.0).  
 EXECUTE.

\*\*\*\* Labeln der Eingabefelder, die zur Prüfung der Daten benötigt werden.

\*\*\*\*\*

VARIABLE LABELS  
 EF401NEU 'EF401NEU Überwiegender Lebensunterhalt'.  
 EXECUTE.

ADD VALUE LABELS  
 EF401NEU  
     1 'Erwerbs-/Berufstätigkeit'  
     2 'Arbeitslosengeld/-hilfe'  
     3 'Rente, Pension'  
     4 'Unterhalt durch Eltern, Ehemann/-frau oder andere Angehörige'  
     5 'Eigenes Vermögen, Vermietung, Zinsen, Altenteil'  
     6 'Sozialhilfe'  
     8 'Sonstige Unterstützungen (z.B. BAföG)'  
     9 'Elterngeld/Erziehungsgeld'.

EXECUTE.

\*\*\*\* Häufigkeiten für  
 \*\*\*\* Überwiegender Lebensunterhalt (EF401).  
 FREQUENCIES  
 VARIABLES= EF401NEU  
 /ORDER= ANALYSIS.

**EF401NEU Überwiegender Lebensunterhalt**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	Erwerbs-/Berufstätigkeit	37.348.179	45,3	45,9	45,9
	Arbeitslosengeld/-hilfe	3.469.886	4,2	4,2	49,5
	Rente, Pension	21.507.936	26,1	26,1	75,6
	Unterhalt durch Eltern, Ehemann/- frau oder andere Angehörige	20.098.841	24,4	24,4	100,0
	Gesamt	82.424.842	100,0	100,0	

\*\*\*\* Um Erwerbsstatus zu erhalten, wird EF29 (Nichterwerbspersonen) in Anlehnung an die  
\*\*\*\* Publikation umkodiert.

\*\*\*\*\*  
RECODE EF29 (3=4) (ELSE=COPY) INTO EF29NEU.  
EXECUTE.

FORMATS EF29NEU (F1.0).  
EXECUTE.

\*\*\*\* Labeln der Eingabefelder, die zur Prüfung der Daten benötigt werden.

\*\*\*\*\*  
VARIABLE LABELS  
EF29NEU 'EF29NEU Erwerbstyp'.  
EXECUTE.

ADD VALUE LABELS  
EF29NEU

1 'Erwerbstätige'  
2 'Erwerbslose'  
4 'Nichterwerbspersonen'.

EXECUTE.

\*\*\*\* Häufigkeiten für Erwerbstyp (EF29)

FREQUENCIES  
VARIABLES= EF29NEU  
/ORDER= ANALYSIS

**EF29NEU Erwerbstyp**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	Erwerbstätige	41.338.509	50,2	50,2	50,2
	Erwerbslose	1.774.836	2,2	2,2	52,3
	Nichterwerbspersonen	39.311.497	47,7	47,7	100,0
	Gesamt	82.424.842	100,0	100,0	

\*\*\*\* Um die Stellung im Beruf nur von Erwerbstätigen zu.

\*\*\*\* erhalten, wird hier EF117 (Stellung im Beruf) umkodiert in EF117\_NEU.

\*\*\*\*\*  
DO IF (EF29=1).  
RECODE EF117 (1=1) (2=1) (3=3) (4=4) (5=5) (6=6) (8=6) (9=4) (12=5) (14=4) (16=4) (17=5)  
INTO EF117NEU.  
END IF.

EXECUTE.  
 FORMATS EF117NEU (F2.0).  
 EXECUTE.

\*\*\*\* Labeln der Eingabefelder, die zur Prüfung der Daten benötigt werden.

\*\*\*\*\*

VARIABLE LABELS

EF117NEU 'EF117NEU Stellung im Beruf für Erwerbstätige'

EXECUTE.

ADD VALUE LABELS

EF117NEU

1 'Selbständige mit und ohne Beschäftigte'

3 'mithelfende Familienangehörige'

4 'Beamte, Richterinnen und Richter'

5 'Angestellter und kaufmännische/technische Auszubildende'

6 'Arbeiterinnen und Arbeiter, Heimarbeiterin und Heimarbeiter und gewerbliche Auszubildende'.

EXECUTE.

\*\*\*\* \*\* Stellung im Beruf für Erwerbstätige (EF117\_NEU).

FREQUENCIES

VARIABLES= EF117NEU

/ORDER= ANALYSIS.

**EF117NEU Stellung im Beruf für Erwerbstätige**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	1 Selbständige mit und ohne Beschäftigte	4.141.587	5,0	10,4	10,4
	3 mithelfende Familienangehörige	157.087	,2	,4	10,8
	4 Beamte, Richterinnen und Richter	2.030.960	2,5	5,1	15,9
	5 Angestellte und kaufmännische/technische Auszubildende	25.831.008	31,3	64,8	80,7
	6 Arbeiterinnen und Arbeiter, Heimarbeiterinnen und Heimarbeiter und gewerbliche Auszubildende	7.695.911	9,3	19,3	100,0
	Gesamt	39.856.554	48,4	100,0	
Fehlend	System	42.568.289	51,6		
Gesamt		82.424.842	100,0		

## Privathaushalte am Haupt- und Nebenwohnsitz<sup>5</sup>:

\*\*\*\* Alle Fälle auswählen, die Bezugsperson in Privathaushalten sind, um die.

\*\*\*\* Privathaushalte auszählen zu können.  
 IF (EF31=1 AND EF37=1) Privathaushalte=1.  
 FILTER BY Privathaushalte.  
 EXECUTE.  
 WEIGHT BY EF952.  
 EXECUTE.

\*\*\*\* Umkodieren von EF20 (Anzahl der Personen in Privathaushalten) in.  
 \*\*\*\* EF20NEU, um die Haushaltsgröße zu kategorisieren.  
 RECODE EF20  
 (1=1) (2=2) (3=3) (4=4) (0=0) (ELSE=5) INTO EF20NEU.  
 EXECUTE.  
 FORMATS EF20NEU (F1.0).  
 EXECUTE.

\*\*\*\* Labeln und Häufigkeiten.

\*\*\*\*\*

VARIABLE LABELS  
 EF20NEU 'EF20NEU Privathaushalte kategorisiert'.

ADD VALUE LABELS  
 EF20NEU  
           0 'kein Privathaushalt'  
           1 'Einpersonenhaushalte'  
           2 'Zweipersonenhaushalte'  
           3 'Dreipersonenhaushalte'  
           4 'Vierpersonenhaushalte'  
           5 'Fünf- und Mehrpersonenhaushalte'.

EXECUTE.

FREQUENCIES  
 VARIABLES= EF20NEU  
 /ORDER= ANALYSIS.

### EF20NEU Privathaushalte kategorisiert

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	Einpersonenhaushalte	16.832.558	41,1	41,1	41,1
	Zweipersonenhaushalte	13.914.918	34,0	34,0	75,1
	Dreipersonenhaushalte	5.028.269	12,3	12,3	87,3
	Vierpersonenhaushalte	3.806.701	9,3	9,3	96,6
	Fünf- und Mehrpersonenhaushalte	1.377.237	3,4	3,4	100,0
	Gesamt	40.959.684	100,0	100,0	

<sup>5</sup> Vgl. Zahlen in Destatis Fachserie 1, Reihe 3, 2016 [https://www.destatis.de/GPStatistik/servlets/MCR-FileNodeServlet/DEHeft\\_derivate\\_00032758/2010300167004.pdf](https://www.destatis.de/GPStatistik/servlets/MCR-FileNodeServlet/DEHeft_derivate_00032758/2010300167004.pdf) Tab. 1.1 und 1.4

\*\*\*\* Labeln und Häufigkeiten.

\*\*\*\*\*

VARIABLE LABELS  
EF1 'EF1 Bundesland'.

FREQUENCIES  
VARIABLES= EF1  
/ORDER= ANALYSIS.

**EF1 Bundesland**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	Schleswig-Holstein	1.437.184	3,5	3,5	3,5
	Hamburg	973.810	2,4	2,4	5,9
	Niedersachsen	3.941.242	9,6	9,6	15,5
	Bremen	359.446	,9	,9	16,4
	Nordrhein-Westfalen	8.707.279	21,3	21,3	37,6
	Hessen	3.026.605	7,4	7,4	45,0
	Rheinland-Pfalz	1.947.099	4,8	4,8	49,8
	Baden-Württemberg	5.201.012	12,7	12,7	62,5
	Bayern	6.358.399	15,5	15,5	78,0
	Saarland	495.137	1,2	1,2	79,2
	Berlin	1.964.855	4,8	4,8	84,0
	Brandenburg	1.241.333	3,0	3,0	87,0
	Mecklenburg-Vorpommern	838.350	2,0	2,0	89,1
	Sachsen	2.173.699	5,3	5,3	94,4
	Sachsen-Anhalt	1.174.508	2,9	2,9	97,3
	Thüringen	1.119.726	2,7	2,7	100,0
	Gesamt	40.959.684	100,0	100,0	

\*\*\*\* Umkodieren von EF735 entsprechend der Publikation umkodiert.

\*\*\*\*\*

RECODE EF735  
(6=2) (2=2) (8=4) (9=5) (7=2) (ELSE=COPY) INTO EF735NEU.  
EXECUTE.

FORMATS EF735NEU (F1.0).  
EXECUTE.

\*\*\*\* Labeln und Häufigkeiten.

\*\*\*\*\*

VARIABLE LABELS  
EF735NEU 'EF735NEU Familienstand der Bezugsperson'.

ADD VALUE LABELS  
EF735NEU

1 'ledig'  
2 'verheiratet'  
4 'geschieden'  
5 'verwitwet'.

EXECUTE.

FREQUENCIES  
VARIABLES= EF735NEU  
/ORDER= ANALYSIS.

**EF735NEU Familienstand der Bezugsperson**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Pro- zente
Gültig	ledig	12.519.892	30,6	30,6	30,6
	verheiratet	18.853.966	46,0	46,0	76,6
	geschieden	4.779.062	11,7	11,7	88,3
	verwitwet	4.806.653	11,7	11,7	100,0
	Gesamt	40.959.574	100,0	100,0	
Fehlend	System	110	,0		
Gesamt		40.959.684	100,0		

Statistische Ämter des Bundes und der Länder,  
Metadatenreport – Teil II: Produktspezifische Informationen zur Nutzung des Mikrozensus 2016 am  
Gastwissenschaftsarbetsplatz sowie per kontrollierter Datenfernverarbeitung (sog. On-Site-Zugangswege/-  
Nutzungen)

Fotorechte Umschlag: ©artSILENCEcom – Fotolia.com