

# BEIM FRAGEBOGEN-„TÜV“ – EVALUIERUNG DES FRAGEBOGENS DER HAUSHALTSSTICHPROBE DES ZENSUS 2011

David Bretschi, Miriam Pfahl

↳ **Schlüsselwörter:** Fragebogen – Fragebogenevaluierung – Zensus – Haushaltsstichprobe – Qualität

## ZUSAMMENFASSUNG

Die Mitgliedstaaten der Europäischen Union sind verpflichtet, alle zehn Jahre einen Zensus durchzuführen. Somit findet 2021 auch in Deutschland wieder eine registergestützte Volkszählung statt. Die Evaluierung des Fragebogens der Haushaltsstichprobe des Zensus 2011 liefert für die Neukonzeption des Erhebungsinstruments für den Zensus 2021 wertvolle Erkenntnisse. Darüber hinaus erlaubt sie eine Qualitätseinschätzung der Zensusdaten von 2011. Der vorliegende Beitrag stellt die Vorgehensweise und ausgewählte Ergebnisse der Evaluierung vor und formuliert Empfehlungen für die Weiterentwicklung des Fragebogens für den Zensus 2021.

↳ **Keywords:** questionnaire – questionnaire evaluation – census – household sample survey – quality

## ABSTRACT

*The Member States of the European Union are obliged to carry out a census every ten years. Hence, a register-based population census will again be taken in Germany in 2021. An evaluation of the questionnaire for the sample-based household survey of the 2011 Census provides valuable insights for the re-design of the 2021 Census survey instrument. In addition, it permits to assess the quality of the 2011 census data. This paper presents the evaluation procedure and selected results. It also formulates recommendations on refining the questionnaire for the 2021 Census.*



**David Bretschi**

studierte Survey Methodology (M. A.) an der Universität Duisburg-Essen und ist wissenschaftlicher Mitarbeiter am GESIS – Leibniz-Institut für Sozialwissenschaften. Von 2010 bis 2012 sowie 2015 bis 2016 war er im Statistischen Bundesamt im Bereich Zensus tätig und unter anderem für die Evaluierung des Fragebogens der Haushaltsstichprobe zuständig.



**Miriam Pfahl**

studierte Politikwissenschaft, Soziologie und Öffentliches Recht in Mainz. Sie ist wissenschaftliche Mitarbeiterin im Referat Haushaltsstichprobe (Konzeption und Aufbereitung) des Statistischen Bundesamtes. Neben der Online-first-Strategie arbeitet sie an der Weiterentwicklung und dem Pretest der Erhebungsinstrumente für die Haushaltsstichprobe im Zensus 2021.

## 1

### Hintergrund und Zielsetzung

Das registergestützte Zensusmodell Deutschlands sieht neben der Auswertung vorhandener Verwaltungsregister ergänzende Primärerhebungen vor. Zentrales Instrument zur Erhebung der relevanten Informationen ist dabei ein standardisierter Fragebogen. Der Fragebogen der Haushaltebefragung auf Stichprobenbasis (Haushaltsstichprobe) im Zensus 2011 bestand aus 46 Fragen zu verschiedenen Themengebieten.<sup>1</sup> Neben demografischen Angaben wurden Informationen zu Bildung und Ausbildung, Erwerbstätigkeit und Arbeitsort, Wirtschaftszweig, Religion und Migrationshintergrund erhoben. Dabei konnten die Befragten ihre Angaben wahlweise im Rahmen eines persönlich-mündlichen Interviews gegenüber Erhebungsbeauftragten (PAPI) oder als Selbstausfüller tätigen. Neben dem Papierfragebogen (PAP) gab es für Selbstausfüller zusätzlich die Option, über einen Online-Fragebogen (CAWI) auf elektronischem Weg Auskunft zu erteilen.<sup>2</sup>

Damit das Erhebungsinstrument gültige und zuverlässige Ergebnisse liefert, der Fragebogen also das misst, was er messen soll, wurde er nach wissenschaftlichen Standards konstruiert. Ferner wurde der Fragebogen mithilfe eines qualitativen als auch eines quantitativen Pretests auf Verständlichkeit und Handhabbarkeit getestet (Gauckler, 2011). Trotz sorgfältiger Entwicklung und Erprobung zeigt letztlich aber erst der Einsatz im Feld, ob das Erhebungsinstrument funktioniert beziehungsweise wo es Schwachstellen aufweist.

Im Rahmen der Evaluierung wurden das Ausmaß an Unplausibilitäten in den Daten der Haushaltebefragung (vor deren Aufbereitung) sowie mögliche Ursachen hierfür untersucht. Damit wurden zwei Ziele verfolgt:

- › Eine Qualitätseinschätzung der Zensusdaten aus der Haushaltebefragung

Messfehler stellen eine Abweichung zwischen dem wahren Wert eines interessierenden Konstrukts (zum Beispiel dem Erwerbsstatus) und dem gemessenen Wert, so wie er im Datensatz hinterlegt ist, dar. Unstimmigkeiten, die im Zuge der Plausibilitätsprüfungen bei der Datenaufbereitung identifiziert wurden, liefern Anhaltspunkte für mögliche Schwierigkeiten bei der Fragenbeantwortung.<sup>3</sup> Allerdings geben diese Unplausibilitäten nur begrenzt Auskunft über deren Ursachen. Eine Ursachenanalyse ist jedoch wichtig für das zweite Ziel der Evaluierung:

- › Die Weiterentwicklung des Fragebogens für den Zensus 2021

Vor diesem Hintergrund wurde untersucht, inwiefern es den Befragten möglich war, eine Frage korrekt zu beantworten.

Um zu verstehen, warum Fehler im Erhebungsprozess entstehen und wie diese verringert werden können, kann man auf kognitionswissenschaftliche Modelle zurückgreifen.

## 2

### Ablaufmodell der kognitiven Aufgaben von Befragten im Antwortprozess

Geläufig sind hier Modelle, die den mentalen Prozess der Beantwortung von Fragen in vier Schritte unterteilen (Tourangeau und andere, 2000):

- › Schritt 1: Verständnis. Zunächst müssen Befragte die Frage verstehen.

Beispiel Frage 11: Wie viele Personen leben insgesamt in Ihrer Wohnung?

Befragte müssen die Begriffe „Wohnung“ und „leben“ interpretieren und hierfür eine Definition finden.

- 3 Zu bedenken ist jedoch, dass diese erkennbaren Unplausibilitäten lediglich einen nur schwer quantifizierbaren Anteil des Gesamtfehlers bei der Schätzung von Populationsparametern, des Total Survey Errors, darstellen (Biemer/Lyberg, 2003).

1 Der Fragebogen der Haushaltebefragung im Zensus 2011 kann abgerufen werden unter [www.zensus2011.de/](http://www.zensus2011.de/)

2 PAPI = Paper and Pencil Interviewing. Die Befragung erfolgte durch Interviewerinnen und Interviewer anhand eines Papierfragebogens. PAP = Paper and Pencil. Die Bezeichnung steht für die papierbasierten Selbstausfüller-Meldungen.

CAWI = Computer Assisted Web Interviewing. Diese Bezeichnung steht für die elektronischen Selbstausfüller-Meldungen über den Online-Fragebogen (IDEV). Die Internetdatenerhebung im Verbund (IDEV) ist das Online-Meldesystem der amtlichen Statistik.

- › Schritt 2: Abruf. Weiterhin müssen die Befragten die zur Beantwortung erforderlichen Informationen abrufen. Im oben genannten Beispiel müssen sie also im Gedächtnis nach der Anzahl an Personen suchen, die in der zuvor definierten Wohnung leben.
- › Schritt 3: Beurteilung. Die Befragten müssen die gefundenen Informationen schließlich in eine passende Antwort integrieren. Das heißt, sie müssen prüfen, ob die gefundene Anzahl an Personen vollständig ist und als Antwort zu der Frage passt. Gegebenenfalls muss eine Schätzstrategie angewandt werden, um unvollständige Erinnerungen zu ergänzen, oder nach weiteren Informationen zu suchen.
- › Schritt 4: Antwort. Letztlich muss die durch die vorherigen Schritte formulierte Antwort in den vorgesehenen Antwortkategorien abgebildet werden. Dabei darf die Angabe dieser gefundenen Antwort nicht anderen, höher eingeschätzten Zielen entgegenstehen (ein Beispiel dafür wäre die Vermeidung negativer Konsequenzen).

Um eine möglichst geringe Anzahl an Messfehlern zu produzieren, sollte umgekehrt erfüllt sein, dass

- › Befragte eine Frage einfach verstehen und einheitlich interpretieren,
- › sie die relevanten Informationen besitzen und korrekt abrufen können,
- › sie diese Informationen unter Verwendung einer geeigneten Schätz- oder Zählstrategie zu einer genauen Antwort verbinden können und
- › sie die Antwort abbilden können sowie bereit sind, ehrlich zu antworten.

Bei jedem dieser Schritte können Fehler auftreten, die eine genaue Antwort verhindern. Aus diesem Modell lassen sich sieben Probleme bei der Beantwortung von Fragen ableiten, die zu Messfehlern im Erhebungsprozess führen können (Groves und andere, 2013):

- › Die fehlende Aufnahme der relevanten Informationen vor der Befragung
- › Missinterpretation der Frage
- › Das Vergessen relevanter Ereignisse sowie andere Erinnerungsprobleme
- › Fehlerhafte Beurteilungs- oder Schätzstrategien

- › Probleme in der Abbildung einer Antwort
- › Absichtliche Falschangaben
- › Das Scheitern von Instruktionen

Das Ablaufmodell hilft dabei, in systematischer Form über verschiedene Probleme nachzudenken, die dazu führen können, dass ungenaue Antworten erfolgen (Groves und andere, 2013). Auf Grundlage dieses Modells können die Fragen des Fragebogens analysiert und gezielt Maßnahmen zu deren Optimierung ergriffen werden.

## 3

---

### Datengrundlage

---

Für die Untersuchung von Fragebogeneffekten auf Messfehler in den Zensusdaten wurden unterschiedliche Datenquellen herangezogen. Zentrale Datenquelle waren die Ergebnisse der Haushaltebefragung sowie der dazugehörige Fragebogen. Der Datenbestand enthält neben den jeweiligen Angaben der Befragungspersonen auch Metadaten in Form von Qualitätskennzeichen. Diese umfassen Informationen über Unplausibilitäten in den Rohdaten, welche durch Prüfprozesse im Rahmen der Aufbereitung identifiziert und bereinigt werden konnten.<sup>14</sup>

#### 3.1 Quantitative Datenanalyse

---

In einem ersten Schritt wurden die Zensusdaten aus der Haushaltebefragung hinsichtlich des Umfangs und Typs der Unplausibilitäten analysiert. Hierbei lassen sich folgende Fehlertypen unterscheiden:

- › Fehlende Angaben/Nonresponse:

Bei dieser Fehlerart fehlen Angaben einer Person, obwohl sie hätten erfolgen sollen.

---

<sup>14</sup> Weitere Informationsquellen stellen Dokumente mit einem direkten oder indirekten Bezug zum Fragebogen der Haushaltsstichprobe dar. Das sind beispielsweise die Berichte vom quantitativen und qualitativen Pretest, Fragebogen von Vergleichsstatistiken wie dem Mikrozensus, der Zensus im Vereinigten Königreich und den Vereinigten Staaten sowie Empfehlungen der Methodenforschung (Demographische Standards).

› Filterfehler/unnötige Angaben:

Hier sind Angaben vorhanden, obwohl diese hätten nicht erfolgen sollen, weil beispielsweise (von Befragten oder Fragenden) ein Filter übersehen oder missachtet und Folgefragen beantwortet wurden. Im Gegensatz zu fehlenden Angaben haben Filterfehler in der Regel geringere negative Konsequenzen für die Datenqualität, da unnötige Daten im einfachsten Fall aus den Datensätzen entfernt werden können.

› Strukturfehler/Interunplausibilitäten:

Zu diesem Fehlertyp zählen unter anderem Wertebereichsverletzungen (Beispiel: die Anzahl an Personen je Wohnung beträgt mehr als 20) oder Angaben zu einer Frage, die logisch nicht mit den Angaben zu weiteren Fragen übereinstimmen (Beispiel: Geschlecht männlich und in Mutterschutz).

### 3.2 Qualitative Analyse

In einem zweiten Schritt erfolgte die systematische Prüfung jeder einzelnen Frage- und Antwortformulierung des Fragebogens anhand einer Prüfliste, dem sogenannten Question Appraisal System (QAS; Willis/Lessler, 1999). Dieses Instrument zur systematischen Untersuchung von Fragen ist abgeleitet aus dem Modell des kognitiven Beantwortungsprozesses und der Fehlerquellen, die diesen Ablauf beeinträchtigen können. Das QAS besteht aus sieben Prüfschritten, die jeweils verschiedene Hinweise auf potenzielle Schwachstellen und damit Fehlerquellen bei der Fragebogenentwicklung enthalten. Dabei werden die Fragen auf typische Probleme bei der Fragenbeantwortung, wie eine Missinterpretation der Frage, Erinnerungsprobleme durch die Befragten, absichtliche Falschangaben oder das Scheitern in der Befolgung von Fragebogeninstruktionen, untersucht. Durch die Betrachtungsweise einzelner Fragen vermag das QAS allerdings keine Fehler zu identifizieren, die sich auf den gesamten Fragebogen beziehen, wie die Fragenreihenfolge oder Effekte, die sich durch die Reihenfolge der Antwortmöglichkeiten ergeben.

### 3.3 Einschränkungen

Eine zentrale Einschränkung der Aussagekraft durch die Analysen ergibt sich daraus, dass nur ein Teil der Fehler durch die Plausibilitätsprüfungen in der Aufbereitung

ersichtlich wird. Dieser Annahme liegt die Erkenntnis zugrunde, dass bestimmte Messfehler keine erkennbaren Unplausibilitäten erzeugen (Bradburn und andere, 2004; Presser und andere, 2004).<sup>5</sup> Weiterhin ist zu berücksichtigen, dass die beobachteten Fehler neben bestimmten Effekten durch den Fragebogen auch durch weitere Fehlerquellen, wie die Interviewerinnen und Interviewer, die Befragten oder den Befragungsmodus, entstehen können.

## 4

### Ausgewählte Ergebnisse

Die Ergebnisse umfassen eine Darstellung der unterschiedlichen Typen an Unplausibilitäten in den Angaben der Haushaltebefragung, differenziert nach Erhebungsmodus. Neben einer Gesamtbetrachtung der erkennbaren Fehlertypen werden ausgewählte Einzelfallanalysen der Frage- und Antwortformulierung des Haushaltsfragebogens präsentiert und interpretiert.

#### 4.1 Unplausibilitäten nach Fragen

Insgesamt zeigt die Analyse, dass der Anteil der identifizierten Unplausibilitäten in den Daten der Haushaltebefragung gering ist. [↘ Tabelle 1](#) Eine Betrachtung differenziert nach einzelnen Fragen lässt erkennen, dass die

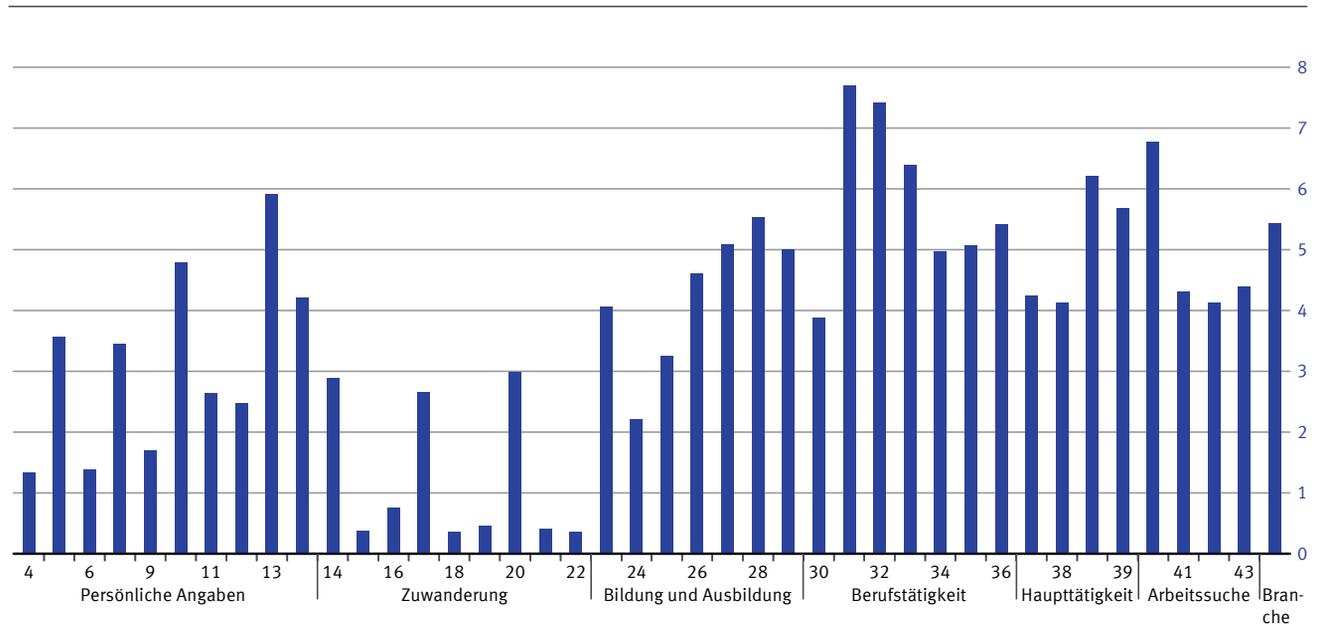
**Tabelle 1**  
Plausibilitäten und Unplausibilitäten beim Fragebogen der Haushaltsstichprobe des Zensus 2011 differenziert nach Erhebungsmodus

	Interview (PAPI)	Online-Fragebogen (CAWI)	Papierfragebogen (PAP)
	%		
Plausibilitäten	97,2	99,7	86,7
Unplausibilitäten	2,8	0,3	13,3
fehlende Angaben	1,1	0,1	6,6
unnötige Angaben	1,0	0,0	3,2
Strukturfehler/ Interunplausibilitäten	0,7	0,2	3,5

5 Es muss umgekehrt auch davon ausgegangen werden, dass die beobachtbaren Unplausibilitäten nur einen Teil der tatsächlich entstandenen Messfehler darstellen.

**Grafik 1**

Identifizierte Unplausibilitäten je Frage<sup>1</sup> der Haushaltsstichprobe des Zensus 2011  
in %



1 Ohne die Fragen 1 bis 3, Frage 8 und Frage 45. Frage 13: unterteilt in Haupt- und Nebenwohnsitz, Frage 39: unterteilt in Postleitzahl und Ort des Arbeitsorts.

2018 - 01 - 0112

Zahl der Unplausibilitäten im Verlauf des Fragebogens steigt. [↪ Grafik 1](#)

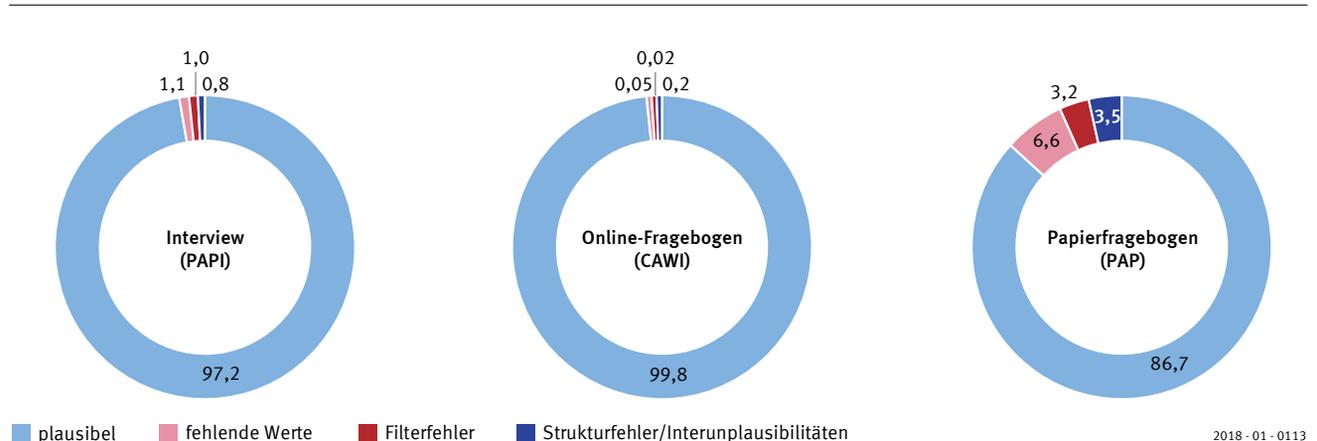
Ursachen für diese Tendenz könnten sein, dass Filterfehler häufiger in späteren Fragen aufgetreten sind, die Themenbereiche komplexer wurden oder möglicherweise die Motivation der Befragten nachgelassen hat.

## 4.2 Zusammenhang zwischen Unplausibilitäten und Erhebungsmodus

Betrachtet man den Zusammenhang zwischen Unplausibilitäten und Erhebungsmodus, sind deutliche Unterschiede je nach Erhebungsweg zu erkennen. [↪ Grafik 2](#)

**Grafik 2**

Plausibilitäten und Unplausibilitäten beim Fragebogen der Haushaltsstichprobe des Zensus 2011 differenziert nach Erhebungsmodus  
in %



2018 - 01 - 0113

Die plausibelsten Daten sind diejenigen, die elektronisch über IDEV eingegangen sind. Hier haben bereits während der Dateneingabe Plausibilitätsprüfungen stattgefunden, die Filterführung wurde automatisch vollzogen und die Vollständigkeit der Eingaben geprüft. So konnten noch im Befragungsprozess etwaige Unstimmigkeiten durch die Auskunftspflichtigen selbst korrigiert werden. Bei den Selbstausfüllern des Papierfragebogens zeigen sich hingegen deutlich höhere Anteile an Unplausibilitäten: im Durchschnitt knapp 13 % je Frage. So ist beispielsweise der Anteil an fehlenden Werten (Nonresponse) bei diesem Erhebungsweg verhältnismäßig hoch. Im Vergleich dazu zeigt sich in den Daten der interviewergeführten Befragungen ein deutlich geringerer Anteil an Unplausibilitäten. Dies könnte auf umfangreiche und aufwendige Schulungen der Interviewerinnen und Interviewer zurückzuführen sein: Diese kompensierten die dem Papierfragebogen fehlende elektronische Unterstützung teilweise dadurch, dass sie die Befragten bei Unverständnis beziehungsweise Unklarheiten beraten konnten. Nonresponse resultiert häufig aus Verständnisschwierigkeiten der Frage durch Befragte. Dass in den IDEV-Daten kaum fehlende Werte enthalten waren, ist allerdings eher auf die technisch geforderte Eingabe einer Angabe zurückzuführen. Die Schlussfolgerung, dass Befragte in diesem Modus alle Fragen sehr viel besser verstanden hätten, lässt sich hieraus nicht ableiten.

### 4.3 Zusammenhang von Unplausibilitäten je Fragebogen und Erhebungsmodus

---

Interessant ist, wie sich die beobachteten Fehleranteile auf einen Fragebogen verteilen. Sind wenige Fragebogen für viele Unplausibilitäten verantwortlich, oder verteilen sich wenige Unplausibilitäten auf viele Fragebogen? Hier zeigt die Analyse, dass bei der elektronischen Datenerhebung ebenso wie bei der Erhebung durch Interviewerinnen und Interviewer eine vergleichsweise geringe Anzahl an Unplausibilitäten je Fragebogen zu erkennen ist. Bei den Selbstausfüllern des Papierfragebogens ist die Anzahl an Fragebogen mit vielen Fehlern hingegen recht hoch. Dies lässt vermuten, dass nicht einzelne Fragen das Problem darstellten. Fehlende Angaben oder bewusste Falschangaben als Zeichen eines Protests oder aber generelle Verständnisschwierigkeiten beispielsweise durch mangelnde Sprachkenntnisse könnten hier

für ursächlich sein. Allerdings lassen sich Moduseffekte nicht von Selektionseffekten trennen. Daher sind auf der Grundlage der Daten keine eindeutigen Rückschlüsse möglich, inwieweit der Erhebungsmodus für die unterschiedlichen Anteile an Unplausibilitäten verantwortlich ist: Personen mit hoher Motivation und hohen kognitiven Fähigkeiten könnten beispielsweise eher den elektronischen Fragebogen genutzt haben.

Die Analyse gibt einen guten Einblick in den Umfang der identifizierbaren Fehler und liefert erste Hinweise, an welchen Stellen im Fragebogen es zu Problemen gekommen ist. Über die Ursachen von Unplausibilitäten und damit Messfehlern erlangt man allerdings keine Kenntnis. Daher erfolgten in einem zweiten Schritt die gezielte Betrachtung des Fragebogens selbst sowie eine Analyse der Antwortfehler.

## 5

---

### Analyse einzelner Fragen des Fragebogens

---

Die Analyse einzelner Fragen nach typischen Antwortfehlern erfolgte zum einen durch gezielte Datenanalyse unter Einbeziehung der Pretestergebnisse zum Zensus 2011, der Demographischen Standards<sup>6</sup> sowie von Vergleichsstatistiken. Die einzelnen Fragen- und Antwortformulierungen wurden zudem anhand der QAS-Checkliste geprüft, um so Anhaltspunkte über potenzielle Probleme zu erhalten. Im Folgenden werden ausgewählte Ergebnisse präsentiert.

#### Fragen 7 und 8: Religionszugehörigkeit und Glaubensbekenntnis

Der größte Anteil an Unplausibilitäten trat bei der Frage nach dem Glaubensbekenntnis auf. Mehr als 51 % der ursprünglichen Angaben wurden im Nachgang der Plausibilitätsprüfung korrigiert. Zu beobachten ist eine hohe

---

<sup>6</sup> Die Demographischen Standards dienen dem Zweck, sozialstrukturelle Erhebungsmerkmale in Bevölkerungsumfragen (Haushalts- und Personenbefragungen) zu vereinheitlichen, um eine größere Vergleichbarkeit zwischen einzelnen Surveys zu ermöglichen. Sie gehen auf die Initiative eines gemeinsamen Arbeitskreises – bestehend aus Vertreterinnen und Vertretern des Statistischen Bundesamtes, der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V. (ASI) und des ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. – zurück (Statistisches Bundesamt, 2016).

**Grafik 3**

Fragen 7 und 8: Religionszugehörigkeit und Glaubensbekenntnis

**7 Welcher Religionsgesellschaft gehören Sie an?**

Römisch-katholische Kirche .....	<input type="checkbox"/>	
Evangelische Kirche .....	<input type="checkbox"/>	
Evangelische Freikirchen .....	<input type="checkbox"/>	
Orthodoxe Kirchen .....	<input type="checkbox"/>	
Jüdische Gemeinden .....	<input type="checkbox"/>	
Sonstige öffentlich-rechtliche Religionsgesellschaft .....	<input type="checkbox"/>	
Keiner öffentlich-rechtlichen Religionsgesellschaft .....	<input type="checkbox"/>	

**8 Zu welcher der folgenden Religionen, Glaubensrichtungen oder Weltanschauungen bekennen Sie sich?**

**Die Beantwortung der Frage ist freiwillig.**

Christentum .....	<input type="checkbox"/>
Judentum .....	<input type="checkbox"/>
Islam	
Sunnitischer .....	<input type="checkbox"/>
Schiitischer .....	<input type="checkbox"/>
Alevitischer .....	<input type="checkbox"/>
Buddhismus .....	<input type="checkbox"/>
Hinduismus .....	<input type="checkbox"/>
Sonstige Religion, Glaubensrichtung oder Weltanschauung .....	<input type="checkbox"/>
Keiner Religion, Glaubensrichtung oder Weltanschauung .....	<input type="checkbox"/>

Anzahl an Filterfehlern, die sich auf den thematischen Kontext der Fragen zurückführen lassen. So haben die Befragten den Filter an Frage 7 häufig nicht beachtet und Frage 8 beantwortet, obwohl dies aufgrund der zuvor getätigten Angabe überflüssig war: Über 1,2 Millionen Befragte (rund 16 %) beantworteten Frage 8, obwohl sie in Frage 7 die Zugehörigkeit zu einer Religionsgesellschaft angegeben hatten. Auch in 25 % der interviewergeführten Fragebogen ist dieser Fehler aufgetreten. Die

Analyse der Fragen- und Antworttexte lässt zudem Verständnisschwierigkeiten vermuten: Insbesondere der Begriff „Religionsgesellschaft“ könnte Interpretationsprobleme hervorgerufen haben.<sup>17</sup> So wird die Bedeutung erst durch die Antwortkategorien verständlicher. Zudem kann die Frage von den Auskunftspflichtigen unterschiedlich interpretiert worden sein. Ist hier die rechtliche Zugehörigkeit zu einer Religionsgesellschaft oder ein Zugehörigkeitsempfinden gemeint? Auch hier könnte erst das Einbeziehen der Folgefrage Aufschluss gegeben haben.

Zu Problemen könnte auch die Formulierung „öffentlich-rechtliche Religionsgesellschaft“ geführt haben. Möglicherweise wussten Befragte und Erhebungsbeauftragte nicht, ob eine Religionsgesellschaft öffentlich-rechtlich ist oder nicht. Zudem bestehen bundesländerspezifische Regelungen, welche Religionsgesellschaften den Status „öffentlich-rechtlich“ haben. Als schwierig erwies sich auch die Einordnung von Muslimen. Von den 230440 Befragten mit Angaben zu einem islamischen Glauben wählten zuvor nur etwa 58 % die korrekte Kategorie „keiner öffentlich-rechtlichen Religionsgesellschaft“ aus. Rund 24 % gaben an, einer „sonstigen“ Religionsgesellschaft anzugehören. Der Anteil an fehlenden Werten liegt in diesem Beispiel bei 16,9 %. Eine Erklärung dieser Fehler könnte darin liegen, dass Befragten und Fragenden nicht bekannt war, dass die islamischen Glaubensrichtungen in Deutschland keine Religionsgesellschaften im öffentlich-rechtlichen Sinne sind.

Verständnisschwierigkeiten könnte in Frage 8 auch das Wort „bekennen“ hervorgerufen haben. „Bekennen“ könnte dahingehend verstanden worden sein, auf Nachfrage eine Zugehörigkeit zuzugeben beziehungsweise einen Glauben öffentlich zu praktizieren. Andererseits könnte der Begriff im Kontext der Frage und des Fragebogens allgemein als ein Zugehörigkeitsempfinden zu einem Kulturkreis interpretiert worden sein. Zudem ist unklar, was Eltern bei einer Proxy-Auskunft für ihre Kinder angeben sollten. Ebenfalls nicht eindeutig ist, wie sich die Kategorie „Sonstige Religion...“ von „Keiner Religion...“ unterscheidet. So könnte es Befragten schwergefallen sein, atheistische Überzeugungen zuzuordnen.

<sup>17</sup> Entsprechende Hinweise hierzu finden sich auch in den Ergebnissen des qualitativen Pretests des Fragebogens.

Unklar ist, ob diese Frage als sensitive Frage wahrgenommen wurde.<sup>8</sup> Dies könnte möglicherweise die vergleichsweise hohe Anzahl an Item-Nonresponse bei den PAP-Selbstaussfüllern erklären (14,1 %).

### Frage 29: Bildungs-/Ausbildungsabschluss

Die Analyse der Frage zum Bildungs- respektive Ausbildungsabschluss offenbart exemplarisch eine Vielzahl potenzieller Probleme, die gleich auf mehrere Fragen zutreffen können. In den Antworten war ein hoher Anteil an Nonresponse sowie Strukturfehlern beziehungsweise

#### Grafik 4

Frage 29: Bildungs-/Ausbildungsabschluss

**29 Welchen höchsten beruflichen Ausbildungs- oder (Fach-)Hochschulabschluss haben Sie?**

• Ordnen Sie bitte im Ausland erworbene  
I Abschlüsse einem gleichwertigen deutschen Abschluss zu.

Anlernausbildung oder berufliches Praktikum von mindestens 12 Monaten .....	<input type="checkbox"/>
Berufsvorbereitungsjahr .....	<input type="checkbox"/>
Lehre, Berufsausbildung im dualen System .....	<input type="checkbox"/>
Vorbereitungsdienst für den mittleren Dienst in der öffentlichen Verwaltung .....	<input type="checkbox"/>
Berufsqualifizierender Abschluss an einer Berufsfachschule/Kollegschule, Abschluss einer 1-jährigen Schule des Gesundheitswesens .....	<input type="checkbox"/>
2- oder 3-jährige Schule des Gesundheitswesens (z.B. Krankenpflege, PTA, MTA) .....	<input type="checkbox"/>
Fachschulabschluss (Meister/-in, Techniker/-in oder gleichwertiger Abschluss) .....	<input type="checkbox"/>
Berufsakademie, Fachakademie .....	<input type="checkbox"/>
Abschluss einer Verwaltungsfachhochschule .....	<input type="checkbox"/>
Fachhochschulabschluss, auch Ingenieur- schulabschluss .....	<input type="checkbox"/>
Abschluss einer Universität, wissenschaftlichen Hochschule, Kunsthochschule .....	<input type="checkbox"/>
Promotion .....	<input type="checkbox"/>

8 Laut Sudman und Bradburn (1982) stellen Fragen nach der Religion (zumindest in den Vereinigten Staaten zur Zeit der Buchveröffentlichung) keine sensitiven Fragen dar. Die Ergebnisse des qualitativen Pretests deuten jedoch eine gewisse Sensitivität der Frage an.

Interunplausibilitäten bei den PAP-Selbstaussfüllern zu beobachten (12,4 % beziehungsweise 6,1 %). In den anderen Modi hingegen kamen kaum Unplausibilitäten vor.

Die Frage ist in ihrer Formulierung für Selbstaussfüller optimiert. In über 80 % der Fälle wurde sie jedoch von einer Interviewerin beziehungsweise einem Interviewer gestellt. Somit standen diese vor der Herausforderung, wie die Frage vorzulesen war: Mit den oder ohne die Angaben in Klammern? Werden alle Antwortkategorien vorgelesen? Werden den Befragten die Fragebogen zum selbstständigen Lesen der Antwortmöglichkeiten gereicht? Es muss davon ausgegangen werden, dass die Interviewerinnen und Interviewer unterschiedliche Vorgehensweisen wählten, wobei eine Vorgehensweise nicht unbedingt besser sein musste als eine andere. Die Abweichung führte dennoch vermutlich zu Interviewer-varianzen.<sup>9</sup> Die Formulierung des Fragentextes sollte künftig so erfolgen, dass Interviewerinnen und Interviewer nicht entscheiden müssen, wie die Frage vorzulesen ist, oder es sollte die Verwendung einer Antwortliste vorgeschrieben werden, die den Befragten gezeigt wird.<sup>10</sup>

Die hohe Anzahl an Antwortkategorien mit den entsprechenden Fachbegriffen sowie die Einordnung einer Wertigkeit könnten zu Doppeleintragungen sowie (je nach Stimulus) zu Primacy- oder Recency-Effekten<sup>11</sup> führen. Letztere sind davon beeinflusst, ob Erhebungsbeauftragte die Kategorien vorlesen oder Befragte diese selbst lesen. Die Frage erfordert aufgrund der hohen Anzahl an Antwortkategorien zudem eine vergleichsweise hohe mentale Anstrengung und Motivation der Befragten, um die Fachbegriffe zu verstehen und den eigenen Status zuzuordnen. Einen Hinweis hierfür liefert möglicherweise der Anteil an Nonresponse unter den PAP-Selbst-

9 Interviewervarianz ist der Anteil an der Gesamtvarianz, die auf den Interviewer beziehungsweise die Interviewerin zurückzuführen ist und die letztlich zu einer Verringerung der Präzision von Schätzungen führen kann. Standardisierte Vorgehensweisen sind ein wesentlicher Grund, den ungewollten Einfluss von Interviewerinnen und Interviewern zu kontrollieren. Es ist folglich sinnvoll, dass man den Standardisierungsgrad erhöht (Lavrakas, 2008).

10 Dies wird im Mikrozensus praktiziert sowie von den Demographischen Standards empfohlen.

11 Primacy-/Recency-Effekte sind Reihenfolgeeffekte, die häufig abhängig von der Präsentation einer großen Anzahl von Antwortkategorien zu beobachten sind. Werden lange Antwortlisten optisch präsentiert, tendieren Befragte häufig dazu, unter den zuerst aufgeführten Kategorien auszuwählen (Primacy-Effekt). Werden die möglichen Antwortkategorien hingegen vorgelesen, ist eher mit einem Recency-Effekt zu rechnen: Befragte wählen tendenziell unter den zuletzt im Gehör gebliebenen Antworten aus (Krosnick/Alwin, 1987).

ausfüllen (12,4 %). Befragte und Erhebungsbeauftragte könnten zudem bei mehreren Abschlüssen Schwierigkeiten haben, die Wertigkeit der Abschlüsse zu bestimmen. Die Rangfolge der Kategorien spiegelt zwar den Hinweis für den „höchsten“ Abschluss, dennoch wird dies möglicherweise unterschiedlich interpretiert. Einen Hinweis auf diese These liefert der vergleichsweise hohe Anteil an Interunplausibilitäten (2,2 %) – wiederum insbesondere bei den PAP-Selbstaussfüllern (6,1 %) –, die zu einem großen Anteil auf Doppeleintragungen zurückgehen. Eine Empfehlung wäre, die Frage in zwei Fragenblöcke aufzuteilen sowie Mehrfachangaben zuzulassen.

#### Frage 40: Arbeitssuche

Bei der Frage nach der Arbeitssuche zeigt sich eine hohe Anzahl an Unplausibilitäten, insbesondere wieder in der Gruppe der Selbstaussfüller des Papierfragebogens. Auch der hohe Anteil an Nonresponse bei den über 71-Jährigen (71 bis 80 Jahre: 4,5 %; 81 Jahre und älter: 8,8%) fällt auf. Eine Ursache könnte darin liegen, dass von dieser Befragtengruppe die Frage als nicht zutreffend interpretiert wurde. Antwortfehler lassen sich insbesondere aufgrund des Referenzrahmens annehmen. Frage 40 erfasst ein zeitvariantes Merkmal: Der zeitliche Referenzrahmen dieser Frage bezieht sich auf die letzten vier Wochen vor dem Zensusstichtag. Ein Stichtagsbezug wird hier aber nicht hergestellt. Somit ist davon auszugehen, dass die Befragten den Frageinhalt auf die letzten vier Wochen vor dem Befragungszeitpunkt und nicht auf die vier Wochen vor dem Stichtag bezogen haben.

#### Grafik 5

Frage 40: Arbeitssuche

**40 Haben Sie in den letzten vier Wochen etwas unternommen, um Arbeit zu finden?**

**i** Gemeint ist z.B. das Lesen von Stellenanzeigen.  
 Bitte kreuzen Sie auch dann „Ja“ an, wenn Sie ...

- ... nach einer Arbeit mit mindestens einer Arbeitsstunde pro Woche (z.B. 400-Euro Job) suchen.
- ... als Schüler/-in oder Student/-in einen Nebenjob suchen.
- ... eine Tätigkeit als Selbstständige/-r anstreben.

Ja .....

Nein, ich habe bereits eine Tätigkeit gefunden. ....

Nein, ich suche keine Arbeit. ....  ► Ende der Befragung.

#### Frage 10: Nichteheliche Lebensgemeinschaften

Der Anteil an Unplausibilitäten liegt bei der Frage nach nichtehelichen Lebensgemeinschaften insgesamt bei 4,8%. Ein sehr hoher Anteil an Nonresponse ließ sich zum einen bei den Selbstaussfüllern des Papierfragebogens, zum anderen bei den unter 10-Jährigen (6,9%) beobachten. Auch hierfür könnten Verständnisschwierigkeiten ursächlich sein. Die Frage ist sehr lang. Komplexität erhält sie zum einen durch die zusätzliche Bedingung einer „Lebensgemeinschaft“, zum anderen durch die Negativabgrenzung „weder...noch“. So müssen Befragte zunächst feststellen, ob sie sich in einer Lebensgemeinschaft befinden. Anschließend gilt es festzulegen, ob auf diese Lebensgemeinschaft „Ehe“ oder „eingetragene Lebensgemeinschaft“ zutrifft, oder keins von beiden. Zudem kann davon ausgegangen werden, dass Interviewerinnen und Interviewer gerade aufgrund dieser hohen Komplexität den Fragentext eigenmächtig angepasst haben, was wiederum zur Interviewervarianz geführt haben könnte. Eine weitere Schwierigkeit könnte mit dem Begriff „Wohnung“ einhergegangen sein: Wohnen Befragte in mehreren Wohnungen gleichzeitig, so ist unklar, welche Wohnung mit der Frage gemeint ist. Zudem fehlt ebenfalls der Hinweis auf den eindeutigen Referenzrahmen – den Zensusstichtag. Dieser findet sich mit Ausnahme der Fragen 23 und 33 vergleichsweise klein gedruckt lediglich auf der ersten Seite des Fragebogens. Die Feldphase der Haushaltsstichprobe hatte letztlich gut drei Monate gedauert, sodass ein Fehlen des entsprechenden Referenzrahmens eine maßgebliche Fehlerquelle darstellen kann.

#### Grafik 6

Frage 10: Nichteheliche Lebensgemeinschaften

**10 Wohnen Sie in Ihrer Wohnung mit einem Partner/einer Partnerin in einer Lebensgemeinschaft zusammen, die weder Ehe noch eingetragene gleichgeschlechtliche Lebenspartnerschaft ist?**

Ja .....

Nein .....

### Fragen 23 bis 25: Bildung und Ausbildung

Der Anteil der beobachteten Unplausibilitäten ist auch in diesen Fragen zur Bildung und Ausbildung moderat (zwischen 2,2 und 4,1%). Nonresponse war in Frage 23 bei 11,4% der Selbstausfüllerbogen zu verzeichnen. So könnte der Begriff „allgemeinbildende Schule“ bei den Befragten Verständnisschwierigkeiten verursacht haben, da sie darunter auch Volkshochschulen, Fachhochschulen und Ähnliches verstanden haben könnten. Möglicherweise führte auch die zeitliche Einschränkung auf die Woche vom 9. bis 15. Mai 2011 zu Interpretationsschwierigkeiten, beispielsweise falls die Befragten aufgrund von Krankheit oder Ausflügen in dieser Woche die Schule nicht besucht haben.

#### Grafik 7

Fragen 23 bis 25: Bildung und Ausbildung

**23 Waren Sie in der Woche vom 9. bis 15. Mai 2011 Schüler/-in einer allgemeinbildenden Schule?**

Ja .....

Nein .....  Weiter mit Frage 26.

**24 Um welche Schule handelte es sich dabei?**

Grundschule .....

Hauptschule .....

Realschule .....

Gymnasium .....

Gesamtschule .....

Sonstige Schule .....

**25 Welche Klasse besuchten Sie?**

Klasse 1 bis 4 .....

Klasse 5 bis 9 oder 10 .....

Klasse 11 bis 13 (gymnasiale Oberstufe) .....

### Frage 30: Erwerbstätigkeit

Die Frage nach der Erwerbstätigkeit hat ebenfalls einen hohen Anteil an Nonresponse bei den Selbstausfüllern des Papierfragebogens (11,3%). Probleme könnten durch Interpretationsschwierigkeiten der Personen entstanden sein, für die aus subjektiver Sicht mehrere Kategorien zutreffen. Der Begriff „überwiegend“ dürfte für viele Befragte schwierig zu deuten gewesen sein – insbesondere dann, wenn mehrere Kategorien zutreffend erschienen. „Überwiegend“ kann dabei als zeitliche Dimension verstanden werden – also als jenen Status, für den die höchste wöchentliche Stundenzahl aufgebracht wird. Der Begriff könnte jedoch ebenso im Sinne von „offiziell“ interpretiert werden – also jener Status, der von den Befragten als bedeutender angesehen wird, auch wenn sie eine geringere Wochenzeit für die entsprechenden Aktivitäten aufgebracht haben. Weitere Schwierigkeiten sind möglicherweise durch den fehlenden Referenzrahmen entstanden.

#### Grafik 8

Frage 30: Erwerbstätigkeit

**30 Was trifft überwiegend auf Sie zu?**

Bitte kreuzen Sie nur eine Antwortmöglichkeit an.

Ich bin erwerbs- bzw. berufstätig (inkl. Auszubildende, Personen in Elternzeit oder Altersteilzeit). .....  Weiter mit Frage 33.

Ich bin Grundwehr-/Zivildienstleistender. ....

Ich bin Schüler/-in. ....

Ich bin Student/-in. ....

Ich bin Rentner/-in, Pensionär/-in. ....

Ich lebe von Einkünften aus Kapitalvermögen, Vermietung oder Verpachtung. ....

Ich bin Hausfrau/-mann oder versorge Kinder und/oder pflegebedürftige Personen. ....

Ich bin arbeitslos. ....

Keine der genannten Auswahlmöglichkeiten (z.B. dauerhaft arbeitsunfähig)

## 6

### Fazit und Empfehlungen für den Zensus 2021

Die Analyse hat gezeigt, dass in den bei der Haushaltsstichprobe erhobenen Rohdaten insgesamt wenige Unplausibilitäten enthalten waren. Während die elektronisch über IDEV eingegangenen Daten den vergleichsweise geringsten Anteil der identifizierten Fehler aufwiesen, sind in der Gruppe der Selbstausfüller des Papierfragebogens die meisten Unstimmigkeiten zu verzeichnen.

Anhand einer gezielten Untersuchung der einzelnen Fragen und Antworten des Fragebogens wurden mögliche Ursachen für die aufgetretenen Unplausibilitäten gesucht. Hieraus ergeben sich für die Weiterentwicklung des Fragebogens für den Zensus 2021 wertvolle Hinweise:

› Anpassung des Fragebogens an den Erhebungsmodus

Um den Interviewerinnen und Interviewern die Aufgabe zu erleichtern, Antwortfehler zu reduzieren und um die Interviewervarianz zu verringern, sollten die Fragen für das persönliche Interview so formuliert sein, dass diese eindeutig und einheitlich vorgetragen werden können. Zudem sollte einheitlich und klar geregelt sein, ob die Befragten Einsichtnahme in die Antwortkategorien bekommen sollen oder nicht. Um Reihenfolgeeffekte zu vermeiden, sollten den Befragten bei einer großen Anzahl von Antwortkategorien Listen vorgelegt werden.

› Sicherstellung eines Stichtagsbezugs bei zeitvarianten Merkmalen

Bei der Abfrage zeitvarianter Merkmale sollte bei der Fragenformulierung unbedingt ein deutlicher Bezug zum Zensusstichtag hergestellt werden.

› Reduktion der Fragenkomplexität

Sofern dies möglich ist, sollten besser mehrere einfache Fragen gestellt werden, statt weniger komplexer Fragen. Es ist also zu prüfen, ob schwierige Sachverhalte gegebenenfalls auf mehrere Fragen aufgeteilt werden können.

› Formulierung von Fragen und Antwortkategorien

Bei der Formulierung der Fragen und Antwortkategorien sollte darauf geachtet werden, einfache Begriffe zu verwenden, die von allen Befragten verstanden werden. Möglicherweise unklare Begriffe sind vorab zu definieren.

Die Erkenntnisse der Evaluierung sollten ebenfalls in die Konzeption der Pretests für den weiterentwickelten Fragebogen einfließen, um mögliche Effekte im Optimalfall iterativ zu testen.

Die Evaluierung hat darüber hinaus gezeigt, wie groß die Bedeutung gut geschulter Interviewerinnen und Interviewer für die Erhebung qualitativ hochwertiger Daten ist. Für Selbstausfüller, die ihre Angaben ohne Hilfestellung durch Erhebungsbeauftragte tätigen, sollte hingegen die elektronische Meldung zum Standardverfahren werden. An dieser Stelle setzt die Online-first-Strategie des Zensus 2021 an. 

## LITERATURVERZEICHNIS

---

Biemer, Paul P./Lyberg, Lars E. *Introduction to Survey Quality*. Hoboken/New Jersey 2003.

Bradburn, Norman M./Sudman, Seymour/Wansink, Brian. *Asking Questions: The Definitive Guide to Questionnaire Design – For Market Research, Political Polls, and Social and Health Questionnaires*. San Francisco 2004.

Gauckler, Britta. [Die Entwicklung des Fragebogens zur Haushaltebefragung des Zensus 2011](#). In: *Wirtschaft und Statistik*. Ausgabe 8/2011, Seite 718 ff.

Groves, Robert M./Fowler Jr., Floyd J./Couper, Mick P./Lepkowski, James M./Singer, Eleanor/Tourangeau, Roger. *Survey Methodology, Second Edition*. Hoboken/New Jersey 2013.

Krosnick, Jon A./Alwin, Duane F. *An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement*. In: *Public Opinion Quarterly*. Jahrgang 51. Ausgabe 2/1987, Seite 201 ff.

Lavrakas, Paul J. (Herausgeber). *Encyclopedia of Survey Research Methods*. Thousand Oaks 2008.

Presser, Stanley/Rothgeb, Jennifer M./Couper, Mick P./Lessler, Judith T./Martin, Elizabeth/Martin, Jean/Singer, Eleanor. *Methods for Testing and Evaluating Survey Questionnaires*. Hoboken/New Jersey 2004.

Statistisches Bundesamt (Herausgeber). *Demographische Standards. Ausgabe 2016*. Band 17 der Schriftenreihe Statistik und Wissenschaft. 6. überarbeitete Auflage. Wiesbaden 2016.

Sudman, Seymour/Bradburn, Norman M. *Asking questions: A Practical Guide to Questionnaire Design*. San Francisco 1982.

Tourangeau, Roger/Rips, Lance J./Rasinski, Kenneth. *The Psychology of Survey Response*. Cambridge 2000.

Willis, Gordon B./Lessler, Judith T. *Question Appraisal System QAS-99*. Rockville 1999.

---

**Herausgeber**

Statistisches Bundesamt (Destatis), Wiesbaden

---

**Schriftleitung**

Dr. Sabine Bechtold

Redaktionsleitung: Juliane Gude

Redaktion: Ellen Römer

---

**Ihr Kontakt zu uns**

[www.destatis.de/kontakt](http://www.destatis.de/kontakt)

---

**Erscheinungsfolge**

zweimonatlich, erschienen im April 2018

Das Archiv aller Ausgaben ab Januar 2001 finden Sie unter [www.destatis.de/publikationen](http://www.destatis.de/publikationen)

---

**Print**

Einzelpreis: EUR 18,- (zzgl. Versand)

Jahresbezugspreis: EUR 108,- (zzgl. Versand)

Bestellnummer: 1010200-18002-1

ISSN 0043-6143

ISBN 978-3-8246-1069-3

---

**Download (PDF)**

Artikelnummer: 1010200-18002-4, ISSN 1619-2907

---

**Vertriebspartner**

IBRo Versandservice GmbH

Bereich Statistisches Bundesamt

Kastanienweg 1

D-18184 Roggentin

Telefon: +49 (0) 382 04 / 6 65 43

Telefax: +49 (0) 382 04 / 6 69 19

[destatis@ibro.de](mailto:destatis@ibro.de)

---

Papier: Metapaper Smooth, FSC-zertifiziert, klimaneutral, zu 61% aus regenerativen Energien

© Statistisches Bundesamt (Destatis), 2018

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.