

# Eine methodische Kritik des Zensus 2022

Gutachten im Auftrag von 287 Gemeinden

## Zusammenfassung

Die Ergebnisse des Zensus 2022 wichen in zahlreichen Gemeinden so stark von den Melderegistern ab, dass mehrere Hundert Gemeinden Widerspruch gegen die Ergebnisse des Zensus eingelegten. Ein Zusammenschluss von 287 Gemeinden hat die Autoren mit einem Gutachten zu den Verfahren des Zensus 2022 beauftragt. Das Statistische Bundesamt hat keine einheitliche Dokumentation des Zensus 2022 vorgelegt, es gab keine Begleitforschung und mit Ausnahme weniger Arbeiten zur sogenannten „Präzisionszielfunktion“ keine wissenschaftlichen Arbeiten speziell zur Methodik des Zensus 2022. Im Rahmen des Gutachtens wurde die Literatur vollständig gesichtet, Dutzende Anfragen an die statistischen Ämter gestellt und schließlich Anfragen nach dem Informationsfreiheitsgesetz gestellt. Zusätzlich wurden auf der Basis anonymisierter Mikrodaten Simulationen, insbesondere zum Hochrechnungsverfahren, durchgeführt. Das Gutachten kritisiert jeden Schritt des Zensus, von der Definition der Population, über das Hochrechnungsverfahren bis zur Wiederholungsbefragung. Besondere Aufmerksamkeit gilt dabei den Nichtstichprobenfehlern und der mangelnden Dokumentation. Der Text stellt das den Gemeinden am 1.1.2026 zur Verfügung gestellte Gutachten mit einzelnen redaktionellen Änderungen dar.

## Abstract

The 2022 census results in many municipalities differed so much from official registers that several hundred filed objections. A consortium of 287 municipalities commissioned the authors to produce an expert report on the 2022 Census procedures. The Federal Statistical Office did not provide any standardised documentation of the 2022 Census; there was no accompanying research and, except for a few studies on the so-called ‘precision target function’, no academic work specifically on the methodology of the 2022 Census. As part of the expert report, the literature was reviewed in full, dozens of enquiries were made to the statistical offices, and requests were submitted under the Freedom of Information Act. The robustness of the population estimators was tested using simulation. The report examines every step of the census, from the definition of the population through the extrapolation procedure to the census coverage survey. Particular attention is given to non-sampling errors and the lack of documentation. Except for minor editorial changes, the text is the report delivered to the local authorities on 1 January 2026.

# Inhaltsverzeichnis

<b>1. Problemstellung</b>	<b>1</b>
<b>2. Unzureichende Dokumentation</b>	<b>2</b>
2.1. Zeitlicher Ablauf der Veröffentlichungen des Statistischen Bundesamtes .	2
2.2. Dokumentationsstandards für einen Zensus . . . . .	5
2.3. Zum Qualitätsbericht des Zensus 2022 . . . . .	6
2.4. Mangelnde Hilfsmerkmale . . . . .	8
2.5. Methodische Veröffentlichungen zum Zensus 2022 . . . . .	9
2.6. Zusammenfassung . . . . .	11
<b>3. Zielpopulation und untererfasste Bevölkerungsgruppen</b>	<b>12</b>
3.1. Populationsdefinition . . . . .	12
3.2. Obdachlose und irreguläre Migration . . . . .	15
3.3. Probleme bei der Erfassung ukrainischer Flüchtlinge . . . . .	17
<b>4. Das Konzept der Präzisionszielfunktion</b>	<b>19</b>
4.1. Konstruktion der Präzisionszielfunktion . . . . .	19
4.2. Quadratische Bézierkurve . . . . .	21
4.2.1. Exkurs: Zur Herleitung der Zielfunktion . . . . .	21
4.2.2. Verlauf der Kurve . . . . .	22
4.3. Zusammenfassung . . . . .	24
<b>5. Methodik der Stichprobenziehung</b>	<b>25</b>
5.1. Zusammenfassung von Gemeinden . . . . .	25
5.2. Schichtung der Stichprobe . . . . .	26
5.3. Das GREG-Modell zur Bestimmung der Stichprobenumfänge . . . . .	28
5.4. Kritische Bemerkungen zum GREG-Modell . . . . .	30
5.5. Zur Strategie der Varianzreduktion . . . . .	31
5.6. Zur Schätzung in kleinen Gemeinden . . . . .	33
5.7. Fallbeispiel des Amtes Röbel/Müritz . . . . .	34
5.8. Zusammenfassung . . . . .	35

---

<b>6. Hochrechnung</b>	<b>37</b>
6.1. Einführendes Beispiel . . . . .	37
6.2. Annäherung an das im Zensus verwendete Modell . . . . .	39
6.3. Das lineare Hochrechnungsmodell GREG . . . . .	40
6.4. Das Konzept des realisierten Standardfehlers . . . . .	42
6.5. Beeinflussung des Standardfehlers durch willkürliche Schichtung . . . . .	45
6.6. Hochrechnung der Über- und Untererfassungen . . . . .	47
6.7. Fallbeispiel der Stadt Hanau . . . . .	49
6.7.1. Anwendung des Regressionsmodells . . . . .	50
6.7.2. Lineare Abhängigkeit der Prädiktoren (Multikollinearität) . . . . .	51
6.8. Erläuterung der Konsequenzen für die Schätzung . . . . .	55
6.9. Zusammenfassung . . . . .	57
<b>7. Zum Verfehlen der Präzisionsziele der Hochrechnung</b>	<b>59</b>
7.1. Kritische Diskussion der Tabellierung . . . . .	61
7.2. Zusammenfassung . . . . .	62
<b>8. Nicht-Stichprobenfehler</b>	<b>63</b>
8.1. Mehrfachfallprüfung . . . . .	66
8.1.1. Record-Linkage und Linkage Bias . . . . .	67
8.1.2. Blocken in der Mehrfachfallprüfung im Zensus 2022 . . . . .	67
8.1.3. Qualitätsmetriken für die Zusammenführung von Datenbanken . . . . .	68
8.1.4. Die Bedeutung der Qualitätsmetriken des Record-Linkage für den Zensus	69
8.2. Auswirkung der Mehrfachfallprüfung auf die Korrektur der Einwohnerzahl	70
8.3. Existenzfeststellung . . . . .	71
8.3.1. Existenzfeststellung durch Proxy-Antworten . . . . .	72
8.3.2. Unvollständige Existenzfeststellungen . . . . .	72
8.3.3. Effekt unvollständiger Existenzfeststellungen auf die Hochrechnung . . . .	73
8.3.4. Probleme in der Erhebungsteilübergreifenden Plausibilisierung im Referenzdatenbestand . . . . .	74
8.3.5. Probleme durch die Softwareunterstützung der Existenzfeststellungen . .	76
8.4. Die Feldarbeit durch die Erhebungsbeauftragten . . . . .	77
8.4.1. Erhebungsbeauftragte als Fehlerquelle . . . . .	78
8.4.2. Exkurs: Interviewereffekte . . . . .	79
8.5. Festlegung von Haupt- und Nebenwohnsitzen . . . . .	83
8.6. Umgang mit Prozessfehlern . . . . .	83

---

<b>9. Wiederholungsbefragung</b>	<b>85</b>
9.1. Der Unterschied zwischen der Wiederholungsbefragung und einem PES .	86
9.2. Zur Durchführung der WDH . . . . .	86
9.3. Veröffentlichung der Ergebnisse der WDH . . . . .	87
9.4. Ergebnisse der Wiederholungsbefragung . . . . .	89
9.5. Zusammenfassung . . . . .	90
<b>10. Zusammenfassung: Wesentliche Kritikpunkte des Zensus 2022</b>	<b>91</b>
<b>11. Schlussbemerkungen</b>	<b>92</b>
<b>Anhang</b>	<b>93</b>
<b>A. Kurzbiografien der Gutachter</b>	<b>94</b>
A.1. Prof. Dr. Rainer Schnell . . . . .	94
A.2. Prof. Dr. Rainer Lenz . . . . .	94
A.3. Diplom-Volkswirt Rolf Schmidt . . . . .	95
<b>Literatur</b>	<b>96</b>

# 1. Problemstellung

Der Stichtag für den Zensus 2022 war der 15.5.2022. Nach mehrmaliger Verschiebung der Veröffentlichung der Ergebnisse wurden am 25.6.2024 die Einwohnerzahlen veröffentlicht. Wie schon beim Zensus 2011 waren die Ergebnisse für viele Gemeinden überraschend. Im Laufe des Jahres 2025 erteilten mehr als 270 Gemeinden den Gutachtern den Auftrag, die wesentlichen Kritikpunkte an der Konzeption, der Durchführung und den Berechnungen des Zensus zusammenzufassen. Die vertragliche Vereinbarung datiert vom 4.11.2025.

Der vorliegende Bericht wurde im Zeitraum vom 1. September 2025 bis zum 31. Dezember 2025 verfasst. Während dieses Bearbeitungszeitraums erfolgte ein fortlaufender Austausch mit verschiedenen Statistischen Landesämtern, dem Statistischen Bundesamt (Destatis) sowie dem europäischen Statistikamt Eurostat.

Trotz wiederholter und intensiver Bemühungen war eine vollständige Beschaffung der für die Untersuchung relevanten Unterlagen nicht möglich. Zu den nicht zugänglichen Materialien zählen primär Programmcodes von Destatis sowie anonymisierte Mikrodaten einzelner Gemeinden. Nach gegenwärtigem Erkenntnisstand wäre für eine vollständige methodische Dokumentation und kritische Analyse des Zensus ein Arbeitsaufwand von mindestens drei vollen Person Jahren erforderlich. Auf Grundlage der verfügbaren Datengrundlage beschränkt sich das Gutachten daher auf die Darstellung der als wesentlich erachteten Prozesse und methodischen Problemfelder.

Eine umfassende Analyse aller betroffenen Gemeinden würde die Durchführung spezifischer Einzeluntersuchungen auf Gemeindeebene voraussetzen – unter anderem die Überprüfung und Validierung der Sonderbereichslisten, die Analyse der Neubautätigkeit sowie eine Überprüfung der kommunalen Mikrodaten auf Vollständigkeit und Konsistenz. Bei einer Anzahl von 279 betroffenen Gemeinden konnte eine solche Detailanalyse im Rahmen der vorhandenen Ressourcen und Zeitlimits nicht realisiert werden. Wir haben Teile der Hochrechnung und Varianzabschätzungen sowie Studien zu den Effekten systematischer Fehler anhand eines vollständig anonymisierten Datensatzes mit den Ziel-1 relevanten Variablen einer Gemeinde simuliert, um Abschätzungen trotz des Mangels an zugänglicher Information zu ermöglichen. Zusammen mit einer detaillierten Literatursuche war so eine Identifikation der zentralen strukturellen Problembereiche und methodischen Schwachstellen möglich.

## 2. Unzureichende Dokumentation

Der Artikel 4, Absatz 2 der EU-Verordnung 1260/2013 legt fest:

»Die Mitgliedstaaten können die Gesamtbevölkerung nach Absatz 1 anhand der sich dort rechtmäßig aufhaltenden oder registrierten Bevölkerung schätzen, wobei wissenschaftlich fundierte, hinreichend dokumentierte und öffentlich verfügbare statistische Schätzmethoden verwendet werden.«

Bedeutsam sind die Begriffe a) *wissenschaftlich fundiert*, b) *hinreichend dokumentiert* und c) *öffentlich verfügbare* Schätzmethoden. Keines dieser Kriterien wird – nach Maßstäben wissenschaftlicher Veröffentlichungen – von der Dokumentation des Zensus 2022 in Deutschland erfüllt.

Der Statistische Verbund (also die kooperative Zusammenarbeit zwischen dem Statistischen Bundesamt und den Statistischen Ämtern der Länder) hat bis Ende 2025 ca. 100 Veröffentlichungen zur Methodik des Zensus 2022 vorgelegt.

Die Veröffentlichungen sind verstreut, es gibt keine Monografie oder auch nur eine Liste, welcher Verfahrensschritt im Detail wo dokumentiert wurde.<sup>1</sup>

### 2.1. Zeitlicher Ablauf der Veröffentlichungen des Statistischen Bundesamtes

Die detailliertesten Darstellungen der Ermittlung der Einwohnerzahl wurden am 12.8.2024, 16.4.2025 und 16.12.2025, also *nach* der Veröffentlichung der Einwohnerzahlen veröffentlicht (siehe Abbildung 2.1).<sup>2</sup>

Die Abbildung 2.1 zeigt den zeitlichen Verlauf der Erhebungen, der Veröffentlichung der Ergebnisse und der tatsächlich verwendeten Formeln der Hochrechnung. An diesem zeitlichen Ablauf sind vor allem drei Fakten bemerkenswert:

---

<sup>1</sup> Dies gilt auch für den Zensus 2011. Die ausführlichste Darstellung mit knapp 70 Seiten (Statistische Ämter des Bundes und der Länder, 2015) enthält für die Einwohnerzahl weder relevante Formeln, noch Code, noch technische Details.

<sup>2</sup> Die erste Veröffentlichung von Bretschgi et al. (2024a) enthält keine technischen Details.



Abbildung 2.1.: Datum der Erhebungen, der Veröffentlichung der Einwohnerzahl, Versand der Bescheide und der beiden Veröffentlichungen mit den Formeln der Hochrechnung.

1. Zwischen dem Ende der Primärdatenerhebung am 19.11.2022 und der Veröffentlichung der Einwohnerzahl im Rahmen einer Pressekonferenz am 25.6.2024 lagen 584 Tage, also ein Jahr mehr als ursprünglich beabsichtigt.
2. Die erste Veröffentlichung des tatsächlich verwendeten Hochrechnungsverfahrens erfolgte am 16.04.2025, also 295 Tage nach der Bekanntgabe der Einwohnerzahlen und damit mehr als vier Monate nach dem Versand der meisten Datenblätter an die Gemeinden.
3. Erst durch Peters und Lorentz (2025) am 16.12.2025 wurden Details des Hochrechnungsverfahrens für kleine Gemeinden veröffentlicht, also mehr als ein Jahr nach dem Versand der Datenblätter.

Der erste Punkt ist bemerkenswert, da das prinzipielle Verfahren der Hochrechnung bereits beim Zensus 2011 verwendet wurde. Eine zufriedenstellende Erklärung für die

---

se Verspätung um ein Jahr hat das Statistische Bundesamt auch auf Nachfrage nicht abgegeben; verwiesen wird lediglich auf qualitätssichernde Maßnahmen. Da nach dem 19.11.2022 keine Daten mehr erhoben oder geliefert wurden, kann es sich nur um zusätzliche, nicht dokumentierte Bereinigungen oder nicht dokumentierte Veränderungen des Hochrechnungsverfahrens handeln.<sup>3</sup>

Bedeutsamer aus Sicht der Gemeinden könnten die beiden anderen Punkte sein: Aufgrund der Veröffentlichungszeitpunkte des Hochrechnungsverfahrens konnte keine Gemeinde das Hochrechnungsverfahren oder die Mitteilung der amtlichen Einwohnerzahl zum Zeitpunkt des Versands der Datenblätter nachvollziehen. Das gilt sowohl rein mathematisch in Hinsicht auf das Verfahren als auch praktisch – aufgrund der nicht verfügbaren Mikrodaten – zur Reproduktion der Ergebnisse.

Darüber hinaus sind immer Programmier- und Ausführungsfehler möglich. Aus diesem Grund wird bei wissenschaftlichen Publikationen zunehmend von Fachverbänden, Forschungsförderungseinrichtungen und wissenschaftlichen Zeitschriften gefordert, dass sowohl Daten als auch die Programme für unabhängige Nachprüfungen prinzipiell immer verfügbar sein müssen. Für so folgenreiche Berechnungen wie den Zensus wären bei den Programmen schon aus Sicht der Informatik unabhängige Prüfungen notwendig, falls – wie in diesem Fall – keine Referenzimplementierung existiert. Das Statistische Bundesamt hat kein einziges Programm und keine einzige Zeile Programmcode veröffentlicht.<sup>4</sup>

Ob die Veröffentlichung von Einwohnerzahlen ohne vorherige detaillierte Dokumentation mit den einschlägigen gesetzlichen Regelungen im Bereich der amtlichen Statistik und den europäischen Vorgaben zur Rolle statistischer Behörden vereinbar ist, kann im Rahmen dieses Gutachtens nicht beurteilt werden. Beurteilt werden kann jedoch, dass die vorliegende Form der Dokumentation der einzelnen Prozessschritte des Zensus weder allgemein anerkannten wissenschaftlichen Standards der Nachvollziehbarkeit und Transparenz genügt, noch dem Selbstverständnis der Mehrzahl anderer nationaler Statistikämter in Europa hinsichtlich Dokumentation, Datenzugang und externer Überprüfbarkeit.

---

<sup>3</sup> Mit Datum vom 14.11.2025 teilt das Statistische Bundesamt mit, dass nach dem Ende der Feldphase keine Existenzfeststellungen mehr zulässig waren.

<sup>4</sup> Auf Anfrage der Gutachter nach Programm-Code teilt uns das Statistische Bundesamt (GZ 120111/812662; GZ120111/809986) mit, dass die Weitergabe von Programmcode gegen Urheberrechte verstoßen und ein Sicherheitsrisiko darstellen würde (wobei auch auf Nachfrage nicht angegeben wurde, worin denn dieses Sicherheitsrisiko besteht). Abgesehen davon bestünde kein Anspruch auf Herausgabe des Programmcodes.

---

## 2.2. Dokumentationsstandards für einen Zensus

Die Vereinten Nationen empfehlen, den gesamten Zensus in einer eigenen Publikation zu dokumentieren. Dies umfasst die Erstellung von Aufzeichnungen über die Erhebung und Verarbeitung der Daten. Der Abschnitt im UN-Handbuch (United Nations Department of Economic and Social Affairs, 2021, S. 228, 5.3.2.7) lautet:

»One of the most important reports in the publication programme is the administrative report, which is a record of the entire census undertaking, including problems encountered and their solutions. The report may include the following topics: a brief history of the census in the country, legal basis for conducting the census, budget requirements and control, census committees and their activities, census organization and personnel structure, quality control procedures, census calendar, census cartographic work, development and design of the questionnaires, enumeration methodology of each census, field organization, manual editing and coding, data-processing development and organization, data capture, computer editing and imputation procedure, computer hardware and software used, census evaluation, publication and data dissemination programme«

Eine solche Dokumentation existiert für den Zensus 2022 nicht.<sup>5</sup> Die einzige zusammenfassende Dokumentation des Zensus 2022 ist der Qualitätsbericht. Weit geringere Anforderungen wurden im Gesetz über den registergestützten Zensus im Jahre 2011 in § 17 (Bewertung der Qualität der Zensusergebnisse) in Absatz 5 formuliert:

»Das Statistische Bundesamt erstellt im Benehmen mit den statistischen Ämtern der Länder bis zum 31. Dezember 2015 einen Qualitätsbericht über die Durchführung des Zensus und dessen Ergebnisse. In dem Bericht ist darzustellen, wie die Qualitätsvorgaben des § 7 Absatz 1<sup>6</sup> erfüllt worden sind. Insbesondere ist darzustellen

1. von welchen Annahmen bei der Entwicklung des Stichprobenverfahrens ausgegangen worden ist und inwieweit sie durch die Ergebnisse der Stichprobenerhebung bestätigt worden sind,

---

<sup>5</sup> Andere Länder demonstrieren, wie detailliert eine solche Dokumentation aussehen kann. Das ONS dokumentiert z.B. die Unterschiede zwischen dem Zensusergebnis und dem Ergebnis des Postenumeration-Surveys auf der Seite <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/measuresshowingthequalityofcensus2021estimates> (abgerufen am 9.11.2025)

<sup>6</sup> In diesem Absatz waren die angestrebten Präzisionsziele für Gemeinden oberhalb und unterhalb der 10 000-Einwohner-Grenze geregelt.

- 
2. nach welchen wissenschaftlichen Standards das Stichprobenverfahren ausgestaltet worden ist,
  3. ob und inwieweit die Ergebnisse der Stichprobe Anlass gaben, das Hochrechnungsverfahren zur Sicherung der Ergebnisqualität an neue Erkenntnisse anzupassen.

Für diesen Qualitätsbericht stellen die statistischen Ämter der Länder jeweils für ihren Zuständigkeitsbereich dem Statistischen Bundesamt Qualitätsberichte über die Durchführung des Zensus bis spätestens zum 1. März 2015 zur Verfügung. Diese Berichte enthalten insbesondere einen Bericht über die Schulung und Aufgabenerledigung der Erhebungsbeauftragten sowie die Ergebnisse der Überprüfungen nach den Absätzen 2 und 3.«

Selbst ein solcher, verglichen mit den Empfehlungen der Vereinten Nationen durchaus überschaubarer Qualitätsbericht liegt zum Zensus 2022 nicht vor, da der Gesetzgeber auf eine Übernahme des § 17 Abs. 5 (ZensG 2011) in das Zensusgesetz 2022 verzichtet hat.

## 2.3. Zum Qualitätsbericht des Zensus 2022

Am 20.11.2024 legte das Statistische Bundesamt den Qualitätsbericht zum Zensus 2022 vor. Die beiden Abschnitte über Stichprobenfehler und nicht stichprobenbedingte Fehler (Nonsampling-Errors, siehe Kapitel 8) sind ungekürzt als Kopie in der Abbildung 2.2 wiedergegeben. Beide Fehlerarten werden in diesem Gutachten ausführlich in den jeweiligen Kapiteln diskutiert.

An dieser Stelle soll nur auf folgende Punkte hingewiesen werden:

1. Der Abschnitt 4.3 verweist für die – in anderen Ländern zur Qualitätskontrolle als unabdingbar betrachteten – Post-Enumeration-Surveys (siehe Kapitel 9) auf Ziffer 1.8.2, was der Abschnitt 1.8.2 sein könnte. Dieser Abschnitt enthält kein Wort zur Wiederholungsbefragung. Die Wiederholungsbefragung wird im Qualitätsbericht nicht mehr erwähnt. Es gibt auch keine andere Textveröffentlichung des Statistischen Bundesamtes zu den Ergebnissen der Wiederholungsbefragung.
2. Der Verweis auf das Gutachten von Küchenhoff (2014) ist irreführend. Das Gutachten wurde für den Zensus 2011 geschrieben, nicht für den Zensus 2022. Der Satz »Das Gutachten bescheinigt dem Stichprobenverfahren (und dem Stichproben-Zufallsfehler) ein klares Design, hinreichende Validierung sowie Durchführung nach wissenschaftlichen Standards« bezieht sich eindeutig auf das Stichprobenverfahren,

---

### 1.8.2 Qualitätsbewertung

Die Qualität der hochgerechneten Ergebnisse aus der Haushaltsstichprobe ist über den Standardfehler quantifiziert. Die Qualität der hochgerechneten Zusatzmerkmale aus der Haushaltsstichprobe wird über ein Mindestfallzahlkriterium approximativ beschrieben. Fallzahlen kleiner oder gleich 211 müssen als zu unzuverlässig angesehen werden (approximativer Standardfehler  $\geq 15\%$ ) und werden daher nicht ausgewiesen.

(a) Abschnitt 1.8.2

### 4.3 Nicht-Stichprobenbedingte Fehler

Hierzu sei auf das Gutachten von Prof. Dr. Helmut Küchenhoff zum Zensus 2011 (zur Verwaltungsrechtssache Stadt Bremerhaven gegen Freie Hansestadt Bremen) vom 7. Oktober 2014 (siehe Methodenpapiere 8.2) sowie auf die Wiederholungsbefragung (siehe Ziffer 1.8.2) verwiesen. Das Gutachten bescheinigt dem Stichprobenverfahren (und dem Stichproben-Zufallsfehler) ein klares Design, hinreichende Validierung sowie Durchführung nach wissenschaftlichen Standards. Weitere potenzielle Fehlerquellen sind im Zensus nicht quantifiziert.

(b) Abschnitt 4.3

Abbildung 2.2.: Kopien der Abschnitte zu Nichtstichprobenfehlern im Qualitätsbericht zum Zensus 2022 in voller Länge (Statistisches Bundesamt, 2024)

also auf den Stichprobenfehler.<sup>7</sup> Auffällig ist aber, dass die kritischen Anmerkungen aus dem Gutachten (Küchenhoff, 2014) zu Nonsampling-Fehlern im Zensus 2011 nicht zitiert werden (siehe Kapitel 8). Löscht man aus diesem Absatz des Qualitätsberichts die irrelevanten Satzteile, dann verbleibt nur der letzte Satz:

3. »Weitere potenzielle Fehlerquellen sind im Zensus nicht quantifiziert« (Statistisches Bundesamt, 2024) Das Wort »potenziell« ist ebenfalls irreführend. Der Duden definiert »potenziell« als »möglich (im Gegensatz zu wirklich)«. Diese Fehlerquellen sind aber *immer* vorhanden und immer wirksam: »Die systematischen Fehler sind im allgemeinen mindestens von der gleichen Bedeutung wie die zufälligen Stichprobenfehler« (Statistisches Bundesamt, 1960, S. 51).

Zusammenfassend kann über den Qualitätsbericht festgestellt werden, dass er weder die Ergebnisse der eigentlich zur Qualitätskontrolle vorgeschriebenen Wiederholungszählung berichtet noch irgendeine Quantifizierung der Nicht-Stichprobenfehler vorgenommen wurde. Nach Ansicht der Gutachter verstoßen diese beiden Tatsachen sowohl gegen die EU-Vorschriften zum Zensus als auch gegen den »Verhaltenskodex für europäische Statistiken«, dem sich alle Statistischen Ämter der Europäischen Union verschrieben haben (European Commission. Statistical Office of the European Union., 2018). Dies betrifft unter anderem die Grundsätze:

**6.4** Informationen zu den verwendeten Datenquellen, Methoden und Verfahren sind öffentlich zugänglich.

---

<sup>7</sup> Genau genommen ist der Einschub »(und dem Stichproben-Zufallsfehler)« falsch: Ein Stichproben-Zufallsfehler hat kein Design, keine Validierung und wird auch nicht durchgeführt.

---

**12.2** Stichprobenfehler und Nicht-Stichprobenfehler werden gemessen und systematisch gemäß den europäischen Standards dokumentiert.

**15.6** Die Nutzerinnen und Nutzer werden fortlaufend über die Methodik der statistischen Prozesse, einschließlich der Verwendung und Integration von Verwaltungsdaten und sonstigen Daten, informiert.

Das Statistische Bundesamt hat für den Zensus 2022 keinen Verfahrensschritt öffentlich so dokumentiert, dass auch nur für eine einzige – gegebenenfalls fiktive – Gemeinde der Prozess im Detail von unabhängigen Wissenschaftlern nachvollzogen werden kann.

## 2.4. Mangelnde Hilfsmerkmale

Das Bundesstatistikgesetz legt in §10 fest, dass Hilfsmerkmale Angaben sind, die der technischen Durchführung von Bundesstatistiken dienen. Das Zensusgesetz 2022 zählt in §6 die Merkmale auf, die als Hilfsmerkmale betrachtet werden. Dazu gehören Anschrift, Name und Geburtsdatum. In der Fachliteratur werden solchen Merkmale als »Schlüsselmerkmale« oder »Quasi-Identifikatoren« (QIDs) bezeichnet (Christen et al., 2020; Hundepool et al., 2025; Lenz et al., 2006). Die QIDs sind außerhalb der Statistischen Ämter von Bund und Ländern nicht zugänglich. Allerdings werden in der amtlichen Statistik in Deutschland auch Daten, die der Datenerhebungsprozess selbst generiert (in der internationalen Fachliteratur: *paradata*, Kreuter (2013)) wie die Dauer eines einzelnen Interviews, die Zahl der Kontaktversuche bis zu einem Kontakt zu einer Zielperson oder eine Interviewnummer, als Hilfsmerkmale betrachtet. Entsprechend werden Hilfsmerkmale, QIDs und Parادات immer so früh wie möglich gelöscht.<sup>8</sup>

Werden aber Parادات und pseudonymisierte QIDs gelöscht, dann ist die Qualitätsbeurteilung einer Erhebung im Nachhinein nur zu kleinen Teilen möglich.<sup>9</sup> Moderne Methodenforschung ist ohne solche Daten kaum denkbar.<sup>10</sup>

---

<sup>8</sup> Es ist davon auszugehen, dass entsprechende Daten für den Zensus bereits gelöscht wurden. Das bedeutet, dass selbst das Statistische Bundesamt einen großen Teil der technischen Prozesse des Zensus 2022 nicht mehr reproduzieren könnte. Aus methodischer Sicht ist das unakzeptabel.

<sup>9</sup> Auch unabhängige Wissenschaftler oder Gutachter könnten solche Analysen innerhalb des Forschungszentrums des Statistischen Bundesamtes mit (je nach Fragestellung anonymisierten oder pseudonymisierten) Daten durchführen.

<sup>10</sup> Dies erklärt auch, warum es in Deutschland keine Methodenforschung – jenseits reiner Stichprobenziehung und Hochrechnung mit simulierten Daten – zu den Methoden des Zensus gibt: Die Datensätze der amtlichen Statistik sind für Methodologen weitgehend uninteressant. Versuche von Wissenschaftlern (Forschungsgruppe Zensus, u.a. die Professoren Rässler, Münnich, Gabler, Rendtel und Schnell), in Zusammenarbeit mit dem Statistischen Bundesamt systematisch Methodenforschung zum Zensus

---

Daher wurde bereits bei den Anhörungen zum Zensusgesetz im Bundestag von Münnich (2019) empfohlen, die Hilfsmerkmale nicht zu löschen («... Löschung dieser Daten (...) verhindert damit eine geeignete Post-Analyse des Zensus. (...) Eine Löschung nach maximal vier Jahren würde diese Forschung möglicherweise verhindern.» Noch vor der Datenerhebung des Zensus 2022 stellte Prof. Münnich dann fest (Bleninger et al., 2020): »Für den Zensus 2011 wurde aber bedauerlicherweise gesetzlich festgelegt, dass jedwede Hilfsmerkmale in angemessener Zeit gelöscht werden müssen. (...) Damit besteht keine geeignete Basis für Nachuntersuchungen zu Nichtstichprobenfehlern im Zensus 2011.«

Mit den Parادات des Zensus 2022 passiert nun exakt das Gleiche wie beim Zensus 2011: Das Statistische Bundesamt hat keine entsprechenden Analysen durchgeführt (oder zumindest nicht veröffentlicht), diese Daten nicht für unabhängige Analysen im Haus zur Verfügung gestellt, und möglicherweise bereits gelöscht.<sup>11</sup>

## 2.5. Methodische Veröffentlichungen zum Zensus 2022

Das Statistische Bundesamt hat für den Zensus 2011 und den Zensus 2022 jeweils nur die Arbeitsgruppe von Prof. Münnich mit Arbeiten zum Hochrechnungsverfahren betraut, eine unabhängige Evaluation dieser Studien hat nicht stattgefunden. Die wissenschaftliche Diskussion um die Modifikation des Zensus 2022 gegenüber dem Zensus 2011 beschränkt sich auf die Beiträge von Bleninger et al. (2020), die sich – mit Ausnahme des darin enthaltenen Kommentars von Küchenhoff und wenigen Sätzen von Krämer zu nicht quantifizierten Effizienzgewinnen – im Wesentlichen nur einem sehr speziellen Aspekt des Hochrechnungsverfahrens (der Spline-Glättung, siehe Kapitel 4) widmen. Die veröffentlichten Diskussionen um die Ergebnisse des Zensus 2011 befassten sich nahezu ausschließlich mit dem Bruch des Hochrechnungsverfahrens bei 10 000 Einwohnern (Christensen et al., 2015; Rendtel, 2015).

Die wenigen methodischen Arbeiten des Statistischen Bundesamtes zur Stichprobenkonstruktion, zur Hochrechnung oder der Mehrfachfallprüfung wurden nicht in einer

---

2022 zu betreiben, sind aufgrund der Bedenken der Juristen des Statistischen Bundesamts gescheitert, letztmalig 2012.

11 Im Zusammenhang mit der Überprüfung der Ergebnisse des Zensus wird von statistischen Ämtern häufig das sogenannte Rückspielverbot herangezogen (als Beispiel: <https://statistik.hessen.de/unsere-zahlen/zensus/faq-zum-zensus-2022/faq-informationen-fuer-hessische-kommunen>, zuletzt abgerufen am 21.12.2025.) Werden Analysen der Zensusdaten in den Forschungsdatenzentren der Länder mit pseudonymisierten Daten durchgeführt, wird nicht gegen das Rückspielverbot verstoßen. Damit wären auch Analysen von und mit Hilfsmerkmalen (anonymisiert/pseudonymisiert) möglich.

---

wissenschaftlichen Zeitschrift mit Peer-Review veröffentlicht.<sup>12</sup> In den vergangenen Jahren ist es Standard geworden, dass die verwendeten Programme und Daten mindestens den Reviewern zugänglich gemacht werden, damit unabhängige Prüfungen und Rechnungen zeigen können, dass die Ergebnisse reproduzierbar sind (National Academies of Sciences & Medicine, 2019b, 2020). Dies gilt auch für die Ergebnisse der amtlichen Statistik in den USA und dem Vereinigten Königreich (National Academies of Sciences & Medicine, 2019a).

Die Veröffentlichungen des Statistischen Bundesamtes sind nicht peer-reviewed, wenn diese in der Zeitschrift des Statistischen Bundesamtes (WISTA) erscheinen. Daher genügen diese Texte nicht den Kriterien, die bei internationalen Publikationen in Fachzeitschriften üblich sind.<sup>13</sup> Die daraus resultierenden Uneinheitlichkeiten der Notation und Begrifflichkeiten, Dokumentationslücken und – ohne Sourcecode – nicht nachvollziehbare Verfahren verunmöglichen Replikationen, selbst der Hochrechnungsverfahren.

Da Fragen an das Statistische Bundesamt nur über den zentralen Auskunftsdienst gestellt werden können, gibt es keinen fachwissenschaftlichen Dialog unter Kollegen, sondern eine über Juristen vermittelte indirekte Kommunikation, die zumeist über mehrfache Anfragen letztlich zu allgemein gehaltenen, immer noch mehrdeutigen Antworten ohne technische Details führt.

Weiterhin sind die Statistischen Ämter des Bundes und der Länder keine unabhängigen Forschungseinrichtungen. So ist beispielsweise das Hessische Statistische Landesamt eine obere Landesbehörde im Geschäftsbereich des Hessischen Ministerpräsidenten und das Statistische Bundesamt eine Bundesbehörde im Geschäftsbereich des Bundesinnenministeriums. Entsprechend wird ein Zensus explizit nicht als Forschungsprojekt, sondern als reiner Verwaltungsvorgang verstanden.<sup>14,15</sup>

---

12 Bei einer Peer-Review wird der Artikel von Experten des jeweiligen Fachgebietes sorgfältig daraufhin geprüft, ob die Aussagen des Textes durch die vorgestellten Fakten gerechtfertigt sind. Dazu gehört die Beurteilung, ob die verwendeten Methoden geeignet sind, das Ziel der Arbeit zu erreichen. Dafür ist es wesentlich, dass die Beschreibung der Methoden so detailliert ist, dass eine Replikation der Untersuchung möglich wäre. Schließlich werden die statistischen Analysen im Hinblick auf ihre Angemessenheit und Nachvollziehbarkeit überprüft.

13 Dazu gehören unvollständige Angaben, Bezug auf unzugängliche Materialien usw.

14 Unter dem GZ 120111/812662; GZ120111/809986 teilt uns das Statistische Bundesamt mit: »Insbesondere möchten wir darauf hinweisen, dass der Zensus kein Forschungsprojekt im Sinne des Kodex zur Sicherung guter wissenschaftlicher Praxis ist, sondern ein Verwaltungsvorgang, auf den der Kodex nicht anwendbar ist.«

15 Als Beispiel: Die Ergebnisse des Zensus in Hinsicht auf die Einwohnerzahl werden den Gemeinden jeweils in einem »Datenblatt zur Ermittlung der amtlichen Einwohnerzahl« mitgeteilt. Für alle Gemeinden liegt dieses Datenblatt als PDF vor und enthält u.a. Einzelheiten zur Stichprobe und Berechnung der Einwohnerzahl. Die Gemeinden behandeln die Datenblätter unterschiedlich: Manche Gemeinden legen diese frei zugänglich in das Internet, manche übermitteln diese auf Anfrage, manche

---

Eine detaillierte externe Qualitätssicherung der operativen Verfahren und einzelner Fachveröffentlichungen der amtlichen Statistik gibt es daher in Deutschland nicht.

## 2.6. Zusammenfassung

Das Statistische Bundesamt hat beim Zensus 2022 ebenso wie beim Zensus 2011 auf eine detaillierte Prozessdokumentation verzichtet. Die Zählergebnisse des Zensus 2011 führten zu einer Klage vor dem Bundesverfassungsgericht – hierbei fokussierten sich die Kläger auf das offensichtlich bestehende Problem der unterschiedlichen Behandlung von Gemeinden unterhalb bzw. oberhalb der Grenze von 10 000 Einwohnerinnen und Einwohnern. Weitere Aspekte des Zensus 2011 wurden im Rahmen der Klage nicht thematisiert. Das Statistische Bundesamt konzentrierte sich daher im Wesentlichen darauf, diese Problematik mittels eines gleitenden Übergangs durch eine mathematische Bézierfunktion (siehe hierzu unsere Ausführungen in Kapitel 4.2) zu adressieren. Weitere potenzielle methodische Schwächen wurden nicht dokumentiert oder bearbeitet.

Auch für den Zensus 2022 liegt damit keine Dokumentation vor, die wissenschaftlichen Ansprüchen genügt. Die mehr als 270 Widersprüche und Klagen gegen die Zensusergebnisse legen nahe, dass das Statistische Bundesamt durch unabhängige Dokumentation oder Begleitforschung erhebliche zeitliche, finanzielle und personelle Ressourcen hätte einsparen können.<sup>16</sup>

---

verweisen auf die Landesämter (die dann verweigern), manche kündigen die Berechnung von Gebühren bis zu 100 Euro für die Einsicht in das PDF an. Die Gesamtheit dieser Datenblätter ist nicht öffentlich zugänglich. Sowohl das Statistische Bundesamt als auch alle Landesämter, die wir angefragt haben, haben die Überlassung der Datenblätter verweigert, auch bei Anfragen nach dem jeweiligen Informationsfreiheitsgesetz. Die Autoren halten ein solches Verhalten nicht für vereinbar mit dem »European Code of Practice« und auch nicht mit dem Anspruch von Destatis, wissenschaftliche Forschung zu betreiben oder wenigstens zu unterstützen.

<sup>16</sup> Eine solche Begleitforschung wäre im Rahmen eines DFG-Projekts im Statistischen Bundesamt ohne Weiteres – und für Destatis sogar kostenneutral – möglich gewesen. Drei Postdocs (DFG-Richtsatz 2025: ca. 90 000 Euro pro Jahr) hätten über drei Jahre eine unabhängige Gesamtdokumentation und methodische Begleitforschung erstellen können. Für Projektkosten deutlich unter einer Million Euro (etwa 0.067 % der Zensuskosten) hätte man Zweifeln an der Methodik oder den Ergebnissen frühzeitig begegnen können.

## 3. Zielpopulation und untererfasste Bevölkerungsgruppen

Die »Verordnung (EU) Nr. 1260/2013 des Europäischen Parlaments und des Rates vom 20. November 2013« ([2013](#)) besagt in Artikel 4:

(1) Für die Zwecke einer Beschlussfassung im Rat mit qualifizierter Mehrheit übermitteln die Mitgliedstaaten der Kommission (Eurostat) Daten über die Gesamtbevölkerung auf nationaler Ebene zur Bezugszeit gemäß Artikel 2 Buchstabe c innerhalb von acht Monaten nach Ablauf des Bezugsjahres.

(2) Die Mitgliedstaaten können die Gesamtbevölkerung nach Absatz 1 anhand der sich dort rechtmäßig aufhaltenden oder registrierten Bevölkerung schätzen, wobei wissenschaftlich fundierte, hinreichend dokumentierte und öffentlich verfügbare statistische Schätzmethoden verwendet werden.

Diese Verordnung besitzt Implikationen auf mehreren Ebenen. In diesem Kapitel befassen wir uns nur mit der Definition der zu schätzenden Population.

### 3.1. Populationsdefinition

In Europa wurden für die letzte Zensusrunde je nach Land sechs unterschiedliche Definitionen der Population verwendet (Wirtz, [2022](#)):

1. gewöhnlicher Wohnsitz,
2. rechtmäßig wohnhaft,
3. registrierte Gebietsansässige,
4. gewöhnlich wohnhaft und registriert,
5. gewöhnlich wohnhaft und rechtmäßig und
6. gewöhnlich wohnhaft und dauerhaft.

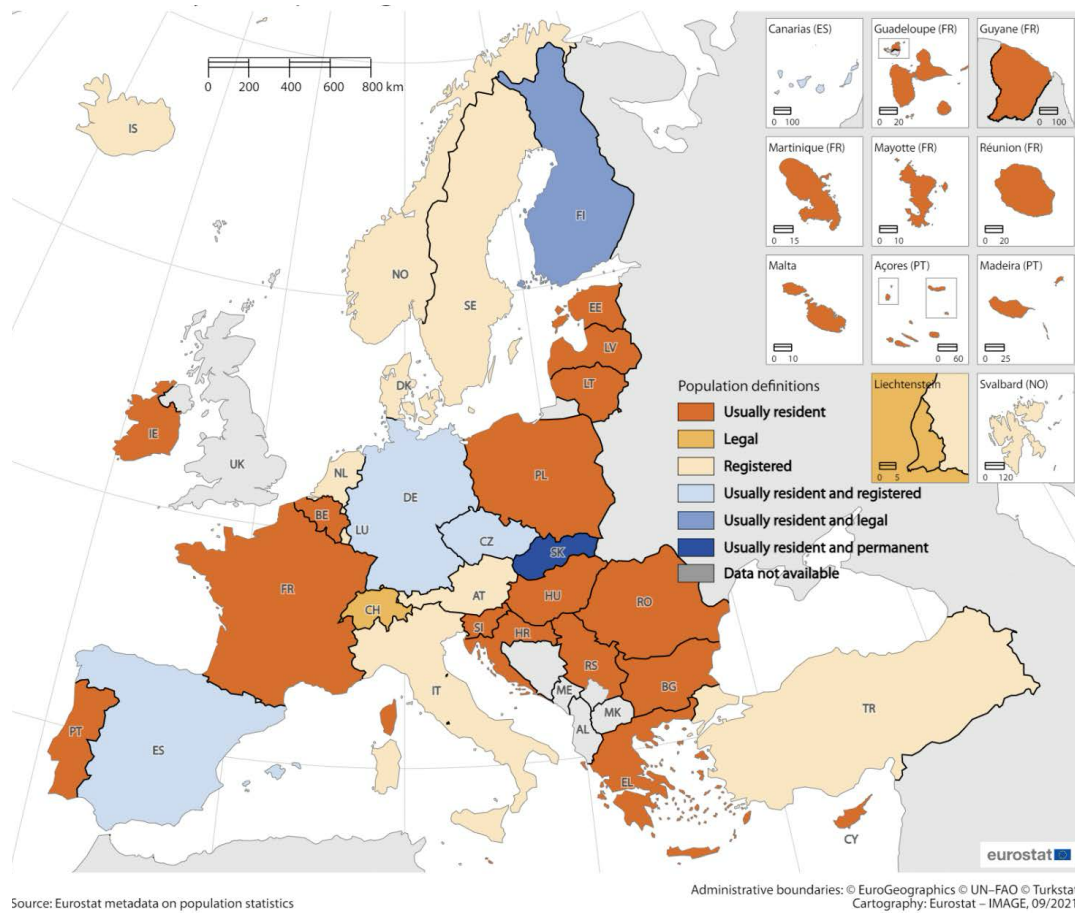


Abbildung 3.1.: Definitionen der Population der Zensen in Europa (Wirtz, 2022)

---

Deutschland verwendet die Definition vier («gewöhnlich wohnhaft und registriert»), die sonst nur Spanien, Luxemburg sowie die Tschechische Republik verwenden (siehe Abbildung 3.1). Die große Mehrheit der Länder in Europa verwendet hingegen die von Eurostat empfohlene Definition »gewöhnlicher Wohnsitz«.<sup>1</sup> Das Zensusgesetz 2022 legt hingegen in §2 (2) fest:

»Der übliche Aufenthaltsort einer Person ist der Ort, an dem sie nach den melderechtlichen Vorschriften mit nur einer alleinigen Wohnung oder mit ihrer Hauptwohnung gemeldet sein sollte.«

Eurostat erlaubt zwar diese Definition, empfiehlt diese aber nicht. Mit der empfohlenen Definition würde ein Registerabzug für die Feststellung der Einwohnerzahl allein genügen.

Da es keine objektiven Maßstäbe für die Zuverlässigkeit des Zensus 2022 gibt, ist prinzipiell unklar, ob die Ergebnisqualität eines Registerabzugs auch ohne Bereinigung schlechter wäre als die des Zensus 2022. Vermutlich ließe sich die Qualität eines solchen Abzugs durch Deduplizierung und einen hochfrequenten Lebenszeichenansatz verbessern, aber dies hängt von Details der Implementierung beider Verfahrensschritte ab.<sup>2</sup> Eine Beurteilung der Ergebnisqualität einer der Varianten eines Registerzensus ist mangels Daten derzeit nicht möglich.<sup>3</sup>

Die Definitionen des Zensusgesetzes sowie die zeitlichen Versäumnisse des Gesetzgebers in der Vorbereitung der Zensen 2011 und 2022 dürften die gesamten Kosten dieser Zensen verdoppelt haben: Die Sekundärziele des Zensus hätten sich für praktische Zwecke ausreichend mit dem Mikrozensus erreichen lassen, das Primärziel mit den Registerabzügen.<sup>4</sup>

---

1 Eurostat definiert als Bevölkerung mit gewöhnlichem Wohnsitz »... die Zahl der Einwohner eines bestimmten Gebiets am 1. Januar des betreffenden Jahres ... Gewöhnlicher Wohnsitz ist der Ort, an dem eine Person normalerweise ihre tägliche Ruhezeit verbringt, ungeachtet vorübergehender Abwesenheit zum Zwecke der Erholung, des Urlaubs, des Besuchs von Freunden und Verwandten, der Geschäftstätigkeit, der medizinischen Behandlung oder der religiösen Pilgerfahrt« (<https://ec.europa.eu/eurostat/de/web/population-demography/demography-population-stock-balance/methodology>, abgerufen am 1.12.2025).

2 Bei einem »Sign-of-Life«-Ansatz werden Fehlbestände und Karteileichen in den Registern durch den Abgleich mit einer hohen Anzahl anderer Register verglichen: Werden dort über einen längeren Zeitraum weniger Signale verzeichnet als unter einem zuvor experimentell festgelegten Schwellenwert, werden Personen im Melderegister abgemeldet. Dieser Ansatz setzt neben entsprechenden rechtlichen Regeln vor allem einen einheitlichen Identifikator und eine hohe Zahl von beteiligten Registern voraus.

3 Dies ist vor allem auf das Fehlen eines unabhängigen »Post-Enumeration Surveys« zurückzuführen, siehe hierzu Kapitel 9.

4 Zwar könnte man argumentieren, dass Ziel-2 Daten auch auf der Ebene von Gemeinden und darunter

---

## 3.2. Obdachlose und irreguläre Migration

Die Verordnung (EG) Nr. 763/2008 des Europäischen Parlaments und des Rates legt in §1 als Ziel von Volks- und Wohnungszählungen die »Bereitstellung umfassender Daten über die Bevölkerung und die Wohnungssituation im Abstand von zehn Jahren« fest. In §2 (a) wird Bevölkerung »als die Bevölkerung an ihrem üblichen Aufenthaltsort zum Stichtag« definiert. Als »üblicher Aufenthaltsort« wird

»der Ort, an dem eine Person normalerweise ihre täglichen Ruhephasen verbringt, ungeachtet vorübergehender Abwesenheit zu Zwecken der Erholung, des Urlaubs, des Besuchs von Freunden und Verwandten, zu geschäftlichen Zwecken, zu medizinischer Behandlung oder religiöser Pilgerfahrt«

festgelegt. Dabei sind nur die Personen als übliche Einwohner des betreffenden geographischen Gebiets zu betrachten,

(1) »die vor dem Stichtag mindestens 12 Monate ununterbrochen an ihrem üblichen Aufenthaltsort gelebt haben«

oder

(2) »die während der letzten 12 Monate vor dem Stichtag an ihrem üblichen Aufenthaltsort mit der Absicht eintrafen, sich dort mindestens ein Jahr aufzuhalten«.

Können die

»beschriebenen Umstände nicht festgestellt werden, so bedeutet ›üblicher Aufenthaltsort‹ den Ort des rechtmäßigen oder eingetragenen Wohnsitzes.«

---

verfügbar sein sollen. Dann würde man erwarten, dass diese Art Daten der Zensen 2011 und 2022 auch für wissenschaftliche Veröffentlichungen oder Planungszwecke verwendet werden. Die Daten des Zensus 2022 sind zu neu, um bereits in Fachzeitschriften verwendet worden zu sein. Allerdings finden sich nur vereinzelte inhaltliche wissenschaftliche Arbeiten, die auf Gitterzellen-Daten oder Aggregationen unterhalb der Gemeindeebene des Zensus 2011 basieren. Ein detaillierter Nachweis, wie Ziel-2 Daten aus 2011 unterhalb der Gemeindeebene für politische Entscheidungen oder Planungszwecke genutzt wurden, findet sich nicht in der Literatur. Die EU-Verordnung Verordnung (EG) Nr. 1201/2009 der Kommission vom 30. November 2009 zur Durchführung der Verordnung (EG) Nr. 763/2008 des Europäischen Parlaments und des Rates über Volks- und Wohnungszählungen in Bezug auf die technischen Spezifikationen für die Themen sowie für deren Untergliederungen schreibt die Gemeindeebene (LAU-2) nur für die Bevölkerungszahl vor.

---

Danach sollten also weder verdeckt Obdachlose noch irreguläre aufhältige Personen ausgeschlossen werden. Die Durchführungsverordnung der EU 2017/543 der Kommission vom 22. März 2017 legt zur Begriffsbestimmung des »üblichen Aufenthaltsortes« fest, dass der Ort der Zählung als üblicher Aufenthaltsort von Obdachlosen, Nichtsesshaften, Landstreichern und Personen, die keinen üblichen Aufenthaltsort besitzen, gilt. Also gehören diese Personen zur Bevölkerung am üblichen Aufenthaltsort.

Ebenso legt diese Verordnung fest, dass irregulär aufhältige Personen oder Personen ohne gültige Ausweispapiere (ebenso wie Asylbewerber und Personen mit Flüchtlingsstatus) mitgezählt werden:<sup>5</sup>

»Dabei ist nicht beabsichtigt, diese Personen besonders zu unterscheiden, sondern dafür zu sorgen, dass sie bei der Zählung nicht ausgelassen werden.«

Genau das geschieht aber bei verdeckt wohnungslosen Personen und bei irregulär aufhältigen Personen durch die Art der Durchführung des Zensus in Deutschland. Im Gegensatz zu vielen anderen Ländern gilt dies auch für die Wiederholungszählungen.

Versucht man die Größe dieser Populationen zu ermitteln, stößt man in Deutschland seit Jahrzehnten auf die gleichen Probleme (Schnell, 1991), wobei sich die Datenlage zumindest für die Wohnungslosen durch das Wohnungslosenberichterstattungsgesetz (WoBerichtsG) verbessert hat. Das WoBerichtsG schreibt in §3(2) die Erhebung der Zahl der Personen vor, denen »(...) wegen Wohnungslosigkeit Räume zu Wohnzwecken überlassen oder Übernachtungsgelegenheiten zur Verfügung gestellt worden sind«. Da dadurch nur ein Teil der Wohnungslosigkeit erfasst wird, sieht das WoBerichtsG in §8 (3) ergänzende Erhebungen zu Personen vor, die »(...) ohne jede Unterkunft obdachlos sind«. Der Bericht von 2024 (Gesellschaft für innovative Sozialforschung und Sozialplanung e.V. (GISS) & Kantar Public, 2024) gibt für die Referenzwoche (1. – 7. 2. 2024) insgesamt 107 705 verdeckt wohnungslose Personen an.

Die Datenlage für irregulär aufhältige Personen ist trotz der Größe und der politischen Relevanz des Themas deutlich unvollständiger. Die neuesten verfügbaren Schätzungen (Kierans & Kraler, 2025) für diese Population sind die mehr als 10 Jahre alten Schätzungen von Vogel (2015): 180 000 – 520 000 Personen.

Wählt man die Mitte dieses Intervalls und addiert die verdeckt Obdachlosen, dann entspricht diese Untergrenze der nicht erfassten Personen der Einwohnerzahl einer der größten 20 Städte Deutschlands.

---

<sup>5</sup> Sofern sie die Kriterien für den üblichen Aufenthaltsort in dem Land erfüllen. Dies sind aber nur die geforderten Aufenthaltszeiten von einem Jahr, wobei schon die Absicht des Aufenthaltes ausreicht. Sollte also ein Kontakt prinzipiell hergestellt werden können, wäre die Absicht oder die Dauer ermittelbar. Durch den deutschen rechtlichen Rahmen des Zensus diese Ermittlung zu verhindern, läuft der Intention der EU zuwider.

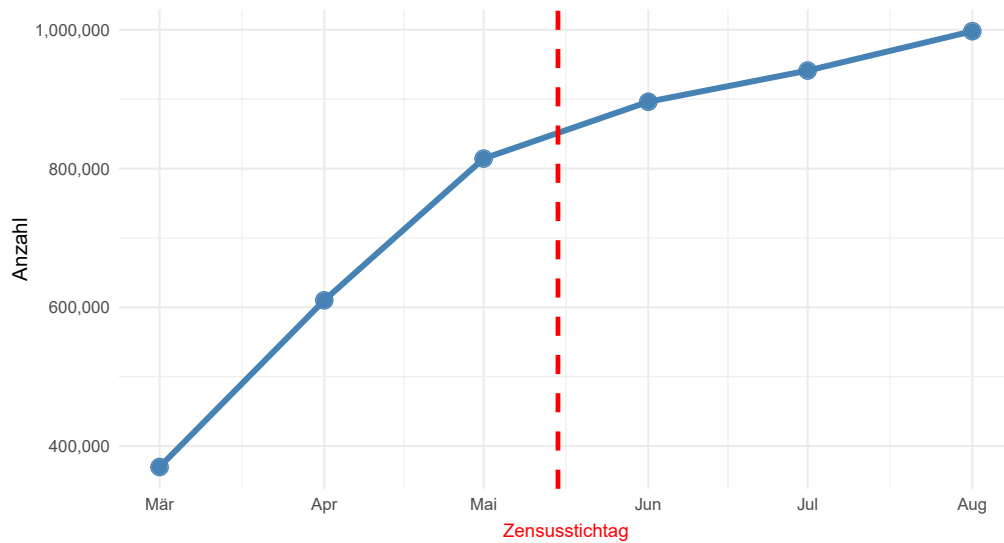


Abbildung 3.2.: Die Entwicklung der Zahl ukrainischer Flüchtlinge nach den Daten des Ausländerzentralregisters in den ersten Monaten 2022.

### 3.3. Probleme bei der Erfassung ukrainischer Flüchtlinge

Am 24.2.2022 begann Russland den Krieg gegen die Ukraine. In der Folge stieg die Anzahl der Flüchtlinge rasch an (Abbildung 3.2). Zum Zensusstichtag waren bereits mehr als 800 000 ukrainische Flüchtlinge in Deutschland beim AZR gemeldet (genauer: 814 385).<sup>6</sup> Destatis weist in der Zensusdatenbank zum Stichtag 454 460 Personen mit ukrainischer Staatsangehörigkeit aus. Selbst falls alle diese Personen Flüchtlinge gewesen wären, fehlen gegenüber den AZR-Daten 359 925 Personen, also 44.2 % (entsprechend ca. 0.4 % der durch den Zensus geschätzten Gesamtbevölkerung von 82 719 540 Personen).

Laut Durchführungsverordnung der EU 2017/543 der Kommission vom 22. März 2017 sollen Flüchtlinge durch die EU-Zensen erfasst werden. Das Statistische Bundesamt hat 2025 ein Methodenpapier zur Erfassung der Flüchtlinge veröffentlicht (Vorndran, 2025).<sup>7</sup>

<sup>6</sup> Datenquelle der AZR-Daten: Mediendienst Integration (2025). Datenquelle der Destatiszahlen: Zensusdatenbank, <https://ergebnisse.zensus2022.de/datenbank/online/statistic/2000S/table/2000S-2004/search/s/VWtyYWluZQ==>, Datum 14.12.2025.

<sup>7</sup> Das Papier gibt auf dem Titelblatt 2024, die Meta-Daten des PDFs aber 2025 an.

---

Vorndran (2025) erläutert:

»Grundsätzlich werden Personen, die sonst im Ausland wohnen, meldepflichtig, wenn sie sich länger als drei Monate in Deutschland aufhalten, dies gilt auch für ukrainische Schutzsuchende.«

Und dann weiter:

»Gemäß der Art ihrer Unterbringung wurden Schutzsuchende aus der Ukraine wie folgt bei der Einwohnerzahlermittlung berücksichtigt:

- Die Schutzsuchenden waren in bestehenden Flüchtlingsunterkünften (z. B. Erstaufnahmeeinrichtungen, Ankerzentren) untergebracht. In diesen Einrichtungen werden die Menschen registriert und dadurch melderechtlich erfasst. Beim Zensus 2022 fand eine Erhebung an diesen Unterkünften statt . . . .
- Die Schutzsuchenden waren vorübergehend, also nur kurzfristig, in provisorisch eingerichteten Sammelunterkünften (z. B. Turnhallen) untergebracht, bevor sie in eine private Wohnung/Gebäude oder eine andere Unterkunft umzogen, um dort für längere Zeit zu wohnen. In diesen Notunterkünften mussten die Schutzsuchenden sich zunächst nicht anmelden. Beim Zensus 2022 fand an diesen Anschriften deshalb auch keine Erhebung statt.«

Danach folgen Ausführungen über private Unterbringungen, bei denen die Zensusrelevanz davon abhing, ob bei der Befragung angegeben wurde, ob die Unterbringung vorübergehend sei oder nicht. Dies dürfte im Einzelfall eine schwierige Frage gewesen sein. Zusammenfassend stellt Vorndran (2025) fest:

»Die Überprüfung der methodischen Ausgestaltung des Zensus hatte ergeben, dass für die Erfassung der Schutzsuchenden aus der Ukraine keine methodischen Anpassungen an den vorhandenen Konzepten vorzunehmen sind.«

Für diesen Satz wird keine Erklärung gegeben: Weder wird die Art der Überprüfung angegeben, noch quantifiziert, noch eine Quelle angegeben. Besonders wichtig für die Beurteilung der Vollständigkeit des Zensus ist der zuvor zitierte Satz, dass an provisorisch eingerichteten Sammelunterkünften keine Erhebungen stattfanden. Ob dies dem Melderecht entspricht oder nicht, ist nicht Gegenstand dieses Gutachtens. Diese Entscheidung widerspricht aber vermutlich der Intention der Durchführungsverordnung der EU 2017/543.

## 4. Das Konzept der Präzisionszielfunktion

Nach dem Urteil des Bundesverfassungsgerichts vom 19. September 2018 wurde vom Statistischen Bundesamt (ursprünglich für den Zensus 2021) eine für alle Gemeinden einheitliche Erhebungs- und Schätzmethodik angestrebt (Burgard et al., 2019). Um auf Vollerhebungen weitgehend zu verzichten, wurde dann eine Funktion gesucht, die einen gleitenden Übergang von kleinen zu großen Gemeinden erlaubte. Hierzu wurde das aus der Mathematik bekannte Konzept der Bézierfunktion verwendet.

Die resultierende Funktion wurde von Burgard et al. (2020) und in der Folge in den Veröffentlichungen des Statistischen Bundesamtes als »Präzisionszielfunktion« bezeichnet. Diese Funktion beschreibt den maximal tolerierbaren Standardfehler in Abhängigkeit von der Gemeindegröße.

### 4.1. Konstruktion der Präzisionszielfunktion

Während sich die Vorgabe einer Obergrenze von 0.5% für den relativen Standardfehler bei Gemeinden mit mehr als 10 000 Einwohnern mit der Gesetzgebung vor dem Zensus 2011 deckt, wurde diese Zielsetzung für kleinere Gemeinden gelockert und es wurden im Vorfeld entsprechende Änderungen am Zensusgesetz vorgenommen. Dieses schreibt nun für Gemeinden mit weniger als 1 000 Einwohnern die Einhaltung eines konstanten absoluten Standardfehlers von 15 Personen vor. Die Festlegung eines absoluten Standardfehlers von 15 als Obergrenze für kleine Gemeinden kann bei der Hochrechnung einer Stichprobe zu großen relativen Abweichungen der geschätzten von den tatsächlichen Einwohnerzahlen führen. Das in Bretschi und Lorentz (2019) vorgetragene Argument, dass sich solche Abweichungen in beide Richtungen bewegen können, nutzt solchen Gemeinden wenig, bei denen das Pendel in die falsche Richtung ausgeschlagen hat. Stellt man die Zielfunktion relativ und nicht absolut dar, so dürfen die tolerierten relativen Standardfehler bei einem konstant gehaltenen absoluten Standardfehler von 15 für sehr kleine Gemeinden tatsächlich sehr groß werden.

---

Die Abbildung 4.1 zeigt den Verlauf des angestrebten<sup>1</sup> relativen Standardfehlers für Gemeinden mit weniger als 1 000 Einwohnern. Unterhalb von 150 Einwohnern liegt der angestrebte relative Standardfehler bereits bei über 10 %. Es ist davon auszugehen, dass selbst diese lockere Präzisionszielvorgabe nur in seltenen Fällen eingehalten wurde. Um sie dennoch zumindest formal einhalten zu können, wurde ein Weg gefunden, die hohen Standardfehler besonders bei kleinen Gemeinden nicht ausweisen zu müssen, indem man kleine Gemeinden mit benachbarten Gemeinden zu Verbänden zusammengefasst und die Standardfehler im höheren Aggregat ausgewiesen hat. Auf diese Thematik werden wir später in Kapitel 5 näher eingehen.

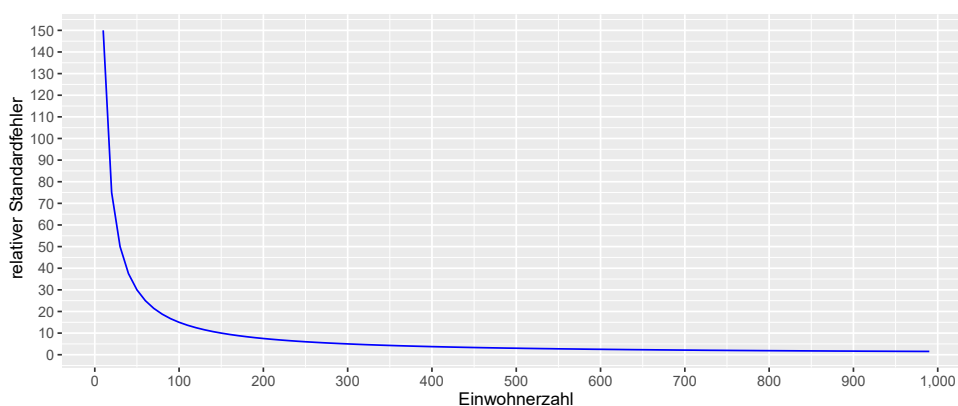


Abbildung 4.1.: Angestrebter relativer Standardfehler im Intervall 10 – 1 000 Einwohner

Für Gemeinden ab 10 000 Einwohnern blieb es bei der 0.5 % -Vorgabe, was für eine fiktive Gemeinde mit genau  $x = 10\,000$  Einwohnern einen absoluten Standardfehler von  $y(x) = 50$  bedeutete.

Zur Form der Kurve im Bereich der Einwohnerzahlen zwischen 1 000 und 10 000 gab es seitens des Gesetzgebers keine Vorgaben, es war also im Zensusprojekt eine geeignete Funktionsvorschrift vorzuschlagen, die im nachfolgenden Abschnitt besprochen wird.

---

<sup>1</sup> Das Adjektiv »angestrebt« wird auch in den Datenblättern zum Zensusergebnis an die Gemeinden verwendet und rührt daher, dass das Regressionsmodell zur Bestimmung der Stichprobenumfänge mit Annahmen an die (a priori unbekannt) Korrelation zwischen Registerinformation und Zensusstichprobe arbeiten muss. Aus diesem Grund nimmt der Begriff des angestrebten Standardfehlers im Zensusgesetz einer Verfehlung der Ziele jede Konsequenz.

---

## 4.2. Quadratische Bézierkurve

In Klink und Lorentz (2022, S. 15) findet sich folgende Funktionsgleichung, in der drei Fälle unterschieden werden:

$$y(x) = \begin{cases} 15 & \text{für } 0 \leq x < 1\,000, \\ \frac{7}{1000}x - \frac{28}{10\sqrt{50}}\sqrt{x-200} + \frac{96}{5} & \text{für } 1\,000 \leq x < 10\,000 \text{ und} \\ \frac{5}{1000}x & \text{für } x \geq 10\,000. \end{cases} \quad (4.1)$$

Der erste und dritte Fall der Funktion stammen aus den im Zensusgesetz formulierten Zielvorgaben für Gemeinden mit weniger als 1 000 und mehr als 10 000 Einwohnern. Der zweite Fall benötigt zum Verständnis etwas Mathematik.

### 4.2.1. Exkurs: Zur Herleitung der Zielfunktion

Die vertraute Darstellung mittels der kanonischen Standardbasis wurde durch eine korrekte Transformation aus der Parameterdarstellung einer Bézierkurve erhalten. Letztere erfolgt üblicherweise via sogenannter Bernsteinpolynome und sieht im vorliegenden quadratischen Modell folgendermaßen aus:

$$y(t) = (1-t)^2P_0 + 2t(1-t)P_1 + t^2P_2,$$

wobei  $P_0 = (1\,000, 15)$ ,  $P_1 = (3\,000, 15)$  und  $P_2 = (10\,000, 50)$  die drei Kontrollpunkte bezeichnen und der frei wählbare Parameter  $t$  das reelle Intervall  $[0, 1]$ , sprich alle reellen Zahlen von 0 bis 1 durchläuft. Zur Verdeutlichung des zweiten Falls in 4.1 setze man in die Gleichung  $t = 0$  ein. Man landet dann beim Punkt  $P_0$ , also dem linken Rand des Intervalls. Setzt man  $t = 1$  ein, so landet man entsprechend am rechten Rand bei Punkt  $P_2$ . Plottet man den zweiten Fall der Funktionsgleichung, so erhält man Abbildung 4.2:

Die in Abbildung 4.2 dargestellte quadratische Bézierkurve ist allerdings kein an die Beobachtungen angepasstes Modell. Der dritte für eine Bézierkurve benötigte Kontrollpunkt  $P_1$  wird nämlich als Schnittpunkt der Tangenten an den anderen beiden Kontrollpunkten (Nahtstellen) ermittelt, wobei die Steigungen der Tangenten – am linken Rand Null, am rechten Rand 0.005 – durch das Zensusgesetz vorgegeben sind.

Die Präzisionszielfunktion ist stetig differenzierbar und streng monoton. Die in Burgard et al. (2020) formulierte Vorgabe der Linkskrümmung kann aber aus vorgenannten Gründen nicht global eingehalten werden. Die Glattheit einer Funktion ist ohnehin ein relativer Begriff und trifft im vorliegenden Falle nicht für die zweiten und höhere Ableitungen an den beiden Nahtstellen  $x = 1\,000$  und  $x = 10\,000$  zu. Würde die Linkskrümmung bei Gemeinden mit mehr als 10 000 Einwohnern beibehalten, hieße das eine

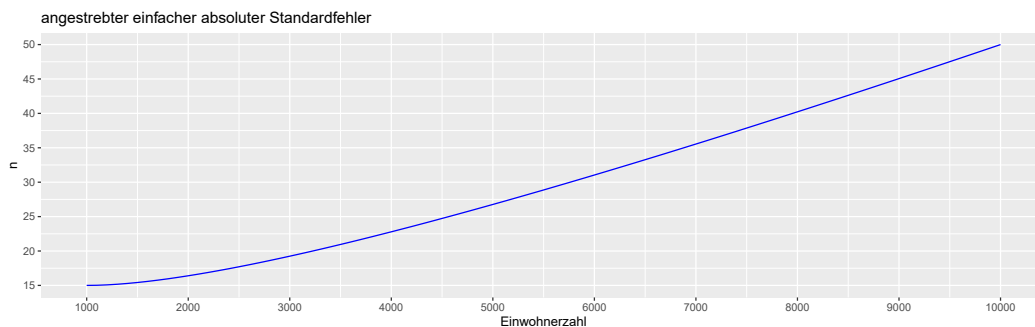


Abbildung 4.2.: Die in der Präzisionszielfunktion verwendete Bézierkurve im Intervall 1 000 – 10 000 Einwohner

kontinuierliche Verschärfung der Präzisionsvorgaben. Ähnliches lässt sich zu Gemeinden im Bereich der 1 000 Einwohner beobachten, wobei Gemeinden oberhalb der Schwelle eine bessere Präzision erwarten dürfen.

Der Einsatz von Bézierkurven ist besonders im Kontext der Computergrafik beliebt, da die Bernsteinpolynombasis eine individuelle Gestaltung der Kurve durch ein manuelles Verschieben der Kontrollpunkte erlaubt. Ein solcher Vorzug besteht aber in der vorliegenden Anwendung nicht, ist doch wie zuvor beschrieben der Verlauf der Kurve bei Auswahl eines quadratischen Modells mit fester Vorgabe der beiden Randpunkte und dortigen Tangenten redundant, wie auch in Burgard et al. (2020) festgestellt wird: »Dies ermöglicht eine einfache Umsetzung als Präzisionszielfunktion unabhängig von a priori zu wählenden Parametern.«

#### 4.2.2. Verlauf der Kurve

Es lässt sich eine approximative Linearität der quadratischen Bézierfunktion feststellen, siehe Abbildung 4.3. Berechnet man die tatsächliche Differenz zwischen Bézierkurve und linearer Funktion, so erhält man die Kenngrößen aus Tabelle 4.1.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Differenz	-3.888	-3.636	-2.902	-2.564	-1.681	0.000

Tabelle 4.1.: Differenz zwischen Bézierkurve und linearer Funktion

Die bei Gemeinden mit ca. 4 000 Einwohnern beobachtete maximale absolute Abweichung liegt folglich bei weniger als vier Personen. Ob nun eine lineare oder quadratische Funktion verwendet wird, hat daher auf praktische Ergebnisse wenig Auswirkung.

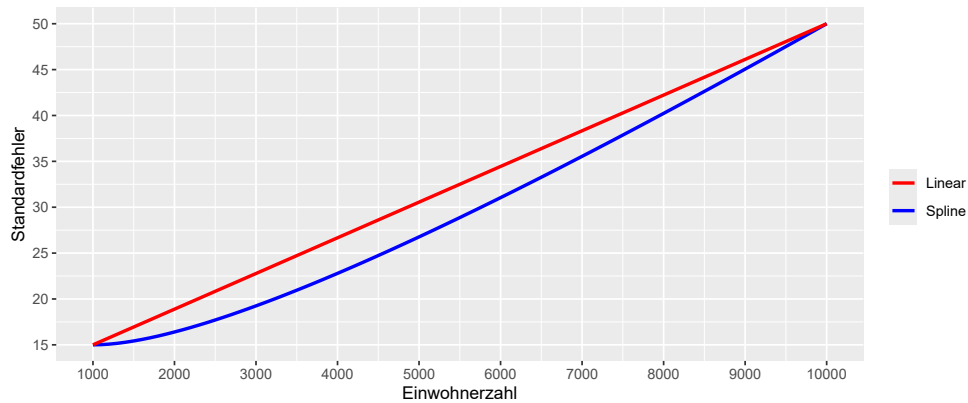


Abbildung 4.3.: Die in der Präzisionszielfunktion verwendete Bézierkurve im Vergleich zu einer linearen Funktion auf dem Intervall 1 000 - 10 000 Einwohner

Nachfolgende Abbildung 4.4 zeigt den angestrebten relativen Standardfehler für Gemeinden zwischen 900 und 11 000 Einwohnern. Ab 1 500 Einwohnern sinkt der angestrebte relative Standardfehler rasch unter die Marke von 1 % und erreicht das Minimum von 0.5 % schließlich bei der 10 000-Einwohner-Markierung.

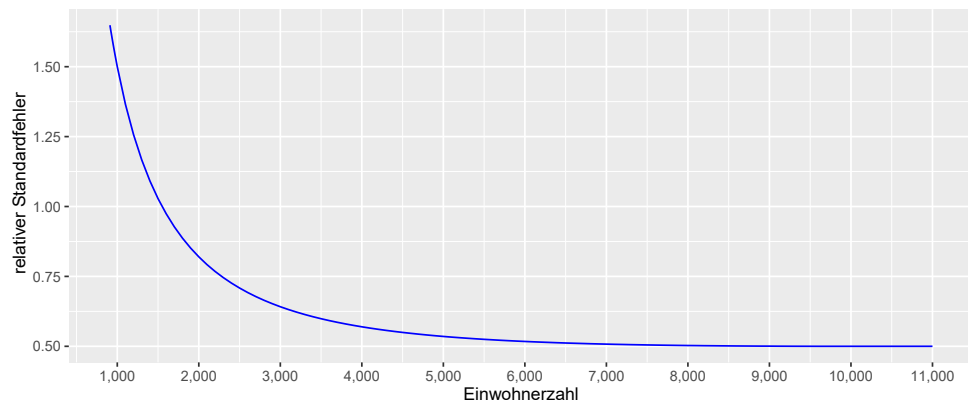


Abbildung 4.4.: Angestrebter relativer Standardfehler im Intervall 900 – 11 000 Einwohner

---

Die relative Darstellung in Abbildung 4.4 ist der Gleichung 4.1 vorzuziehen. Nicht zuletzt, da bei der Veröffentlichung der Ergebnisse des Zensus 2022 an die Gemeinden stets die relativen Standardfehler weitergegeben wurden. Und zwar sowohl die durch die Zielvorgaben angestrebten, als auch die kolportierten »tatsächlich realisierten Standardfehler«, mit denen wir uns in Kapitel 6.4 näher beschäftigen werden.

Kritisch anzumerken wäre, dass die Definition der Präzisionszielfunktion neben den Nahtstellen  $x = 1\,000$  und  $x = 10\,000$  andere erhebungsmethodische Schwellen nicht berücksichtigt. Zum Beispiel wurden Gemeinden mit mehr als 2 000 Einwohnern hinsichtlich der Stichprobenumfänge und Hochrechnung oftmals anders eingestuft als Gemeinden zwischen 1 000 und 2 000 Einwohnern.<sup>2</sup> Auch haben Gemeinden mit weniger als 200 Anschriften eine besondere Behandlung erfahren (siehe Kapitel 5). Die von Bretschi und Lorentz (2019) getroffene Aussage »Die Lockerung des Präzisionsziels soll beim Zensus 2021 nicht sprunghaft, sondern gleitend erfolgen« entspricht nicht der erhebungstechnischen Realität.

### 4.3. Zusammenfassung

Durch die Festlegung eines absoluten Standardfehlers von 15 bei Gemeinden mit weniger als 1 000 Einwohnern und von 50 bei Gemeinden mit 10 000 Einwohnern war der Kurvenverlauf einer quadratischen Bézierfunktion als Präzisionszielfunktion eindeutig festgelegt. Hätte man stattdessen einfach eine lineare Funktion gewählt, wäre der Funktionsgraph nahezu identisch mit geringem Effekt auf Stichprobenschwankungen.

Die gesamte Diskussion um die mathematischen Details der Präzisionszielfunktion ist aus praktischer Sicht für Schätzungen der Einwohnerzahl nicht relevant und kann bei einer Kritik oder Rechtfertigung des Zensus ignoriert werden.

---

<sup>2</sup> Dies zielt auf die Möglichkeit, dass einige Länder entscheiden konnten, ob sie ihre Gemeinden zu Gemeindeverbänden (*GV-Option*) oder sogenannten Gemeindeverbandsresten (*GV-Rest-Option*) zusammenfassen wollten, worauf wir im folgenden Kapitel 5 eingehen werden.

## 5. Methodik der Stichprobenziehung

Die Schätzung der Einwohnerzahlen geschieht separat auf Ebene der sogenannten *Sampling-Points*. Der Begriff Sampling-Point, nachfolgend im Gutachten stets mit SMP abgekürzt, definiert eine regionale Basiseinheit, aus der eine (Teil-)Stichprobe gezogen wird. Dies können (kleine) Gemeinden, Städte oder gar Stadtteile von Metropolen mit mehr als 200 000 Einwohnern sein. Außerdem bestand für einige Bundesländer die Option einer Zusammenfassung von Gemeinden zu Verbänden.

### 5.1. Zusammenfassung von Gemeinden

Zehn Bundesländern, deren Landesrecht die Struktur der Gemeindeverbände erlaubt, wurde im Zuge der Gesetzgebung zum Zensus 2022 die Möglichkeit eingeräumt, zwischen folgenden drei Optionen zu wählen: keine Zusammenfassung von Gemeinden, Option *GV* oder Option *GV-Rest*. Bei der Option *GV* werden Gemeinden eines gemäß Landesrecht existierenden Gemeindeverbands zu einem SMP zusammengefasst. Bei der Option *GV-Rest* werden Gemeinden eines Gemeindeverbands mit mehr als 2 000 Einwohnern als separate SMP behandelt und die restlichen beteiligten Gemeinden zu einem SMP zusammengefasst. Mit diesen beiden Optionen wurde das Ziel verfolgt, die gerade bei kleinen Gemeinden trotz Lockerung der Präzisionsvorgaben immer noch beachtlichen erforderlichen Stichprobensätze und damit auch den bundesweiten Gesamtstichprobenumfang auf etwa 10,3 Millionen zu reduzieren, siehe Bretsch et al. (2024a). Die Entscheidung für oder wider die Optionen *GV* oder *GV-Rest* fiel in den Ländern sehr unterschiedlich aus. Für manche Bundesländer bestand eine solche Option aus rein administrativen Gründen erst gar nicht.

Mit der Option *GV* konnten Gemeindeverbände gebildet werden. Diese Option wurde von den Bundesländern Niedersachsen, Rheinland-Pfalz und Thüringen gezogen. Die Option *GV-Rest*, wie oben erläutert bestehend aus einer Zusammenfassung von verbands- oder amtsangehörigen Gemeinden mit weniger als 2 000 Einwohnern, wurde von den Bundesländern Mecklenburg-Vorpommern und Schleswig-Holstein gezogen. In Mecklenburg-Vorpommern etwa gehört die Mehrzahl der Gemeinden (684 von 724) zu einem Amt und nur ungefähr jede sechste davon zählt mehr als 2 000 Einwohner. In

---

den verbleibenden Bundesländern definiert eine Gemeinde stets ihren eigenen SMP. Die Einwohner dieser Gemeinden wurden also unabhängig voneinander erhoben und hochgerechnet. Den Bundesländern Baden-Württemberg, Bayern, Brandenburg, Sachsen und Sachsen-Anhalt standen aus administrativen Gründen ebenfalls die beiden Optionen *GV* und *GV-Rest* offen, sie haben sich aber dagegen entschieden.

Zusammengefasst gibt es in Deutschland 10 786 Gemeinden, aus denen 6 499 SMP gebildet wurden, siehe Lorentz und Zwick (2025):

1. Große Stadtteile (mehr als 200 000 Einwohner) noch größerer Gemeinden mit mehr als 400 000 Einwohnern,
2. Gemeinden mit weniger als 400 000 Einwohnern,
3. Niedersachsen, Rheinland-Pfalz und Thüringen: Option *GV*,
4. Mecklenburg-Vorpommern und Schleswig-Holstein: Option *GV-Rest*.

## 5.2. Schichtung der Stichprobe

Die oberste Ebene der Schichtung ist streng genommen der SMP, denn es wird je SMP eine eigene Stichprobe gezogen.

Wie bei den Vorarbeiten zum Zensus 2011 wird der Stichprobenziehung eine Klassifizierung nach Anschriftengrößen zugrunde gelegt. In 2011 wurden für größere Gemeinden acht Anschriftengrößenklassen (sogenannte Oktile) gebildet, welche durch eine Unterteilung der aufsteigend sortierten Liste in mit möglichst gleicher Anzahl an Bewohnern besetzte Intervalle entstehen. In 2022 wurde die Anzahl der ebenfalls mit gleicher Bewohnerzahl zu besetzenden Schichten variabel an die Größe der SMP angepasst mit bis zu 16 Schichten, deren Anzahl von der Größe des jeweiligen SMP abhängt.

Die Zuweisung einer Schichtenzahl zu einem SMP mit gegebener Einwohnerzahl findet sich in Tabelle 5.1 (Klink & Lorentz, 2022, Tab.2, S.84). Beispielsweise werden bei einem SMP mit laut Melderegister 187 Bewohnern vier hinsichtlich der Bewohnerzahl möglichst gleich besetzte Schichten gebildet. Die Auswahlätze für die Schichten wurden im Intervall von 5 bis 50 % festgelegt.

Daneben existieren noch jeweils eine Schicht der Wohnheime (Auswahlatz von 100 %), der Gemeinschaftsunterkünfte (Auswahlatz von 100 %) und der sogenannten Nullanschriften (Auswahlatz von 20 %). Letztere sogenannte *Nullschicht* beinhaltet weder in Haupt- noch Nebenwohnsitz bemeldeten Wohnraum. In der Praxis enthält die

---

### Festlegung der Anzahl der Anschriftengrößenklassen je Sampling Point

Anzahl gemeldeter Bewohnerinnen und Bewohner (Haupt- und Nebenwohnsitze)	Anzahl Anschriftengrößenklassen
0	
1 bis 49	1
50 bis 99	2
100 bis 149	3
150 bis 199	4
200 bis 299	5
300 bis 399	6
400 bis 599	7
600 bis 799	7
800 bis 999	9
1 000 bis 1 499	10
1 500 bis 1 999	11
2 000 bis 2 499	12
2 500 bis 4 999	13
5 000 bis 9 999	14
10 000 bis 24 999	15
25 000 und mehr	16

Abbildung 5.1.: Tabelle 2 aus Klink und Lorentz (2022)

---

Nullschicht jedoch aus unerklärlichen Gründen einen beträchtlichen Anteil bemeldeter Adressen.<sup>1</sup>

### 5.3. Das GREG-Modell zur Bestimmung der Stichprobenumfänge

Unter dem Gebot der Gleichbehandlung wurden anders als beim Zensus 2011 im neuen Zensus grundsätzlich alle Gemeinden, als SMP oder Bestandteil eines solchen, im Design der Haushaltsstichprobe berücksichtigt. Der Zensus 2022 wurde außerdem flankiert durch die im vorherigen Kapitel diskutierte Präzisionszielfunktion. Nun stellt sich die anspruchsvolle Frage, in welchem Umfang die Stichprobe erhoben werden muss, um die formulierten Präzisionsziele erreichen zu können.

Die Grundlage zur Bestimmung der Stichprobenumfänge für einzelne Gemeinden bzw. SMP bildet wie bereits beim Zensus 2011 ein allgemeines lineares Regressionsmodell. Die Kernidee der Modellierung besteht in einer freien Hochrechnung (in der Statistik als Horvitz-Thompson-Schätzer bekannt) der Erhebungsergebnisse aus der Stichprobe zum Zensus 2011 auf Ebene der SMP, welche additiv durch ein lineares Regressionsmodell ergänzt wird. Das lineare Modell, in allen dargestellten Varianten seit dem Zensus 2011 mit GREG bezeichnet (für Generalised REGression) und in diversen Publikationen rund um den Zensus zitiert, verwendet Hilfsinformationen zur Bemeldung von Adressen im Melderegister.<sup>2</sup> Dabei wird innerhalb der Registerdaten der von der Hochrechnung der Stichprobe auf die Auswahlgesamtheit geschätzte Fehler zur Korrektur der freien Hochrechnung verwendet. In der Fachliteratur wird bei solchen Schätzungen meist von einer gebundenen Hochrechnung gesprochen, da die Hochrechnung an ein Hilfsmerkmal gebunden ist. Da es an dieser Stelle aber noch nicht um die Hochrechnung geht, sondern vielmehr um eine adäquate Bestimmung der Stichprobenumfänge VOR der Erhebung, musste aus nachvollziehbaren Gründen die Zensusstichprobe von 2011 als Hilfsinformation verwendet werden.

---

<sup>1</sup> Die Gemeinde Hanau bat am 5.12.2025 das Hessische Statistische Landesamt um Erklärung, wieso eine Nullschicht gemeldete Adressen enthalten kann, da dies in der Literatur und im Datenblatt nicht erläutert wird. Mit einer Email vom 17.12.2025 lehnte das HSL mit Hinweis auf einen anhängigen Rechtsstreit zum Zensus 2022 eine Auskunft zu dieser rein methodischen Frage ab.

<sup>2</sup> Das hierzu verwandte GREG-Modell zur Hochrechnung der Stichprobe wird in Kapitel 6.3 besprochen. Bei beiden in den Zensus-Veröffentlichungen mit GREG abgekürzten Modellen – eines zur Bestimmung der Stichprobenumfänge und zur Hochrechnung – wird ohne weitere Differenzierung stets auf die sehr umfangreiche Monographie Särndal (1992) verwiesen. Gleichwohl es natürlich Gemeinsamkeiten gibt, wird doch jeweils die freie Hochrechnung durch Registerinformation korrigiert.

---

Zum weiteren Verständnis müssen einige technische Begriffe eingeführt werden. Mit dem Totalwert wird in der Statistik die Aufsummierung aller beobachteten Werte eines bestimmten Merkmals in der Grundgesamtheit bezeichnet. In unserem Falle geht es hierbei um das *Zielmerkmal*  $Y$  der Einwohnerzahlen an Adressen, welches in der Zensusstichprobe 2022 zu erheben und anschließend auf SMP-Ebene hochzurechnen war.

Neben den Daten der Zensusstichprobe 2011 standen den statistischen Ämtern die Bewohnerzahlen der zugehörigen Adressen aus dem Melderegister für die gesamte Bevölkerung zur Verfügung. Die *Hilfsgröße*  $X$  beschreibe die eingetragenen Bewohnerzahlen an Adressen im Melderegister. Es ist wichtig festzuhalten, dass die Stichprobeneinheiten *nicht Personen*, sondern alle bewohnbaren *Adressen* sind. In den meisten Formeln in diesem Kontext wird eine bestimmte Adresse mit  $k$  bezeichnet. Die Bewohnerzahl der Adresse  $k$  wird mit  $x_k$  (Melderegister) und  $y_k$  (Zensusstichprobe von 2011) bezeichnet.

Die folgende Hilfskonstruktion wurde schon im Zensus 2011 verwendet. Da für die alte Stichprobe aus dem Zensus 2011 alle Paare  $(x_k, y_k)$  verfügbar sind, ist der Zusammenhang zwischen  $X$  und  $Y$  schon vor der neuen Erhebung zumindest abschätzbar. Je größer die Korrelation zwischen  $X$  und  $Y$ , desto mehr wird sich die Totalwertschätzung für  $Y$  (in diesem Abschnitt mit  $\hat{\tau}_Y$  bezeichnet) derjenigen von  $X$  (mit  $\hat{\tau}_X$  bezeichnet) angleichen. Wird durch  $\hat{\tau}_X$  der via Hochrechnung der Stichprobenwerte erhaltene Totalwert über- oder unterschätzt, so sollte dies tendenziell auch für  $\hat{\tau}_Y$  gelten. Zentral für ein Gelingen der Hochrechnung war also die Prämisse, dass ein starker Zusammenhang zwischen den Daten der Melderegister und den Einwohnerzahlen aus der neuen Zensusstichprobe pro Adresse im Jahr 2022 besteht.

Die Formel für den verwendeten kombinierten Totalwertschätzer  $\hat{\tau}_Y$  lautet<sup>3</sup>:

$$\hat{\tau}_{kR} = \sum_{h=1}^H N_h \cdot (\bar{y}_h + (\bar{X}_h - \bar{x}_h) \cdot \beta)$$

In der ersten Summe  $\sum_{h=1}^H N_h \cdot \bar{y}_h$  findet sich die freie Hochrechnung, in der restlichen Summe die oben skizzierte Korrektur derselben. Der Koeffizientenschätzer  $\beta$  wird nicht wie üblich über die Methode der kleinsten Quadrate bestimmt, sondern sinnvollerweise via Minimierung der Varianz  $\text{var}(\hat{\tau})$ , da im vorliegenden Falle aufgrund der Schichtung keine reine Zufallsstichprobe gezogen wird. Man erhält dann

$$\beta_{kR,min} = \frac{\sum_{h=1}^H \gamma_h^2 \frac{1-f_h}{n_h} S_{y_x h}}{\sum_{h=1}^H \gamma_h^2 \frac{1-f_h}{n_h} S_{x h}^2},$$

---

<sup>3</sup> Dies sind die Formeln (1-3) bei Burgard et al. (2019)

---

wobei gilt:  $\gamma_h = \frac{N_h}{N}$  und  $f_h = \frac{n_h}{N_h}$ .

Die Varianz des Totalwertschätzers ergibt sich dann als

$$V(\hat{\tau}_{kR}) = \sum_{h=1}^H N_h^2 \cdot \frac{1 - f_h}{n_h} \cdot (S_{yh}^2 - 2\beta S_{xyh} + \beta^2 S_{xh}^2).$$

Dabei bezeichnen  $S_{yh}^2$  und  $S_{xh}^2$  die Varianzen von  $X$  und  $Y$  in Schicht  $h$  und  $S_{xyh}$  die entsprechende Kovarianz.

## 5.4. Kritische Bemerkungen zum GREG-Modell

Aus unserer Sicht wurde der Zusammenhang zwischen Hilfs- und Zielgröße überschätzt. Dies war bereits im Vorfeld des Zensus 2011 der Fall, wo man die Ergebnisse aus dem Zensustest 2001 zur Generierung der Stichprobenumfänge heranziehen musste. So schrieben Berg und Bihler (2014) im Zusammenhang mit der Verfehlung der Präzisionsziele im Zensus 2011:

»Der Grund liegt in erster Linie darin, dass das Bestimmtheitsmaß der Regression mit den tatsächlichen Stichtagsdaten deutlich niedriger ist als der Planungswert aus dem Zensustest, oder mit anderen Worten, dass der statistische Zusammenhang zwischen den existenten und den gemeldeten Personen nicht so stark ist, wie man durch den Zensustest annehmen musste. Gemäß der Formel hängt die Fehlervarianz nämlich sehr stark von der Streuung der Residuen und damit vom Bestimmtheitsmaß der Regression ab.«

Besonders kritikwürdig ist die Entscheidung für einen über alle Schichten hinweg konstanten gehaltenen Regressionskoeffizienten  $\beta$ . In der Fachliteratur wird hier von einer kombinierten Schätzung gesprochen. In der vom Forscherteam im Zensusprojekt mehrfach zitierten Quelle Krug et al. (2001, S. 197) wird aber zu Recht festgestellt, dass eine kombinierte Regressionsschätzung nur dann sinnvoll ist, wenn die Gruppen (hier Schichten) ähnliche Koeffizienten aufweisen:

»Hierbei ist außerdem vorauszusetzen, dass die tatsächlichen Regressionskoeffizienten der einzelnen Schichten nicht zu stark voneinander abweichen. Das Schichtungsmerkmal sollte – wie bei den anderen gebundenen Hochrechnungsarten – möglichst nicht als Bezugsmerkmal gewählt werden.«

Beide Voraussetzungen werden jedoch im vorliegenden Modell nicht erfüllt:

- 
1. Es ist keinesfalls davon auszugehen, dass die Regressionskoeffizienten der einzelnen Schichten nah beieinander liegen.
  2. Als Schichtungs- und Bezugsmerkmal dient jeweils die Anschriftengröße in den Melderegisterdaten.

In solchen Fällen empfehlen Krug et al., 2001 ausdrücklich die Durchführung einer separaten Regressionsschätzung:

»Sollen die Stichprobenergebnisse nach Gruppen gegliedert werden, so werden im Auswahlplan die Schichten danach ausgerichtet worden sein. In diesem Fall empfiehlt sich die Anwendung einer separaten Regressionsschätzung.«

Es wird also eine separate Schätzung der Schichten angeraten oder ein Modell, in dem jede Schicht mit einem eigenen Koeffizienten versehen wird.

## 5.5. Zur Strategie der Varianzreduktion

Ein zentrales Instrument bei der Generierung der Stichprobenumfänge auf SMP-Ebene ist das in Krug et al. (2001) vorgestellte und in Burgard et al., 2020 adaptierte Konzept der Mittelung von Varianzreduktionsfaktoren.

Der via Korrektur der freien Hochrechnung erzielbare Effizienzgewinn ist mittels Korrelation zwischen Hilfsgröße  $X$  und Zielmerkmal  $Y$  darstellbar, die bei geeigneter Zerlegung der Varianz des GREG-Schätzers im Faktor  $(1 - \varrho^2)$  aufgeht. Und zwar gilt

$$V(\hat{\tau}_{kR}) = \sum_{h=1}^H N_h^2 \cdot \frac{1 - f_h}{n_h} \cdot S_{yh}^2 \cdot (1 - \varrho^2)$$

mit

$$(1 - \varrho^2) = 1 - \frac{\left(\sum_{h=1}^H \gamma_h^2 \frac{1-f_h}{n_h} S_{yhx}\right)^2}{\left(\sum_{h=1}^H \gamma_h^2 \frac{1-f_h}{n_h} S_{yh}^2\right) \cdot \left(\sum_{h=1}^H \gamma_h^2 \frac{1-f_h}{n_h} S_{xh}^2\right)}$$

Im Grunde beschreibt die Formel die gewichtete Summe der Varianzen von  $Y$  auf Ebene der Schichten  $h$ , wobei der Faktor  $(1 - \varrho^2)$  die Information des Zusammenhangs zwischen  $X$  und  $Y$  bündelt und aus der Summe herausgelöst werden kann. Der Effizienzgewinn von freier zu gebundener Hochrechnung kann damit durch das Bestimmtheitsmaß des Regressionsmodells beschrieben werden. Da es sich aufgrund der Schichtung der Stichprobe nicht um eine einfache Zufallsstichprobe handelt, ist die Interpretation der nun auf SMP-Ebene  $d$  bestimmten Faktoren  $(1 - \vartheta_d^2)$  etwas schwieriger. Nichtsdestotrotz taugen diese Faktoren

---

als Zusammenhangsmaße zwischen Hilfs- und Zielgröße. Ein schwächerer Zusammenhang zwischen  $X$  und  $Y$  bedeutet eine höhere Varianz und muss später entsprechend mit einer größeren Stichprobe kompensiert werden.

Wie im vorherigen Abschnitt erläutert, basierten die Schätzungen auf der Zensusstichprobe von 2011. Das Konzept sah zunächst eine Bestimmung des Gesamtstichprobenumfangs vor unter Verwendung der auf SMP-Ebene bestimmten Varianzreduktionsfaktoren. Beim Herunterbrechen des Gesamtstichprobenumfangs auf die einzelnen SMP – und im nächsten Schritt, sofern ein  $GV$  oder ein  $GV$ -Rest gebildet wurde, auf die darin beteiligten Gemeinden – wurden die SMP-spezifischen Varianzreduktionsfaktoren durch einen global über alle SMP gemittelten Faktor ersetzt.

Zu befürchten war eine tendenzielle Untererhebung kleinerer Gemeinden. Diese Problematik war den Autoren der Studie Burgard et al., 2020 bewusst, als sie feststellten, dass ein Großteil der Varianz auf die Schicht mit den größten Anschriften zurückgeht und Länder mit überproportional vielen kleinen Gemeinden<sup>4</sup> eine schlechtere Schätzqualität aufweisen. Zum Ausgleich wurden den SMP-spezifischen Stichprobenumfängen daher variable Sicherheitsaufschläge zgedacht, die sich im Mittel auf etwa 40 % des im Modell berechneten Umfangs beliefen.

Das Konzept der Varianzreduktion bedurfte daher im Vorfeld des Zensus 2022 ganz sicher einer kritischen Diskussion und empirischen Untersuchung. Es finden sich in Burgard et al. (2020) jedoch keine konkreten Ergebnisse, in welcher Größenordnung eine Verbesserung der Schätzung realistisch zu erwarten war. Bei der Mittelung der Varianzreduktionsfaktoren über alle SMP hinweg wird ebenfalls nicht berücksichtigt, dass sowohl die Anzahl als auch die Definition der Schichten durch die variable Klassierung sehr SMP-spezifisch gestaltet wurden.

Die Varianzreduktionsfaktoren basieren auf einem ungeeigneten Modell, was die Qualität der Schätzung negativ beeinflusst und damit auch eine angemessene Schätzung der erforderlichen Stichprobenumfänge verhindert. Zwar wird in Krug et al., 2001 eine Mittelung dieser Faktoren im kombinierten Regressionsmodell empfohlen, jedoch war wie zuvor erläutert dieses Modell hier nicht die passende Wahl.

---

<sup>4</sup> Mit Blick auf die uns vorliegenden Informationen sind dies vor allem Mecklenburg-Vorpommern, Rheinland-Pfalz, Schleswig-Holstein und Thüringen

---

## 5.6. Zur Schätzung in kleinen Gemeinden

Bei Gemeinden mit maximal 100 Anschriften wurde laut Bretschi et al. (2024a, S. 25) mangels belastbarer Hochrechnung eine Konvexkombination<sup>5</sup> aus Stichprobenziehung und Übernahme des Melderegisterwerts angewendet, um die Schätzung zu stabilisieren. Das kann formal dargestellt werden durch

$$\lambda \cdot H_g + (1 - \lambda) \cdot R_g,$$

wobei wir mit  $H_g$  den aus der Stichprobe hochgerechneten Wert und mit  $R_g$  den Wert des betreffenden Registerauszugs bezeichnen. Der Parameter  $\lambda$  kann dabei Werte zwischen 0 (keine Anschrift in der Stichprobe) und 1 (bei 100 Anschriften) annehmen.

Die Literaturquellen zur Durchführung der Schätzung für kleine Gemeinden sind überaus widersprüchlich (vgl. Bretschi und Lorentz, 2019, Bretschi et al., 2024a, Peters und Lorentz, 2025). Zusammenfassend könnte laut den Publikationen für eine kleine Gemeinde eines der folgenden Ereignisse eingetreten sein:

- Vollerhebung mit Korrektur der Unter- und Übererfassungen<sup>6</sup>
- Hochrechnung der Stichprobe, wobei die Gewichte des Modells zuvor auf übergeordneter SMP-Ebene *GV* oder *GV-Rest* ermittelt wurden
- Hochrechnung der Stichprobe, wobei die Gewichte des Modells zuvor auf Gemeindeebene ermittelt wurden
- Konvexkombination von aktuellem Registerauszug und Hochrechnung der Stichprobe, wobei die Gewichte des Modells zuvor auf übergeordneter SMP-Ebene *GV* oder *GV-Rest* ermittelt wurden
- Konvexkombination von aktuellem Registerauszug und Hochrechnung der Stichprobe, wobei die Gewichte des Modells zuvor auf Gemeindeebene ermittelt wurden
- Übernahme des aktuellen Melderegisterauszugs

Insgesamt stellen wir fest, dass sich die Publikationen zur Vorgehensweise bei der Stichprobenziehung und Hochrechnung der Einwohnerzahlen von kleinen Gemeinden widersprechen und keine klare Linie erkennen lassen. Dabei folgen die verschiedenen

---

<sup>5</sup> Eine Konvexkombination ist eine gewichtete Summe, bei der alle Gewichte nicht-negativ sind und sich zu 1 addieren.

<sup>6</sup> Nach Auskunft des Statistischen Bundesamtes vom 21.11.2025 gab es 33 Gemeinden, bei denen eine Vollerhebung durchgeführt wurde.

---

Strategien in Teilen weniger inhaltlichen Argumenten, sondern sind vielmehr administrativen Rahmenbedingungen geschuldet. Ob die Schätzung mit einem großen oder kleinen relativen Standardfehler behaftet ist, hängt nicht von der speziellen Einwohner- und Anschriftenkonfiguration der Gemeinde ab, sondern vielmehr davon, ob das jeweilige Bundesland über Gemeindeverbände verfügt und sich dazu entschlossen hat, diese Struktur bei der Durchführung des Zensus zu nutzen.

Ähnliches gilt für die schwankende Genauigkeit bei der Ermittlung des Standardfehlers (siehe Kapitel 6.4) und die durch die Präzisionszielfunktion vorgegebene Obergrenze, welche für Gemeindeverbände kleinere Stichprobenumfänge auf Gemeindeebene erlaubt. Im Gegensatz dazu sind die Gutachter der Meinung, dass sich die Präzisionsziele immer und ohne Abstriche an der Größe einer Gemeinde orientieren sollten. Wir halten es für wichtig, den Gemeinden ganz unabhängig davon, wie die Daten erhoben wurden (*GV*, *GV-Rest* oder keins von beiden), mitzuteilen, inwieweit das formulierte Präzisionsziel erreicht oder verfehlt wurde. Eine Ausweisung etwa von Standardfehlern oder Konfidenzintervallen auf übergeordneter SMP-Ebene ist für eine Gemeinde wertlos.

Damit ist natürlich die Frage, wie die Stichprobenumfänge für kleine Gemeinden am Ende tatsächlich bestimmt wurden, noch nicht geklärt. Zur Beantwortung betrachten wir beispielsweise die SMP-spezifischen Datenblätter des Amtes Röbel/Müritz, woraus sich implizit die Vorgehensweise erahnen lässt.

## 5.7. Fallbeispiel des Amtes Röbel/Müritz

Für die kleinste Gemeinde des Amtes Röbel/Müritz (Mecklenburg Vorpommern), Groß Kelle, wurden laut Datenblatt 94 Personen geschätzt bei einem angestrebten relativen Standardfehler von 0.51 % und einem »tatsächlich realisierten Standardfehler« von 1.33 %.<sup>7</sup> Es deutet einiges darauf hin, dass kleine Gemeinden wie Groß Kelle im übergeordneten Amt Röbel/Müritz (Option *GV-Rest*) verortet und anhand der für den *GV-Rest* gezogenen Stichprobe hochgerechnet wurden, also nur eine Handvoll Anschriften aus der Kleinstgemeinde in die Berechnungen eingegangen sind. Für diese These spricht auch, dass allen anderen Gemeinden des *GV-Rest* identische angestrebte und tatsächlich realisierte Standardfehler zugeschrieben wurden. Tiefere Einblicke in die Stichprobenziehung bei diesem *GV-Rest* waren bedauerlicherweise nicht möglich, da den Gemeinden hier keine Information zur Stichprobenziehung oder zur Schichtung nach Anschriftengrößenklassen zur Verfügung stand.

Das Amt Röbel/Müritz besteht insgesamt aus der gleichnamigen Stadt mit ca. 5 000 Einwohnern (eigener SMP), der Gemeinde Rechlin (*vor* dem Zensus noch mehr als 2 000

---

<sup>7</sup> Eine präzise Definition des Begriffs findet sich in Kapitel 6.4.

---

Einwohner, daher ebenfalls ein eigener SMP) und einem *GV-Rest* aus 17 kleineren und Kleinstgemeinden. Auch bei der Stadt Röbel/Müritz wird laut Datenblatt der angestrebte relative Standardfehler von 0.53 % mit dem tatsächlich realisierten Standardfehler von 1.13 % deutlich verfehlt. Bei der Gemeinde Rechlin, die durch die Neuschätzung der Einwohnerzahlen gut 12 % weniger Einwohner zählt, wird ein angestrebter relativer Standardfehler von 0.79 % einem tatsächlich realisierten Standardfehler von 1.5 % gegenübergestellt. Bei den 17 kleineren Gemeinden zeigt sich wie oben erwähnt ein angestrebter relativer Standardfehler von 0.51 % und ein tatsächlich realisierter Standardfehler von 1.33 %.

Ebenfalls in den Datenblättern ausgewiesen werden die 95 % -Konfidenzintervalle zur Einwohnerzahlschätzung für den *GV-Rest*, aber nicht für die einzelnen Gemeinden, denen ein solches – ohnehin einfach aus dem Stichprobenumfang und dem realisierten Standardfehler ermittelbares – Intervall (im vorliegenden Falle 11896 bis 12548 Einwohner) nichts nutzt. Die angestrebte Präzisionsschwelle von 0.51 % wurde am *GV-Rest* festgemacht, konnte aber weder für den *GV-Rest* noch für die beteiligten Gemeinden eingehalten werden. Dennoch hat durch die Zusammenfassung der Gemeinden eine Reduzierung des Standardfehlers für kleinere Gemeinden stattgefunden, der auf Gemeindeebene absolut (mit einer Obergrenze von 15) hätte gemessen werden sollen und sich in relativer Darstellung deutlich erhöhen würde.

## 5.8. Zusammenfassung

Die Einwohnerzahlen von vergleichbaren kleinen Gemeinden oder gar Kleinstgemeinden in verschiedenen Bundesländern können auf sehr unterschiedliche Art geschätzt werden, was dem Gleichheitsgrundsatz widerspricht. Die Stichprobenumfänge und damit auch die Präzision bei der Hochrechnung hängen in erster Linie davon ab, ob das jeweilige Bundesland über Gemeindeverbände verfügt und sich gegebenenfalls für eine der beiden Optionen *GV* oder *GV-Rest* entschieden hat.

Die Einführung einer Nullschicht halten wir für sinnvoll, können aber nicht nachvollziehen, warum die Datenblätter der Gemeinden in dieser Schicht einen beträchtlichen Teil bemeldeter Anschriften ausweisen.

Hinsichtlich des GREG-Modells halten wir die Verwendung eines über alle Schichten hinweg konstant gehaltenen Regressionskoeffizienten  $\beta$ , sprich den Ansatz einer kombinierten Schätzung, für besonders kritikwürdig. Von dieser Modellvereinfachung ist aus methodischer Sicht dringend abzuraten, da es keinen Grund zur Annahme gibt, dass die Regressionskoeffizienten der einzelnen Schichten bei getrennter Berechnung nahe beieinander lägen.

---

Außerdem ist zu bemängeln, dass die Schichtung aus der Hilfsgröße selbst gebildet wird. Krug et al., 2001 empfehlen in solchen Fällen die Durchführung einer separaten Regressionsschätzung.

Am Fallbeispiel in Kapitel 5.7 haben wir beispielhaft gezeigt, dass mit der Formulierung eines angestrebten Präzisionsziels und der Schätzung eines relativen Standardfehlers auf Ebene eines Gemeindeverbands die Ergebnisse besser erscheinen, als sie es auf Gemeindeebene sind. Wobei im Fallbeispiel selbst auf SMP-Ebene das angestrebte Präzisionsziel von 0.51 % deutlich verfehlt wurde.

Da die in Kapitel 5.5 ausgiebig diskutierten Varianzreduktionsfaktoren auf einem nach unserer Auffassung schlecht geeigneten Modell basieren, wird die Frage, ob diese Größen sinnvoll zur Verteilung des Gesamtstichprobenumfangs auf die SMP beigetragen haben, letztlich zur Marginalie.

Dass den SMP-spezifischen Stichprobenumfängen zusätzlich variable Sicherheitsaufschläge zugedacht werden mussten, die sich im Mittel auf 40 % gegenüber den durch das Regressionsmodell berechneten Umfängen beliefen, spricht ebenfalls gegen das verwendete Modell.

## 6. Hochrechnung

Ein in der Vorbereitung des Zensus 2011 in Auftrag gegebenes gemeinsames Forschungsprojekt der Universität Trier und GESIS Mannheim sollte neben der Entwicklung eines Stichprobenkonzepts eine Hochrechnungsmethodik empfehlen. Zur Ermittlung der Einwohnerzahl wurde wie für das Stichprobenkonzept eine Variante des verallgemeinerten Regressionsschätzers GREG auf Anschriftenebene vorgeschlagen und auch umgesetzt. Dabei wurde die Anschriftengröße, gemessen an der Zahl der mit Hauptwohnsitz gemeldeten Personen, als zentrale Hilfsinformation verwendet. Zur Fehlerrechnung wurde durch das Forschungsprojekt eine klassische Residualvarianzschätzung empfohlen. Die Modellierung wurde zur Hochrechnung der Zensus 2022 Stichprobe beibehalten.

In diesem Kapitel ist ein Verzicht auf tiefere mathematische Terminologie nicht möglich. Daher versuchen wir, die verwendeten Formeln anhand eines kleinen Beispiels zu erläutern.

### 6.1. Einführendes Beispiel

Zur Motivation des Hochrechnungsmodells betrachten wir ein fiktives Beispiel einer kleinen Gemeinde  $G$  mit 100 registrierten Anschriften. Wir wählen aus dem Melderegister zufällig 10 Anschriften aus, anders ausgedrückt: Wir ziehen eine einfache 10 % ige Stichprobe  $s$ . Mit  $x_s$  sei die Bewohnerzahl laut Register in diesen 10 Anschriften bezeichnet, etwa  $x_s = 12$  (Registerstichprobe). Nun führen wir an denselben 10 Anschriften eine Erhebung durch und stellen dort  $y_s = 15$  Bewohner fest (Zensusstichprobe). Es leben also in der Stichprobe von 10 Anschriften 25 % mehr Bewohner als gedacht. Das lineare Modell lautet hier  $y_s = b \cdot x_s$ . Setzen wir unsere Werte von oben ein, erhalten wir

$$b = \frac{y_s}{x_s} = \frac{15}{12} = 1,25.$$

Die Aufgabe besteht nun darin, die Bewohnerzahl der Gemeinde zu schätzen, oder genauer, die Bewohnerzahl der Zensusstichprobe auf die Grundgesamtheit in  $G$  hochzurechnen. Tatsächlich wissen wir nicht einmal mit Bestimmtheit, wie viele Anschriften es in der Gemeinde tatsächlich gibt, können diese Zahl aber mittels Melderegister annähern.

---

Im einfachsten aller Modelle multiplizieren wir  $y_s$  mit dem reziproken Auswahlatz: Da  $x_s$  eine 10% -Stichprobe darstellt, gilt dies näherungsweise auch für  $y_s$ , der reziproke Auswahlatz ist also der Kehrwert von  $1/10$ , sprich  $d = 10$  (das *Designgewicht*). Mit der sogenannten freien Hochrechnung erhalten wir danach in allererster Schätzung

$$Y_G \approx d \cdot y_s = 10 \cdot 15 = 150 \quad \text{Personen in der Gemeinde } G.$$

Nun versuchen wir, diese erste Schätzung zu verbessern. Dazu verwenden wir wieder das Melderegister, aus dem wir ja bereits die Stichprobe gezogen hatten. Hat das Melderegister eine angemessene Qualität, so enthält es näherungsweise alle Anschriften aus  $G$  und deren korrekte Bewohnerzahlen. Laut Melderegister, aus dem wir 10 Anschriften (Stichprobe  $s$ ) mit  $x_s = 12$  Bewohnern gezogen haben, leben in den 100 registrierten Anschriften insgesamt, sagen wir,  $X_G = 132$  Personen. Mit diesen Informationen sind wir in der Lage, ein verfeinertes Hochrechnungsmodell aufzustellen:

$$Y_G = d y_s + b \cdot (X_G - d x_s).$$

Wenn wir die Werte von oben einsetzen, ergibt sich

$$Y_G = 10 \cdot 15 + \frac{15}{12} \cdot (132 - 120) = 150 + \frac{15}{12} \cdot 12 = 165.$$

Dieses Ergebnis macht durchaus Sinn: Da der Wert  $X_G$  den aus  $x_s$  hochgerechneten Wert um 10% übersteigt (132 statt 120), sollte auch  $Y_G$  den aus  $y_s$  hochgerechneten Wert für die Zensusstichprobe um 10% übersteigen (165 statt 150).

Nun rechnen wir dasselbe Modell mit anderen Zahlen durch. Angenommen, wir beobachten in der Zensusstichprobe genauso viele Personen wie in der zugehörigen Registerstichprobe, also  $y_s = x_s = 12$ . Dann gilt  $b = y_s/x_s = 1$  und damit

$$Y_G = 10 \cdot 12 + 1 \cdot (132 - 120) = 132 = X_G.$$

Nehmen wir nun abschließend an, die Registerstichprobe  $x_s$  lasse sich fehlerfrei auf  $X_G$  hochrechnen, zum Beispiel seien  $x_s = 12$  und  $X_G = 120$ . Dann würde die Korrektur durch das Register überflüssig und der Korrekturterm auf der rechten Seite verschwinden. Die geschätzte Einwohnerzahl in  $G$  ergäbe sich dann als freie Hochrechnung:  $Y_G = 120$ .

---

## 6.2. Annäherung an das im Zensus verwendete Modell

Im Zensus wurde ein umfangreicheres Modell aufgestellt, das anstelle von  $X_G$  mehrere Variablen verwendet. Wir nähern unser Beispiel daher diesem Modell an. Da im Melderegister auch Angaben über Geschlecht, Alter und Nationalität der Einwohner erfasst sind, können wir die Zahl  $X_G$  durch einen Vektor

$$\mathfrak{X}_G = (X_{G,1}, X_{G,2}, \dots, X_{G,11})$$

sogenannter Hilfsgrößen ersetzen:

$X_{G,1}$  = Anzahl der gemeldeten Personen insgesamt

$X_{G,2}$  = Anzahl der gemeldeten männlichen Personen

$X_{G,3}$  = Anzahl der gemeldeten Personen im Alter von 0 bis 5 Jahre (Altersklasse 1)

$X_{G,4}$  = Anzahl der gemeldeten Personen im Alter von 6 bis 17 Jahre (Altersklasse 2)

$X_{G,5}$  = Anzahl der gemeldeten Personen im Alter von 18 bis 24 Jahre (Altersklasse 3)

$X_{G,6}$  = Anzahl der gemeldeten Personen im Alter von 25 bis 29 Jahre (Altersklasse 4)

$X_{G,7}$  = Anzahl der gemeldeten Personen im Alter von 30 bis 39 Jahre (Altersklasse 5)

$X_{G,8}$  = Anzahl der gemeldeten Personen im Alter von 40 bis 49 Jahre (Altersklasse 6)

$X_{G,9}$  = Anzahl der gemeldeten Personen im Alter von 50 bis 59 Jahre (Altersklasse 7)

$X_{G,10}$  = Anzahl der gemeldeten Personen im Alter von 60 bis 64 Jahre (Altersklasse 8)

$X_{G,11}$  = Anzahl der gemeldeten deutschen Personen.

Entsprechend ersetzen wir  $x_s$  durch einen Vektor  $\mathfrak{x}_s = (x_{s,1}, x_{s,2}, \dots, x_{s,11})$ . Anstelle des Koeffizienten  $b$  tritt nun ein Vektor  $\beta = (\beta_1, \dots, \beta_{11})$ . Das lineare Modell lautet dann

$$y_s = \beta \cdot \mathfrak{x}_s^T = \beta_1 \cdot x_{s,1} + \beta_2 \cdot x_{s,2} + \dots + \beta_{11} \cdot x_{s,11} = \sum_{j=1}^{11} \beta_j \cdot x_{s,j}.$$

Die Koeffizienten dieses Modells werden klassisch via »Methode der Kleinsten Quadrate« geschätzt und mit  $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_{11})$  bezeichnet.

---

Die so erhaltenen Werte verwenden wir danach für die Hochrechnung der Einwohnerzahl von  $G$  und erhalten entsprechend

$$Y_G = d y_s + b \cdot (X_G - d x_s) = \sum_{k \in s_d} d_k y_k + \sum_{j=1}^{11} \hat{\beta}_j \left( \sum_{k \in U_d} x_{kj} - \sum_{k \in s_d} d_k x_{kj} \right).$$

Um die Schätzung zu verbessern, könnte man vor der Ziehung der Stichprobe eine Einteilung der Registeradressen nach Anschriftengrößenklassen vornehmen, zum Beispiel:

*Klasse 1* : Anschriften mit einem Bewohner

*Klasse 2* : Anschriften mit mindestens zwei Bewohnern

Jede Klasse, im Folgenden Schicht genannt, habe dabei ihren eigenen Auswahlatz bzw. ihr eigenes Designgewicht bzgl. der zu ziehenden Stichprobe. Damit kann auch jeder aus dem Register gezogenen Anschrift  $k$  das Designgewicht  $d_k$  der betreffenden Schicht zugeordnet werden. Zum Beispiel bedeutet  $d_1 = 8$ , dass die Anschrift  $k = 1$  acht Anschriften der Gemeinde repräsentiert. Die Hochrechnungsformel sieht dann so aus:

$$\begin{aligned} Y_G &= \sum_{k \in s} d_k y_k + \hat{\beta} \begin{pmatrix} \sum_{k \in G} \mathcal{X}_{G,1} - \sum_{k \in s} d_k \mathcal{X}_{s,1} \\ \sum_{k \in G} \mathcal{X}_{G,2} - \sum_{k \in s} d_k \mathcal{X}_{s,2} \\ \vdots \\ \sum_{k \in G} \mathcal{X}_{G,11} - \sum_{k \in s} d_k \mathcal{X}_{s,11} \end{pmatrix} \\ &= \sum_{k \in s} d_k y_k + \sum_{j=1}^{11} \hat{\beta}_j \left( \sum_{k \in G} x_{kj} - \sum_{k \in s} d_k x_{k,j} \right), \end{aligned}$$

wobei  $x_k$  und  $y_k$  die Bewohner der Anschrift  $k$  im Melderegister und in der Zensusstichprobe bezeichnen. Entsprechend sind  $x_{k,j}$  die Bewohner an Anschrift  $k$  bzgl. einer der elf Hilfsgrößen.

### 6.3. Das lineare Hochrechnungsmodell GREG

Das Statistische Bundesamt hat erst sechs Monate nach der Mitteilung der amtlichen Einwohnerzahlen mit der Veröffentlichung von Lorentz und Zwick (2025) eine erste technische Dokumentation zur Hochrechnung vorgelegt.<sup>1</sup> Wir haben in einer Reihe von

---

<sup>1</sup> Der Aufsatz erschien am 16. April 2025, die Datenblätter waren z.B. in Thüringen ab dem 14.10.2024 verschickt worden ([https://statistik.thueringen.de/presse/2024/pr\\_287\\_24.pdf](https://statistik.thueringen.de/presse/2024/pr_287_24.pdf)).

---

Anfragen an das Statistische Bundesamt versucht, die Unklarheiten zu beseitigen, aber viele Details konnten nicht abschließend geklärt werden.<sup>2</sup>

Die Modellierung beim Zensus 2011 wurde zur Hochrechnung der Haushaltsstichprobe zum Zensus 2022 weitgehend beibehalten. Im Modell werden die vorgenannten elf Hilfsgrößen über Geschlecht, Herkunft und Lebensalter der an den Anschriften gemeldeten Personen verwendet. Das Modell impliziert die nach oben offene Klasse aller Personen ab einem Alter von 65 Jahren.<sup>3</sup>

Wie im Einführungsbeispiel werden also elf Hilfsgrößen verwendet, es wird damit – im Gegensatz zum GREG-Modell zur Bestimmung der Stichprobenumfänge (siehe Kapitel 5.3) – ein Vektor  $\beta = (\beta_1, \dots, \beta_{11})$  von Koeffizienten geschätzt. Die Formel der Totalwertschätzung für die Bevölkerungszahl lautet in Anlehnung an Särndal et al. (1992):

$$\hat{t}_{y,d,GREG} = \sum_{k \in s_d} d_k y_k + \sum_{j=1}^J \hat{\beta}_j \left( \sum_{k \in U_d} x_{kj} - \sum_{k \in s_d} d_k x_{kj} \right), \quad (6.1)$$

wobei in unserem Falle  $J = 11$  gilt und  $\hat{\beta}_j$  die  $j$ -te Komponente des Vektors  $\hat{\beta}$  der geschätzten Regressionskoeffizienten bezeichnet. Die erste Summe auf der rechten Seite von Gleichung 6.1 beschreibt die freie mit den reziproken Inklusionswahrscheinlichkeiten  $d_k$  gewichtete Hochrechnung.<sup>4</sup> Rechts davon findet sich die Korrektur durch das Melderegister, in der die Stichprobenelemente ebenfalls mit  $d_k$  gewichtet sind.

Der Koeffizientenvektor  $\hat{\beta}$  muss vorher aus der multiplen Regression von  $Y$  auf  $X$  geschätzt werden, in der die  $x_k$  entsprechend ihrer Schichtzugehörigkeit gewichtet werden. Hierbei handelt es sich um einen klassischen Kleinste-Quadrate-Schätzer mit Wichtung,

---

2 Am 17.11.2025 teilt uns das Statistische Bundesamt (GZ 120111/818674) mit, dass weder das Programm zur Berechnung der Stichprobenumfänge, noch die Programme zur Hochrechnung und zur Schätzung der realisierten Standardfehler im Source-Code zur Verfügung gestellt werden.

3 Beim Zensus 2011 war der Vektor der Prädiktoren des Hochrechnungsmodells etwas länger und umfasste insgesamt 13 Komponenten, siehe Berg und Bihler (2014), laut Münnich et al. (2013) sogar 20 Komponenten. Dem gegenüber stehen, wie oben beschrieben, 11 Komponenten im gleich gestalteten Modell des Zensus 2022. Es gibt für das Zensusmodell keine uns bekannte Analyse zur optimalen Anzahl an Prädiktoren, auch mit Blick auf die in Kapitel 6.7.2 diskutierte Problematik der Multikollinearität.

4 Im Zensus wurden die Designgewichte erst nach Löschung der Ausfälle, z.B. der Aussteuerung eines oder mehrerer Gebäude, berechnet (Hessisches Statistisches Landesamt, 2025). Sollten diese Ausfälle systematisch erfolgt sein (z.B. dadurch, dass vereinzelte Großgebäude im Vergleich zu anderen Großgebäuden andere Zusammensetzungen der Bewohner, anderes Meldeverhalten oder eine andere Befragbarkeit aufweisen), dann wären die als »stichprobenneutral« bezeichneten Ausfälle für die Hochrechnung nicht neutral.

---

die Formel lautet

$$\hat{\beta} = \left( \sum_{k \in s_d} d_k \mathbb{X}_k \mathbb{X}_k^T \right)^{-1} \sum_{k \in s_d} d_k \mathbb{X}_k y_k. \quad (6.2)$$

Unter bestimmten Voraussetzungen an das Regressionsmodell ist dieser Schätzer als bester linearer erwartungstreuer Schätzer einzustufen. Hierzu werden in der Statistik Annahmen an das Modell getroffen, im vorliegenden Falle muss insbesondere für die Störgrößen gelten, dass sie unabhängig und identisch verteilt sind. Nach unserer Auffassung besteht hier jedoch keine Unabhängigkeit (Stichwort Interviewereffekte, vgl. Kapitel 8.4.2) und die Varianz hängt ganz offensichtlich von der Anschriftengröße ab (Stichwort Heteroskedastie).

Der Totalwertschätzer aus Gleichung 6.1 kann abschließend in schlankerer Form dargestellt werden, die rechentechnische Vorteile besitzt:

$$\hat{t}_{y,d,GREG} = \sum_{k \in s_d} d_k g_k y_k = \sum_{k \in s_d} w_k y_k, \quad (6.3)$$

wobei

$$g_k := 1 + \left( \sum_{i \in U_d} \mathbf{x}_i - \sum_{i \in s_d} d_i \mathbb{X}_i \right)^T \left( \sum_{i \in s_d} d_i \mathbb{X}_i \mathbb{X}_i^T \right)^{-1} \mathbb{X}_k \quad (6.4)$$

und das Gewicht der Anschrift  $k$  durch  $w_k := d_k g_k$  gegeben ist.

## 6.4. Das Konzept des realisierten Standardfehlers

In den Datenblättern, welche die Gemeinden zur Ermittlung der amtlichen Einwohnerzahl im Rahmen des Zensus 2022 erhalten haben, wird ein »tatsächlich realisierter Standardfehler« angegeben. Diese in der Literatur unübliche Bezeichnung suggeriert, dass es sich um einen auf übliche Weise geschätzten Standardfehler handelt. Das ist hier aber nicht der Fall.

In einem mühsamen Austausch mit dem Auskunftsdienst von Destatis haben wir in Erfahrung gebracht, dass dieser »tatsächlich realisierte Standardfehler« im Wesentlichen durch den mit dem Regressionsmodell verknüpften Residualvarianzschätzer berechnet wird. Hier handelt es sich um die bei Lorentz und Zwick (2025) mit  $\hat{V}(\hat{t}_{y,d,GREG})$  bezeichnete Größe. Darauf basierend bestimmen Lorentz und Zwick (2025) einen modellbasierten relativen Standardfehler für die geschätzte Einwohnerzahl durch:

$$\text{rse}(\hat{t}_{y,d,GREG}) := \frac{\left( \hat{V}(\hat{t}_{y,d,GREG}) \right)^{\frac{1}{2}}}{\hat{t}_{y,d,GREG}}. \quad (6.5)$$

---

Um via Formel 6.5 den realisierten Standardfehler schätzen zu können, braucht man neben dem Totalwertschätzer der Hochrechnung (Nenner) aus Gleichung 6.1 vor allem die Varianzschätzung im Zähler.

Die Varianz des Totalwertschätzers kann, unter Berücksichtigung aller Anschriften  $k = 1, \dots, n_d$  der zugehörigen Stichprobe  $s_d$ , folgendermaßen geschätzt werden:

$$\hat{V}(\hat{t}_{y,d}) = \frac{N_d^2}{n_d} \left(1 - \frac{n_d}{N_d}\right) \cdot S_{Y \rightarrow X}^2, \quad (6.6)$$

wobei  $n_d$  die Anzahl der Beobachtungen in der Stichprobe und  $N_d$  die Grundgesamtheit aller Beobachtungen im SMP  $d$  bezeichne.

Den Aufbau der Formel 6.6 findet man in einschlägigen Lehrbüchern über Stichprobentheorie wie Cochran (1977), Lohr (2022) und Särndal et al. (1992). Interessant ist die Ausgestaltung der Größe  $S_{Y \rightarrow X}^2$ , sprich der Fehlervarianz, die bei der Regression von  $Y$  auf  $X$  zu beobachten ist. Diese wird im Folgenden *Residualvarianz* genannt und beschreibt, wie stark die Residuen einer Regression um ihren Mittelwert streuen.

Zur Motivation der Residualvarianz betrachten wir daher zunächst den Begriff des Residuums einer Beobachtung (lateinisch für ‘Rest’), welches sich als Differenz zwischen dem beobachteten und dem via Regression geschätzten Wert einer Stichprobeneinheit  $k$  ergibt, also  $e_k := y_k - f(x_k)$ . Eine umfassende Residualanalyse – die nicht erfolgt ist oder zumindest in den Fachaufsätzen zum Zensus nicht veröffentlicht wurde – ist essentiell zur Bewertung eines Modells, um Fehler identifizieren zu können, die nicht durch das Modell erklärt werden.

Zur Schätzung  $S_{Y \rightarrow X}^2$  muss beachtet werden, dass wir es mit einer multiplen Regression zu tun haben. Anstelle des erklärenden Merkmals  $X$  tritt ein Vektor von Merkmalen  $X_i, i = 1, \dots, J$ , im vorliegenden Falle  $J = 11$ . Durch die Hinzunahme sinnvoller erklärender Variablen  $X_i$  in ein multiples Regressionsmodell werden sukzessive die Residuen verkleinert. Der Preis dafür ist eine Verringerung der Freiheitsgrade des Modells, da nunmehr  $J$  Koeffizienten  $\beta_i, i = 1, \dots, J$ , zu schätzen sind. Damit gelangen wir zu folgender Schätzung der Residualvarianz:

$$S_{Y \rightarrow X}^2 := \frac{1}{n_d - J - 1} \cdot \sum_{k \in s_d} (v_k - \bar{v}_d)^2, \quad (6.7)$$

wobei  $v_k$  die in der Stichprobe beobachteten gewichteten Residuen auf Anschriftenebene und  $\bar{v}_d$  deren arithmetisches Mittel bezeichne.

In Lorentz und Zwick (2025) wird dagegen die Verringerung der Freiheitsgrade bei der Varianzschätzung nicht berücksichtigt. Außerdem wird die Varianz zusätzlich nach Anschriftengrößenklassen  $h$ , sprich schichtweise geschätzt und danach über alle Schichten

---

$h$  aufsummiert:

$$\hat{V}(\hat{t}_{y,d,GREG}) = \sum_{h=1}^H \frac{N_{d,h}^2}{n_{d,h}} \left(1 - \frac{n_{d,h}}{N_{d,h}}\right) \frac{1}{n_{d,h} - 1} \sum_{k \in s_{d,h}} (v_k - \bar{v}_{d,h})^2, \quad (6.8)$$

Vereinfacht ausgedrückt haben wir es also mit einer gewichteten Summe  $\sum_{h=1}^H \lambda_h \cdot S_h^2$  der Varianzen je Schicht zu tun. Eine Herleitung von Gleichung 6.8 wird aber erst dann möglich, wenn man gewisse Unabhängigkeitsannahmen unterstellt.<sup>5</sup>

Von besonderer Bedeutung ist die nicht thematisierte Abhängigkeit der Beobachtungen (und damit der Residuen) erstens innerhalb einer Anschrift und zweitens innerhalb der vom selben Erhebungsbeauftragten bearbeiteten Anschriften (siehe hierzu Kapitel 8.4.2). Beides sind technisch gesehen Nicht-Stichprobenfehler (Kapitel 8), die sich unmittelbar auf die Qualität der Regressionsschätzung auswirken (vgl. hierzu auch unsere Ausführungen in Kapitel 6.3) und im Rahmen des Zensus nicht quantifiziert und damit nicht in den Berechnungen berücksichtigt wurden. Als Konsequenz sind die geschätzten Varianzen kleiner und die zu schätzenden Konfidenzintervalle größer.<sup>6</sup>

Darüber hinaus können z.B. die vernachlässigten Varianzen der Koeffizienten  $\hat{\beta}_j$  zur Residualvarianz beitragen, insbesondere dann, wenn eine starke lineare Abhängigkeit der erklärenden Merkmale vorliegt. Dies führt zu einer Unterschätzung von  $\hat{V}(\hat{t}_{y,d})$  und wird wahrscheinlicher bei kleinen Stichprobenumfängen. Auf diese Problematik kommen wir in Kapitel 6.7 zurück.

Bei Lorentz und Zwick (2025) findet sich wie gesagt keine Herleitung zur Begründung der zusätzlichen Schichtung in Formel 6.8. Selbst wenn man innerhalb der Schichten eine Homogenitätsannahme trifft, wird die in der statistischen Varianzanalyse als Varianz zwischen den Gruppen (hier: Schichten) bezeichnete Größe unterschlagen, welche grob skizziert von der Größenordnung  $\sum_h (\bar{v}_{d,h} - \bar{v}_d)^2$  ist, wobei  $\bar{v}_d$  das arithmetische Mittel der  $\bar{v}_{d,h}$  bezeichnet. Außerdem wird Formel 6.8, wie oben besprochen, nicht dem Umstand der Reduktion der Freiheitsgrade im multiplen Modell gerecht. Die darin geschätzte Varianz muss daher, einmal ungeachtet der in Frage stehenden Schichtung, gemäß Formel 6.7 mit dem Faktor  $(n_{d,h} - 1)/(n_{d,h} - J - 1)$  korrigiert werden, was eine weitere Vergrößerung von  $\hat{V}(\hat{t}_{y,d})$  und damit der je SMP berechneten relativen Standardfehler zur Folge hat, ganz besonders bei kleinen Gemeinden bzw. Stichprobenumfängen  $n_{d,h}$ .

---

<sup>5</sup> Auf unsere Bitte hin, die getroffenen Grundannahmen der Regressionsschätzung zu nennen, wurden uns vom Destatis-Auskunftsdienst am 27.11.2025 gleich zwei Quellen genannt, nämlich Deville und Sarndal (1992) und Särndal et al. (1992), mit einem Gesamtumfang von 715 Seiten, begleitet von dem Satz: »Für ein tieferes Verständnis sollten die Texte in Gänze durchgearbeitet werden, deshalb wird hier auch nicht auf einzelne Seiten der angegebenen Literatur verwiesen.«

<sup>6</sup> Auf ein verwandtes Problem hatte Küchenhoff (Bleninger et al., 2020) bereits vor dem Zensus 2022 hingewiesen.

---

Zusammenfassend können wir festhalten, dass die nicht thematisierten Unabhängigkeitsannahmen, zusammen mit der Nichtberücksichtigung der Komplexität des Modells (Multikollinearität, Reduktion der Freiheitsgrade) und der Nicht-Stichprobenfehler, zu einer Unterschätzung der »tatsächlich realisierten Standardfehler« führen.

## 6.5. Beeinflussung des Standardfehlers durch willkürliche Schichtung

In zahlreichen Gemeinden, deren Datenblätter eine Schichtungstabelle enthalten, zeigt sich eine nicht trennscharf vorgenommene Schichtung. Das bedeutet, dass zwei Gebäude mit identischer Bewohnerzahl im Register in zwei verschiedenen Schichten gezählt werden konnten, also z.B. kann innerhalb einer Gemeinde ein Gebäude mit 74 Einwohnern zu Schicht 15 oder Schicht 16 gezählt werden. Diese ungewöhnliche Art der Schichtung erlaubt bei der Fehlerrechnung unerwartete Spielräume.

Die im vorherigen Abschnitt kritisierte Schichtung in Formel 6.8 verwendet die jeweiligen absoluten Häufigkeiten der Anschriften in den Schichten und der Stichprobe. Da die bis zu 17 Schichten (inklusive Schicht Null der bewohnbaren, aber unbemeldeten Anschriften) anhand der laut Register zu erwartenden Einwohnerzahl auf SMP-Ebene ausgewogen wurden, nimmt die Anzahl der Anschriften von Schicht 1 bis 16 naturgemäß kontinuierlich ab.

Hierzu wird die Schichtung variabel vorgenommen und dies wie oben angesprochen in aller Regel nicht trennscharf. Die Obergrenze einer Schicht ist also meist die Untergrenze der nächsthöher gelegenen. Beispiel: die erste Schicht beinhalte Anschriften mit 1 – 2 Bewohnern, die zweite Anschriften mit 2 – 5, die dritte Anschriften mit 5 – 10 Bewohnern usw.. Wie nun die Anschriften mit etwa 5 Bewohnern auf die beiden Schichten verteilt werden, ist dem Zufall überlassen oder gar eine subjektive Entscheidung.

Da die Zahl der in der Befragung festgestellten Bewohner in der Regel variiert, steigt die Varianz bei steigender Schichtnummer. Hier könnte man der Versuchung erliegen, die Beobachtungen im Sinne einer Minimierung der Fehlervarianz zu sortieren.

Zur Verdeutlichung der Problematik betrachten wir ein konstruiertes Beispiel einer kleinen Gemeinde, bei der 20 Anschriften mit 38 Bewohnern in die Stichprobe gelangt seien. Es werden drei Schichten gebildet wie in Tabelle 6.1 dargestellt, die Beobachtungen samt Schichtung finden sich in Tabelle 6.2.

---

Schicht	Einwohner
1	1 – 2
2	2 – 3
3	3 – 5

Tabelle 6.1.: Einwohnerzahl je Schicht

	Schicht 1						Schicht 2					Schicht 3								
$x_i$	1	1	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	4	
$y_i$	0	0	1	1	1	1	1	1	0	2	0	2	2	3	4	2	3	4	4	5
$\tilde{y}_i$	0	0	1	1	1	1	1	1	4	2	0	2	3	3	0	5	3	4	4	2

Tabelle 6.2.: Beispieltabelle mit  $x_i$ ,  $y_i$  und  $\tilde{y}_i$

Mit den  $x_i$  (erste Zeile in Tabelle 6.2) sind die Bewohnerzahlen der Anschriften  $i$  im Melderegister bezeichnet. Die zugehörigen erhobenen Beobachtungen  $y_i$  werden in der zweiten Zeile aufsteigend sortiert. Schichtweise aufsummiert ergeben sich danach teilweise große Abweichungen zwischen Register und festgestellter Einwohnerzahl:

**Schicht 1:** 12 versus 8

**Schicht 2:** 13 versus 13

**Schicht 3:** 13 versus 16

Bei einer »fehleroptimierten« Sortierung  $\tilde{y}_i$  (Vergleich der ersten und dritten Zeile der Tabelle) sind dagegen keine Abweichungen zwischen Register und festgestellter Einwohnerzahl zu verzeichnen (‘12 versus 12’, ‘13 vs. 13’ und ‘13 vs. 13’).

Man kann also schon vor der Erhebung und Hochrechnung erahnen, dass später eine objektive Bewertung des Fehlermaßes der Residualvarianz unmöglich sein wird. Obwohl es bei der Residualvarianz natürlich auf die Sortierung der Residuen ankommt, und nicht auf die Sortierung der abhängigen Variablen wie im Beispiel. Klink und Lorentz (2022) schreiben hierzu:

»Die Aufteilung der bemeldeten Normalanschriften auf die Anschriftengrößenklassen erfolgte folgendermaßen: Die bemeldeten Normalanschriften wurden

---

gemäß ihrer Bemeldung aufsteigend angeordnet und dann so auf die Anstiftengrößenklassen aufgeteilt, dass diese annähernd die gleiche Gesamtbemeldung aufweisen.«

Die nicht-trennscharfe Definition der Schichten wird hier nicht thematisiert. Auch im Widerspruchsbescheid des Hessischen Statistischen Landesamtes (HSL) an die Stadt Hanau (Hessisches Statistisches Landesamt, 2025) finden wir keine befriedigende Antwort:

»Um die Genauigkeit der Stichprobenergebnisse innerhalb der Erhebungsgebiete (hier: Stadt Hanau) zu erhöhen, wurde eine Schichtung eingeführt, bei der in allen Erhebungsgebieten die Anstiften in verschiedene Größenklassen mit jeweils eigenem Auswahlatz bzw. Hochrechnungsfaktor eingeteilt wurden. Die Schichtzugehörigkeit einzelner Anstiften ergibt sich aus der Schicht-ID, die dem Datenblatt entnommen werden kann.«

Was die Eindeutigkeit der Zuordnung einer Anstift zu einer Schicht angeht, ist die Auffassung des HSL wie am Beispiel dargestellt nicht zutreffend. Auch was den Hochrechnungsfaktor angeht, ist diese Aussage schlicht falsch. Wie in Kapitel 6.3 diskutiert, werden die Hochrechnungsfaktoren an elf Hilfsgrößen (zu Alter, Geschlecht und Nationalität) und nicht an der Anstiftengröße festgemacht.

## 6.6. Hochrechnung der Über- und Untererfassungen

In Münnich und Gabler (2012) ist auf Seite 94 nachzulesen, dass die in der Zensusstichprobe festgestellten Über- und Untererfassungen durch das GREG-Modell hochgerechnet werden. Für deren Anzahl, grob gesprochen auch als Fehlbestände und Karteileichen bezeichnet, gelten folgende Beziehungen

$$\begin{aligned}\hat{\tau}_F &= \hat{\tau}_Z - \hat{\tau}_P \\ \hat{\tau}_K &= \tau_R - \hat{\tau}_P\end{aligned}$$

Dabei bezeichnet  $\hat{\tau}_F$  die Schätzung der Anzahl der Fehlbestände, die sich als Differenz aus dem geschätzten Totalwert  $\hat{\tau}_Z$  der Einwohnerzahl und der geschätzten Anzahl der paarigen Personen (gemeldete und an den Anstiften festgestellte Bewohner)  $\hat{\tau}_P$  ergibt. Der Schätzer  $\hat{\tau}_K$  für die Anzahl der Karteileichen ergibt sich als Differenz aus dem dublettenbereinigten Registerwert und der geschätzten Zahl der paarigen Personen  $\hat{\tau}_P$ .

Die Schätzer auf der linken Seite der Gleichungen sind keine eigenständigen Hochrechnungen. Wendet man eine mathematische Funktion auf einen Schätzer an, so ist das

---

Ergebnis auch ein Schätzer. Hochgerechnet werden streng genommen nur der Totalwert von  $Y$ , sprich  $\hat{\tau}_Z$ , und der Totalwert der paarigen Personen  $\hat{\tau}_P$ . Was Letzteren angeht, verstehen wir nicht, wie die Hochrechnung erfolgte.

Erinnern wir uns an das einführende Beispiel in Kapitel 6.1. Im Modell konnten wir  $x_S$  an  $X_G$  ausrichten, also die Kenntnis über den Totalwert im Register ausnutzen. Ein solches Regulativ steht uns für die paarigen Personen leider nicht zur Verfügung. Wir kennen zwar durch die Erhebung die Anzahl der paarigen Personen in der Stichprobe, etwa  $x_s^P = 10 = y_s^P$ , uns ist aber die Anzahl der paarigen Personen in der Gemeinde unbekannt. Fehleranfällig wäre die Hochrechnung der paarigen Personen via Multiplikation mit dem Designgewicht:  $\hat{\tau}_P = d \cdot x_s^P = 10 \cdot 12 = 120$ . Bretschi et al. (2024a, S. 25) schreiben: »Für die Schätzung der existenten und paarigen Personen wird – wie bereits im Zensus 2011 – ein verallgemeinerter Regressionsschätzer (GREG) genutzt.« Weitere Informationen werden dort nicht gegeben.

Wichtig ist der Punkt, dass für die Berechnung der paarigen Personen keine Registerinformationen vorliegen können. Daher ist es naheliegend, dass die gleichen Koeffizienten der Totalwertschätzung der Einwohnerzahl für die Schätzung der paarigen Personen verwendet wurden.<sup>7</sup> Das bedeutet, es wird der gleiche Mechanismus für die Zahl der festgestellten und die Zahl der paarigen Personen verwendet.<sup>8</sup>

Nehmen wir an, wir haben zwei benachbarte Gebäude derselben Anschriftengrößenklasse. Das erste Gebäude werde von Studierenden bewohnt, das zweite vornehmlich von älteren Menschen. Umzüge und damit Personenänderungen (auch unabhängig vom Melderegister) sind dann in Gebäude 1 höher als in Gebäude 2. Für die Schätzung der paarigen Personen wird nun der Mechanismus für die Gesamtzahl der Einwohner der Anschriftengrößenklasse verwendet.

Ein anderes Problem der Hochrechnung erwähnen Bretschi et al. (2024a, S. 13):

»In Einzelfällen werden dadurch gemeindespezifische Schätzer für die Anzahl an existenten und paarigen Personen erstellt, die eine oder beide der folgenden Plausibilitätsbedingungen nicht erfüllen:

- geschätzte Anzahl an paarigen Personen  $\leq$  Anzahl an gemeldeten Personen und
- geschätzte Anzahl an paarigen Personen  $\leq$  geschätzte Anzahl an existenten Personen.

---

<sup>7</sup> Das legen die Formeln 10, 12, 13 bei Lorentz und Zwick (2025) nahe.

<sup>8</sup> In Peters und Lorentz (2025) (Erscheinungsdatum 16.12.2025) wurden einige Details einer algorithmischen Anpassung dieser Größen für Gemeinden innerhalb von Gemeindeverbänden beschrieben, die zuvor nicht erwähnt wurden.

---

In diesen Fällen erfolgt eine algorithmische Anpassung der gemeindespezifischen Schätzer für die Anzahl an existenten und paarigen Personen, die sowohl die Konsistenz der gemeindespezifischen Schätzer zu den Schätzern auf Sampling-Point-Ebene als auch die Erfüllung der beiden aufgeführten Plausibilitätsbedingungen gewährleistet.«

Einmal ungeachtet der Frage, worin genau die »algorithmische Anpassung« besteht, stellt eine Erfüllung der beiden Ungleichungen natürlich eine Minimalanforderung an Plausibilität dar. Nach unserer Auffassung sollte eine Verletzung dieser Ungleichungen mit größerer Wahrscheinlichkeit nur bei kleinen Gemeinden auftreten, was eine systematische Verschiebung der Einwohner(zahlen) von kleinen zu größeren Gemeinden innerhalb eines SMP zur Folge hat.<sup>9</sup> Es ist daher von Interesse, wie viele SMP tatsächlich betroffen waren und wie viele Personen durch die nachträgliche Anwendung des Algorithmus bei der Hochrechnung den einzelnen Gemeinden verloren gingen oder zugeschrieben wurden. Dies wäre eine einfach zu erstellende, sinnvolle Qualitätsmetrik, die aber nicht veröffentlicht wurde. Ohne eine Quantifizierung ist dieser Effekt in der Wirkung auf die Einwohnerzahlen der Gemeinden nicht beurteilbar.

## 6.7. Fallbeispiel der Stadt Hanau

Um die Auswirkung der Schichtung auf die Fehlervarianz der Regressionsschätzung zu untersuchen, führen wir eine Simulation durch. Hierzu verwenden wir eine anonymisierte Einwohnermeldedatei der Stadt Hanau (AGS 6435014) mit Stand vom 31.10.2025. Daneben steht uns eine ebenfalls von der Stadt Hanau bereitgestellte Schichtungstabelle zum Zensus 2022 zur Verfügung.

Zur Nachahmung der Regressionsschätzung im Rahmen der Zensus-Hochrechnung werden die Daten zunächst gemäß der Schichtungstabelle in 16 Schichten mit entsprechender Anchriftenzahl partitioniert. Vor der Schichtung werden die Anchriften dazu nach ihrer Bemeldung aufsteigend sortiert, wobei Anchriften mit gleich großer Bemeldung bei

---

<sup>9</sup> In einem Widerspruchsbescheid teilt das Hessische Statistische Landesamt (Hessisches Statistisches Landesamt, 2025) weitere, bislang nicht dokumentierte Details mit: »Zur Aufdeckung vermeintlich unplausibler Anchriften wurden verschiedene Prüfkategorien definiert. Die Prüfkategorien lauteten: ›Keine paarigen Personen, aber sowohl Karteileichen als auch Fehlbestände‹, ›Keine paarigen Personen und nur Fehlbestände‹, ›Keine paarigen Personen und nur Karteileichen‹, ›Deutlich mehr existente Personen als laut MR zu erwarten‹, ›Deutlich weniger existente Personen als laut MR zu erwarten‹ bzw. ›Hoher Anteil nicht paariger Datensätze‹«. Wie mit diesen Kategorien verfahren wurde und wie häufig diese Ereignisse eintraten, ist nicht dokumentiert. Bedeutsam wären Untersuchungen zur Ursache lokaler Häufungen dieser Probleme.

---

der Sortierung in zufällige Reihenfolge gebracht werden. Allein bei der Darstellung der Nullschicht muss improvisiert werden, hier haben wir uns für eine einfache Zufallsstichprobe in passendem Umfang entschieden. Nicht nur die Verteilung der Anschriften auf die 17 Schichten, sondern auch die der Bewohner stimmt im Ergebnis weitgehend mit der Verteilung im Datenblatt überein.

### 6.7.1. Anwendung des Regressionsmodells

Zur Anwendung des Regressionsmodells wird zunächst die Stichprobe simuliert. Auf Registerseite ziehen wir je Schicht dieselbe Anzahl an Anschriften wie in der realisierten Stichprobe gemäß Schichtungstabelle. Die Zensusstichprobe bzw. die dazu passenden Erhebungsdaten stehen uns natürlich nicht zur Verfügung. Allerdings können wir aus der Schichtungstabelle die durchschnittliche Bemeldung je Schicht berechnen via Quotient aus den Spalten *Anzahl festgestellter Personen* und *Anzahl der Anschriften*. Diese Werte verwenden wir schichtweise als Parameter  $\lambda_h$  einer  $\text{Poisson}(\lambda_h)$ -Verteilung.

Für die Durchführung der Regression stehen uns damit zum einen je Anschrift  $k$  die Bewohnerzahlen aus dem Melderegister für die elf Hilfsgrößen (vgl. Kapitel 6.3) zur Verfügung:  $x_k = (x_{k,1}, x_{k,2}, \dots, x_{k,11})$ . Zum anderen liegt uns zu jeder Anschrift  $k$  aus der Zensusstichprobe eine simulierte Bewohnerzahl  $y_k$  vor. Wir haben uns aus inhaltlichen Gründen – aufgrund des Vorhandenseins einer Nullschicht – für eine Regression mit konstantem Term entschieden, da es uns für die Modellierung sinnvoller erschien als ein Verzicht auf denselben.<sup>10</sup>

Wir beginnen mit der Berechnung des Varianzschätzers  $\hat{V}(\hat{t})$  ohne Verwendung der Schichtung:

$$\hat{V}(\hat{t}) = N^2 \cdot \frac{1 - \frac{n}{N}}{n} \cdot \frac{1}{n-1} \sum_{k \in s} (v_k - \bar{v})^2. \quad (6.9)$$

Des Weiteren schätzen wir  $\hat{V}_H(\hat{t})$  unter Einbezug der Schichtung:

$$\hat{V}_H(\hat{t}) = \sum_{h=1}^H N_h^2 \cdot \frac{1 - \frac{n_h}{N_h}}{n_h} \cdot \frac{1}{n_h - 1} \sum_{k \in s_{d,h}} (v_k - \bar{v}_{d,h})^2. \quad (6.10)$$

Eine direkte Bestimmung der relativen Standardfehler gemäß Formel 6.5 ist hier nicht nötig. Uns interessiert allein das Verhältnis von  $rse(\hat{t})$  und  $rse_H(\hat{t})$ . In der Simulation

---

<sup>10</sup> Mit Blick auf die Publikationen zum Zensus – das Regressionsmodell bei Berg und Bihler (2014) enthält eine Konstante, das Modell bei Lorentz und Zwick (2025) hingegen nicht – und widersprüchliche Informationen des Auskunftsdienstes von Destatis konnten wir nicht abschließend klären, für welchen Weg (Modellierung mit oder ohne Konstante) man sich am Ende entschieden hat.

---

ergibt sich für den Quotienten

$$\frac{rse(\hat{t})}{rse_H(\hat{t})} = \sqrt{\frac{\hat{V}(\hat{t})}{\hat{V}_H(\hat{t})}} \approx \sqrt{6.36} \approx 2.52. \quad (6.11)$$

Die Fehlervarianz erhöht sich also bei einem Verzicht auf die Schichtung um mehr als das 6-fache und der relative Standardfehler etwa um das 2.5-fache. Da bei der Berechnung der Standardfehler in den Datenblättern Formel 6.10 anstelle von Formel 6.9 angewendet wurde, ist der dort berichtete Standardfehler zu klein, es sind die tatsächlichen Konfidenzintervalle breiter und das Ausmaß der Verfehlung der Präzisionsziele deutlich größer. Hier kommt die auf Seite 44 erwähnte implizite Unterstellung von Unabhängigkeitsannahmen bei der Herleitung von Formel 6.8 zum Tragen.

### 6.7.2. Lineare Abhängigkeit der Prädiktoren (Multikollinearität)

Lässt sich ein Vektor als Vielfaches eines anderen darstellen, werden die beiden Vektoren in der linearen Algebra als *kollinear* bezeichnet. Bei einer größeren Menge von Vektoren ist von *linearer Abhängigkeit* die Rede. In der Statistik spricht man dagegen bereits dann von *Multikollinearität* zwischen den erklärenden Merkmalen einer Regression, wenn eine solche – meist ungewollte – Beziehung zumindest näherungsweise besteht.

In der Folge nennen wir zwecks besserer Lesbarkeit die erklärenden Merkmale *Prädiktoren*. Im Falle dreier Merkmale bedeutet Multikollinearität

$$X_3 \approx \lambda_0 + \lambda_1 X_1 + \lambda_2 X_2,$$

mit geeigneten reellen Koeffizienten  $\lambda_i$ . *Näherungsweise*, in der Gleichung mit  $\approx$  symbolisiert, ist natürlich ein relativer Begriff. Als Maß zur Bestimmung der Multikollinearität und damit zur Zuverlässigkeit von Regressionsmodellen verwenden Statistiker sogenannte *Varianzinflationsfaktoren*  $VIF_i = 1/(1 - R_i^2)$ . Sie beschreiben, wie gut ein Prädiktor  $X_i$  durch die anderen Prädiktoren  $X_j$  ( $j \neq i$ ) vorhergesagt werden kann. Hierzu führt man eine multiple lineare Regression innerhalb der Menge der Prädiktoren durch, mit einem ausgewählten  $X_i$  als Zielgröße, erklärt durch alle anderen  $X_j$  ( $j \neq i$ ), und erhält das Bestimmtheitsmaß  $R_i^2$ . Liegt das Bestimmtheitsmaß  $R_i^2$  nahe bei 1, so wird der Faktor  $VIF_i$  sehr groß.

In der statistischen Gemeinschaft wird ein VIF von 5 oder größer als kritisch und ein VIF von 10 oder größer als ernst zu nehmender Indikator für Multikollinearität eingestuft. Diese Faustregel gibt einen Hinweis darauf, dass die Varianzen der Koeffizienten  $\beta_i$  deutlich größer sind als bei Unabhängigkeit zu erwarten wäre. Beträgt der  $VIF_i$  zum Beispiel 5,

so ist die *Varianz* von  $\beta_i$  *fünfmal größer* als in der Situation ohne Multikollinearität. Damit wird schnell klar, warum der Varianzinflationsfaktor seinen Namen trägt. Es wird insgesamt schwierig bis unmöglich, den Einfluss der einzelnen  $X_i$  auf die Zielvariable  $Y$  zu bestimmen, da die Prädiktoren ähnliche Informationen enthalten bzw. einzelne Koeffizienten an Aussagekraft verlieren und in den anderen aufgehen. Große Varianzen bei an sich sinnvollen Prädiktoren sind daher ein typischer Indikator für Multikollinearität.

Hohe Multikollinearität erhöht also die Varianz des Totalwertschätzers, wobei der Effekt der Multikollinearität auf gebundene Hochrechnungen in der Literatur bekannt ist.<sup>11</sup>

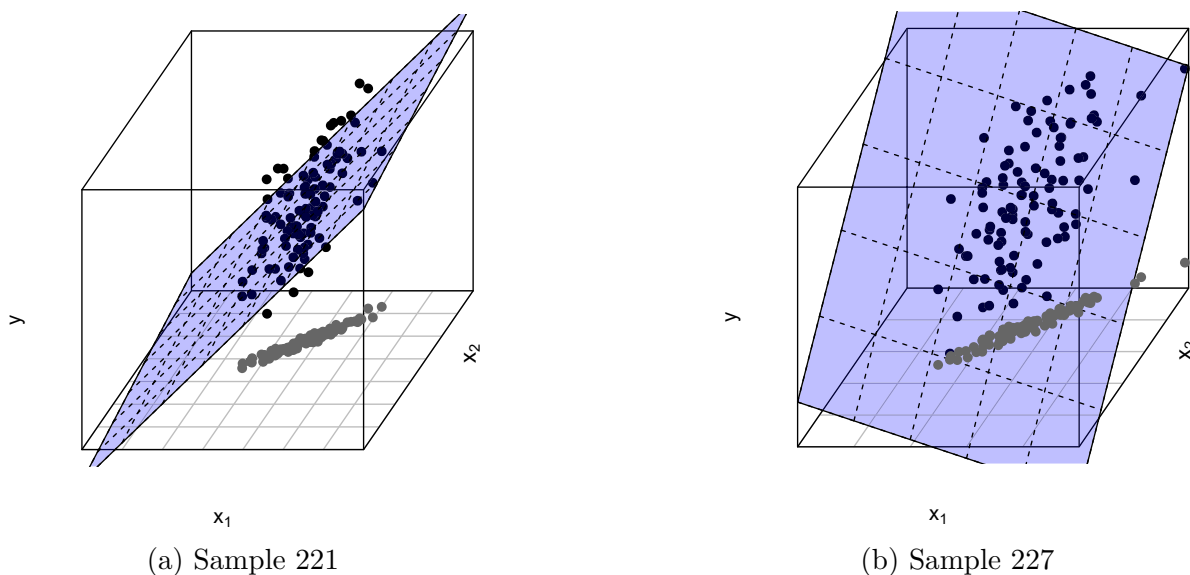


Abbildung 6.1.: Beispiel für instabile Schätzungen durch Multikollinearität. Zwei Stichproben aus derselben Grundgesamtheit mit hoch korrelierten Prädiktoren.

Schon eine oder wenige Beobachtungen können eine große Veränderung der Regressionsfunktion bewirken. Dies lässt sich graphisch in drei Dimensionen am Beispiel zweier

<sup>11</sup> So schreiben Knobelspies und Münnich (2008, S. 344) in einer Arbeit über Variablenselektion bei gebundener Hochrechnung: »Man erkennt, dass entgegen der Erwartungen und der zuvor aufgezeigten Ergebnisse die sehr umfangreichen Modelle nicht notwendigerweise am besten abschneiden. Vielmehr neigen diese in Einzelfällen zu sehr schlechten oder gar irrationalen Schätzungen«. Etwas später im Text findet sich: »Neben der eigentlichen Modellierung und Schätzung kann ebenso das Argument einer stabilen Varianzschätzung eine Rolle spielen. Die enormen Ausreißer bei den Schätzverteilungen wird man sicher in der Praxis erkennen und daraufhin ein besser geeignetes Regressionsmodell auswählen«. Es gibt keine Hinweise darauf, dass Destatis für jede Gemeinde eine Regressionsdiagnostik durchgeführt hat. Ausreißer blieben daher unbemerkt.

---

Merkmale  $X_1$  und  $X_2$  darstellen. Bei hoch korrelierten Merkmalen werden die Schätzungen instabil. Die Abbildungen 6.1a und 6.1b zeigen das *Schwanken* der Hyperebene bei starker Korrelation der Prädiktoren.<sup>12</sup> Die Neigung der Hyperebene verändert sich bei nahezu gleichbleibenden Residuen drastisch, obwohl es sich um Stichproben aus derselben Grundgesamtheit und dasselbe Modell handelt. Bei identischen Ausprägungen der Prädiktoren kann es einmal zu einer hohen und einmal zu einer niedrigen Vorhersage kommen. Multikollinearität ist also kritisch für die Vorhersage.

Im vorliegenden Falle der Registerdaten der Stadt Hanau beobachten wir in der Tat eine sehr hohe Multikollinearität zwischen den elf Prädiktoren im linearen Regressionsmodell. Tabelle 6.3 enthält die empirischen VIF-Werte für die elf Prädiktoren. Mit Blick auf die in der Tabelle absteigend sortierten VIF-Werte überrascht es wenig, dass das Merkmal der anschriftbezogenen Personenzahl den größten Wert von 46.67 zeigt. Im Mittel über alle elf Merkmale beobachten wir immerhin noch einen Wert von 9.02. Dies deutet darauf hin, dass die Varianzen der Koeffizientenschätzer  $\beta_i$  weit größer ausfallen als es ohne Multikollinearität der Fall wäre, die Koeffizienten also an Aussagekraft verlieren. Der im Datenblatt ausgewiesene realisierte Standardfehler ist daher zu klein.

Merkmal $i$	VIF <sub><math>i</math></sub>
Personen gesamt	46.67
Geschlecht	23.77
Nationalität	5.45
Alter von 30 bis 39	4.59
Alter von 40 bis 49	3.42
Alter von 6 bis 17	3.18
Alter kleiner als 6	2.71
Alter von 18 bis 24	2.61
Alter von 50 bis 59	2.58
Alter von 25 bis 29	2.53
Alter von 60 bis 64	1.68
<b>Arithm. Mittel</b>	9.02

Tabelle 6.3.: Varianzinflationsfaktoren zum Melderegister der Stadt Hanau

Die Multikollinearität hat zudem eine gewisse Hebelwirkung zur Folge, indem Ausreißer die Regressionsfunktion stärker anziehen, was zu einer veränderten Hochrechnung bei nahezu gleichbleibender Fehlervarianz führt. Damit einhergehend sind Vorhersagen an abseits gelegenen Punkten besonders fehleranfällig:

---

<sup>12</sup> Die Idee zu dieser Abbildung verdanken wir einem Vortrag von Johnson (2014).

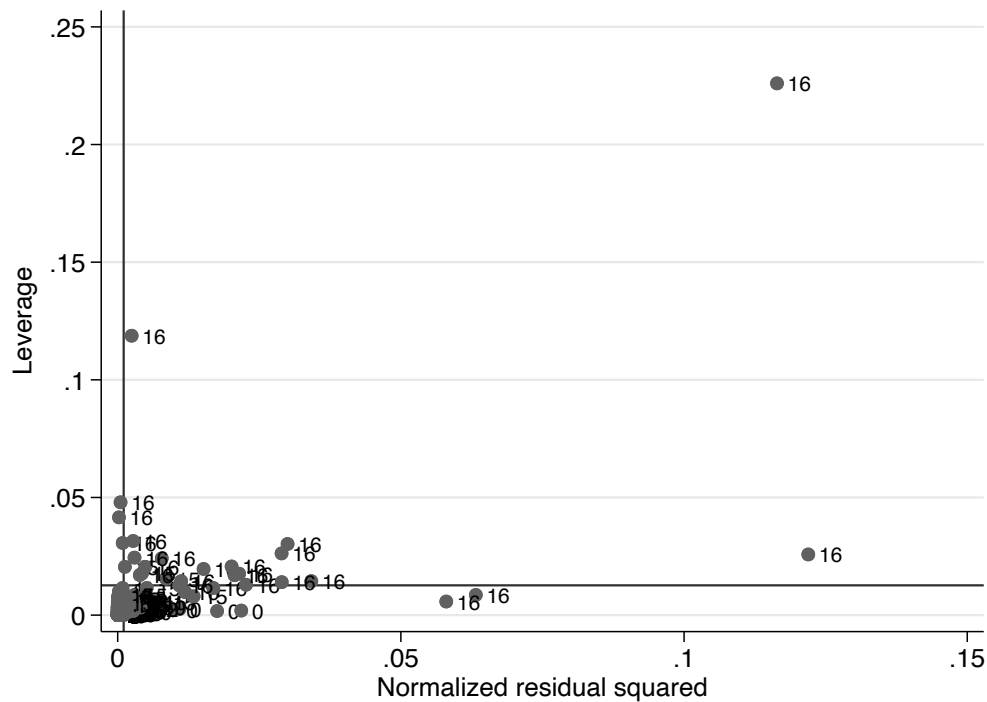


Abbildung 6.2.: Regressionsdiagnostik für das Hochrechnungsmodell anhand einer Stichprobe aus der Einwohnermeldedatei Hanau. Schichtung und Auswahlatz entsprechen dem Datenblatt der Gemeinde. Die Zahlen im Plot geben die Schichtnummern der Beobachtungen an.

»Conversely, predictions made at outlying points have large errors. This is not of much consequence, as it is unwise to make a prediction at an outlying point. We cannot be sure that the model holds at points remote from the observed data.« Seber und Lee (2003, S. 261)

Anhand von Abbildung 6.2 erkennt man gut, dass die großen Residuen (Abszisse) und vor allem alle Beobachtungen mit großem Einfluss auf die Schätzung (Leverage, Ordinate) aus der oberen Schicht Nummer 16 stammen. Einige davon liegen nahe der  $y$ -Achse, was ein kleines Residuum bedeutet. Die Kombination aus kleinem Residuum und großer Hebelwirkung lässt sich geometrisch so deuten, dass sich die Hyperebene der Regressionsfunktion eng an die Beobachtung schmiegt, die Beobachtung also die Neigung der Ebene beeinflusst.

---

## 6.8. Erläuterung der Konsequenzen für die Schätzung

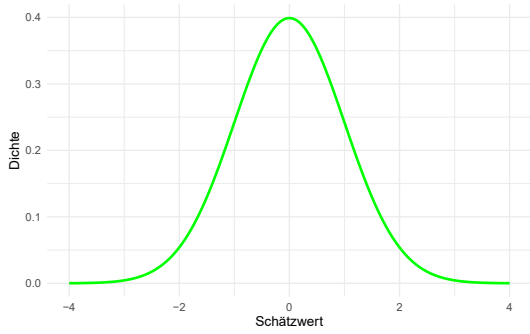
Selbst die vom Statistischen Bundesamt angegebenen realisierten Standardfehler sind aufgrund der erörterten Probleme konservative Schätzungen. In den Datenblättern wird den Gemeinden mitgeteilt, dass mit einer Sicherheit von 95 % die tatsächliche Einwohnerzahl in einem Intervall läge. Die Grundlage für die Berechnung dieses Intervalls ist die diskutierte Varianzschätzung. Da die Varianzen unterschätzt werden, sind diese Intervalle empirisch mit Sicherheit größer als die ausgewiesenen Intervalle. Es muss betont werden, dass diese Intervalle durch systematische Fehler, die das Statistische Bundesamt nicht untersucht hat (siehe Kapitel 8) nochmals deutlich vergrößert werden.<sup>13</sup>

Um den Effekt von systematischen Fehlern und vergrößerten Standardfehlern zu veranschaulichen, zeigt die Abbildung 6.3 die vier möglichen Kombinationen von systematischen Fehlern (Bias) und vergrößerten Standardfehlern. In 6.3a ist die angestrebte Verteilung von Schätzungen dargestellt, was bedeutet: kleine Streuung und mittlerer Fehler von Null. In Abbildung 6.3b ist zusätzlich zur angestrebten Verteilung die Verteilung von Schätzungen mit einem größeren Standardfehler, die aber keinen systematischen Fehler besitzen, enthalten. Die Streuung der Schätzungen ist größer, aber der Mittelwert der Fehler ist immer noch Null. Allerdings werden stark abweichende Ergebnisse häufiger vorkommen als bei der angestrebten Verteilung. Die Abbildung 6.3c zeigt zusätzlich zur angestrebten Verteilung die Verteilung von Schätzungen mit einem systematischen Fehler, in diesem Beispiel mit negativem Vorzeichen. Die Streuung der Schätzungen ist unverändert, aber der Mittelwert der Fehler nicht mehr Null. Daher werden die Fehler häufiger unterschätzt als bei der angestrebten Verteilung. In Abbildung 6.3d beobachten wir zusätzlich zur angestrebten Verteilung die Verteilung von Schätzungen mit einem systematischen Fehler mit negativem Vorzeichen und einem größeren Standardfehler. Diese Kombination führt zu einem besonders häufigen Vorkommen starker Unterschätzungen.

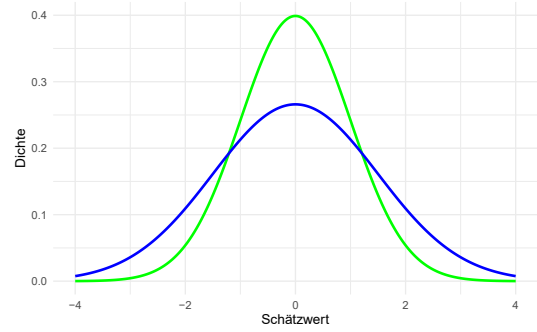
Die Planungen des Zensus gingen von einer Verteilung wie 6.3a aus. Wie wir im nachfolgenden Kapitel sehen werden, speziell mit Blick auf Tabelle 7.1, waren die Standardfehler beim Zensus 2022 deutlich größer als geplant. Da das Statistische Bundesamt nicht von systematischen Fehlern ausgeht, würde die Verteilung (Symmetrie unterstellt) der Abbildung 6.3b entsprechen. Eine entsprechende Verteilung der geschätzten Standardfehler («realisierte Standardfehler») wurde vom Statistischen Bundesamt nicht veröffentlicht. Wir gehen aber davon aus, dass die Standardfehler deutlich größer waren als die Veröffent-

---

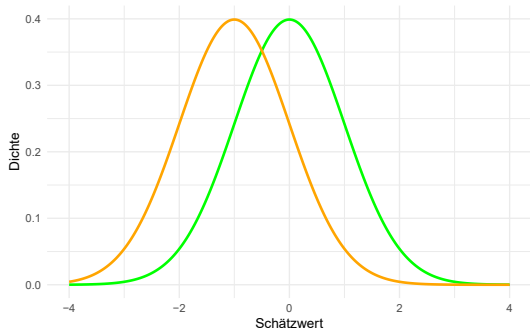
<sup>13</sup> Dass das Hochrechnungsverfahren mindestens in Einzelfällen zu problematischen Ergebnissen führte, bestätigt das Hessische Statistische Landesamt (Hessisches Statistisches Landesamt, 2025) mit dem Hinweis, dass im Rahmen der »Ergebniserstellung . . . bei einzelnen demografischen Daten Auffälligkeiten festgestellt wurden. Im Zuge dessen wurde die Hochrechnung nachjustiert, um die bestmögliche Ergebnisqualität zu erreichen.«



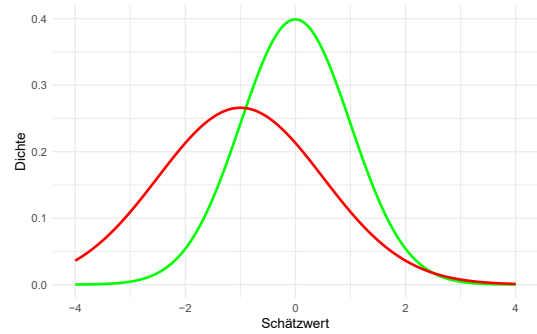
(a) Unverzerrte Schätzungen



(b) Schätzungen mit größerem Standardfehler



(c) Schätzungen mit systematischem Fehler



(d) Schätzungen mit systematischem Fehler und größerem Standardfehler

Abbildung 6.3.: Veranschaulichung: Der Einfluss größerer Standardfehler und eines systematischen Fehlers auf Schätzungen

---

lichung Lorentz und Zwick (2025) erkennen lässt. Weiterhin gehen wir von zusätzlichen systematischen Fehlern aus. Damit könnte die tatsächliche Verteilung der Schätzungen eher der Abbildung 6.3d entsprechen.<sup>14</sup> In der Gesamtbetrachtung hätten wir es also mit einer verzerrten Schätzung mit unterschätzter Varianz zu tun.

## 6.9. Zusammenfassung

Die Qualität der Schätzung der Einwohnerzahl wird in den Veröffentlichungen des Statistischen Bundesamtes allein an dem »tatsächlich realisierten Standardfehler« gemessen. In den Publikationen zum Zensus wird dieser in der Literatur unübliche Begriff nicht erläutert.

Ein für den SMP festgelegtes Hochrechnungsmodell wird unverändert auf alle beteiligten Gemeinden angewendet. Insbesondere in Kleinstgemeinden sind häufig, wenn überhaupt, nur wenige Prädiktoren besetzt. Die im Modell unberücksichtigte, aufgrund von Multikollinearität teils erhebliche, Variabilität der Koeffizienten  $\hat{\beta}_i$  kann dazu führen, dass Einwohnerzahlen deutlich unter- oder überschätzt werden. Ursächlich ist, dass die für die Hochrechnung verwendeten Regressionskoeffizienten  $\hat{\beta}_i$  für den SMP insgesamt geschätzt werden und daher für kleine Gemeinden unangemessene Ergebnisse möglich sind.<sup>15</sup>

Obwohl das zugrundeliegende Regressionsmodell auf den gewichteten Beobachtungen gerechnet wurde und damit auch in den Koeffizientenvektor  $\hat{\beta}$  implizit die Anschriftengröße eingeht, handelt es sich bei Formel 6.1 nicht um eine geschichtete Hochrechnung. Daher war die Schichtung in Formel 6.8 nach unserer Auffassung unzulässig. Mit der Folge, dass eine deutlich kleinere Residualvarianz und damit auch ein kleinerer relativer Standardfehler ausgewiesen wurde. Die Beispielrechnung in Kapitel 6.7 zeigt die erhebliche Größenordnung der Korrektur dieses Fehlermaßes bei Verzicht auf die Schichtung.

Die der Schichtung zugrundeliegende Homogenitätsannahme ist besonders in der oberen, nach oben offenen Schicht nicht erfüllt. Dies wird exemplarisch an der Gemeinde Mainaschaff mit rund 1 100 Einwohnern deutlich, deren Gemeindegebiet durch eine Vielzahl kleiner sowie wenige sehr große Anschriften geprägt ist. Im Rahmen der Regressions-schätzung entfalten die großen Anschriften eine erhebliche Hebelwirkung und

---

14 Ohne entsprechende Untersuchungen wird die Abwesenheit systematischer Fehler unterstellt. Zu den empirischen Hinweisen auf das Vorliegen systematischer Fehler siehe das Kapitel 8.

15 Der Totalwertschätzer des Hochrechnungsmodells wird damit zwar dem Plausibilitätsanspruch gerecht, dass die Hochrechnungsergebnisse für Teilgebiete konsistent mit dem Ergebnis für das Gesamtgebiet sind (siehe hierzu auch Särndal et al. (1992, S. 400)). Dies geht aber zu Lasten der Schätzgenauigkeit in den kleinen Gemeinden, zumal der Standardfehler nur auf SMP-Ebene ausgewiesen wird.

---

beeinflussen die Regressionsfunktion in nicht unerheblichem Umfang. Eine Über- oder Untererfassung dieser großen Anschriften führt infolgedessen dazu, dass im Zuge der Hochrechnung auch die Beobachtungen in den unteren Anschriftengrößenklassen überproportional nach unten bzw. oben korrigiert werden. Hochrechnung und Fehlerrechnung können also aus folgenden Gründen hinterfragt werden:

1. Die Annahmen des der Hochrechnung zugrundeliegenden Regressionsmodells werden nicht formuliert. Daher sind die Residuen  $e_k$  und daraus abgeleitete Größen wie  $\hat{V}(\hat{t})$  und die von Destatis in den Datenblättern der Gemeinden angegebenen Grenzen der Konfidenzintervalle wenig aussagekräftig.
2. Die Partitionierung nach Anschriftengrößenklassen wird nicht trennscharf vorgenommen, was den Gepflogenheiten in der Statistik widerspricht.
3. Die Größe  $\hat{V}(\hat{t})$  wird im Gegensatz zum Totalwertschätzer  $\hat{t}$  nach Anschriftengrößenklassen geschichtet. Durch die damit verbundene Gruppierung der Residuen wird die Varianz deutlich unterschätzt.
4. Die Größe  $\hat{V}(\hat{t})$  wird nicht dem komplexen Design des Regressionsmodells gerecht. Zum einen mit Blick auf den Verlust an Freiheitsgraden, zum anderen mit Blick auf die vernachlässigten Varianzen der geschätzten Modellkoeffizienten  $\hat{\beta}_i$ .
5. In den Mitteilungen an die Gemeinden wird basierend auf  $\hat{V}(\hat{t})$  und  $\hat{t}$  ein modellbasierter Schätzer als realisierter Standardfehler dargestellt. Dieser wird nicht für die Gemeinden separat, sondern zum Nachteil vieler Gemeinden allein für den übergeordneten SMP geschätzt.
6. Die den Gemeinden mitgeteilten 95 %-Konfidenzintervalle auf SMP-Ebene um den Totalwertschätzer der Form  $\hat{t} \pm 2\sqrt{\hat{V}(\hat{t})}$  stiften eher Verwirrung denn einen Nutzen.
7. Es mangelt an einer Modelldiagnostik. Die Abhängigkeit unter den Prädiktoren des Hochrechnungsmodells hat negative Auswirkungen auf die Zuverlässigkeit der Hochrechnung samt Fehlerrechnung. Notwendige Maßnahmen zur Abschwächung der Multikollinearität werden nicht diskutiert.
8. Das Hochrechnungsmodell ist insgesamt unausgewogen in Hinsicht auf unterschiedliche Gemeindegrößen innerhalb eines Verbundes, da bei kleineren Gemeinden teilweise nur wenige Anschriften in die Stichprobe gelangen.
9. Die Qualität der Hochrechnungsergebnisse wird ausschließlich an den – wie dargestellt deutlich unterschätzten – relativen Standardfehlern gemessen, wohingegen Nicht-Stichprobenfehler (siehe Kapitel 8) bei der Fehlermessung keine Rolle spielen.

## 7. Zum Verfehlen der Präzisionsziele der Hochrechnung

Die in Kapitel 4 besprochene Präzisionszielfunktion sollte dazu dienen, die Stichprobengröße so zu wählen, dass die Einwohnerzahlen mit vorgegebener Genauigkeit geschätzt werden konnten. Wir möchten in Erinnerung rufen, dass nach der aufwendigen Bestimmung der Stichprobenumfänge (siehe Kapitel 5.3) diese noch einmal um 40 % erhöht wurden.<sup>1</sup> Mit dieser Stichprobengröße sollte sich also die vorgegebene Genauigkeit erreichen lassen. Operational bedeutet das üblicherweise, dass maximal 5 % der Schätzungen außerhalb der Zufallsschwankungen liegen sollten.

Am 16.04.2025 erschien mit Lorentz und Zwick (2025) ein Artikel in der Zeitschrift WISTA (Wirtschaft und Statistik) des Statistischen Bundesamtes, der unter anderem die berechneten relativen Standardfehler für den Zensus 2022 in einer Tabelle zusammenfasste (siehe Abbildung 7.1).

Es muss beachtet werden, dass es sich hier nicht um empirische Werte handelt, sondern um die in Kapitel 6.4 bereits ausführlich diskutierten sogenannten »tatsächlich realisierten Standardfehler«.

---

<sup>1</sup> So schreiben Klink und Lorentz (2022): »Aufgrund der Unsicherheit in den Varianzreduktionsfaktoren, die für die Ex-ante-Präzisionsabschätzung eine wichtige Einflussgröße darstellen, wurde der ermittelte Gesamtstichprobenumfang für die Gesamtheit der bemeldeten Normalanschriften um 40 % erhöht. Dieser Aufschlagfaktor war eine wissenschaftliche Empfehlung der Universität Trier aufgrund von durchgeführten Simulationsrechnungen.«

**Tabelle 2**

Ermittlung der Einwohnerzahl im Zensus 2022: Gemeinden/Stadtteile mit Standardfehler in den angegebenen Grenzen nach Gemeindegrößenklassen

	Präzisionsziel eingehalten		Standardfehler zwischen dem 1- und einschließlich dem 1,2-Fachen des Präzisionsziels		Standardfehler zwischen dem 1,2- und einschließlich dem 2-Fachen des Präzisionsziels		Standardfehler mit mehr als dem Doppelten des Präzisionsziels		Insgesamt	
	Anzahl	%	Anzahl	%	Anzahl	%	Anzahl	%	Anzahl	%
Gemeinden mit ... weniger als 10 000 Einwohnerinnen und Einwohnern	1 325	28,5	518	11,1	1 832	39,4	980	21,1	4 655	100
10 000 bis 50 000 Einwohnerinnen und Einwohnern	95	5,9	135	8,4	948	59,0	428	26,7	1 606	100
50 000 bis 100 000 Einwohnerinnen und Einwohnern	31	27,7	31	27,7	47	42,0	3	2,7	112	100
mehr als 100 000 Einwohnerinnen und Einwohnern	96	76,2	19	15,1	11	8,7	0	0,0	126	100
Insgesamt	1 547	23,8	703	10,8	2 838	43,7	1 411	21,7	6 499	100

Abbildung 7.1.: Tabelle 2 aus Lorentz und Zwick, 2025

Das wichtigste Ergebnis der Tabelle ist die Tatsache, dass die Präzisionsziele nur in 23.8 % der SMP (nicht Gemeinden!) eingehalten, also in beachtlichen 76.2 % der SMP verfehlt wurden. In 21.7 % der SMP sogar um den Faktor 2. Es wurde hierzu keine Liste der Standardfehler aller Gemeinden veröffentlicht. Eine solche Liste würde detailliertere Analysen erlauben.

Bemerkenswert ist z.B., dass für Rheinland-Pfalz die Präzisionsziele in 91.7 % der Gebietskörperschaften nicht erreicht wurden.<sup>2</sup> Ein relativer Standardfehler über 2 % ist bei den uns vorliegenden Datenblättern keine Seltenheit. Auch die maximalen Abweichungen sind von besonderem Interesse. In den uns vorliegenden Datenblättern zeigt die Gemeinde Albersdorf in Schleswig-Holstein bei einem angestrebten relativen Standardfehler von 0.59 % einen tatsächlich realisierten Standardfehler von 2.34 %, also knapp das Vierfache.

Warum das Präzisionsziel in der Mehrheit der Gemeinden nicht erreicht wurde, lässt sich anhand einer einzigen Tabelle ohnehin kaum beantworten. Die Tabelle weist aber weitere Schwächen auf.

<sup>2</sup> <https://www.statistik.rlp.de/nachrichten/nachrichtendetailseite/zensus-2022-wesentliche-ergebnisse-der-zaehlung-liegen-vor>, abgerufen am 18.11.2025. Dort findet sich: »So wurde in 156 von 170 Gebietskörperschaften (kreisfreie Städte, verbandsfreie Gemeinden und Verbandsgemeinden) die im Zensusgesetz vorgesehene Güte der festzustellenden amtlichen Einwohnerzahl nicht erreicht.«

---

## 7.1. Kritische Diskussion der Tabellierung

An Tabelle 7.1 sind mehrere Dinge auffällig, vor allem aber die viel zu grobe Gliederung, die eine inhaltliche Analyse aus sechs Gründen nicht zulässt:

1. Es sind keine Aussagen für sehr kleine Gemeinden (unter 100 Anschriften) möglich, obwohl diese aufgrund ihrer uneinheitlichen Sonderbehandlung (siehe Kapitel 5.6) von praktischem Interesse sind.
2. Es sind keine Aussagen für Gemeinden um 1 000 Einwohner möglich, obwohl sich ab einer Größe von 1 000 Einwohnern die Präzisionsziele und damit auch die Stichprobenauswahlsätze ändern (vgl. Kapitel 4).
3. Es sind nur mit Beschränkung Aussagen für Gemeinden um 2 000 Einwohner möglich, da bei der Option *GV-Rest* Gemeinden mit wenigstens 2 000 Einwohnern – im Gegensatz zur Option *GV* – separat behandelt wurden.
4. Es sind keine Aussagen für Gemeinden um 10 000 Einwohner möglich, obwohl sich auch hier Präzisionsziele und Stichprobenauswahlsätze ändern.
5. Das für viele Großstädte relevante Intervall liegt zwischen 90 000 und 110 000 Einwohnern. Von dieser Einordnung hängt die Einordnung zur Großstadt und damit die Höhe der finanziellen Zuwendungen in besonderem Maße ab.
6. Da Stadtteile mit mehr als 200 000 Einwohnern innerhalb einer Gemeinde mit mehr als 400 000 Einwohnern einen eigenen SMP bilden, wird das Ausmaß der Erreichung der Präzisionsziele zwischen Gemeinden mit Stadtteilen oberhalb 200 000 Einwohnern von Gemeinden ohne diese Einteilung abweichen.

Zusammenfassend ist diese Tabelle für statistische Analysen von geringem Wert. Die im Zensusgesetz festgelegten Spezifika (länderspezifische Regelungen zur Zusammenfassung von Gemeinden, zensuspezifische Stadtteile bei Gemeinden über 400 000 Einwohnern), die bei der Stichprobenziehung und Hochrechnung eingeführten Grenzen (100, 1 000, 2 000 und 10 000 Einwohner) und weitere relevante Grenzen (z.B. das Intervall um 100 000 Einwohner) lassen sich in ihrem Effekt auf die Genauigkeit mit den in Tabelle 7.1 willkürlich gewählten Intervallgrenzen nicht analysieren. In allen genannten Intervallen müssen zudem die kritischen Punkte in der Mitte der Intervalle liegen und nicht an den Intervallgrenzen.

Wünschenswert wäre eine Liste aller berechneten und angestrebten Standardfehler für alle Gemeinden anstelle der Tabelle. Es gibt keinen erkennbaren Grund, warum eine solche Liste nicht veröffentlicht wurde.

---

## 7.2. Zusammenfassung

Die Präzisionsziele des Zensus 2022 waren gegenüber denen des Zensus 2011 bereits gelockert. In 2011 bestand das Ziel in der Einhaltung der Obergrenze von 0.5 % für den relativen Standardfehler. In 2022 wurde diese Genauigkeit nur noch für Gemeinden mit wenigstens 10 000 Einwohnern gefordert.

Trotz dieser Lockerung des Präzisionsziels und einer sehr einseitigen Fehlerrechnung – der relative Standardfehler wurde in den Berechnungen stark unterschätzt (siehe unsere Ausführungen in Kapitel 6) und Nicht-Stichprobenfehler gar nicht erst berücksichtigt (siehe nachfolgendes Kapitel 8) – war das Ausmaß der verfehlten Präzisionsziele beim Zensus 2022 weit höher als beim Zensus 2011.

Im Zensus 2011 wiesen laut Statistisches Bundesamt (2016) knapp 63 Prozent der Gemeinden einen relativen Standardfehler größer als 0.5 % auf. Im Zensus 2022 wurden bei gut 76 Prozent der Gemeinden die Präzisionsziele verfehlt, in 21.7 % der Gemeinden sogar um den Faktor 2. Ein Präzisionsziel, das in der Mehrheit der Gemeinden und zum Teil erheblich verfehlt wurde, deutet auf Fehler im Hochrechnungsverfahren, der Methodik allgemein oder in der Durchführung hin. Alle diese Erklärungen sind für die Abweichungen der beobachteten Einwohnerzahlen von der Fortschreibung relevant.

## 8. Nicht-Stichprobenfehler

In einem klassischen Lehrbuch für die amtliche Statistik schreibt das Statistische Bundesamt (1960, S.49):

»Grundsätzlich sind zwei Arten von Fehlern zu unterscheiden (...):  
*Zufallsfehler* sind Abweichungen, die darauf zurückzuführen sind, dass nicht alle Einheiten der untersuchten Gesamtheit, sondern nur die nach dem Zufallsprinzip ausgewählten Einheiten aus der Gesamtheit tatsächlich für die Statistik herangezogen werden.

*Systematische Fehler* sind alle übrigen Abweichungen, die auf falschen Angaben der Befragten oder der Interviewer, Fehlern bei der Abgrenzung der Gesamtheit und bei der Zuordnung von Einheiten zu bestimmten Teilen der Gesamtheit, Ausfall von zu befragenden Einheiten sowie auf Fehlern bei der Erhebung und bei der Aufbereitung beruhen (...).«

In der neueren Fachliteratur werden diese beiden Fehlerarten als »sampling errors« und »non-sampling errors« bezeichnet. Die Notwendigkeit, bei der Beurteilung der Ergebnisse immer beide Fehlerquellen zu berücksichtigen, hat unter der Bezeichnung »Total Survey Error« oder TSE-Modell eine umfangreiche wissenschaftliche Literatur hervorgebracht.<sup>1</sup> Unter dem Begriff »Non-Sampling-Errors« werden in der Literatur zahlreiche verschiedene Fehlerarten zusammengefasst (siehe Abbildung 8.1).

Die Abbildung zeigt die große Heterogenität der Fehlerquellen, von der Messung des falschen Konstrukts bis zu Fehlern in den Datentransformationsanweisungen an die Auswertungsprogramme. Entsprechend schwierig ist die Quantifizierung jeder einzelnen Fehlerquelle.

---

<sup>1</sup> Standardwerke sind z.B. Biemer und Lyberg (2003), Groves (2004), Lessler und Kalsbeek (1992) und Weisberg (2005). Neuere Arbeiten finden sich über Biemer et al. (2017).

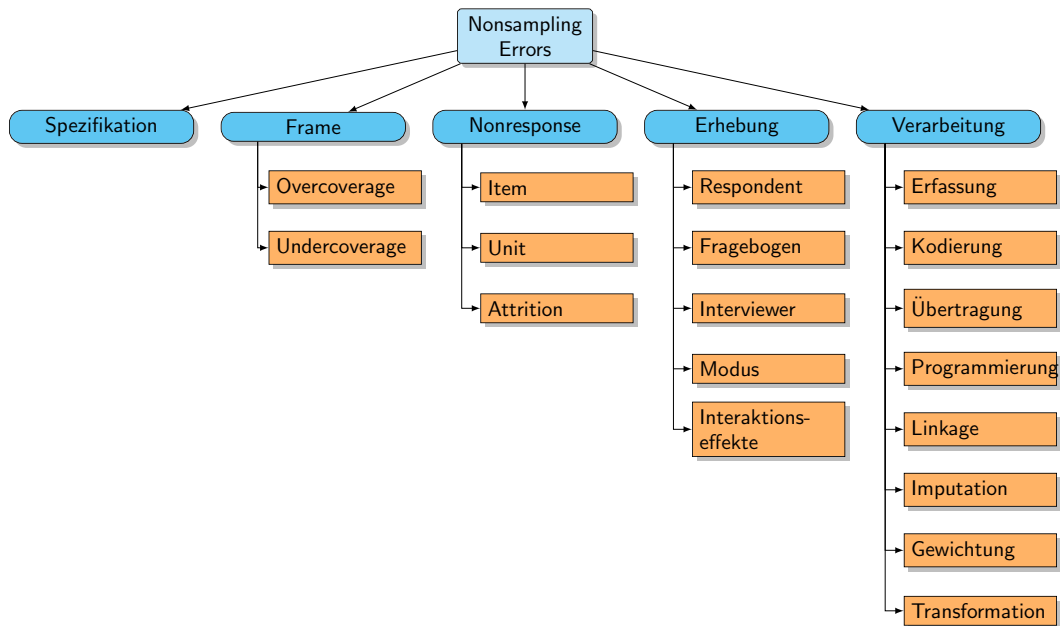


Abbildung 8.1.: Formen von Nicht-Stichprobenfehlern (Schnell, 2025; Tanur, 2011)

In dem Lehrbuch des Statistischen Bundesamtes heißt es wenige Zeilen später:

»Die Erfahrung zeigt, dass gerade systematische Fehler oft erheblich unterschätzt werden (...). So wird im Zusammenhang mit Stichprobenverfahren nicht selten von deren Ungenauigkeit gesprochen, wobei nur an die Zufallsfehler gedacht wird. Darüber werden die systematischen Fehler, die bei Totalstatistiken (infolge der Erfassung und Bearbeitung einer technisch nur schwer zu handhabenden großen Masse von Unterlagen) meist größer als bei Stichproben sind, völlig vergessen.«

Und schließlich (S.51):

»Die systematischen Fehler sind im allgemeinen mindestens von der gleichen Bedeutung wie die zufälligen Stichprobenfehler. Sie lassen sich quantitativ nur durch Kontrollerhebungen feststellen (...).« Mit einzelnen Ausnahmen sei »(...) nur wenig über die Höhe der systematischen Fehler bekannt. Hier größere Klarheit zu gewinnen, ist eine der wichtigsten künftigen Aufgaben.«

In Hinsicht auf den Zensus hat das Statistische Bundesamt zu keinem der Non-Sampling-Fehler eine Quantifizierung veröffentlicht (siehe Seite 7).

---

Dies ist vor dem Hintergrund des vom Statistischen Bundesamt im Qualitätsbericht des Zensus 2022 (Statistisches Bundesamt, 2024) zitierten Gutachtens von Helmut Küchenhoff (Küchenhoff, 2014) zu bewerten. Küchenhoff hat in seinem Gutachten zum Zensus 2011 auf einige zentrale Aspekte von Nonsampling-Fehlern im Zensus 2011 hingewiesen. Da der Qualitätsbericht zum Zensus 2022 zwar dieses Gutachten zitiert, aber eben nicht die kritischen Aspekte zu Nonsampling-Errors, sollen drei der Anmerkungen von Küchenhoff zitiert werden:

- a »Insgesamt fehlen Untersuchungen zum Thema Erhebungsfehler. Daher können diese kaum quantifiziert werden. Es wäre aus meiner Sicht wichtig gewesen, diesen Aspekt nicht getrennt von der Problematik des ›Sampling Errors‹ zu betrachten und auch in die Überlegungen bei dem Zensus Test mit einzubeziehen.«
- b »Nach dem Zensus leben in Deutschland deutlich wenige [sic!] Ausländer als bisher angenommen. Dies wurde mit Überraschung aufgenommen. Als Grund dafür wurde insbesondere das Meldeverhalten und in der Folge die Fehlerhaftigkeit der Melderegister angegeben. Neben diesem plausiblen Grund könnte auch eine Ursache dafür in einem Messfehler bei der Zensus-Befragung liegen, d.h. in einer systematischen Untererfassung der Ausländer(innen).«
- c »Da aber keine unabhängigen Kontrollzählungen (›Post Enumeration Surveys‹) durchgeführt wurden und eine Auswertung der Wiederholungsbefragung nicht vorliegt, ist eine Quantifizierung des Messfehlers und dessen Wirkung [auf, RS/RL/RS] die Güte der Schätzung der Einwohnerzahl nicht möglich. Es ist aber davon auszugehen, dass der Gesamtfehler bei der Schätzung der Einwohnerzahl beim Zensus höher ist als der jeweils ausgewiesene relative Standardfehler.«

Jeder dieser Sätze galt für den Zensus 2011 und gilt erneut für den Zensus 2022. Es ist bemerkenswert, dass das Statistische Bundesamt keinen der genannten Aspekte durch Studien, Maßnahmen oder Veröffentlichungen für den Zensus 2022 aufgegriffen hat. Das Gutachten von Küchenhoff (2014) im Abschnitt 4.3 des Qualitätsberichts zum Zensus 2022 (Statistisches Bundesamt, 2024) als positiv für den Zensus 2022 zu zitieren, ist daher irreführend. Das Zitat im Qualitätsbericht ist eingedenk der Tatsache, dass Küchenhoff in 2020 (Bleninger et al., 2020) die kritischen Punkte nochmals unterstrich, schwer zu rechtfertigen.

---

In der neueren Stellungnahme forderte Küchenhoff (Bleninger et al., 2020) unter anderem:

1. Messfehleranalysen mit Hilfe der Erfahrungen aus dem Zensus 2011,
2. Einbeziehen der Messfehler in die Simulationen zur Schätzgenauigkeit,
3. Neubewertung der zu erreichenden Genauigkeit,
4. Entwicklung von Strategien zur Korrektur der Messfehler insbesondere unter Nutzung der Registerdaten und
5. Überdenken der Strategie, welche die Melderegistervariablen nur als Hilfsvariablen verwendet.

Küchenhoff warnt in seinem Kommentar (Bleninger et al., 2020) explizit, dass ohne diese Untersuchungen die Ansprüche an die Genauigkeit, die an die Zensusergebnisse gestellt werden, einer detaillierten Prüfung nicht standhalten können. Im Folgenden werden wir einige Aspekte dieser Nonsampling-Fehler näher diskutieren.

## 8.1. Mehrfachfallprüfung

Ein bedeutender Teil der Datenverarbeitung des Personendatenbestands ist die Mehrfachfallprüfung. Die Mehrfachfallprüfung sollte die Melderegisterlieferungen aller Gemeinden zu einem Datensatz zusammenführen, so dass jede zum Zensusstichtag in Deutschland lebende Person mit genau einem einwohnerzahlrelevanten Melderegisterdatensatz im Personenbestand enthalten sein würde (Wittmaack, 2025). Für den Zensus 2022 wurde die Mehrfachfallprüfung neu programmiert. Technische Einzelheiten dazu finden sich nur in der Publikation von Wittmaack (2025). Weder werden die Algorithmen detailliert dargestellt, noch der Programmcode zur Verfügung gestellt. Ebenso wenig gibt es technische Details zur Validierung des Programms. Damit kann weder die Zuverlässigkeit der Algorithmen, noch die korrekte technische Umsetzung noch die Leistung des Programms beurteilt werden.

Aus Sicht der Informatik ist die Mehrfachfallprüfung ein Record-Linkage-Problem, auch wenn das Statistische Bundesamt diesen Begriff selten verwendet. Um die Probleme zu erläutern, muss etwas ausgeholt werden.

---

### 8.1.1. Record-Linkage und Linkage Bias

Record-Linkage ist in der Statistik der übliche Begriff, um das Zusammenführen von Daten derselben Einheit aus verschiedenen Datenbanken zu bezeichnen (Christen, 2012). Bei der Zusammenführung können Fehler durch die fälschliche Zusammenführung verschiedener Personen (falsch Positive) oder durch die fälschliche Nichtzusammenführung von Daten derselben Person (falsch Negative) entstehen. Durch diese beiden Fehler können statistische Analysen zu fehlerhaften Ergebnissen führen. Solche Verzerrungen statistischer Analysen werden als »Linkage-Bias« bezeichnet.

Linkage-Bias entsteht vor allem dann, wenn bestimmte Merkmale, die zur Zusammenführung verwendet werden (wie Name, Geburtsdatum, Adresse), häufiger zu falsch Positiven oder falsch Negativen führen. Das einfachste Beispiel ist Geschlecht: Erfolgt die Zusammenführung über den Nachnamen und ändert sich dieser durch Heirat bei Frauen häufiger, dann werden die Daten von Frauen seltener erfolgreich zusammengeführt.

Zu den Merkmalen, für die am häufigsten Linkage-Bias in der Literatur nachgewiesen wurde, gehören Geschlecht und ethnische Zugehörigkeit (Bohensky, 2016).<sup>2</sup> In Simulationsstudien (Schnell & Weiand, 2023) und empirischen Analysen auf der Basis des Hamburger Schülerregisters (Brändle et al., 2025) konnte gezeigt werden, dass Linkage-Bias auch in administrativen deutschen Datenbanken häufiger bei Ausländern auftritt als bei Deutschen. Entsprechend werden Daten von Ausländern schlechter verknüpft als bei Deutschen.<sup>3</sup>

### 8.1.2. Blocken in der Mehrfachfallprüfung im Zensus 2022

Ein technisches Detail bei Record-Linkage-Verfahren ist das sogenannte »Blocken«. Hierbei wird die Zahl möglicher Ähnlichkeitsberechnungen zwischen zwei Datenbeständen verringert, indem eine Blocking-Variable zur Vorauswahl überhaupt relevanter Fälle verwendet wird (Lenz, 2006). Als Beispiel kann man z.B. das Geburtsjahr als Blockingvariable verwenden und nur Personen mit gleichem Geburtsjahr untereinander vergleichen. Blocken setzt voraus, dass die Blocking-Variable vollkommen fehlerfrei und für jeden Fall

---

<sup>2</sup> Die Erfassung ethnischer Zugehörigkeit erfolgt in den meisten Studien zum Record-Linkage entweder über die Staatsangehörigkeit oder über die relative Häufigkeit von Namen in Subgruppen, die zuvor als ethnische Gruppe bezeichnet wurden.

<sup>3</sup> Die Ursachen hierfür sind häufigere Umzüge, Probleme bei der Datenerfassung unvertrauter Namen sowie Unterschiede in den Häufigkeitsverteilungen bei Subgruppen von Ausländern. So sind die Namen »Wu«, »DaSilva« oder »Kim« in den jeweiligen Subgruppen deutlich häufiger als die häufigsten deutschen Namen in Deutschland. Die Algorithmen zur Zusammenführung von Daten sind gegenüber den Unterschieden in solchen Häufigkeitsverteilungen empfindlich (Schnell & Weiand, 2023; Weiand, 2025).

---

im Datensatz vorhanden ist. In der Praxis verwendet man daher häufig eine experimentell festgelegte Abfolge verschiedener Blocking-Variablen, wie z.B. zuerst Geburtsjahr, dann Geburtsort, dann Adressinformationen. Eine solche Abfolge (eine »Blockingstrategie«) sollte auch den Fall berücksichtigen, dass die Blocking-Variablen für einen Teil der Datensätze nicht vorhanden sind. Die experimentelle Entwicklung einer Blocking-Strategie ist nicht trivial.

Wittmaack (2025) beschreibt das Blocken in der Mehrfachfallprüfung so:

»Um die Anzahl der paarweisen Abgleiche innerhalb der Obergruppen noch weiter zu reduzieren, wurde der Geburtsort<sub>standardisiert</sub> als weitere, gleitende Einschränkung des Suchraums herangezogen. Innerhalb jeder Obergruppe musste sich der erste Buchstabe des Geburtsort<sub>standardisiert</sub> gleichen oder mindestens einer der abzugleichenden Datensätze wies einen leeren Geburtsort<sub>standardisiert</sub> auf. Dies reduzierte die durchgeführten Abgleiche um 93 % (... )«

Das bedeutet, dass bei unterschiedlichen Anfangsbuchstaben des Geburtsortes keine Möglichkeit bestand, zusammengehörende Personen zu finden.

### 8.1.3. Qualitätsmetriken für die Zusammenführung von Datenbanken

Zur Beurteilung, wie erfolgreich Daten aus verschiedenen Registern für einen Zensus zusammengeführt werden konnten, ist die Analyse der Matching-Statistiken notwendig (Baffour & Valente, 2012, S. 130). Dabei ist die Zahl der erfolgten Verknüpfungen **nicht** von Interesse, da eine vollständig falsche Zuordnung ja eine zwar falsche, aber vollständige Zuordnung erbringen kann. Dies ist aber die einzige Zahl, die das statistische Bundesamt berichtet.

#### Übliche Qualitätsmetriken

Record-Linkage besteht aus der maschinellen Klassifikation potentieller Paare von Datensätzen aus Datenbanken. In der internationalen technischen Literatur zum Record-Linkage wird zwischen verschiedenen Formen von Klassifikationsergebnissen unterschieden (siehe Abbildung 8.2):

**richtig positive (rp)** Dies sind korrekt klassifizierte Paare.

**falsch positive (fp)** Dies sind fälschlich als Paar klassifizierte Paare.

**falsch negative (fn)** Dies sind übersehene Paare.

**richtig negative (rn)** Dies sind korrekt klassifizierte nicht zusammengehörende Paare.

---

		tatsächlicher Zustand	
		Match	Non-Match
Klassifikation	Link	richtig positiv	falsch positiv
	Non-Link	falsch negativ	richtig negativ

Abbildung 8.2.: Klassifikation potentieller Paare im Record-Linkage

Man möchte Klassifikationen, bei denen der Anteil der richtig-positiven möglichst nahe an Eins liegt. Gleichzeitig soll der Anteil der falsch-positiven möglichst nahe an Null liegen. Zur Beurteilung einer maschinellen Klassifikation im Record-Linkage werden in der Regel die beiden Maßzahlen *Precision* und *Recall* verwendet, die als

$$\text{Precision} = \frac{rp}{rp + fp} \quad (8.1)$$

$$\text{Recall} = \frac{rp}{rp + fn} \quad (8.2)$$

definiert sind (Christen, 2012). Die gleichzeitige Maximierung von Precision und Recall ist bei realen, fehlerbehafteten Datensätzen schwierig.

#### 8.1.4. Die Bedeutung der Qualitätsmetriken des Record-Linkage für den Zensus

Das Statistische Bundesamt hat keinerlei Ergebnisse zu Match-Statistiken wie Precision oder Recall veröffentlicht, sondern nur den Anteil verknüpfter Paare (100%). Diese 100% ließen sich auch durch eine zufällige Zuordnung von Paaren erreichen und ist damit zur Beurteilung der Klassifikationen des Record-Linkage im Zensus nahezu wertlos.<sup>4</sup> Damit stellte sich die Frage, ob das Statistische Bundesamt überhaupt eine Qualitätskontrolle für das Record-Linkage vorgenommen hat.

---

<sup>4</sup> Läge die Zahl unter 100% läge entweder ein Programmierfehler oder unzureichende Information für eine Klassifikation vor, die nicht durch das Programm berücksichtigt wurde. Damit ist der Anteil der verknüpften Fälle eine triviale Diagnoseziffer, aber kein Qualitätsmaß.

---

Wie problematisch ein Datenbankabgleich mit nur wenigen Merkmalen ist, lässt sich am Beispiel der Erprobung des Identitätsabrufs im Rahmen der Registermodernisierung erkennen. Bei diesem Abruf übermitteln berechnete öffentliche Stellen »... aus den bei ihnen gespeicherten Daten mindestens den Familiennamen, den Wohnort, die Postleitzahl sowie das Geburtsdatum der Person in einem automatisierten Verfahren unter Nutzung des Standards XBasisdaten an das BVA« (Klein, 2022). In einem Test des Verfahrens gelang es dem BVA, von den 2.4 Millionen Einträgen im Waffenregister ca. 89.5 % der Fälle zuzuordnen, mit zusätzlichen Vornamen dann 90.0 % (Bundesverwaltungsamt, 2023). Gemessen an den Maßstäben für die Entwicklung von Linkage-Verfahren (Christen, 2012; Christen et al., 2020) sind das unzureichende Ergebnisse.

Bei großen Datensätzen und nur wenigen Quasi-Identifikatoren (wie Name, Vorname, Geschlecht und Geburtsdatum) sind (nachweisbar korrekte) Zuordnungen unter 90 % erwartbar (Li et al., 2006; Schnell, 2019b; Schnell & Weiland, 2023). Für den Zensus 2022 bedeutet dies, dass die Kernmerkmale für das Ziel-1 keinesfalls ausreichen, um eine eindeutige Personenzuordnung zu erreichen.<sup>5</sup>

## 8.2. Auswirkung der Mehrfachfallprüfung auf die Korrektur der Einwohnerzahl

Vergleicht man die Datenblätter einer Vielzahl von Gemeinden, dann fällt auf, dass viele Gemeinden als Ergebnis der Mehrfachfallprüfung empfindliche Einwohnerverluste erlitten. Da ein solcher Verlust einer Person im Regelfall zu einem Gewinn bei einer anderen Gemeinde hätte führen können, stellt sich die Frage, welche Gemeinden von solchen Bereinigungen besonders profitiert haben. Also bräuchte man für jeden Fall des Registerdatenbestandes einen Pfad, der die Effekte auf die Gemeinden wiedergibt. Als grobe Näherung könnte man für jede Gemeinde eine einfache Gewinn/Verlust-Zahl für jede andere Gemeinde erstellen. Dies ließe sich auch durch Aggregation vergrößern. Als Beispiel haben die Autoren eine Gemeinde mit besonders hohem Studierendenanteil gewählt: Marburg. Die Nachfrage bei Destatis vom 05.11.2025, ob eine Auswertung an einer Beispielgemeinde (hier Marburg) möglich wäre, wurde mit dem Hinweis als nicht realisierbar verneint. Das mag den derzeitigen faktischen Stand der Dinge widerspiegeln, aber einen vollständig anonymisierten Datensatz dieser gemeindebezogenen Register-

---

<sup>5</sup> Es muss bedacht werden, dass Verzeigerungen in den Melderegistern (Querverweise auf Ehe- oder Lebenspartner, Eltern-Kind-Beziehungen; Geschwister sind nicht verzeigert) nur für Teilmengen der Population und dann nicht in allen Registern vorliegen. Diese Merkmale sind außerhalb der für die Einwohnerzahl nicht relevanten Haushaltegenerierung für maschinelle Abgleiche weitgehend unbrauchbar.



## Fachanwendung Übersicht

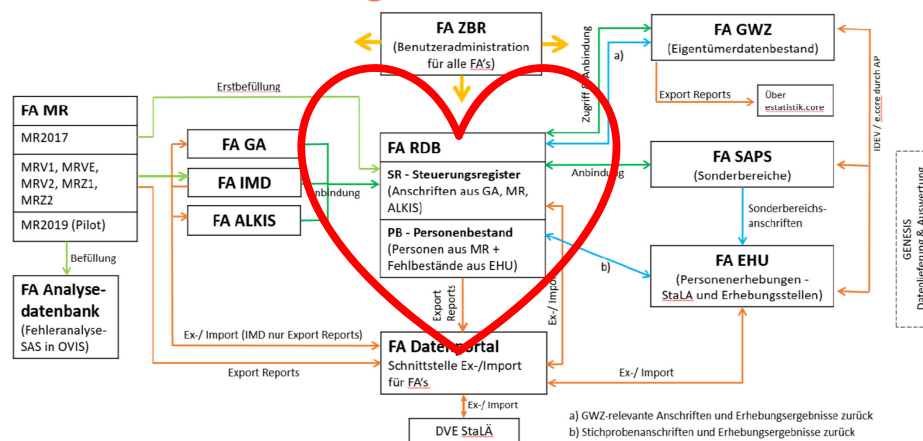


Abbildung 8.3.: Übersicht über die Fachanwendungen im Zensus. Quelle: Warschofsky und Wagenknecht (2021). Hervorhebung des RDB im Original.

bewegungen hätte Destatis im Sinne einer Transparenz für die Gemeinden erstellen können.

### 8.3. Existenzfeststellung

Die Datenerhebung der Existenzfeststellung wird auf einer Destatis-Seite (Statistisches Bundesamt, 2025) folgendermaßen beschrieben:

»Zur Bestimmung der Bevölkerungszahl stellten die Erhebungsbeauftragten zunächst vor Ort fest, welche Personen an der besuchten Anschrift wohnen. Dafür wurde unter anderem nach Familiennamen und Vornamen, Geburtsdatum, Geschlecht, Staatsangehörigkeit und Kontaktdaten wie Haupt- oder Nebenwohnsitz aller Haushaltsmitglieder gefragt. Die Beantwortung der Fragen konnte von einer erwachsenen Person stellvertretend für alle anderen Haushaltsangehörigen übernommen werden, sofern die betreffenden Angaben bekannt waren.«

---

Die Daten wurden anschließend über das Erhebungsunterstützungssystem (FA EHU) eingepflegt.<sup>6</sup>

### 8.3.1. Existenzfeststellung durch Proxy-Antworten

Bei der Existenzfeststellung waren Proxy-Auskünfte ausdrücklich zugelassen.<sup>7</sup> Eine Proxy-Auskunft bedeutet, dass diese Auskünfte auch von Familienangehörigen, Vertrauenspersonen, Betreuern oder anderen an der Wohnanschrift Anwesenden erteilt werden konnten.<sup>8</sup>

Hierzu ist anzumerken, dass Proxy-Auskünfte in der Survey-Methodologie als mit höheren Fehlerquoten behaftet angesehen werden als Selbstauskünfte (Schnell, 2019a). Zweitens ist es nicht unwahrscheinlich, dass Proxy-Auskünfte für unklare Wohnverhältnisse intentional verweigert wurden (Tourangeau et al., 2012). Drittens sind Auskünfte zu Wohnverhältnissen (in der Fachliteratur: »Household roster« oder »rules of residence« (Martin, 1999, 2007)) mit großen Interpretationsspielräumen behaftet. Daher sind Effekte durch Erhebungsbeauftragte bei solchen Fragen erwartbar (Schnell & Kreuter, 2005; West & Blom, 2016).

Daher könnten insbesondere bei Verständnis- und Sprachproblemen existierende Personen nicht korrekt erfasst worden sein. Dies kann ein Erklärungsansatz für die hohen Einwohnerverluste bei Ausländern sein. Dies ließe sich empirisch überprüfen, indem die Haushaltszusammensetzungen und das Merkmal »Proxy-Auskunft« zusammen mit einem Identifikator für die Erhebungsbeauftragten analysiert werden. Entsprechende Analysen wurden nicht veröffentlicht.

### 8.3.2. Unvollständige Existenzfeststellungen

In der Haushaltsstichprobe wurden die Kernmerkmale (Nachname, Vorname, Geschlecht, Geburtsdatum) und gegebenenfalls weitere Ziel-1-Merkmale erfragt.<sup>9</sup> Konnten diese

---

6 Das ist die einzige auffindbare Übersicht über die Fachanwendungen im Rahmen des Zensus 2022. Eine Fachanwendung in der Sprechweise des Statistischen Bundesamtes ist ein spezialisiertes IT-System, das fachlich definierte statistische Prozesse abbildet und unterstützt. Fachanwendungen bestehen aus mehreren Programmen samt ihrer Schnittstellen, die zusammen ein Problem innerhalb eines statistischen Workflows lösen sollen. Soweit uns bekannt, gibt es keinerlei Veröffentlichungen zu Fachanwendungen des Statistischen Bundesamtes – und damit auch keine unabhängigen Tests oder Validierungen.

7 Die Tatsache einer Proxy-Auskunft wurde auf dem Haushaltsbogen vermerkt.

8 Dies gilt auch für die zeitlich späteren Fälle, die durch die Erhebungsstellen geklärt wurden.

9 Bretschki et al. (2024b, S. 22) schreiben: »Voraussetzung für die Bestätigung einer Existenz ist das vollständige Vorliegen der Kernmerkmale Name, Vorname, Geburtsdatum, Geschlecht.«

---

Merkmale bis zum Abschluss der Erhebungszeit nicht vollständig erhoben werden, handelt es sich – nach unserem Wissen – um eine unvollständige Existenzfeststellung.<sup>10</sup>

### 8.3.3. Effekt unvollständiger Existenzfeststellungen auf die Hochrechnung

Da es keine Veröffentlichungen zur statistischen Berücksichtigung dieser Fälle gab, erfragten die Gutachter die Vorgehensweise.<sup>11</sup> Das Statistische Bundesamt antwortete:

»Anschriften, an denen die Existenzfeststellung nicht abgeschlossen werden konnten, wurden als solche gekennzeichnet. Diese Anschriften wurden dann aus der Hochrechnung zur Korrektur von Karteteilen und Fehlbeständen ausgesteuert. Eine Publikation gibt es dazu nicht.«

Damit stellt sich die Frage, welchen Effekt unvollständige Existenzfeststellungen auf die Hochrechnungen haben. Das ist im Detail nicht dokumentiert.<sup>12</sup> Betrachtet man den Einfluss der Anschriftengröße auf die Wahrscheinlichkeit, dass die gesamte Anschrift aus der Hochrechnung ausgeschlossen wird, dann ist es offensichtlich (Abbildung 8.4), dass mit steigender Anschriftengröße die Wahrscheinlichkeit, dass die Adresse vollständig erfasst wird, rasch sinkt.

Das bedeutet, dass größere Gebäude eher aus der Hochrechnung ausgeschlossen werden. Damit steigt in jedem Fall aufgrund der verringerten Stichprobengröße der tatsächliche Standardfehler der vorhergesagten Werte zunächst für diese Schicht, aber auch für alle anderen Gebäude. Sollten die ausgeschlossenen Gebäude eine andere Bewohnerdynamik besitzen als die berücksichtigten Gebäude, werden auch die Übererfassungen und Untererfassung systematisch über- bzw. unterschätzt.

---

10 Die Schulungsunterlagen für Erhebungsbeauftragte sind unveröffentlicht und werden auf Anfrage nicht zur Verfügung gestellt.

11 Anfrage am 22.10.2025, Antwort des Auskunftsdienstes am 5.11.2025, GZ 491600/818495.

12 Hierzu erläutern die Datenblätter an die Gemeinden, dass in dem als »Korrektur III« bezeichneten Schritt die Ergebnisse der Haushalbefragung mit den Positionen 13 und 14 korrigiert werden.

Position 13 enthält die aus der Haushaltsstichprobe hochgerechneten Übererfassungen (Personen, die an dieser Anschrift gemeldet sind, für die aber in der Haushalbefragung keine Existenz festgestellt werden konnte). Diese Übererfassungen werden auf den Gesamteinwohnerbestand des Erhebungsgebietes hochgerechnet und dieser entsprechend korrigiert (= Abzug der Personen).

Position 14 enthält die aus der Haushaltsstichprobe hochgerechneten resultierenden Untererfassungen (Personen, die an diesen Anschriften nicht gemeldet sind, aber bei der Haushalbefragung als Bewohner angetroffen wurden). Diese Personen werden für die Einwohnerzahlermittlung mitgezählt und auf den SMP hochgerechnet (= Hinzunahme der Personen).

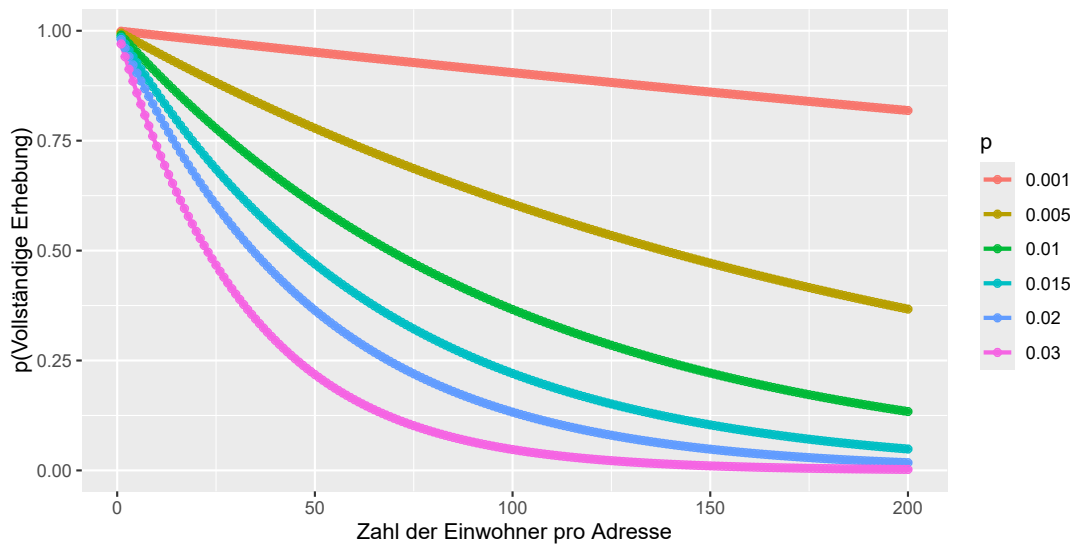


Abbildung 8.4.: Wahrscheinlichkeit für eine vollständige Erhebung nach Anschriftengröße unter Annahme der Unabhängigkeit für unterschiedliche Ausfallwahrscheinlichkeiten einer Zielperson.

Wie sich das auf die Einwohnerzahlen auswirkt, müsste pro Sampling-Point (SMP) getrennt untersucht werden. Dazu sind aber die kleinräumigen Daten, einschließlich der Hilfsgrößen des Hochrechnungsmodells pro Adresse notwendig. Ohne solche Daten kann über die Richtung des Effekts keine Aussage gemacht werden. Entsprechend kann auch das Statistische Bundesamt nicht von einer notwendigerweise unverzerrten Schätzung sprechen. Prinzipiell kann man Beispiele konstruieren, bei denen der systematische Ausfall eines Gebäudes mit besonderer Populationsdynamik die Ergebnisse des Hochrechnungsverfahrens für einen SMP wesentlich verändert. Dieses Problem wird in keiner Veröffentlichung des Statistischen Bundesamtes oder in den Datenblättern thematisiert.

### 8.3.4. Probleme in der Erhebungsteilübergreifenden Plausibilisierung im Referenzdatenbestand

In den SMP erfolgte die Feststellung der Existenz der gemeldeten Personen durch Befragung mindestens einer Person des Zielhaushaltes. Die Kernmerkmale mussten nach der Eingabe in das EHU dann mit dem Registerbestand abgeglichen werden. In der Abbildung 8.3 gibt es einen Pfad (b) zwischen dem Personenbestand (PB) innerhalb

---

der FA RDB und der FA EHU. Dieser in keiner Veröffentlichung des Statistischen Bundesamtes explizit erwähnte oder im Detail beschriebene Vorgang ist der zentrale Teil der Existenzfeststellung.

Dies war Bestandteil der erhebungsteilübergreifenden Plausibilisierung im Referenzdatenbestand (Bretnütz et al., 2024; Hüsgen-Brodhäger, 2024).<sup>13</sup> Über die Probleme bei der Identifikation einer Person über den Merkmalskatalog der Ziel-1 Erhebung oder allein der Kernmerkmale in diesem Schritt gibt es keine Daten und keine Veröffentlichungen.

Hier geht es um die Frage, ob eine erfasste Person im Datenbestand enthalten ist oder nicht. Das ist ein Record-Linkage-Prozess, also der maschinelle Abgleich zweier Datenbanken (Christen, 2012). Record-Linkage ohne (scheinbar) eindeutige Identifikatoren wie Sozialversicherungsnummern sind immer fehlerbehaftete Prozesse (Christen & Schnell, 2023). Es gibt für diesen zentralen Baustein des Zensus keinerlei Publikationen, veröffentlichte Programme oder Qualitätskennzeichen.

Fehler im Record-Linkage ohne exakte Identifikatoren sind unvermeidlich. Wie bei jedem Record-Linkage-Projekt stellen sich die Fragen nach dem Ausmaß dieser Fehler (siehe Kapitel 8.1.4) und dem resultierenden Linkage-Bias (siehe 8.1.1). Beide Fragen hätten sich durch eine manuelle Überprüfung und anschließende Validierung anhand einer nachgeschalteten Klärung mit den betroffenen Personen in einer Stichprobe klären lassen. Veröffentlichungen zu einer solchen unentbehrlichen Fehlerkontrolle gibt es nicht. Sollte dies nicht erfolgt sein, so wäre das grob fahrlässig und gegen alle Kunstregeln. Sollten die Identifikatoren alle bereits gelöscht sein, ließe sich auch keine nachträgliche Überprüfung durchführen.<sup>14</sup>

Es ist insbesondere unklar, wie sichergestellt wurde, dass zwei in Kernmerkmalen übereinstimmende Datensätze auch tatsächlich dieselbe Person sind. Dieses Problem falsch-positiv paariger Personen wurde ebenso wenig diskutiert wie das Problem nicht eindeutig zuordenbarer Paare. Als Extremfälle gelten in der Literatur Zwillinge.

Solche Zuordnungsprobleme werden unmittelbar relevant für die Einwohnerzahl, da für Personen, die sowohl in der Erhebung als auch im Melderegister zählungsrelevant vorhanden sind, geklärt werden muss, ob die Information zum Wohnungsstatus im weiteren

---

13 Die Darstellungen des Statistischen Bundesamtes lassen den Ablauf keineswegs so klar erkennen, wie dieser Abschnitt nahelegt. Daher sind Fehler in unserer Darstellung möglich, aber eben auf das Fehlen aller Programmdokumentationen oder Ablaufpläne zurückzuführen.

14 Auf einer auf den 6.7.2025 datierten Seite teilt Destatis (Statistisches Bundesamt, 2025) mit: »An dieser Stelle möchten wir darauf hinweisen, dass zu Anträgen hinsichtlich Ihrer Rechte nach der DS-GVO als betroffene Person zum Zensus 2022 keine Informationen mehr zur Verfügung gestellt werden können, weil die personenidentifizierenden Daten im Zensus 2022 entsprechend den Löschvorschriften zwischenzeitlich bereits gelöscht wurden.« Der von der Stadt Köln erwirkte Stopp der Löschungen könnte dies zumindest für Köln verhindert haben (es ist uns im Detail nicht bekannt, welche Daten nicht gelöscht wurden).

---

Verfahren aus der Primärerhebung oder aus dem Melderegister zu nutzen ist (Bretschi et al., 2024a, S. 14). Da für den Zensus 2022 der Wohnungsstatus bei allen primärstatistischen Personenerhebungen aus der Registerinformation verwendet wird, bedeutet dies, dass eine in der Stichprobe festgestellte existente Person, die laut Melderegister an dieser Anschrift mit Nebenwohnsitz gemeldet ist, beim Zensus an dieser Anschrift nicht in die Einwohnerzahl einfließt (unabhängig davon welchen Wohnungsstatus sie bei der Befragung angibt).

### **8.3.5. Probleme durch die Softwareunterstützung der Existenzfeststellungen**

In der Feldarbeit wurden von einer Reihe von statistischen Landesämtern (Bayern, Berlin-Brandenburg, Statistik Nord, Mecklenburg-Vorpommern) mobile Endgeräte, zumindest partiell zur Existenzfeststellung bei der Haushaltsstichprobe eingesetzt. Die Existenzfeststellung erfolgte dort mithilfe der Endgeräte; auf Wunsch der Haushalte konnten auch die weitergehenden Fragen zu Bildung, Erwerbstätigkeit und Migration unmittelbar erfasst werden. In den meisten Ländern kamen für die Existenzfeststellung ausschließlich Papierfragebögen zum Einsatz. Die Angaben zur Existenzfeststellung mussten daher später händisch erfasst werden.

Bereits in der Phase der Feldarbeit kam es zu zahlreichen Beschwerden über die Funktionalitäten der IT-Instrumente. Im Juli 2022 haben der Deutsche Städtetag und der Deutsche Landkreistag in einem Schreiben an das Bundesministerium des Innern und für Heimat auf erhebliche Defizite hingewiesen und die Befürchtung von Beeinträchtigungen geordneter Abläufe und von Ergebnisverzerrungen geäußert. In einem länderübergreifenden Erfahrungsaustausch wurde von kommunalen Erhebungsstellen festgestellt, dass es insbesondere bei der zentralen IT-Anwendung des Erhebungsunterstützungssystems (EHU) zu Performance-Problemen, Programmabbrüchen und Schwierigkeiten bei der Bearbeitung von Haushalten gekommen ist. So konnten z.B offensichtlich unrichtige Angaben nicht nachträglich korrigiert werden. Insgesamt 95 % der Erhebungsstellen in Baden-Württemberg berichten von Problemen mit dem Erhebungsunterstützungssystem und 66 % bezeichnen das EHU als die größte Herausforderung des Zensus 2022 (Gehr, 2024).

Diese Softwareprobleme hatten Konsequenzen für die Mahnverfahren. Einerseits wurden ungerechtfertigte Mahnverfahren durchgeführt, da sich die Erfassung bereits vorliegender Existenzfeststellungen verzögerte und andererseits wurden auf Empfehlungen verschiedener Landesämter Erinnerungs- und Mahnverfahren zum Teil erst später

---

durchgeführt.<sup>15</sup> So konnten sich die Softwareprobleme des EHU auf den Umfang der abgeschlossenen Existenzfeststellungen auswirken.

Nimmt man z.B. eine Gemeinde mit hohem Ausländeranteil oder einem hohen Anteil an Studierenden, dann werden zwei Kontaktversuche für einen hohen Anteil der Auskunftspflichtigen nicht ausreichen. Dann werden in diesen Gemeinden die zeitlichen Abläufe durch Probleme mit der EHU besonders gedrängt. Also ist in solchen Gemeinden mit einem hohen Anteil nicht abgeschlossener Existenzfeststellungen zu rechnen. Daher wirken sich die technischen Probleme selektiv auf Gemeinden aus, die resultierenden Fehler werden nicht gleichmäßig auf die Gemeinden verteilt. Da es solche Probleme bereits beim Zensus 2011 gab (Küchenhoff, 2014), kamen diese Probleme nicht unerwartet.

Über den gesamten Prozess der zeitlichen Abläufe, der Mahnverfahren und Maßnahmen und das Ausmaß abgeschlossener Existenzfeststellungen existieren keine Veröffentlichungen. Der Einfluss dieser Probleme auf die Einwohnerzahlen kann nicht beurteilt werden. Im Umkehrschluss bliebe die Beweislast, dass diese Probleme keinen Einfluss auf die Einwohnerzahlen haben, bei den Landesämtern bzw. dem Statistischen Bundesamt.

## 8.4. Die Feldarbeit durch die Erhebungsbeauftragten

Die Datenerhebung in den Erhebungsbezirken erfolgte durch die Erhebungsbeauftragten. Deren Arbeitsbelastung war erheblich. Insgesamt scheinen ca. 100 000 Erhebungsbeauftragte eingesetzt worden zu sein. Bei 10.3 Millionen Befragten entspräche dies im Mittel 103 Befragten pro Erhebungsbeauftragtem. In Mecklenburg-Vorpommern wurden von 19 Erhebungsstellen ca. 1 600 Interviewer zur Erhebung von rund 293 000 Personen in der Haushaltebefragung eingesetzt. Dies sind im Mittel 183 Personen pro Erhebungsbeauftragtem.

Die Erhebungsbeauftragten sollten die zugewiesenen Adressen (Erhebungsbezirke) aufsuchen, Kontakt zu Auskunftspflichtigen aufnehmen, wobei Anzahl und Nachnamen der gemeldeten Personen den Erhebungsbeauftragten durch die sogenannte Namensliste bekannt waren. Für die bekannten Personen und eventuell weitere dort wohnende Personen wurden die Kernmerkmale (Namen, Vornamen, Geschlecht und Geburtsdatum) aufgenommen.

Die Erhebungsbeauftragten waren überwiegend zwischen dem 15.5.2025 und vermut-

---

<sup>15</sup> So schrieb der Städtetag in einem Schreiben an das Innenministerium: »Entsprechend haben bis vor kurzem verschiedene Statistische Landesämter – aufgrund gravierender Systemfehler und -mängel – davon abgeraten, Erinnerungs- und Mahnverfahren einzuleiten« (Deutscher Städtetag & Deutscher Landkreistag, 2022)

---

lich dem 15.8.2025 tätig.<sup>16</sup> Das bedeutet, die reine Feldzeit beträgt maximal 12 Wochen, zusammen mit der verbleibenden Zeit für die Tätigkeit der Erhebungsstellen also maximal 27 Wochen.

Den Erhebungsbeauftragten waren mindestens zwei Kontakte vorgeschrieben. Konnte eine Erhebung an einer Anschrift nicht bis zum Ende der Tätigkeit des Erhebungsbeauftragten abgeschlossen werden, sollte die Erhebungsstelle übernehmen. Bei sozialwissenschaftlichen Erhebungen sind deutlich mehr Kontaktversuche üblich, daher ist zu vermuten, dass die Erhebungen der Erhebungsbeauftragten häufig unvollständig blieben. Damit lag die Verantwortung für die Vervollständigung bei den Erhebungsstellen.

Die Befragung von Gehr (2024) in allen Erhebungsstellen Baden-Württembergs zeigt, dass keineswegs alle Erhebungsstellen die Erhebungen abschließen konnten. In den großen Gemeinden gaben mehr als 5 % der Erhebungsstellen an, lediglich maximal 85 % der Existenzfeststellungen abgeschlossen zu haben. Aus welchen Gründen auch immer, wurden z.B. Postzustellungsurkunden nur sehr begrenzt eingesetzt. Darüber hinausgehende Maßnahmen, wie Zwangsgeldverfahren wurden anscheinend nur in Einzelfällen angestrengt; Zahlen dazu sind nicht veröffentlicht.

Im Vergleich zu den anderen Arbeitsphasen wurde für die Feldarbeit der Kommunen nur wenig Zeit eingeplant. Zum Vergleich: Bei großen Standarderhebungen der Sozialwissenschaften wie dem ALLBUS oder dem SOEP liegen die reinen Feldzeiten zwar in der gleichen Größenordnung von 6–7 Monaten, aber das sind reine Feldzeiten, die Stichproben umfassen nur wenige Tausend Personen.

### **8.4.1. Erhebungsbeauftragte als Fehlerquelle**

Methodisch ist zu berücksichtigen, dass ein erheblicher Teil der Erhebungsbeauftragten erstmals eingesetzt wurde und der Zensus nur alle zehn Jahre durchgeführt wird. Dadurch konnten typische Lerneffekte, wie sie bei regelmäßig wiederholten Erhebungen auftreten, nur eingeschränkt wirksam werden. Unterschiede in den persönlichen Voraussetzungen, Vorerfahrungen und Kompetenzen der Erhebungsbeauftragten stellen eine potenzielle Quelle für systematische Erhebungsfehler dar.

Die Schulungsmaßnahmen wurden teilweise flexibel an lokale Bedingungen angepasst, was zu einer heterogenen Schulungsqualität führen konnte. Die Vorgaben des Statistischen Bundesamtes konnten von den Landesämtern und Gemeinden individuell modifiziert werden. Dies betraf insbesondere die Dauer der Schulungen, die Handhabung mobiler Endgeräte sowie die Regelungen zu Aufwandsentschädigungen. Solche Unterschiede in

---

<sup>16</sup> Es ist möglich, dass diese Datumsangaben in Einzelfällen inkorrekt sind. Das Statistische Bundesamt hat keinen Zensuskalender, der alle relevanten Informationen enthält, veröffentlicht.

---

der Umsetzungspraxis können zu methodischen Verzerrungen bei der Datenerhebung führen. Entsprechend sind Interviewereffekte erwartbar.

### 8.4.2. Exkurs: Interviewereffekte

Interviewereffekte sind eine besondere Form von Nonsampling-Fehlern. Interviewereffekte können beim Auffinden einer Adresse, beim Herstellen eines Kontaktes zu Personen, durch die Art der Befragung, die Protokollierung der Antworten und durch Fehler bei der Übermittlung der Antworten auftreten. Die Untersuchung von Interviewereffekten hat eine lange Tradition in der Methodologie.<sup>17</sup> In der Zensusstichprobe und bei der Wiederholungsbefragung wurden die Interviewer als Erhebungsbeauftragte bezeichnet. Um deren Effekte zu erläutern, muss zunächst dargestellt werden, wie die Adressen auf die Erhebungsbeauftragten verteilt wurden.

#### Zuteilung der Adressen auf die Erhebungsbeauftragten

Da keine zentral erstellten Vorlagen zu Erhebungsstellen und Erhebungsbeauftragten zugänglich sind,<sup>18</sup> verwenden wir als Beispiel das Bundesland Hessen.

Das Handbuch für Erhebungsstellen (Statistisches Landesamt Hessen, 2021, S. 38) beschreibt, dass die Arbeitsbelastung von der Erhebungsstelle für die Erhebungsbeauftragten von der Voreinstellung 150 Auskunftspflichtige (AP) pro Erhebungsbeauftragten (EB) zwischen 100 und 200 im Erhebungsunterstützungssystem (EHU) eingestellt werden konnte. Anscheinend wurden dann durch das Erhebungsunterstützungssystem zwei Listen erstellt: Eine Liste der Erhebungsbeauftragten und eine Liste der Erhebungsbezirke (EB)(Statistisches Landesamt Hessen, 2021, S. 60).<sup>19</sup>

Die Zuweisung der Erhebungsbeauftragten zu den Erhebungsbezirken erfolgt dann manuell. Dazu wurden die beiden Listen auf dem Bildschirm nebeneinander angezeigt. Die Liste der Erhebungsbeauftragten enthielt zusätzliche Informationen über die jeweilige

---

17 Eine Reihe neuerer Arbeiten und Hinweise auf diese Forschungstradition finden sich in dem von Olson et al. (2020) herausgegebenen Sammelband.

18 Die Handbücher für Erhebungsstellen und Erhebungsbeauftragte sind nicht öffentlich und wurden vom Statistischen Bundesamt und IT.NRW auf Anfrage nicht zur Verfügung gestellt. Uns liegt das Handbuch für Erhebungsstellen aus Hessen (Statistisches Landesamt Hessen, 2021) und das Handbuch für Erhebungsbeauftragte aus Hannover (LSN, 2022) vor.

19 Ein EHB entspricht einer Anschrift. Bei Anschriften mit vielen erwarteten Personen, kann eine Anschrift *durch die Erhebungsstelle* in mehrere EHBs geteilt werden (Statistisches Landesamt Hessen, 2021, S. 90). Dadurch, dass die Erhebungsstelle diese Entscheidung fällen *kann*, werden weitere Unterschiede zwischen den EHBs und auch zwischen den EHST entstehen. Entsprechend steigt die Varianz durch diesen Nonsampling-Error weiter.

---

freie Kapazität und bestimmte Sprachkenntnisse. Weiterhin konnte die Liste der EHBs z.B. nach Postleitzahlen sortiert werden, um die Anfahrten der Erhebungsbeauftragten zu minimieren.

Aus rein erhebungstechnischer Sicht ist das eine nachvollziehbare Option. Allerdings entstehen hierdurch zwei Probleme:

1. Die EHSTs können unterschiedliche maximale Arbeitsbelastungen für die Erhebungsbeauftragten wählen. Hierdurch vergrößern sich die Standardfehler durch eine weitere Varianzquelle.
2. Durch die Möglichkeit, dass den Erhebungsbeauftragten räumlich nahe beieinanderliegende Adressen zugeteilt werden können, entstehen zusätzliche Klumpeneffekte, die die Standardfehler erheblich vergrößern können.

Bedeutsam ist vor allem das zweite Problem. In der Literatur werden diese Nonsampling-Fehler als »Interviewer-Varianz« (Groves et al., 2009; Schnell & Kreuter, 2005) oder »clustering by interviewer« (Elliott & West, 2015) bezeichnet. Am einfachsten lässt sich der Effekt von räumlichen (oder interviewerbedingten) Klumpungen durch den Begriff des »Design-Effekts« quantifizieren.

Als Designeffekt wird das Verhältnis der Varianz eines Schätzers für eine komplexe Stichprobe<sup>20</sup> zur Varianz eines Schätzers einer Statistik in einer einfachen Zufallsstichprobe (»Simple Random Sample«, SRS) verstanden (Kish, 1995):

$$\text{deft} = \sqrt{\frac{\text{var}(b)}{\text{var}(b)_{SRS}}}. \quad (8.3)$$

Die Stichprobe im Zensus ist schon durch die Schichtung nach Gebäudegrößen keine einfache Zufallsstichprobe und durch die Überlappung der Schichten kompliziert sich das Design weiter. Durch die manuelle Zuweisung der Erhebungsbeauftragten zu den Adressen ist das Design der Stichprobe innerhalb eines SMP daher eine komplexe Stichprobe.

Je größer der Designeffekt wird, desto größer wird der tatsächliche Standardfehler und damit das Konfidenzintervall. Für einen beliebigen Parameter  $\theta$  ergibt sich ein Konfidenzintervall (CI) dann als

$$CI = \hat{\theta} \pm z_{\alpha/2} \cdot \text{deft} \cdot \text{SE}(\hat{\theta}),$$

wobei

---

<sup>20</sup> Komplex bedeutet hier nichts anderes, als das es keine einfache Zufallsstichprobe ist, sondern etwas anderes.

---

$\hat{\theta}$  : Schätzer des Parameters  $\theta$ ,

$z_{\alpha/2}$  : z-Wert für das gewünschte Konfidenzniveau (z.B.  $z_{0.025} = 1.96$  für ein 95 %-CI),

$SE(\hat{\theta})$  : Standardfehler des Schätzers  $\hat{\theta}$  und

deft : Design-Effekt.

Bei einem Designeffekt (deft) von 2.0 verdoppelt sich der Standardfehler und damit die Breite des Konfidenzintervalls. Effekte durch Interviewer in dieser Größenordnung sind – ja nach Komplexität der Fragestellung – nicht selten (Schnell & Kreuter, 2005). Entsprechend sind tatsächliche Konfidenzintervalle häufig breiter, als diese bei Betrachtung der Sampling-Fehler allein zu sein scheinen.

Die Größe des Designeffekts lässt sich nur nach einer Datenerhebung berechnen und nur dann, wenn man die Zuweisung der Interviewer (in diesem Fall: der Erhebungsbeauftragten) zu den Befragten (in diesem Fall: Zu den Erhebungsbezirken) kennt.<sup>21</sup>

Der Designeffekt kann aber grob abgeschätzt werden. Eine einfache Möglichkeit geht auf Kish (1965) zurück. Wenn  $\bar{b}$  die durchschnittliche Größe eines Clusters (»Klumpen«, hier: Erhebungsbezirk) und ICC die Ähnlichkeit der Objekte im Cluster bezeichnet, dann kann deft mit

$$\text{deft} = \sqrt{1 + \text{ICC} \cdot (\bar{b} - 1)} \quad (8.4)$$

geschätzt werden.

In unserem Zusammenhang bezeichnet  $b$  die durchschnittliche Größe eines Erhebungsbezirks, das wären in Hessen ca. 18 Adressen.<sup>22</sup> Das Maß für die Ähnlichkeit ICC ist der sogenannte »Intraklassenkorrelationskoeffizient«. Wenn man Gleichung 8.4 umstellt, erkennt man, dass ein ICC  $\approx 0.176$  zu einer Verdopplung der Konfidenzintervalle führt. Das ist ein vergleichsweise hoher Wert, wobei aber bei komplizierten Fragen auch größere Werte beobachtet wurden (Crossley et al., 2021; Schnell & Kreuter, 2005).

Wächst die Arbeitsbelastung der Interviewer, dann sind für deutlich kleinere ICC Verdopplungen der Konfidenzintervallbreite zu beobachten (siehe Abbildung 8.5).

Daher sind die Zahl der den Erhebungsbeauftragten zugeteilten Gebäude sowie die Abhängigkeit der Feststellungen der Erhebungsbeauftragten keineswegs irrelevante technische Details, sondern zentral für die Interpretation der Ungenauigkeiten der Schätzungen in jeder einzelnen Gemeinde.

---

<sup>21</sup> Solche Analysen wären innerhalb der Landesämter möglich, solange die Hilfsmerkmale nicht gelöscht werden. Daher sollte zumindest eine Pseudonymisierung der EHBs und EBs erfolgen.

<sup>22</sup> Zahl der Adressen/Zahl der Erhebungsbeauftragten  $\approx 144\,000/8\,000=18$ .

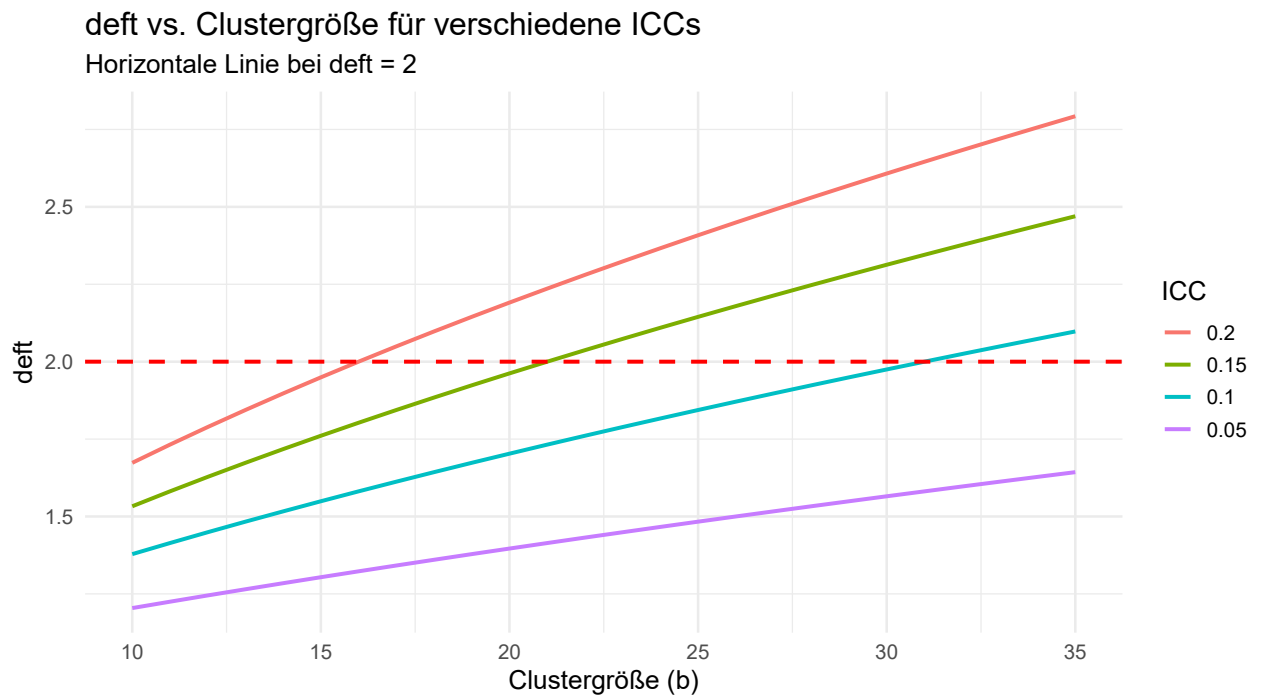


Abbildung 8.5.: Designeffekt (deft) in Abhängigkeit von Homogenität (ICC) der Fälle innerhalb der Adressen eines Erhebungsbeauftragten und der Fallzahl: Bei hoher Arbeitsbelastung wird ein deft > 2 auch bei kleinen ICCs erreicht. Die Breiten der Konfidenzintervalle verdoppeln sich dann.

---

## 8.5. Festlegung von Haupt- und Nebenwohnsitzen

Ziel 1 des Zensus ist die Feststellung der amtlichen Einwohnerzahl. Das bedeutet, dass jede existente Person nur einmal, mit Hauptwohnsitz oder alleiniger Wohnung, zu zählen ist. Hat eine Person mehrere Wohnsitze, muss also festgelegt werden, welcher Wohnsitz zählungsrelevant ist. Um dies zu erreichen, wurde im Zensusgesetz die Mehrfachfallprüfung angeordnet. Im Zensus 2022 erfolgte die Mehrfachfallprüfung rein maschinell und der zählungsrelevante Wohnsitz wurde regelbasiert festgelegt. Dabei wurden die Informationen nur aus den Melderegistern genutzt. Falls eine der infrage kommenden Anschriften in der Haushaltsstichprobe war und die Person einen anderen Hauptwohnsitz, als maschinell zugeordnet angab, war das Befragungsergebnis nicht hochrechnungsrelevant.

Da Melderegister nicht durchgängig aktuell und fehlerfrei sind, kann auch eine Festlegung von Haupt- und Nebenwohnsitzen, die ausschließlich auf den dort enthaltenen Informationen beruht, nicht fehlerfrei erfolgen. In der Summe kann dies bundesweit zu einer durchaus relevanten Anzahl fehlerhaft zugeordneter Hauptwohnungen führen.

Die im Rahmen der Stichprobenerhebung gewonnenen Auskünfte bleiben für die Hochrechnung unberücksichtigt; hochgerechnet wird ausschließlich der im Zuge der Mehrfachfallprüfung festgelegte Wohnsitz. Dadurch werden die tatsächlichen Lebensverhältnisse in bestimmten Konstellationen (z. B. bei Wochenendpendlern oder Personen mit getrennter Lebensführung) nicht adäquat abgebildet.

Da zudem in diesem Prozessschritt keine Einbindung der Erhebungsstellen erfolgte – etwa durch die Berücksichtigung von Wahllisten für die Teilmenge der Wahlberechtigten oder durch eine Kontaktaufnahme mit Personen mit mehreren Wohnsitzen – standen den Erhebungsstellen keine Möglichkeiten zur Korrektur zur Verfügung.

## 8.6. Umgang mit Prozessfehlern

Jeder Datenanalytiker weiß, dass in einem Datenanalyseprozess hunderte von Fehlern durch Fehler in Programmen, falsche Eingaben, Verwechslungen von Dateien, Verarbeiten veralteter Dateien, Versionskonflikte etc. möglich sind. Entsprechend schwierig ist die reine Reproduzierbarkeit statistischer Analysen (Xiong & Cribben, 2023). Eine einzelne Analysegruppe bemerkt solche Probleme selten allein: Die Schwierigkeiten zeigen sich erst bei einem unabhängigen Reproduktionsversuch oder eklatanten Widersprüchen zu anderen Daten.

Man kann einzelne besonders hohe Abweichungen zwischen der Fortschreibung und der hochgerechneten Einwohnerzahl einer Gemeinde als einen solchen Widerspruch eines Analyseergebnisses mit anderen Ergebnissen interpretieren. Aus statistischer Sicht ist eine

---

solche Beobachtung zunächst nur ein Ausreißer (outlier). Für Statistiker sind Ausreißer von besonderem Interesse, da diese häufig entweder auf Modell- oder Datenfehler hinweisen.

Ein Statistisches Landesamt oder das Statistische Bundesamt sollte dann bemüht sein, durch Outlier-Analysen zu klären, ob solche Prozess- oder Datenfehler vorlagen.

Nach unserem Wissen wurden solche Analysen nicht publiziert und entsprechende Fragen nicht fachlich beantwortet.<sup>23</sup>

Die notwendigen Überprüfungen der Berechnungen oder Eingabedaten kann nur das Statistische Bundesamt durchführen, da nur dieses die entsprechenden Daten besitzen darf. Die Tatsache, dass eine Gemeinde mangels dieser Daten keine Fehler nachweisen kann, kann nicht als Argument genutzt werden, dass es keine Fehler gäbe. Diese Argumentationsfigur wird in wissenschaftlichen Kontexten zumeist mit dem Satz »Absence of evidence is not evidence of absence« (Altman & Bland, 1995) kommentiert.

---

<sup>23</sup> Beispiel: In einem Widerspruchsbescheid des Hessischen Statistischen Landesamtes (Hessisches Statistisches Landesamt, 2025) wird die Möglichkeit von einfachen Prozessfehlern oder Datenerhebungsfehlern in keiner Weise diskutiert oder der Widerspruch zum Anlass genommen, die Berechnungen zu überprüfen oder Outlier-Analysen durchzuführen. Stattdessen wird ausschließlich darauf hingewiesen, dass das HSL das Verfahren für verfassungsrechtlich zulässig hält. Es scheint wichtiger, dass ein Ergebnis rechtskonform ist, als dass es inhaltlich korrekt ist. So schreibt das HSL: »Dabei wurde im Zensus 2022 zur Hochrechnung ein im Rahmen des Zensus 2011 bereits höchstrichterlich anerkanntes Verfahren eingesetzt, das zwei Ansätze miteinander verbindet.« Dass der gemeinte GREG-Schätzer auf Annahmen beruht, die schlicht nicht erfüllt sein können, wird ebenso wenig thematisiert wie die Tatsache, dass auch ein perfekter Schätzer mit fehlerhaften Daten zu irreführenden Ergebnissen führen muss.

## 9. Wiederholungsbefragung

Es entspricht internationalen Standards, Zensusergebnisse durch unabhängige Kontrollzählungen zu überprüfen. Die Vereinten Nationen empfehlen daher in ihrem Handbuch für die Durchführung eines Zensus die anschließende Erhebung eines »Post-Enumeration Surveys« (PES) (United Nations Secretariat, 2010).<sup>1</sup> Bei einem PES wird kurz nach einem Zensus eine **unabhängige** Stichprobe (meist aus geografischen Flächen) gezogen und für die ausgewählten Gebiete dann die Bevölkerungszahl vollkommen unabhängig vom Zensus ermittelt.

	nicht im	
	PES	PES
im Zensus	$n_{11}$	$n_{10}$
nicht im Zensus	$n_{01}$	$n_{00}$

Tabelle 9.1.: Vergleich eines Zensus mit einem Post-Enumeration Survey (PES)

Die Zelle  $n_{00}$  enthält die Fälle, die sowohl vom Zensus als auch vom PES nicht erfasst werden. Diese Zahl kann aber – bei Unabhängigkeit von PES und Zensus – ebenso wie die anderen Zahlen der Tabelle geschätzt werden. Daher lassen sich sowohl eine geschätzte Bevölkerungszahl  $\hat{N}$  als auch Fehlerraten berechnen.<sup>2</sup>

Die Voraussetzung für die Berechnung ist die vollständige Unabhängigkeit der Erhebungen, sonst sind die Annahmen für die Herleitung der Schätzungen nicht gegeben.

<sup>1</sup> In manchen Ländern wird diese Art der Erhebung auch als *Census Coverage Survey* bezeichnet.

<sup>2</sup> Die geschätzte Bevölkerungszahl  $\hat{N}$  ergibt sich als

$$\hat{N} = \frac{(n_{11} + n_{10})(n_{11} + n_{01})}{n_{11}}. \quad (9.1)$$

---

## 9.1. Der Unterschied zwischen der Wiederholungsbefragung und einem PES

In Deutschland gab es weder 2011 noch 2022 einen PES, sondern sogenannte »Wiederholungsbefragungen« (WDH). Das Zensusgesetz 2022 legt in §22 fest:

» (1) Zur Prüfung der Qualität der in der Haushaltsstichprobe und den Erhebungen an Adressen mit Sonderbereichen ermittelten Einwohnerzahl sind repräsentative Wiederholungsbefragungen durch das zuständige statistische Landesamt durchzuführen. Auswahlseinheiten sind die nach §12 ausgewählten Adressen und die nach §14 erfassten Adressen mit Sonderbereichen, an denen keine Gemeinschaftsunterkünfte bestehen. Es ist ein Auswahlatz von höchstens 4 Prozent der an den nach §12 ausgewählten Adressen und den nach §14 erfassten Wohnheimen wohnenden Personen zugrunde zu legen.«

Das Entscheidende im §22 ist die unscheinbare Formulierung »an den nach §12 ausgewählten Adressen«. Damit sind Erhebungen nur noch an bereits bekannten Adressen möglich. Durch diese Festlegung können nicht erfasste Adressen nicht mehr nachträglich festgestellt werden. Technisch werden solche Adressen als »Undercoverage« bezeichnet (siehe Kapitel 8). Die WDH kann kein Undercoverage von Adressen entdecken. Damit ist die Wiederholungsbefragung kein PES.

Da dieser Punkt so wichtig ist, soll nochmals betont werden, dass bei der Wiederholungsbefragung des Zensus in Deutschland nur Personen an bereits ermittelten Adressen gezählt werden. Wurden die Adressen unvollständig erhoben, dann hat eine Person an einer nicht erfassten Adresse keine Chance, in die Wiederholungsbefragung aufgenommen zu werden. Folglich kann mit der Wiederholungsbefragung die Qualität der Zensusergebnisse prinzipiell nicht unabhängig validiert werden.

## 9.2. Zur Durchführung der WDH

Die Wiederholungserhebung erfolgte *nicht* unabhängig vom Zensus. Zwar wurden anscheinend im Gegensatz zum Zensus 2011 nicht mehr die Nachnamen aus dem Zensus an die Erhebungsbeauftragten der Wiederholungszählung übermittelt, aber zumindest in einigen Erhebungsgebieten die Anzahl der im Zensus ermittelten Personen an die Erhebungsbeauftragten der Wiederholungszählung.<sup>3</sup>

---

<sup>3</sup> Bedauerlicherweise war es uns nicht möglich, eine Dokumentation zum Prozedere bei der Wiederholungsbefragung von Destatis oder einzelnen Landesämtern zu erhalten. Da die Schulungsunterlagen

---

Weiterhin wurden teilweise (auch dies ist nicht flächendeckend dokumentiert) dieselben Personen für den Zensus und die Wiederholungsbefragung eingesetzt.<sup>4</sup> Selbst wenn die Erhebungsbeauftragten andere Erhebungsbezirke in der WDH als im Zensus bearbeitet haben, sind WDH und Zensus nicht unabhängig.

Schließlich war die WDH sowohl personell als auch zeitlich sehr begrenzt. Zum Beispiel hatten in Niedersachsen die Erhebungsbeauftragten mehrheitlich eine Bearbeitungszeit von 4 Wochen. Das Landesamt für Statistik Niedersachsen (LSN) erwähnt lediglich zwei terminlich angekündigte Kontaktversuche innerhalb dieses Zeitraums (LSN, 2023). Dabei waren die von den Erhebungsbeauftragten zu bearbeitenden Flächen in einigen Ländern durchaus groß: Görner (2024) gibt an, dass jede Person pro Verteilungsrunde nicht mehr als 250 zu Befragende im Umkreis von maximal 50 km um den angegebenen Wohnsitz zugeteilt bekommen sollte.

### 9.3. Veröffentlichung der Ergebnisse der WDH

Es gibt keine Dokumentation der Ergebnisse der Wiederholungsbefragung. Der Qualitätsbericht des Zensus (Statistisches Bundesamt, 2024) (siehe Abbildung 2.2) verweist in Abschnitt 4.3 auf Ziffer 1.8.2. Dieser Abschnitt enthält kein Wort zur Wiederholungsbefragung. Die Wiederholungsbefragung wird im Qualitätsbericht dann nicht mehr erwähnt. Es gibt auch keine andere Textveröffentlichung des Statistischen Bundesamtes zu den Ergebnissen der Wiederholungsbefragung. Auf unsere erste Nachfrage<sup>5</sup> teilt das Statistische Bundesamt mit:

»Derzeit liegen keine Veröffentlichungen zur Wiederholungsbefragung vor, und es sind auch keine weiteren Veröffentlichungen geplant.«

Auf unsere Nachfrage unter Hinweis darauf, dass das Zensusgesetz 2022 in §22 Wiederholungsbefragungen zur Qualitätsbewertung durch das zuständige statistische Landesamt vorsieht, antwortet das Statistische Bundesamt:<sup>6</sup>

---

der Wiederholungsbefragung nicht veröffentlicht wurden, lässt sich dies nur durch Berichte einzelner Erhebungsstellen belegen.

4 So betrachtet z.B. das LSN den Einsatz derselben Personen als »besonders vorteilhaft« und berichtet von einem Anteil von rund 3/4 der Erhebungsbeauftragten der Wiederholungsbefragung, die bereits in der Ziel-1-Stichprobe eingesetzt wurden (LSN, 2023).

5 Anfrage vom 3.11.2025, Antwort vom 6.11, GZ 120111/819044

6 Unsere Anfrage am 6.11.2025, Antwort des Statistischen Bundesamts am 21.11.2025, GZ 120111/819314.

---

»Dieser Berichtspflicht ist Destatis nachgekommen und kann bei Eurostat nachvollzogen werden: <https://ec.europa.eu/CensusHub/selectHyperCubeDataSets?qhc=true&clearSession=true&language=en> (Domain: Census 2021 QHC; Merkmal COV\_IND und hier under-coverage und over-coverage auswählen).«

Auf der Seite von Eurostat war diese Information nicht auffindbar. Es gelang auch Eurostat an zwei Werktagen nicht, diese Information auf ihrer eigenen Seite zu finden.<sup>7</sup> Danach berichtete Eurostat, dass es ein technisches Problem gäbe, welches den Zugang zu diesen Daten verhindere.

Dieser Sachverhalt ist vor dem Hintergrund zu sehen, welche Datenerfordernisse Eurostat (2019) an die Ergebnisse der Zensusrunde 2021 im Abschnitt 5.2 in Hinsicht auf die Qualität des Zensus stellt:<sup>8</sup>

»Member States must provide a coverage assessment of their census population bases at national level, evaluating the incidence of record imputations and deletions, as well as the estimated resulting over- and under-coverage, in other words the efforts made so that the ‘census population’ matches the ‘target population’ (quality hypercube QHC1). In the other quality hypercube (QHC2), the focus is on measuring the data sources of several selected topics as well as the incidence of imputed and missing information, in other words the degree to which the information on a topic is based on real observations.«

Dem Buchstaben der Guidelines mag Destatis dieser Verpflichtung nachgekommen sein, dem Sinn nach aber in keiner Weise: Eine Qualitätsbeurteilung ohne Verlinkung von Destatis auf einer nur indirekt mit Schwierigkeiten abrufbaren Eurostat-Seite zu veröffentlichen, widerspricht den Regeln des »Code of Practice«. Ebenso ist die Tatsache, dass das Zensusgesetz keinerlei Konsequenz infolge einer gegebenenfalls unzureichenden Qualität festlegt und auch die Veröffentlichung nicht explizit vorschreibt, schwer nachvollziehbar.<sup>9</sup>

---

7 Eurostat Anfrage am 24.11, Gz: EN39152

8 <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-18-010>

9 Bereits in 2011 mangelte es an Dokumentationen zur Erhebung und auch zur Wiederholungsbefragung, was zuvor in (Küchenhoff, 2014) bemängelt wurde: »Da aber keine unabhängigen Kontrollzählungen ›Post Enumeration Surveys‹ durchgeführt wurden und eine Auswertung der Wiederholungsbefragung nicht vorliegt, ist eine Quantifizierung des Messfehlers und dessen Wirkung die Güte der Schätzung der Einwohnerzahl nicht möglich.«

## 9.4. Ergebnisse der Wiederholungsbefragung

Auf unsere dritte Anfrage an Destatis wurde eine Excel-Datei mit einer Tabelle geliefert, die dann auch wieder auf der Eurostat-Seite zugänglich war.<sup>10</sup> Die Tabelle 9.1 ist diese Tabelle in voller Länge, ohne jede Kürzung. Auf der Seite bei Eurostat oder bei Destatis gibt es keine weiteren Publikationen. Wir möchten zwei Dinge betonen:

Data on		DF_QHC1_2021 by GEO, SEX, AGE, COV_IND		
GEOGRAPHICAL AREA	Germany			
COVERAGE INDICATOR	Under-coverage (estimated)			
Age/Sex	Total	Female	Male	
Total	1274121	623094	651027	
under 15 years	210514	111348	99166	
15 to 29 years	340788	159941	180847	
30 to 49 years	327153	143577	183576	
50 to 64 years	233877	112669	121208	
65 to 84 years	140804	81143	59661	
85 years and over	20985	14416	6569	
GEOGRAPHICAL AREA	Germany			
COVERAGE INDICATOR	Over-coverage (estimated)			
Age/Sex	Total	Female	Male	
Total	3585352	1786796	1798556	
under 15 years	561239	283981	277258	
15 to 29 years	558845	252929	305916	
30 to 49 years	1006611	489379	517232	
50 to 64 years	839979	411123	428856	
65 to 84 years	516203	283243	232960	
85 years and over	102475	66141	36334	

Abbildung 9.1.: Die Tabelle zur Wiederholungsbefragung vom Eurostat CensusHub

1. Diese Tabelle ist das **einzig**e Ergebnis von Destatis zur Wiederholungsbefragung.
2. Die Beschaffung dieser Tabelle hat mehrere Wochen gedauert und die wiederholte Inanspruchnahme des Auskunftsdienstes von Destatis und des Helpdesks von Eurostat erfordert.

<sup>10</sup> <https://ec.europa.eu/CensusHub/selectHyperCubeDataSets?qhc=true&clearSession=true&language=en> Dann: Domain: Census 2021 QHC; Merkmal COV\_IND und hier under-coverage und over-coverage auswählen.

---

Beide Tatsachen sind schwer mit den Ansprüchen des »European Code of Practice« vereinbar. Der Grundsatz 15 lautet:

»Die europäischen Statistiken werden klar und verständlich präsentiert, in geeigneter und benutzerfreundlicher Weise veröffentlicht und sind zusammen mit einschlägigen Metadaten und Erläuterungen entsprechend dem Grundsatz der Unparteilichkeit verfügbar und zugänglich.«

Es gibt keinerlei methodische Erläuterungen, wie diese Zahlen zustande gekommen sind. Eine Zahl ohne Kontext und ohne Erläuterung ihres Zustandekommens ist statistisch wenig bedeutsam und aus methodischer Sicht unakzeptabel. Dass für diese kaum wirklich veröffentlichte Tabelle eine Befragung von mehr als 400 000 Personen notwendig war, ist schwer vermittelbar. Es wäre daher eine detaillierte Dokumentation des Prozesses und der Ergebnisse (z.B. eine getrennte Auswertung nach Bundesland oder Ausländern) notwendig, um irgendetwas aus der WDH lernen zu können. Der gesetzliche Auftrag sah explizit »Prüfung der Qualität der in der Haushaltsstichprobe und den Erhebungen an Anschriften mit Sonderbereichen ermittelten Einwohnerzahl« vor. Die Tabelle 9.1 ist aus methodischer Sicht keine Prüfung der Qualität der Einwohnerzahl.

## 9.5. Zusammenfassung

Aufgrund der oben erwähnten mehrfachen Verletzung der Unabhängigkeitsannahmen (Wiederholung statt PES durch Verwendung gleicher Adressen, Nennung der Zahl der zu findenden Personen, Überschneidung beim Personal) sind die Ergebnisse der WDH gegenüber einem unabhängigen PES ohnehin Unterschätzungen und international kaum vergleichbar. Die einzige Auszählung als Tabelle 9.1 genügt aus methodischer Sicht nicht als Ergebnis der gesetzlich vorgeschriebenen Beurteilung der Qualität der Einwohnerzahl. Die Tabelle wurde nicht in einer Veröffentlichung des Statistischen Bundesamts oder der Homepage von Destatis veröffentlicht, sondern unverlinkt auf einer nicht direkt abrufbaren Seite auf Eurostat, was nach Ansicht der Gutachter dem »European Code of Practice« widerspricht.

# 10. Zusammenfassung: Wesentliche Kritikpunkte des Zensus 2022

1. Die Dokumentation ist so unzureichend, dass weder die Gemeinden noch unabhängige Wissenschaftler die Durchführung, Korrekturen und Hochrechnungen nachvollziehen können (siehe Kapitel 2).
2. Die beiden Modelle zur Generierung der Stichprobenauswahlsätze (siehe dazu Kapitel 5.8) und zur Hochrechnung (siehe Kapitel 6.9) sind nicht aufeinander abgestimmt. Der geschätzte Stichprobenumfang ist zu klein für die komplexe Spezifikation des Hochrechnungsmodells.
3. Das Hochrechnungsverfahren führt bei kleinen Gemeinden zu lediglich »plausiblen« Schätzungen zulasten tatsächlich nachgewiesener Personen (siehe Kapitel 6.6).
4. Die Fehlerberechnungen des Hochrechnungsmodells basieren auf optimistischen Annahmen, die nicht erfüllt waren (siehe Kapitel 6.4). Entsprechend sind die Schätzungen der Einwohnerzahlen mit deutlich größerer Unsicherheit behaftet, als es die Datenblätter zur Ermittlung der amtlichen Einwohnerzahl suggerieren.
5. In der Mehrfachfallprüfung und in der »erhebungsteilübergreifenden Plausibilisierung« (EÜPL) werden Record-Linkage-Verfahren verwendet, über deren Linkage Bias nichts bekannt ist (siehe Kapitel 8.3.4).
6. Es wurde auf die Quantifizierung aller Nicht-Stichprobenfehler verzichtet (siehe Kapitel 8). Damit führen die Fehlerberechnungen prinzipiell zu Unterschätzungen: Die tatsächlichen Fehlermargen sind wesentlich größer als die ausgewiesenen.
7. Die Wiederholungsbefragung erlaubt über ihre Konzeption keine unabhängige Beurteilung der Qualität der Zensusergebnisse und wurde weder detailliert analysiert noch angemessen publiziert (siehe Kapitel 9.5).

# 11. Schlussbemerkungen

Die zentralen Probleme des Zensus gehen nicht auf die Mitarbeiter der Statistischen Landesämter oder des Statistischen Bundesamtes zurück, sondern auf die gesetzlichen Rahmenbedingungen.

Der damalige Präsident der »Royal Statistical Society«, Sir David Spiegelhalter, hat in seiner Antrittsrede (Spiegelhalter, 2017) die Rolle des Vertrauens in statistische Ergebnisse thematisiert. Unter Bezug auf die Arbeiten von O’Neill (2003) stellt Spiegelhalter fest, dass Vertrauen nicht einfach erwartet werden kann, sondern Vertrauenswürdigkeit demonstriert werden muss. Dies erfordert mehr als eine einfache Transparenzforderung, sondern die Notwendigkeit, dass Informationen über die Herstellung von Daten

1. zugänglich,
2. verständlich,
3. überprüfbar und
4. nützlich

sein müssen.<sup>1</sup> Wie wir gezeigt haben, wird weder die Transparenzforderung noch eines dieser vier Kriterien durch die Dokumentation des Zensus 2022 erfüllt.

---

<sup>1</sup> Im Original: »accessible, intelligible, assessable, and useable«; diese Kriterien stammen aus einem Report der Royal Society (2012).

# Anhang

# A. Kurzbiografien der Gutachter

## A.1. Prof. Dr. Rainer Schnell

Rainer Schnell promovierte 1986 über Missing-Data-Schätztechniken in der empirischen Sozialforschung an der Ruhr-Universität Bochum im Fachgebiet Sozialwissenschaften. 1996 erfolgte die Habilitation an der Universität Mannheim über Nonresponse in Bevölkerungsumfragen. Von 1996 bis 2008 war er Professor für Methoden der empirischen Politik- und Verwaltungsforschung an der Universität Konstanz. Nach seiner Berufung war er von 2008 bis 2025 Professor für Methoden der empirischen Sozialforschung an der Universität Duisburg-Essen. Die Tätigkeit dort wurde von 2015 bis 2017 unterbrochen durch seine Tätigkeit als Direktor des Centre for Comparative Social Surveys an der City University London. Zusätzlich war er von 2017 bis 2020 »Adjunct Professor« an der »Faculty of Health Science« der Curtin University in Perth (Australien). Seit 2025 ist er Seniorprofessor an der Universität Duisburg-Essen. Rainer Schnell war 2016 Gastwissenschaftler am »Isaac Newton Institute for Mathematical Studies«, Cambridge (UK) und 2023 am »National Institute of Informatics« (NII) in Shonan (Japan). Er war für die amtliche Statistik in Luxemburg, dem Vereinigten Königreich und Deutschland als Gutachter bzw. Auftragnehmer tätig. Gemeinsam mit der amtlichen Statistik der Niederlande hat er an Projekten zur Verknüpfung von Datenbanken mit Erhebungen der amtlichen Statistik gearbeitet. Schnell ist Mitglied des Gutachtergremiums für den Wissenschaftspreis des Statistischen Bundesamtes und des wissenschaftlichen Beirats des Viktimisierungssurveys des Bundeskriminalamtes. Sein Arbeitsschwerpunkt ist der Einfluss psychologischer und sozialer Bedingungen der Datenerhebungen auf die Ergebnisse statistischer Berechnungen sowie mathematischer Verfahren ihrer Korrektur, vorrangig durch die Verknüpfung mit populationsabdeckenden Datenbanken durch Record-Linkage.

## A.2. Prof. Dr. Rainer Lenz

Rainer Lenz promovierte im Jahr 2002 am Fachbereich Mathematik der TU Kaiserslautern über Algebraisch abgeschlossene Klassen graphentreuer Funktionen. Danach arbeitete er vier Jahre im Statistischen Bundesamt und übte parallel dazu diverse Lehraufträge zu

---

Mathematik und Statistik an verschiedenen Hochschulen aus, bis er im Jahr 2006 einem ersten Ruf an die Hochschule Mainz folgte. Er war zudem im Jahr 2009 als Gastprofessor für das Fachgebiet »Financial Decision Analysis« an der Leeds Metropolitan University tätig, ehe er im Jahr 2010 einen weiteren Ruf an die HTW des Saarlandes annahm. Dort wurde ihm die Leitung des Forschungsinstituts für Diskrete Mathematik und Angewandte Statistik übertragen. Gleichzeitig habilitierte er sich an der Statistischen Fakultät der TU Dortmund am Lehrstuhl für Wirtschafts- und Sozialstatistik und bietet seitdem als Privatdozent bis heute regelmäßig Lehrveranstaltungen in der Statistischen Fakultät an. Nach der Annahme eines weiteren Rufs im Jahr 2016 ist er an der TH Köln als Professor für Methoden der Angewandten Statistik tätig. Der Schwerpunkt seiner Forschungstätigkeit liegt im Bereich der Angewandten Statistik und Mathematischen Optimierung. Neben der Veröffentlichung von über 100 Fachbeiträgen berät er seit 20 Jahren Eurostat als nationaler Sachverständiger und Unternehmen der Privatwirtschaft bei der Lösung von Problemen der Datenintegration und -hochrechnung sowie der Umsetzung der DSGVO in die Praxis. Überdies wird er regelmäßig als Gutachter von einer Reihe internationaler Fachzeitschriften bestellt. In den vergangenen 25 Jahren hat er in zahlreichen nationalen und internationalen Forschungsprojekten in verschiedenen Funktionen als Kooperationspartner, Projektleiter, wissenschaftlicher Berater und wissenschaftlicher Beirat mitgewirkt.

### **A.3. Diplom-Volkswirt Rolf Schmidt**

Nach der Ausbildung zum Industriekaufmann folgte das Studium der Wirtschaftswissenschaften an der Universität Duisburg-Essen mit dem Abschluss als Diplom-Volkswirt im März 1978. Von 1978 bis 2021 war Rolf Schmidt beim Landesbetrieb für Information und Technik (IT.NRW, früher Statistisches Landesamt) in der amtlichen Statistik in unterschiedlichen Aufgabenbereichen beschäftigt. Zunächst war er als Referent für Methodik, Erhebung und Aufbereitung tätig und für Analysen in unterschiedlichen Wirtschafts-, Bevölkerungs- und Verwaltungsstatistiken verantwortlich. Im Großprojekt Volkszählung 1987 war er zuständig für die Organisation und den Personaleinsatz der ca. 1 000 Beschäftigten der Zählung.

Ab 2005 war er Fachbereichsleiter für Datenerhebungen und das Unternehmensregister an den Standorten Düsseldorf, Oberhausen und Paderborn. Von 2010 bis zum 31.03.2017 war Rolf Schmidt Geschäftsbereichsleiter für die amtliche Statistik in Nordrhein-Westfalen. Nach dem Ausscheiden aus dem aktiven Dienst ist Rolf Schmidt seit dem Wintersemester 2017/2018 in der Lehre über amtliche Statistik an der Universität Duisburg-Essen, der Ruhruniversität Bochum sowie der TU Dortmund tätig.

# Literatur

- Altman, D. G., & Bland, J. M. (1995). Absence of evidence is not evidence of absence. *BMJ*, 311(7003), 485. <https://doi.org/10.1136/bmj.311.7003.485>
- Baffour, B., & Valente, P. (2012). An evaluation of census quality. *Statistical Journal of the IAOS*, 28(3-4), 121–135. <https://doi.org/10.3233/SJI-2012-0752>
- Berg, A., & Bihler, W. (2014). Das Hochrechnungsverfahren zur Ermittlung der Einwohnerzahl im Zensus 2011. *Wirtschaft und Statistik*, 229–235.
- Biemer, P. P., de Leeuw, E., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L. E., Tucker, N. C., & West, B. T. (Hrsg.). (2017). *Total Survey Error in Practice*. Wiley.
- Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to Survey Quality*. Wiley.
- Bleninger, S., Fürnrohr, M., Kiesl, H., Krämer, W., Küchenhoff, H., Burgard, J. P., Münnich, R., & Rupp, M. (2020). Kommentare und Erwiderung zu: Qualitätsziel-funktionen für stark variierende Gemeindegrößen im Zensus 2021. *ASTA Wirtschafts- und Sozialstatistisches Archiv*, 14(1), 67–98.
- Bohensky, M. (2016). Bias in Data Linkage Studies. In K. Harron, H. Goldstein & C. Dibben (Hrsg.), *Methodological Developments in Data Linkage* (S. 63–82). Wiley.
- Brändle, T., Weiland, S. V., & Schnell, R. (2025). Perspektiven für ein bundesweites Bildungsverlaufsregister. Eine Analyse des Zentralen Schülerregisters Hamburg. *DDS – Die Deutsche Schule*, 2025(3), 192–204. <https://doi.org/10.31244/dds.2025.03.04>
- Bretnütz, H., Estatico, S., Walther, S., Hüsgen-Brodhacker, F., Lange, K., Kleber, B., & Haberzettl, B. (2024). Aufbereitung und Plausibilisierung der primärstatistischen Erhebungsteile im Zensus 2022. *WISTA – Wirtschaft und Statistik*, (6), 51–59.
- Bretschi, C., & Lorentz, K. (2019). Präzisionsziele für die Ermittlung der Einwohnerzahl bei der Haushaltsstichprobe im Zensus 2021. *Wirtschaft und Statistik, Sonderheft Zensus 2021*, 12–22.
- Bretschi, C., Seibel, S., Vorndran, I., & Pfahl, M. (2024a). Ermittlung der Einwohnerzahl im Zensus 2022. *WISTA Wirtschaft und Statistik*, 2024(6), 17–28.
- Bretschi, C., Seibel, S., Vorndran, I., & Pfahl, M. (2024b). Ermittlung der Einwohnerzahl im Zensus 2022, Methodenpapier, 12.8.2024, Version 1.
- Bundesverwaltungsamt. (2023). *Erprobung Identitätsdatenabruf (IDA) mit dem Nationalen Waffenregister* (Kurzbericht). Bundesverwaltungsamt. Köln.

- 
- Burgard, J., Münnich, R., & Rupp, M. (2019). Die Entwicklung des Stichprobenkonzepts für den Zensus 2021. *Wirtschaft und Statistik, Sonderheft Zensus 2021*, 23–35.
- Burgard, J., Münnich, R., & Rupp, M. (2020). Qualitätszielfunktionen für stark variierende Gemeindegrößen im Zensus 2021. *AStA Wirtschafts- und Sozialstatistisches Archiv*, 14(1), 5–65.
- Christen, P. (2012). *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Springer.
- Christen, P., Ranbaduge, T., & Schnell, R. (2020). *Linking Sensitive Data: Methods and Techniques for Practical Privacy-Preserving Information Sharing*. Springer.
- Christen, P., & Schnell, R. (2023). Thirty-Three Myths and Misconceptions About Population Data: From Data Capture and Processing to Linkage. *International Journal of Population Data Science*, 8(1). <https://doi.org/10.23889/ijpds.v8i1.2115>
- Christensen, B., Christensen, S., Hoppe, T., & Spandel, M. (2015). Everything counts! *AStA Wirtschafts- und Sozialstatistisches Archiv*, 9(3-4), 215–232.
- Cochran, W. G. (1977). *Sampling Techniques* (3. Aufl.). Wiley.
- Crossley, T. F., Schmidt, T., Tzamourani, P., & Winter, J. K. (2021). Interviewer Effects and the Measurement of Financial Literacy. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 184(1), 150–178.
- Deutscher Städtetag & Deutscher Landkreistag. (2022). Zensus 2022: Brief an das Bundesministerium des Innern und für Heimat, 14.7.2022.
- Deville, J.-C., & Sarndal, C.-E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, 87(418), 376–382.
- Elliott, M. R., & West, B. T. (2015). “Clustering by Interviewer”: A Source of Variance That Is Unaccounted for in Single-Stage Health Surveys. *American Journal of Epidemiology*, 182(2), 118–126. <https://doi.org/10.1093/aje/kwv018>
- European Commission. Statistical Office of the European Union. (2018). *Verhaltenskodex für europäische Statistiken: für die nationalen statistischen Ämter und Eurostat*. Publications Office.
- Eurostat. (2019). *EU legislation on the 2021 population and housing censuses. Explanatory Notes*. Publications Office of the European Union. <https://doi.org/doi:10.2785/513433>
- Gehr, L. *Eine kritische Betrachtung der Methoden und Probleme des Zensus 2022*. Bachelorarbeit, HVF, Ludwigsburg, 2024. [https://opus-hslb.bsz-bw.de/frontdoor/deliver/index/docId/5234/file/Gehr\\_Laura\\_PM\\_2023\\_BA.pdf](https://opus-hslb.bsz-bw.de/frontdoor/deliver/index/docId/5234/file/Gehr_Laura_PM_2023_BA.pdf)
- Gesellschaft für innovative Sozialforschung und Sozialplanung e.V. (GISS) & Kantar Public. (2024). *Empirische Untersuchung zum Gegenstand nach §8 Abs. 2 und 3*

- 
- WoBerichtsG (Forschungsbericht). Bundesministerium für Arbeit und Soziales. Berlin. <https://www.bmas.de/>
- Görner, I. (2024). Zensus 2022: Wiederholungsbefragung: Aufgaben der Erhebungsstelle. *Bayern in Zahlen, 2024* (7), 30–37.
- Groves, R. M. (2004). *Survey Errors and Survey Costs*. Wiley.
- Groves, R. M., Fowler, F. J., Couper, M. P., Lebkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2. Aufl.). Wiley.
- Hessisches Statistisches Landesamt. (2025, 23. Juni). Feststellung der Einwohnerzahl im Rahmen des Zensus 2022, Widerspruchsbescheid, ZC-JUR06/25.
- Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Lenz, R., Naylor, J., Nordholt, E. S., Seri, G., Wolf, P.-P. D., Tent, R., Młodak, A., Gussenbauer, J., & Wilak, K. (2025). *Handbook on Statistical Disclosure Control, Second Edition*. Centers of Excellence.
- Hüsgen-Brodhäcker, F. (2024). *Erhebungsteilübergreifende Plausibilisierung im Referenzdatenbestand des Zensus 2022, Methodenpapier 17.2.2024*. Statistisches Bundesamt.
- Johnson, P. E. (2014). *Multicollinearity in Regression, Präsentation*, University of Kansas, Center for Research Methods und Data Analysis.
- Kierans, D., & Kraler, A. (Hrsg.). (2025). *Handbook on Irregular Migration Data: Concepts, Methods and Practices*. University of Krems Press.
- Kish, L. (1965). *Survey Sampling*. Wiley.
- Kish, L. (1995). Methods for Design Effects. *Journal of Official Statistics*, 11(1), 55–77.
- Klein, M. (2022). Registermodernisierung – ein Überblick Europäische Perspektive inklusive [newspaper]. *www.egovernment.de*. <https://www.egovernment.de/europaeische-perspektive-inklusive-a-51a5f2058ba5de272e61939d3c453319>
- Klink, S., & Lorentz, K. (2022). Auswahlplan und Stichproben-Hauptziehung für den Zensus 2022. *WISTA*, 1, 13–24.
- Knobelspies, M., & Münnich, R. (2008). Variablenselektion bei gebundener Hochrechnung. *Austrian Journal of Statistics*, 34(3/4), 335–347.
- Kreuter, F. (Hrsg.). (2013). *Improving Surveys with Paradata: Analytic Uses of Process Information*. Wiley.
- Krug, W., Nourney, M., & Schmidt, J. (2001). *Wirtschafts- und Sozialstatistik: Gewinnung von Daten* (6. Aufl.). Oldenbourg.
- Küchenhoff, H. (2014). Gutachten zur Verwaltungsrechtssache Stadt Bremerhaven gegen Freie Hansastadt Bremen.
- Lenz, R. (2006). Measuring the disclosure protection of micro aggregated business microdata – an analysis taking as an example the German Structure of Costs Survey. *Journal of Official Statistics*, 22(4), 681–710.

- 
- Lenz, R., Rosemann, M., Vorgrimler, D., & Sturm, R. (2006). Anonymising Business Micro Data – Results of a German Project. *Schmollers Jahrbuch, Journal of Applied Social Science Studies*, 126(4), 635–651.
- Lessler, J. T., & Kalsbeek, W. D. (1992). *Nonsampling Error in Surveys*. Wiley.
- Li, B., Quan, H., Fong, A., & Lu, M. (2006). Assessing record linkage between health care and Vital Statistics databases using deterministic methods. *BMC Health Services Research*, 6(1), 48. <https://doi.org/10.1186/1472-6963-6-48>
- Lohr, S. L. (2022). *Sampling: Design and Analysis* (3. Aufl.). CRC Press.
- Lorentz, K., & Zwick, M. (2025). Die Anschriftenstichprobe des Zensus 2022: Ermittlung der Einwohnerzahl in den Sampling Points. *WISTA*, 2025(2), 17–29.
- LSN. (2022). *Handbuch für Erhebungsbeauftragte: Durchführung der Erhebung, Anlagen*. Landesamt für Statistik Niedersachsen.
- LSN. (2023). Wenn die Interviewerin oder der Interviewer zweimal klingelt. Verfügbar 18. November 2025 unter <https://magazin.statistik.niedersachsen.de/die-wiederholungsbefragung-des-zensus-2022-in-niedersachsen>
- Martin, E. (1999). Who knows who lives here? within-household disagreements as a source of survey coverage error. *Public Opinion Quarterly*, 63(2), 220–236. <https://doi.org/10.1086/297712>
- Martin, E. (2007). Strength of attachment: Survey coverage of people with tenuous ties to residences. *Demography*, 44(2), 427–440. <https://doi.org/10.1353/dem.2007.0014>
- Mediendienst Integration. (2025). Zahl der Flüchtlinge aus der Ukraine in Deutschland. Verfügbar 13. Dezember 2025 unter <https://mediendienst-integration.de/fluechtlinge/ukrainische-fluechtlinge-in-deutschland/wie-viele-fluechtlinge-aus-der-ukraine-leben-in-deutschland>
- Münnich, R. (2019). Schriftliche Stellungnahme zum Entwurf eines Gesetzes zur Durchführung des Zensus im Jahr 2021, Bundestagsdrucksache 19/8693.
- Münnich, R., & Gabler, S. (2012). *Stichprobenoptimierung und Schätzung im Zensus 2011*. Statistisches Bundesamt.
- Münnich, R., Gabler, S., Ganninger, M., Burgard, J. P., & Kolb, J.-P. (2013). *Validierungsprojekt zum deutschen Zensus 2011: Abschlussbericht* (Techn. Ber.) (Version vom 6. März 2013). Universität Trier.
- National Academies of Sciences, E., & Medicine. (2019a). *Methods to foster transparency and reproducibility of federal statistics: Proceedings of a workshop* (M. L. Cohen, Hrsg.). The National Academies Press. <https://doi.org/10.17226/25305>
- National Academies of Sciences, E., & Medicine. (2019b). *Reproducibility and replicability in science*. The National Academies Press. <https://doi.org/10.17226/25303>
- National Academies of Sciences, E., & Medicine. (2020). *Enhancing scientific reproducibility in biomedical research through transparent reporting: Proceedings of a*

- 
- workshop* (T. Wizemann, S. Ncha, A. W. Gee & C. Shore, Hrsg.). The National Academies Press. <https://doi.org/10.17226/25627>
- Olson, K., Smyth, J. D., Dykema, J., Holbrook, A. L., Kreuter, F., & West, B. T. (Hrsg.). (2020). *Interviewer Effects from a Total Survey Error Perspective*. CRC Press.
- O’Neill, O. (2003). *A question of trust*. Cambridge University Press.
- Peters, F., & Lorentz, K. (2025). Die Einwohnerzahlermittlung für verbandsangehörige Gemeinden im Zensus 2022. *WISTA*, (6), 38–47.
- Rendtel, U. (2015). Warum im Zensus die Ergebnisse der Stichprobenmethode keine Benachteiligung der großen Gemeinden darstellen. Eine Detektivarbeit. *ASTA Wirtschafts- und Sozialstatistisches Archiv*, 9(3-4), 233–249.
- Royal Society (Hrsg.). (2012). *Science as an open enterprise: open data for open science*. The Royal Society.
- Särndal, C.-E., Swensson, B., & Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer.
- Schnell, R. (1991). Wer ist das Volk? Zur faktischen Grundgesamtheit bei ‘allgemeinen Bevölkerungsumfragen’: Undercoverage, Schwererreichbare und Nichtbefragbare. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 43(1), 106–154.
- Schnell, R. (2019a). *Survey-Interviews: Methoden standardisierter Befragungen* (2. Aufl.). Springer VS.
- Schnell, R. (2019b, 3. Dezember). *Eignung von Personenmerkmalen als Datengrundlage zur Verknüpfung von Registerinformationen im Integrierten Registerzensus* (WP-GRLC-2019-01). DuEPublico. Duisburg. <https://doi.org/10.17185/DUEPUBLICO/49551>
- Schnell, R. (2025). *Grundkurs Survey-Methodology*, Universität Duisburg-Essen, Vorlesungsmanuskript.
- Schnell, R., & Kreuter, F. (2005). Separating Interviewer and Sampling-Point Effects. *Journal of Official Statistics*, 21(3), 389–410.
- Schnell, R., & Weiland, S. V. (2023). Microsimulation of an Educational Attainment Register to Predict Future Record Linkage Quality. *International Journal of Population Data Science*, 8(1), 01.
- Seber, G. A., & Lee, A. J. (2003). *Linear Regression Analysis* (2. Aufl.). Wiley.
- Spiegelhalter, D. (2017). Trust in Numbers. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 180(4), 948–965. <https://doi.org/10.1111/rssa.12302>
- Statistische Ämter des Bundes und der Länder. (2015). Zensus 2011: Methoden und Verfahren. [https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaeetze\\_Archiv/2015\\_06\\_MethodenUndVerfahren.pdf](https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaeetze_Archiv/2015_06_MethodenUndVerfahren.pdf)
- Statistisches Bundesamt. (1960). *Stichproben in der amtlichen Statistik*. Kohlhammer.
- Statistisches Bundesamt. (2016). *Zensus 2011: Qualitätsbericht*.

- 
- Statistisches Bundesamt. (2024, 20. November). *Zensus 2022* (Qualitätsbericht).
- Statistisches Bundesamt. (2025). Haushaltebefragung – Methoden im Zensus 2022 [Datiert auf den 6.7.2025; Zugriff am 27. Dezember 2025]. <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Zensus2022/Methoden/Befragungen/befragungen-haushaltebefragung.html>
- Statistisches Landesamt Hessen. (2021). *Zensus 2022: Handbuch für die Erhebungsstellen in Hessen – Vorbereitung der Erhebung*.
- Tanur, J. M. (2011). Nonsampling Errors in Surveys. In *International Encyclopedia of Statistical Science* (S. 988–991). Springer.
- Tourangeau, R., Kreuter, F., & Eckman, S. (2012). Motivated underreporting in screening interviews. *Public Opinion Quarterly*, 76(3), 453–469. <https://doi.org/10.1093/poq/nfs033>
- United Nations Department of Economic and Social Affairs. (2021, 18. Oktober). *Handbook on the Management of Population and Housing Censuses: Revision 2*. United Nations. <https://doi.org/10.18356/9789210601467>
- United Nations Secretariat. (2010). *Post Enumeration Surveys: Operational Guidelines* (Techn. Ber.). United Nations Secretariat, Department of Economic und Social Affairs, Statistics Division.
- Verordnung (EU) Nr. 1260/2013 des Europäischen Parlaments und des Rates vom 20. November 2013. (2013). <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32013R1260>
- Vogel, D. (2015). *Update report Germany: Estimated number of irregular foreign residents in Germany (2014)* (Techn. Ber.). Database on Irregular Migration. <https://irregular-migration.net/reports>
- Vorndran, I. (2025). Umgang mit Schutzsuchenden aus der Ukraine im Zensus 2022. Verfügbar 13. Dezember 2025 unter [https://www.zensus2022.de/DE/Wie-funktioniert-der-Zensus/\\_dokumente/Methodenpapier-Einwohnerzahl-Zensus2022.pdf](https://www.zensus2022.de/DE/Wie-funktioniert-der-Zensus/_dokumente/Methodenpapier-Einwohnerzahl-Zensus2022.pdf)
- Warschofsky, M., & Wagenknecht, L. (2021). Der Referenzdatenbestand (RDB) beim Zensus 2022. <https://www.deutsche-statistik.de/veranstaltungen/ags-ost-verband-deutscher-staedtestatistiker/>
- Weiland, S. V. (2025). Vergleich von Record-Linkage-Methoden anhand der Mikrosimulation eines bundesweiten Bildungsverlaufsregisters. *WISTA – Wirtschaft und Statistik*, (4), 95–105.
- Weisberg, H. F. (2005). *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. The University of Chicago Press.

- 
- West, B. T., & Blom, A. G. (2016). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, smw024. <https://doi.org/10.1093/jssam/smw024>
- Wirtz, C. (2022). *Anforderungen und Chancen der Bevölkerungsstatistik im europäischen Kontext* [Vortrag 20.10.2022, Statistisches Bundesamt, Wiesbaden].
- Wittmaack, M. (2025). Mehrfachfallprüfung im Zensus 2022 – Die neue Strategie zur automatisierten Identifikation und Bewertung von Mehrfachfällen. *WISTA Wirtschaft und Statistik*, 2025(1), 127–143.
- Xiong, X., & Cribben, I. (2023). The State of Play of Reproducibility in Statistics: An Empirical Analysis. *The American Statistician*, 77(2), 115–126. <https://doi.org/10.1080/00031305.2022.2131625>

# IMPRINT

## Publisher

German Record-Linkage Center  
Regensburger Str. 100  
D-90478 Nuremberg

## Editor

Rainer Schnell

## Template layout

Christine Weidmann

## All rights reserved

Reproduction and distribution in any form, also in parts,  
requires the permission of the German Record-Linkage Center

## Download

[www.record-linkage.de](http://www.record-linkage.de)

The German Record Linkage Center is funded  
by the German Research Foundation (DFG).

# DuEPublico

Duisburg-Essen Publications online

UNIVERSITÄT  
DUISBURG  
ESSEN

*Offen im Denken*

ub

universitäts  
bibliothek

Dieser Text wird via DuEPublico, dem Dokumenten- und Publikationsserver der Universität Duisburg-Essen, zur Verfügung gestellt. Die hier veröffentlichte Version der E-Publikation kann von einer eventuell ebenfalls veröffentlichten Verlagsversion abweichen.

**DOI:** 10.17185/duepublico/86443

**URN:** urn:nbn:de:hbz:465-20260618-082227-1

Alle Rechte vorbehalten.