

Enderle, Tobias; Kleber, Birgit

Article

Geheimhaltung mit der Cell-Key-Methode im Zensus 2022

WISTA - Wirtschaft und Statistik

Provided in Cooperation with:

Statistisches Bundesamt (Destatis), Wiesbaden

Suggested Citation: Enderle, Tobias; Kleber, Birgit (2024) : Geheimhaltung mit der Cell-Key-Methode im Zensus 2022, WISTA - Wirtschaft und Statistik, ISSN 1619-2907, Statistisches Bundesamt (Destatis), Wiesbaden, Vol. 76, Iss. 6, pp. 82-91

This Version is available at:

<https://hdl.handle.net/10419/309502>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Dr. Tobias Enderle

ist Volkswirt und hat an der Universität Trier zum Thema Imputation promoviert. Im Referat „Statistische Geheimhaltung“ des Statistischen Bundesamtes befasst er sich als Referent mit Fragen der mathematisch-statistischen Geheimhaltungsmethodik und deren Software-Implementierung.

Birgit Kleber

ist Diplom-Soziologin und arbeitet seit 2008 im Statistischen Bundesamt. Für den Zensus betreut sie die Themen Imputation und Geheimhaltung, zunächst für den Zensus 2011 und aktuell im Referat Methodik, Haushalte als Referentin für den Zensus 2022.

GEHEIMHALTUNG MIT DER CELL-KEY-METHODE IM ZENSUS 2022

Tobias Enderle, Birgit Kleber

↳ **Schlüsselwörter:** Volkszählung – informationelle Selbstbestimmung – stochastische Überlagerung – Informationsverlust – Geheimhaltungswerkzeug

ZUSAMMENFASSUNG

Beim Zensus 2022 hat die amtliche Statistik in Deutschland das datenverändernde Geheimhaltungsverfahren der stochastischen Überlagerung nach der Cell-Key-Methode (CKM) für Veröffentlichungen produktiv eingeführt. Das generische Geheimhaltungswerkzeug wurde entwickelt, da mit dem Wunsch, Nutzenden flexible Auswertungsoptionen anzubieten, eine Automatisierung der Geheimhaltung zwingend einhergehen musste. Der Beitrag erläutert zunächst das Verfahren der Cell-Key-Methode und geht auf die Herausforderungen ein, die damit verbunden sind. Anschließend wird die praktisch-technische Umsetzung der Cell-Key-Methode im Kontext des Zensus 2022 dargelegt.

↳ **Keywords:** population census – informational self-determination – stochastic perturbation – loss of information – SDC tool

ABSTRACT

In the 2022 Census, official statistics institutions in Germany introduced the cell key method (CKM) – a perturbative disclosure control method – for data publications. The generic statistical disclosure control (SDC) tool was developed as the desire to offer users flexible evaluation options made the automation of confidentiality essential. This article first describes the cell key method and examines the challenges associated with it. It then explains the practical and technical implementation of the cell key method in the context of the 2022 Census.

1

Einleitung

Um das Grundrecht aller Bürgerinnen und Bürger auf informationelle Selbstbestimmung zu schützen und den Vorgaben des § 16 Bundesstatistikgesetz Rechnung zu tragen, dürfen die Veröffentlichungen des Zensus keine Rückschlüsse auf Einzelfälle ermöglichen. Gleichzeitig sollen die Daten so viele Informationen und Erkenntnisse wie möglich bereitstellen.

Bei einem komplexen und umfangreichen Tabellenprogramm, wie es für den Zensus 2022 veröffentlicht wird, wären die Geheimhaltungsarbeiten mit einem traditionellen Verfahren wie der Zellspernung äußerst zeit- und kostenintensiv: Sperrmuster müssten aufwendig erstellt und abgestimmt werden. Zudem ist Zellspernung mit einem hohen Informations- und Flexibilitätsverlust verbunden. Daher wird im Zensus 2022 – wie beim Zensus 2011¹ – ein datenveränderndes Geheimhaltungsverfahren eingesetzt. Es findet ein post-tabulares stochastisches Überlagerungsverfahren Anwendung, die sogenannte Cell-Key-Methode. Das neue Verfahren ermöglicht eine hohe Flexibilität bei der Erstellung von Auswertungen und eine hohe – vorab festgelegte – Genauigkeit der Ergebnisse.²

↘ Exkurs

Streng genommen findet mit der Cell-Key-Methode im Zensus eine Tabellengeheimhaltung statt. Eine Tabelle wird von Tabellengliederungen aufgespannt. Zur Gliederung von Tabellen werden üblicherweise kategoriale Merkmale verwendet, die die Merkmalsträger selbst kategorisieren (zum Beispiel Region, Alter, Geschlecht, Gebäudeart). In einer Häufigkeitstabelle gibt ein Tabellenfeld die Anzahl der betreffenden Merkmalsträger an (Einwohnerzahl, Anzahl leerstehender Gebäude). In

einer Wertetabelle (siehe hierzu Abschnitt 3.3 „Überlagerung stetiger Merkmale“) enthält ein Tabellenfeld als Zellwert die Merkmalssumme (oder eine daraus abgeleitete Kennzahl) eines quantitativen Wertmerkmals für die betreffenden Merkmalsträger (zum Beispiel Miete) – im Folgenden Wertsomme genannt. Ein Veröffentlichungsergebnis kann demnach auch eine einfache Kennzahl umfassen (zum Beispiel Leerstandsquote oder durchschnittliche Miete in Wiesbaden) und muss nicht als Tabelle dargestellt werden.

Nach den Rahmenbedingungen und Anforderungen an die Geheimhaltung im Zensus in Kapitel 2 beschreibt Kapitel 3 ausführlich die einzelnen Schritte, wie die stochastische Überlagerung nach der Cell-Key-Methode abläuft. Kapitel 4 informiert über die Qualität der Fallzahl-Ergebnisse im Zensus, die technischen Umsetzungen der Cell-Key-Methode erläutert Kapitel 5. Ein kurzes Fazit schließt den Beitrag ab.

2

Rahmenbedingungen und Anforderungen an die Geheimhaltung

Die Cell-Key-Methode wird im Zensus für alle Veröffentlichungen, die auf demografischen Daten, Gebäude- und Wohnungsdaten, Haushaltsdaten und Familiendaten basieren, angewendet. Bei Stichprobenergebnissen, also Tabellen, deren Werte ausschließlich Resultat der Hochrechnung sind, ist die statistische Geheimhaltung durch die Hochrechnung und anschließende Rundung auf ein Vielfaches von 10 sichergestellt.

Aufgrund ihrer rechtlichen Relevanz werden Einwohnerzahlen³ für administrative Gebietseinheiten (Gemeinden, Verbandsgemeinden, Kreise, Regierungsbezirke, Länder und Bund) im Original veröffentlicht. Hierunter fallen auch die Einwohnerzahlen der Bundestagswahlkreise. Die entsprechenden Ergebnisse werden nicht mit der Cell-Key-Methode geheim gehalten und daher in den Erstellungsprozessen von der Überlagerung ausgeschlossen.

1 Beim Zensus 2011 wurde das prä-tabulare Verfahren SAFE (Sichere Anonymisierung für Einzeldaten) verwendet (Statistische Ämter des Bundes und der Länder, 2014).

2 Das Verfahren der stochastischen Überlagerung geht auf das australischen Statistikamt entwickelte Überlagerungsverfahren zurück (Fraser/Wooton, 2016; Thompson und andere, 2013). Das für die amtliche Statistik angepasste Verfahren mit Cell-Keys ist in Giessing (2016) und Enderle/Vollmar (2019) beschrieben. Höhne/Höhninger (2018) geben ebenfalls eine Einführung in die Methode. Die Anwendung der Cell-Key-Methode in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder stellen Setzer und andere (2024) sowie Rothe und andere (2024) dar.

3 Informationen zur Ermittlung der Einwohnerzahl im Zensus 2022 siehe Bretschki und andere (2024).

Eine weitere Besonderheit bei der Nutzung der Cell-Key-Methode als Geheimhaltungsverfahren stellen im Zensus Fallzahlen von 0, 1 und 2 dar. Die 0 ist von der Überlagerung ausgeschlossen, sodass eine nicht vorhandene Beobachtung auch nicht künstlich erzeugt wird. Kleine Fallzahlen können hingegen auf eine 0 verändert werden.¹⁴ Fallzahlen von 1 oder 2 im Original werden immer überlagert und sind in den Zensus-Veröffentlichungen nie enthalten.

3

Stochastische Überlagerung nach der Cell-Key-Methode

Bei der Cell-Key-Methode erfolgt die Geheimhaltung bei der Erzeugung der Ergebnisse (post-tabular). Dabei wird für jedes Originalergebnis (beziehungsweise Tabellenfeld, englisch: cell) ein „zufälliger Überlagerungswert“ deterministisch (das heißt nach vorgegebenen Regeln) ermittelt. Anstelle des Originalergebnisses wird jeweils die Summe aus Originalergebnis und Überlagerungswert veröffentlicht.

3.1 Vom zufälligen Record-Key zur konsistenten Geheimhaltung

Man spricht von einer stochastischen, also zufälligen Überlagerung der Originalergebnisse, da der Überlagerungswert mithilfe einer gleichverteilten Zufallszahl ermittelt wird. Auch wenn es sich um ein post-tabulares Verfahren handelt, müssen die Mikrodaten zunächst erweitert werden. Jedem Datensatz wird eine gleichverteilte Zufallszahl – der sogenannte Record-Key – angefügt. Er liegt im Intervall $[0,1)$ und hat 15 Nachkommastellen. Beim Zensus 2022 erhalten die Einzeldatensätze aller statistischen Einheiten (Personen, Haushalte, Familien, Wohnungen und Gebäude) Record-Keys.

Für die Erstellung von Ergebnissen, die den Bereich der amtlichen Statistik verlassen, werden – parallel zur Bildung eines Originalergebnisses (zum Beispiel die Häufigkeitsauszählung bei Fallzahlentabellen) – anhand der Record-Keys die sogenannten Cell-Keys ermittelt. Für alle Einzeldatensätze beziehungsweise Merkmals-träger, die zum jeweiligen Originalergebnis beitragen (das heißt mit identischen Ausprägungskombinationen) wird die Summe der Record-Keys gebildet. Die aufsummierten Record-Keys werden auf das Ausgangsintervall $[0,1)$ rücktransformiert, das heißt auf die Nachkommastellen beschränkt. Diese Cell-Keys liegen selbst wiederum gleichverteilt im Intervall $[0,1)$. Jeder Ausprägungskombination einer Tabelle ist damit neben ihrem Originalergebnis auch ein Cell-Key zugeordnet. Durch die Vorgehensweise des Aufsummierens der den Merkmalsträgern fest zugewiesenen Record-Keys erhalten logisch identische Ausprägungskombinationen automatisch immer denselben konsistenten Cell-Key.

Um eine ganzheitliche Konsistenz im Zensus zu gewährleisten, wurden die Record-Keys der einzelnen Datenbestände von den Personen ausgehend zudem logisch „vererbt“. Besteht zum Beispiel ein Gebäude aus zwei Wohnungen, so entspricht der Record-Key des Gebäudes dem Cell-Key der beiden Wohnungen. Der Record-Key eines Einpersonenhaushalts entspricht dem Record-Key dieser einen Person.

Im letzten Verfahrensschritt – dem deterministischen „Lookup“ – wird für jedes Originalergebnis anhand des Cell-Keys der jeweilige „Überlagerungswert“ aus einer Überlagerungstabelle (englisch: perturbation table, kurz: ptable) ausgelesen.

Die verwendete Überlagerungstabelle¹⁵ beschreibt die statistikspezifischen Regeln des deterministischen Prozesses (siehe Kapitel 2 „Rahmenbedingungen und Anforderungen an die Geheimhaltung“). Dazu gehören die statistischen Eigenschaften der Überlagerungen wie Unverzerrtheit und konstante Varianz sowie zensuspezifische Parameter des Verfahrens, wie die maxi-

⁴ Um die Schutzwirkung der Geheimhaltung zu gewährleisten, darf keine Unterscheidung zwischen originalen Nullen und auf 0 veränderten Fallzahlen möglich sein. Zudem sollen auf 0 veränderte Werte nicht fälschlich als nicht vorhanden gelesen werden. In den Veröffentlichungen des Zensus werden sowohl im Original fehlende Beobachtungen als auch auf 0 veränderte Werte nicht mit dem Zahlenwert „0“, sondern einheitlich durch das wertersetzende Zeichen Dash „-“ (Legende: „Genau Null oder auf Null geändert“) dargestellt.

⁵ Die Überlagerungstabelle ist eine für IT-Verfahren geeignete Form der sogenannten Übergangsmatrix. Diese enthält die bedingten Übergangswahrscheinlichkeiten für das Eintreten einer bestimmten Überlagerung abhängig von der Größe der Originalhäufigkeit (Enderle/Vollmar, 2019, hier: Seite 92 ff., Abschnitt 4.2). Für die Berechnung wird das für die Statistiksoftware R entwickelte „ptable“-Paket (Enderle, 2023) verwendet. Die Parameter der Überlagerungsmatrix werden dabei statistikspezifisch festgelegt.

male Höhe der Überlagerungen. Die Auswirkungen der festgelegten Überlagerungstabelle auf die Qualität der Veröffentlichungsergebnisse werden im Kapitel 4 „Qualitätsaussagen zur Überlagerung“ aufgegriffen.

3.2 Fallzahlüberlagerung

In Fallzahltabellen werden absolute Häufigkeiten dargestellt. Der Überlagerungswert, der zur jeweiligen Originalhäufigkeit hinzuaddiert wird, stellt die Differenz aus Ziel- und Originalhäufigkeit dar und ist wegen der konsistenten Cell-Keys für logisch identische Ausprägungskombinationen immer identisch.

Die Überlagerungstabelle kann grafisch als Überlagerungstableau dargestellt werden. [Grafik 1](#) Jeder Balken entspricht einer Originalhäufigkeit. Unterschiedliche Farben entsprechen unterschiedlichen Überlagerungswerten und die jeweilige Summe der farbigen Teilbalken entspricht der in der Übergangsmatrix vorgegebenen Wahrscheinlichkeit, mit der es zu der betreffenden Überlagerung der entsprechenden Originalhäufigkeit kommt. Der Lookup-Schritt „liest“ die Überlagerung im Überlagerungstableau in der durch den Originalwert i gegebenen Zeile an der Stelle $p = \text{Cell-Key}$ ab. Die Überlagerungswerte für alle nicht explizit genannten Originalhäufigkeiten (hier: $i > 4$) werden aus der „obersten“ Verteilung (4+) abgelesen.

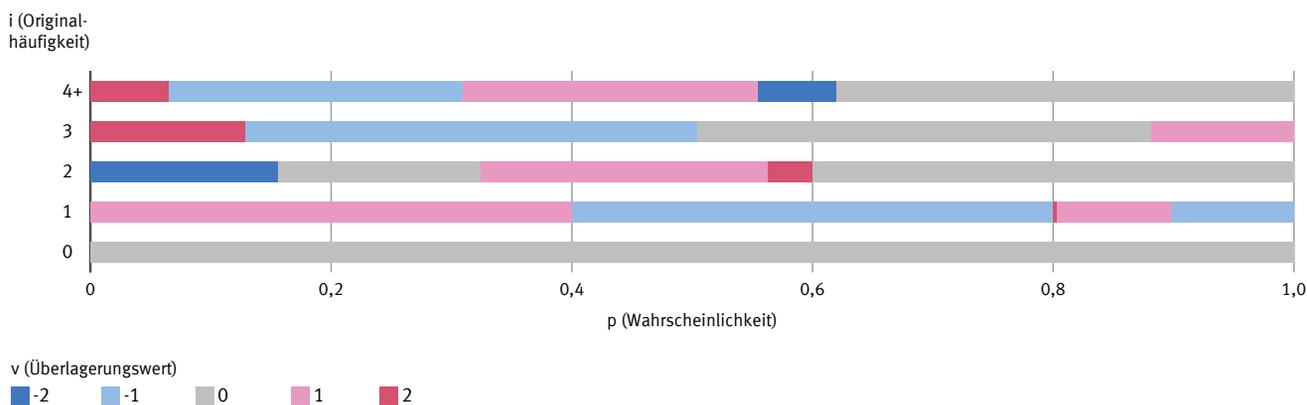
Das folgende Fallzahl-Beispiel illustriert den Lookup-Schritt:

Der Cell-Key für drei leerstehende Wohnungen der Baujahr-Kategorie 1950 bis 1960 sei die Summe der einzelnen Record-Keys $0,95 + 0,40 + 0,75 : 0,1$ (das heißt die Summe der Record-Keys ohne Vorkommastelle). Anhand des in Grafik 1 gegebenen Überlagerungstableaus wird für die Originalhäufigkeit $i = 3$ (Wohnungen) und $p = 0,1$ (Cell-Key) ein Überlagerungswert von $+2$ (dunkelblauer Balken) abgelesen. In diesem konstruierten Beispiel würden in einer Veröffentlichung fünf Wohnungen ausgewiesen werden.

Genauigkeit gegenüber Nicht-Additivität

Durch die konsistenten Cell-Keys liefert das Verfahren konsistente Ergebnisse. Um dies zu gewährleisten, müssen grundsätzlich alle Teilergebnisse gleichbehandelt werden. Dies bedeutet, dass alle Tabellenfelder separat überlagert werden müssen. Das Verfahren unterscheidet damit nicht zwischen unterschiedlichen Tabellenfeldern – wie Innenfeldern und dem Insgesamt. Die jeweiligen Tabellenfelder einer Tabellenzeile oder -spalte nach Geheimhaltung addieren sich daher nicht notwendigerweise zur ausgewiesenen Ingesamt-Summe. Die Summe aus Männern und Frauen zum Beispiel wird separat überlagert und nicht aus den überlagerten Häu-

Grafik 1
Modifiziertes Überlagerungstableau mit permutierten Intervallen



Die Parameter des Beispiels sind so gewählt, dass nach Geheimhaltung keine Fallzahlen von 1 auftreten und die maximale Abweichung ± 2 beträgt. Das Beispiel gibt dabei nicht die im Zensus 2022 verwendete Parametrisierung wieder und dient hier nur zu Veranschaulichungszwecken. Zur Erhöhung der Schutzwirkung sind die Intervalle pro Originalhäufigkeit zufällig angeordnet (permutiert) und teilweise in Subintervalle aufgeteilt.

figkeiten von Männern und Frauen gebildet. Man spricht von der sogenannten Nicht-Additivität⁶.

Das folgende Fallzahl-Beispiel in [Tabelle 1](#) veranschaulicht dies:

Das in der ersten Tabellenzeile ausgewiesene Insgesamt über alle Altersklassen für die Ausprägung „Männlich“ beträgt 175. Werden die zugehörigen Tabellenfelder der ersten Zeile jedoch separat addiert, ergibt deren Summe 173 (= 20 + 31 + 32 + 40 + 50). Das Insgesamt über die komplette Tabelle ist mit 371 ausgewiesen. Werden die Fallzahlen für die Ausprägungen „Männlich“ und „Weiblich“ jedoch separat addiert, ergibt deren Summe 372 (= 175 + 197). Auch bei einer Addition über die Insgesamthäufigkeiten aller Altersklassen (47 + 56 + 71 + 86 + 109 = 369) sowie bei einer Addition über alle Innenfelder der Tabelle (20 + 31 + 32 + 40 + 50 + 25 + 25 + 40 + 45 + 60 = 368) ergeben sich jeweils leichte Abweichungen zum ausgewiesenen Insgesamt-Wert.

Die Nicht-Additivität widerspricht zunächst der Intuition. Sie wird jedoch in Kauf genommen, da hierdurch zwei wichtige Eigenschaften des Verfahrens erst möglich sind:

1. Tabellenübergreifende Konsistenz: Für ein bestimmtes Ergebnis (zum Beispiel Anzahl der unter 18-Jährigen) ist der hinzuaddierte Überlagerungswert und somit das dargestellte Ergebnis immer identisch, unabhängig davon, in welcher anderen Kombination dieses Ergebnis dargestellt wird (zum Beispiel Personen nach Altersklassen und Geschlecht oder Personen nach Altersklassen und Nationalität).

⁶ Untersuchungen auf Basis von regulären Ergebnistabellen des Zensus 2011 haben gezeigt, dass das nachträgliche Wiederherstellen der Additivität bereits bei vergleichsweise simplen Tabellen (bis zu zwei Sachmerkmale bei einem ausgewählten Tabellenset) mit einem erheblichen Genauigkeitsverlust verbunden wäre.

2. Genauigkeit (geringer Informationsverlust): Da alle Ergebnisse getrennt voneinander überlagert werden, wird vermieden, dass sich eine Reihe zufällig gleich gerichteter Überlagerungen in Summen kumulieren und dann im Einzelfall größere Veränderungen zwischen Originalhäufigkeiten und überlagerten Häufigkeiten hervorrufen. Um im Beispiel zu bleiben: Wenn die „unter 18-Jährigen“ in einer Tabelle als Summe der entsprechenden 18 einzelnen Altersjahre (0, 1, 2, ...) dargestellt werden und in allen 18 Altersjahren der Überlagerungswert bei einer Maximalabweichung von $D = 2$ zufällig negativ ausfällt (zum Beispiel immer -2), würde ein als Summe der 18 Abweichungen gebildeter Überlagerungswert (im Beispiel: -36) das Gesamtergebnis deutlich stärker verändern als dies bei einer separaten Überlagerung des Gesamtergebnisses der Fall wäre. Dafür ist lediglich die vorab festgelegte Abweichung von ± 2 möglich.

Das bis hier beschriebene Verfahren zeigt die Geheimhaltung von Fallzahlauswertungen. Der Zensus bietet daneben viele spannende Kennzahlen, deren überlagertes Ergebnis nicht ausschließlich aus der Überlagerung von Fallzahlen resultiert.

Kennzahlen auf Basis von Fallzahlen

Kennzahlen wie Anteile, Verhältnisse, Veränderungsraten oder Salden werden auf Basis der überlagerten Fallzahlen berechnet. Hierbei können zwei unerwünschte Effekte auftreten:

1. Ungenauigkeit (hoher Informationsverlust): Bei kleinen Fallzahlen kann es zu großen absoluten Abweichungen zwischen originaler und der Kennzahl nach Geheimhaltung kommen.
2. Inkonsistenz/Unplausibilität: Da die Originalhäufigkeiten der Terme (zum Beispiel Zähler und Nenner) wie alle anderen Häufigkeiten unabhängig voneinander

Tabelle 1

Kreuztabelle mit Fallzahlen nach Geschlecht und fünf Altersklassen

	Alter					Insgesamt
	unter 18 Jahre	18 bis 29 Jahre	30 bis 49 Jahre	50 bis 64 Jahre	65 Jahre und älter	
	Anzahl					
Männlich	20	31	32	40	50	175
Weiblich	25	25	40	45	60	197
Insgesamt	47	56	71	86	109	371

der überlagert werden, kann es bei Kennzahlen, bei denen die Original-Terme nahe beieinander liegen, zu unplausiblen Werten kommen. Beispiele wären Anteile größer 100 % oder Salden, die statt einer Steigerung einen Rückgang ausweisen.

Zu beiden Effekten wurden geeignete Maßnahmen ergriffen:

- › Kennzahlen, die aufgrund der möglichen – und nicht tatsächlichen – Überlagerung eine hohe Ungenauigkeit⁷ aufweisen können, werden geklammert dargestellt.
- › Kennzahlen behalten auch nach Überlagerung ihre grundsätzliche Aussagekraft durch eine entsprechende Behandlung (zum Beispiel Skalierung) im Zuge der Überlagerung. So behalten Salden beispielsweise ihr Vorzeichen oder Anteile werden nach Überlagerung mit maximal 100 % ausgewiesen.

3.3 Überlagerung stetiger Merkmale

Aus dem Zensus werden auch metrisch skalierte (sowohl stetige als auch quasi stetige) Merkmale in Wertetabellen veröffentlicht. Bei stetigen Merkmalen bestehen bei Anwendung der Cell-Key-Methode grundsätzlich zwei Umsetzungsoptionen: Mit der ersten Option wird eine höhere Genauigkeit der gebildeten Wertsumme selbst erreicht, die zweite Option erzielt demgegenüber eine höhere Genauigkeit bei der Ermittlung von Mittelwerten (Quotient mit Wertsumme als Zähler und Anzahl beitragender Merkmalsträger im Nenner). Im Zensus wurde der Veröffentlichung von Durchschnittswerten eine höhere Priorität eingeräumt als der Veröffentlichung aufsummierter stetiger Merkmale und daher Option 2 umgesetzt.⁸

Die Überlagerung von Wertetabellen erfolgt vom Grundprinzip her zunächst wie die Überlagerung von Fallzahltabellen. Während bei Fallzahltabellen vorrangig kleine Fallzahlen geheim gehalten werden, sind bei Wertetabellen (zumeist) die größten Beitragswerte (zum Bei-

- 7 Übersteigt der aufgrund einer möglichen Überlagerung statistisch zu erwartende relative Fehler der Kennzahl 15 %, so wird das Ergebnis geklammert (Legende: „Aussagewert eingeschränkt, da der Zahlenwert durch das Geheimhaltungsverfahren möglicherweise relativ stark verändert wurde“).
- 8 Beispielsweise ist die durchschnittliche Miete einer Gemeinde eine deutlich stärker nachgefragte Information als die Summe aller Mieten einer Gemeinde zusammen.

spiel größte Miete) einer Wertsumme zu schützen. Bei der Überlagerung stetiger Merkmale müssen jedoch einige grundlegende Anpassungen vorgenommen werden:⁹

1. Die Stärke (Varianz) der Überlagerung korreliert mit der Größe der größten Einzelwerte und wird durch wählbare Parameter optimal an die Daten angepasst.
2. Es wird gewährleistet, dass auch für nicht ganzzahlige Werte Überlagerungswerte bereitgestellt werden. Der Überlagerungswert selbst ist daher metrisch skaliert und wird anhand einer für stetige Merkmale konzipierten Überlagerungstabelle ermittelt.

Mittelwerte werden wie Anteile durch eine Klammerung qualitätsgesichert dargestellt, falls von einem eingeschränkten Interpretationsgehalt nach Geheimhaltung ausgegangen werden kann.

3.4 Sperrung von Ergebnissen

Neben der bereits skizzierten Klammerung von Kennzahlen wurde auch eine Sperrung der Auswertungen umgesetzt, falls Ergebnisse ein zu hohes Aufdeckungsrisiko und/oder einen zu hohen Informationsverlust bergen. Mögliche Sperrungen werden durch das wertschöpfende Zeichen „,“ (Legende: „Zahlenwert unbekannt oder geheim“) dargestellt. Die Prüfung erfolgt für einzelne Tabellenteile separat, wobei die zu prüfenden Tabellenteile anhand ihrer Aggregationsebene voneinander unterschieden werden. Im Falle einer einfachen Kreuztabelle und zwei Merkmalen sind das zum Beispiel vier unterschiedliche Tabellenteile: Innenfelder, Ingesamt-Felder der Zeilen, Ingesamt-Felder der Spalten und Ingesamt-Eckfeld. [↘ Tabelle 2](#) Die Prüfung auf mögliche Sperrungen wird immer mit den Innenfeldern auf der feinsten Ebene begonnen. Weisen diese weder ein hohes Aufdeckungsrisiko noch einen hohen Informationsverlust aus, wird die gesamte Tabelle freigegeben. Werden die Innenfelder bereits als zu sperren bewertet, werden die nächsthöheren Aggregate auf eine mögliche Sperrung hin geprüft. So wird sukzessive bis zum Eckwert (höchste Aggregationsebene) fortgefahren. Die Sperrentscheidung betrifft in der Regel wenige Tabellenteile – meist nur die Innenfelder – einer Abfrage.

9 Giessing/Tent (2019) zeigen Erweiterungen der Cell-Key-Methode für stetige Merkmale.

Tabelle 2

Beispieltable mit 4 Tabellenteilen

		Art des Gebäudes		Insgesamt
		Wohngebäude	Sonstige Gebäude mit Wohnraum	
Gebäudetyp	Freistehendes Haus	.	.	.
	Doppelhaushälfte	.	.	.
	Gereihtes Haus	.	.	.
	Anderer Gebäudetyp	.	.	.
Insgesamt		1 034	206	1 242

Jede Farbe kennzeichnet einen Tabellenteil. Die Innenfelder (dunkelblau) und Ingesamt-Felder des Merkmals „Gebäudetyp“ (hellblau) können gesperrt sein, während die Ingesamt-Felder des Merkmals „Art des Gebäudes“ (hellrot) und das Ingesamt-Eckfeld (dunkelrot) freigegeben werden.

Um mögliche Risiken und zu hohe Informationsverluste zu vermeiden, gleichzeitig aber das Veröffentlichungsprogramm des Zensus nicht zu stark einzuschränken, wurden verschiedene Sperrregeln anhand der Ergebnisse des Zensus 2011 untersucht.¹⁰ Die Einstellungen wurden schließlich so gewählt, dass je Bundesland 98 % der getesteten Tabellen auf Gemeindeebene vollständig angezeigt werden. Deutschlandweit erhalten lediglich 0,1 % der getesteten Tabellen auf Gemeindeebene eine Sperrung und auf Kreisebene können alle Tabellen vollständig veröffentlicht werden.

4

Qualitätsaussagen zur Überlagerung

Der Überlagerungswert, der zu den Originalergebnissen addiert wird und im Mittel 0 ist, nimmt bei Fallzahlentabellen ganzzahlige Werte überwiegend zwischen -2 und +2 an und hat damit auf größere Ergebnisse quasi keinen Einfluss. Der Überlagerungswert kann auch 0 betragen, sodass das Veröffentlichungsergebnis mit dem Originalergebnis übereinstimmt. Die Parameter des Verfahrens (zum Beispiel die maximale Überlagerung, die Varianz der Überlagerung oder die Wahrscheinlichkeiten, dass mit 0 überlagert wird) selbst können aus Gründen der Schutzwirkung nicht veröffentlicht werden.

10 Insgesamt wurden 574 reguläre Tabellen aus dem Veröffentlichungsprogramm von 2011 untersucht. Davon wurden 13 vom Fach- und Veröffentlichungsbereich als zentrale Tabellen definiert und prioritär betrachtet.

Stattdessen können die folgenden Qualitätsaussagen herangezogen werden, um die Qualität der Fallzahl-Ergebnisse des Zensus zu bewerten:

1. Der Absolutbetrag der Abweichungen zwischen überlagerten und originalen Fallzahlen liegt für reguläre Tabellen im Mittel unter 1.
2. Mindestens 65 % der Fallzahlen in den Tabellen bleiben unverändert oder weichen um maximal 1 vom Originalwert ab.
3. Bei höchstens 20 % der Fallzahlen in den Tabellen liegt die Abweichung bei 3 oder mehr.
4. Bei höchstens 10 % der Fallzahlen in den Tabellen liegt die Abweichung bei 4 oder mehr.

Selbstständige Weiterverarbeitungen von überlagerten Ergebnissen (zum Beispiel eigenständige Summenbildung aus überlagerten Fallzahlen statt auf das bereitgestellte überlagerte Ingesamt direkt zuzugreifen) können zu höheren Informationsverlusten beziehungsweise Ungenauigkeiten in der Ergebnisdarstellung führen. Das Veröffentlichungsprogramm des Zensus 2022 ist entsprechend umfangreich gestaltet.

5

Technische Umsetzungen der Cell-Key-Methode

Zur Automatisierung und einheitlichen Durchführung der Geheimhaltung mit der stochastischen Überlagerung nach der Cell-Key-Methode hat das Statistische Bundesamt im Rahmen des Zensus ein flexibles und generisches Geheimhaltungswerkzeug, die Fachanwendung CKM-Geheimhaltung¹¹ entwickelt. Die Verwendung und Anbindung der modularen Fachanwendung an bestehende Werkzeuge der amtlichen Statistik (Analyse- und Auswertungswerkzeuge, Datenbanken und landeseigene Informationssysteme) ermöglichen eine harmonisierte Geheimhaltung und konsistente Veröffentlichung einschließlich Qualitätskennzeichnungen (Klammerungen und Sperrungen). Das Geheimhaltungswerkzeug wird in den Statistischen Ämtern des Bundes und der Länder (als Backend-System) bereits für diverse

11 Die Fachanwendung basiert grundsätzlich auf einer JAVA-Architektur und bietet als Webservice eine REST-Schnittstelle.

Veröffentlichungsprodukte des Zensus 2011 und des Zensus 2022 eingesetzt:

- › Zensusdatenbank,
- › Regionaltabellen,
- › Gitterdaten,
- › Zensus-Atlas und
- › Sonderauswertungen für Dritte.

Geheimhaltung durch Dritte

Neben den Statistischen Ämtern des Bundes und der Länder erhalten auch weitere Einrichtungen – wie die Wissenschaft über die Forschungsdatenzentren (Setzer und andere, 2024; Rothe und andere, 2024) oder die Kommunen mit abgeschotteter Statistikstelle – Einzeldaten aus dem Zensus (§ 16 Absätze 5 und 6 Bundesstatistikgesetz sowie § 32 Absatz 2 Zensusgesetz 2022). Sollen hieraus Ergebnisse veröffentlicht werden, muss sichergestellt sein, dass die Anwendung der Cell-Key-Methode unter Verwendung der Zensus-Parametrisierung erfolgt und zu identischen und konsistenten Ergebnissen führt. Hierzu wird das im Statistischen Bundesamt entwickelte Geheimhaltungswerkzeug auch an Auswertungssysteme der jeweiligen Einrichtungen (zum Beispiel Werkzeuge der KOSIS-Arbeitsgemeinschaften KO.R und DUVA für Kommunen mit abgeschotteter Statistikstelle) angebunden.

Die Geheimhaltung der vom Statistischen Amt der Europäischen Union (Eurostat) veröffentlichten Ergebnisse zum deutschen Zensus 2022 wurde ebenfalls mittels des Geheimhaltungswerkzeuges sichergestellt.

reitgestellt werden, welches eine Automatisierung der Geheimhaltung der Ergebnisse ermöglicht. Diese Umsetzung weist jedoch auch Herausforderungen auf, die allerdings bereits in der Konzeptionsphase erkannt und entsprechend vorbereitet wurden. Im Ergebnis stellt das Verfahren konsistente Ergebnisse mit einer hohen, vorab festgelegten Genauigkeit zur Verfügung. [UU](#)

6

Fazit

Die Cell-Key-Methode zeigt sich – mit einigem Vorlauf – als geeignetes Verfahren zur Geheimhaltung der deutschen Zensusergebnisse. So konnte der Umfang des Veröffentlichungsprogramms – insbesondere hinsichtlich der regionalen Auswertungstiefe – deutlich erhöht werden. Zudem konnte mit der Entwicklung der Fachanwendung CKM-Geheimhaltung ein Auswertungssystem-übergreifendes Instrument geschaffen und be-

LITERATURVERZEICHNIS

Bretschi, Corinna/Seibel, Steffen/Vorndran, Ingeborg/Meyn, Christoph. *Ermittlung der Einwohnerzahl im Zensus 2022*. In: WISTA Wirtschaft und Statistik. Ausgabe 6/2024, Seite 17 ff.

Enderle, Tobias. *pstable: Generation of Perturbation Tables for the Cell-Key Method*. R package version 1.0.0. 2023. [Zugriff am 1. November 2024]. Verfügbar unter: CRAN.R-project.org

Enderle, Tobias/Vollmar, Meike. *Geheimhaltung in der Hochschulstatistik*. In: WISTA Wirtschaft und Statistik. Ausgabe 6/2019, Seite 87 ff.

Fraser, Bruce/Wooton, Janice. *A proposed method for confidentialising tabular output to protect against differencing*. Work session on Statistical Data Confidentiality. Supporting paper. Genf 2005. [Zugriff am 1. November 2024]. Verfügbar unter: unece.org

Giessing, Sarah. *Computational Issues in the Design of Transition Probabilities and Disclosure Risk Estimation for Additive Noise*. In: Domingo-Ferrer, Josep/Peji-Bach, Mirjana (Herausgeber). *Privacy in Statistical Databases*. LNCS (Lecture Notes in Computer Science). 2016. Ausgabe 9867, Seite 237 ff.
DOI: [10.1007/978-3-319-45381-1_18](https://doi.org/10.1007/978-3-319-45381-1_18)

Giessing, Sarah/Tent, Reinhard. *Concepts for generalising tools implementing the cell key method to the case of continuous variables*. In: Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality. Den Haag 2019. [Zugriff am 1. November 2024]. Verfügbar unter: unece.org

Höhne, Jörg/Höninger, Julia. *Die Cell-Key-Methode – ein Geheimhaltungsverfahren*. In: Zeitschrift für amtliche Statistik Berlin Brandenburg. Ausgabe 3+4/2018, Seite 14 ff. [Zugriff am 9. Dezember 2024]. Verfügbar unter: www.statistischebibliothek.de

Rothe, Patrick/Güttgemanns, Volker/Rohde, Johannes/Setzer, Stefanie. *Die Cell-Key-Methode in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. Teil 2: Auswirkungen des neuen Geheimhaltungsverfahrens*. In: WISTA Wirtschaft und Statistik. Ausgabe 3/2024, Seite 45 ff.

Setzer, Stefanie/Rohde, Johannes/Güttgemanns, Volker/Rothe, Patrick. *Die Cell-Key-Methode in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. Teil 1: Vorstellung des neuen Geheimhaltungsverfahrens*. In: WISTA Wirtschaft und Statistik. Ausgabe 3/2024, Seite 31 ff.

Statistische Ämter des Bundes und der Länder. *Nutzer-Kurzinformation zu SAFE*. 2014. [Zugriff am 1. November 2024]. Verfügbar unter: www.zensus2011.de

Thompson, Gwenda/Broadfoot, Stephen/Elazar, Daniel. *Methodology for the Automatic Confidentialisation of Statistical Outputs from Remote Servers at the Australian Bureau of Statistics*. Paper presented at the Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality. Ottawa 2013. [Zugriff am 1. November 2024]. Verfügbar unter: unece.org

RECHTSGRUNDLAGEN

Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) in der Fassung der Bekanntmachung vom 20. Oktober 2016 (BGBl. I Seite 2394), das zuletzt durch Artikel 14 des Gesetzes vom 8. Mai 2024 (BGBl. 2024 I Nr. 152) geändert worden ist.

Gesetz zur Durchführung des Zensus im Jahr 2022 (Zensusgesetz 2022 – ZensG 2022) vom 26. November 2019 (BGBl. I Seite 1851), das durch Artikel 2 des Gesetzes vom 3. Dezember 2020 (BGBl. I Seite 2675) geändert worden ist.

Herausgeber
Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung
Dr. Daniel Vorgrimler
Redaktion: Ellen Römer

Ihr Kontakt zu uns
www.destatis.de/kontakt

Erscheinungsfolge
zweimonatlich, erschienen im Dezember 2024
Ältere Ausgaben finden Sie unter www.destatis.de sowie in der [Statistischen Bibliothek](#).

Artikelnummer: 1010200-24006-4, ISSN 1619-2907

© Statistisches Bundesamt (Destatis), 2024
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.