



**Prof. Dr. Helmut Küchenhoff**

Telefon +49 (0)89 2180-2789  
[kuechenhoff@stat.uni-muenchen.de](mailto:kuechenhoff@stat.uni-muenchen.de)

Institut für Statistik  
Akademiestr. 1/IV  
80799 München

## **Gutachten zur Verwaltungsrechtssache Stadt Bremerhaven gegen Freie Hansastadt Bremen**

### **Im Auftrag des Verwaltungsgerichts der Freien Hansestadt Bremen (4. Kammer)**

Das folgende Gutachten besteht aus 2 Teilen. Im ersten Teil werden allgemeine Überlegungen dazu angestellt, wie eine Fehlerrechnung zur Schätzung der Einwohnerzahl im Rahmen des Zensus 2011 erfolgen kann. Dies ist eine wichtige Grundlage, um die konkrete Frage nach der Größe des Fehlers bei der Schätzung der Einwohnerzahl in Bremerhaven beantworten zu können. Im zweiten Teil des Gutachtens werden dann die von dem Gericht gestellten Fragen einzeln beantwortet. Das Gutachten basiert auf den mir zur Verfügung gestellten Gerichtsakten und auf den in der Literaturliste erwähnten Publikationen. Hinzu kamen persönliche Gespräche mit beteiligten Forschern (Prof Ralf Münnich, Dipl. math. Wolf Bihler, Prof Rainer Schnell, Prof Jörn Christensen und Frau Dipl. Stat. Uta Thien-Seitz).

### **Teil I: Fehlerquellen bei der Schätzung der Einwohnerzahl im Zensus 2011**

Ich beschränke mich auf das hier relevante Verfahren bei Gemeinden mit über 10 000 Einwohnern, bei denen mit Hilfe einer Stichprobe unter Verwendung von Registerdaten die Einwohnerzahl geschätzt wurde. Das Verfahren und das Vorgehen wurden vom Statistischen Bundesamt in vielfältiger Weise dokumentiert (siehe z.B. Anlagen zum Schreiben des statistischen Landesamtes Bremen vom 18.12.2013). Wesentliche methodische Grundlagen sind dazu die Publikationen Berg und Bihler (2011) und Berg und Bihler (4/2014).

Die Fehlerquellen bei Schätzungen, die auf Befragungen basieren, sind sehr vielfältig. In der internationalen Forschung wird nach dem Konzept des sog. „Total

Survey Error“ systematisch nach verschiedenen Fehlerquellen gesucht. Mit Hilfe dieser Strategie, wie sie z.B. in der Publikation von Fuchs (2009) dargelegt ist, werden im Folgenden die einzelnen Aspekte bei der Schätzung der Einwohnerzahl dargestellt. Man unterscheidet grundsätzlich zwischen Aspekten, die die Repräsentativität betreffen und solchen, die den Messfehler betreffen.

## **Repräsentativität**

### ***„Sampling Error“ (Stichprobenfehler)***

Mit den „Sampling Error“ bezeichnet man den Fehler, der dadurch entsteht, dass keine Vollerhebung durchgeführt wurde, sondern eine Zufallsstichprobe gezogen wurde. Dieser Fehler kann durch ein geeignetes Stichprobendesign, eine optimale Schätzstrategie und durch die Wahl des Stichprobenumfangs minimiert werden. Weiter kann die Standardabweichung des Fehlers nach der Ziehung der Stichprobe nach den Regeln der Stichprobentheorie geschätzt werden. Das Design und die Schätzstrategie wurden im Vorfeld des Zensus („Zensus Test“) sehr intensiv und ausführlich untersucht, siehe dazu z.B. Münnich und Gabler (2012) und Münnich et al (2012). Das verwendete Stichprobendesign ist in der Publikation von Berg und Bihler (2011) dargestellt. Dabei wurde die Anschrift als Stichprobeneinheit gewählt. Die hier relevante Zielgröße ist die tatsächliche Anzahl der Bewohner einer Anschrift am Stichtag. Aus der Schätzung der (durchschnittlichen) Anzahl von Bewohnern pro Anschrift ergibt sich dann direkt die Einwohnerzahl. Diese hier sinnvolle Setzung unterscheidet sich von anderen Befragungen, bei denen die Untersuchungseinheiten oft Haushalte oder Personen sind.

Zunächst wurde eine möglichst vollständige Liste von Anschriften erstellt und nach Abzug von sensiblen Sonderanschriften (Krankenhäuser, JVs etc.) ergab sich eine sog. Auswahlgesamtheit (das sind die Anschriften, aus denen die Stichprobe gezogen wird) von insgesamt 19.9 Millionen Anschriften. Diese wurde dann nach räumlichen Kriterien und nach der Größe der Anschriften in sog. Schichten aufgeteilt. Aus diesen Schichten wurde dann eine Zufallsstichprobe gezogen. Die Umfänge der Zufallsstichproben in den Schichten wurden so gestaltet, dass die Genauigkeit der folgenden Schätzung der Einwohnerzahlen optimiert wurde. Das verwendete Design ist aus meiner Sicht sinnvoll und plausibel und auf dem Stand der Wissenschaft.

Zur Schätzung der Einwohnerzahl werden nun die Daten aus den gezogenen Adressen (als wichtigste Größe die ermittelte Zahl von Bewohnern) herangezogen und auf die Auswahlgesamtheit hochgerechnet. Die Daten aus den Registern werden als sog. Hilfsgrößen genutzt. Dabei wird aus den erhobenen Daten ein

Zusammenhang zwischen den Registerdaten und den erhobenen Einwohnerzahlen auf der Ebene der Gemeinden bzw. Stadtteile geschätzt und dieser wird dann benutzt, um die Hochrechnung durchzuführen. Dieser sogenannte allgemeine Regressionsschätzer ist in der Publikation von Berg und Bihler (2014) korrekt beschrieben. Das gilt auch für die Schätzung der (relativen) Standardabweichung des Schätzfehles. In dem Hochrechnungsverfahren wird das komplexe Design in adäquater Weise durch Gewichtung berücksichtigt. Für die Stadt Bremerhaven resultiert daraus nach Angaben aus dem Datenblatt ein relativer Standardfehler von 0,55%. Der relative Standardfehler ist auf Grund der unterschiedlichen Verteilung der Einwohnerzahlen pro Adresse in den einzelnen Städten und Gemeinden unterschiedlich und nur nach der Erhebung genau abschätzbar. Wichtig ist, dass bei dieser Schätzstrategie die Registerdaten nur als Hilfsgröße eingehen. Das hat zur Folge, dass die Schätzung auch bei Fehlern in den Registerdaten valide ist. Entscheidend ist jedoch die Korrektheit der in der Stichprobe erhobenen Werte (siehe die Diskussion unten). Berechnet man ein sog. Konfidenzintervall mit einem Sicherheitsniveau von 95%, so ergibt sich für Bremerhaven ein Schwankungsbereich der Einwohnerzahl von ca.  $\pm 1.1\%$ .

Ein Problem, das im Vorfeld etwas unterschätzt wurde, war das Auftreten von sog. Nulladressen, also Adressen bei denen keine Personen gemeldet sind. Grundsätzlich führt die Einbeziehung solcher Adressen zu keiner Verzerrung in der Schätzung der Einwohnerzahl. Das Merkmal „Zahl der Personen“ hat dann die Ausprägung 0 und geht so in die Hochrechnung ein. Die Problematik der Nulladressen wurde u.a. in einem weiteren Forschungsprojekt ausführlich untersucht (Münnich et al. (2013)). Auch wenn die Autoren zu dem Schluss kommen, dass das ursprüngliche Modell nutzbar ist, erscheint mir ein Regressionsmodell mit einer zusätzlichen Dummyvariable für die Nulladressen inhaltlich sinnvoller zu sein. Insgesamt ist das Vorgehen jedoch sinnvoll und es ist davon auszugehen, dass die Quantifizierung des Schätzfehlers korrekt ist. Das wird auch durch die ausführlichen Simulationsstudien bestätigt, siehe dazu z.B. Münnich und Gabler (2012).

Zu beachten ist allerdings, dass der angegebene Schätzfehler nur den Aspekt der Stichprobenziehung abdeckt. Bei dem Teil der Adressen, bei denen eine Vollerhebung durchgeführt wurde, taucht dieser Fehler nicht auf. Alle anderen im Folgenden zusammengestellten Fehler sind bei dieser Fehlerabschätzung ( $\pm 1.1\%$  im Fall Bremerhaven) nicht berücksichtigt.

## ***„Coverage Error“***

Der „Coverage Error“ beinhaltet die Frage nach der Übereinstimmung der Grundgesamtheit und der Auswahlgesamtheit. Hier geht es also um die Frage der Vollständigkeit und Korrektheit der Liste der Adressen, die in der Auswertung berücksichtigt wurden. Man unterscheidet hierbei zwischen „undercoverage“ (hier: Fehlen von Adressen) und „overcoverage“ (hier: Einbeziehen von Adressen, die nicht zu der Gemeinde gehören). Zu beachten ist, dass die Einbeziehung von Nulladressen hier unproblematisch ist, da diese nicht zu einem systematischen Fehler führen (siehe oben). Fehler durch „overcoverage“ könnte hier nur auftreten, wenn eine Adresse mit Bewohnern einer Gemeinde fälschlicherweise zugeordnet wird. Kritisch könnte allerdings das Fehlen von relevanten Adressen in der Auswahlgesamtheit sein. Daher wurden bei der Erstellung der Auswahlgesamtheit neben dem Anschriften- und Gebäuderegister weitere Quellen hinzugezogen, siehe dazu auch Ziprik (2012) und Hirner und Stiglmayr (2013). Das Vorgehen ist in Berg und Bihler (2011) dargestellt. Generell wurde eine Strategie, eine möglichst vollständige Auswahlgesamtheit zu erhalten, verfolgt. Dabei wurde in Kauf genommen, dass Nulladressen auftreten. Dies ist aus meiner Sicht sinnvoll, da eine unvollständige Auswahlgesamtheit zwangsläufig zu einer systematischen Unterschätzung der Einwohnerzahl geführt hätte, während durch eine adäquate Berücksichtigung der Nulladressen bei der Hochrechnung systematische Fehler vermieden werden können. Weiter wurde eine zusätzliche Erhebung für Neuzugänge durchgeführt. Dabei wurde berücksichtigt, dass zwischen Hauptziehung und Stichtag Änderungen z.B. durch Neubauten auftreten. Das Vorgehen ist in dem Artikel Berg und Bihler (3/2014) ausführlich dokumentiert. Trotz des erheblichen Aufwandes bei der Erstellung der Adressenliste sind natürlich Fehler hier nicht auszuschließen. Probleme könnten eventuell durch die Zuordnung von Adressen zu falschen Gemeinden entstehen und bei der inkorrekten Behandlung von Sonderadressen. Insgesamt sollte der „Coverage Error“ auf Grund der intensiven Vorbereitungsarbeiten und der Hinzuziehung verschiedener Quellen eher gering ausfallen.

## ***Non Response Error***

„Non Response Error“ bedeutet hier, dass man von gewissen Anschriften keine Angaben erhält. In der Survey-Forschung wird diesem Fehlertypus größte Aufmerksamkeit geschenkt. Zu beachten ist allerdings, dass hier die Untersuchungseinheiten die Adressen sind und nicht die Personen. Daher geht es nicht um die Antwortverweigerung einzelner Personen, sondern um das Fehlen von Informationen bezüglich der Einwohnerzahl zu den einzelnen Adressen. Die Quote der Adressen

ohne Auskunft ist auf Grund der Auskunftspflicht nach Berg und Bihler (4/2014) sehr gering (0.67%). Diese Quote wurde dann in den Design- Gewichten entsprechend berücksichtigt. Ein Fehler in der Abschätzung der Einwohnerzahl tritt nur dann auf, wenn sich die Anschriften ohne Angaben von denen mit Angaben erheblich unterscheiden. Dieser mögliche Fehler wird allerdings durch die Schichtung und die Verwendung des Schätzverfahrenes (Regressionsschätzer) reduziert.

Insgesamt sollte der durch „Non Response“ in der Stichprobe zu Stande gekommene Fehler gering sein.

## **Messung**

### ***Messfehler***

Damit sind Fehler bei der Erhebung der Daten gemeint. Da hier die statistischen Einheiten die Adressen sind, ist die relevante Messgröße hier die Einwohnerzahl an einer Adresse. Der Messfehler ist damit die Abweichung der in der Befragung erhobenen Einwohnerzahl von der tatsächlichen Einwohnerzahl. Abweichungen können hier auf verschiedene Weise zu Stande kommen. Einerseits könnten Personen, die an dieser Adresse wohnen, nicht in der Zählung berücksichtigt werden. Andererseits könnten Personen, die nicht an der entsprechenden Adresse wohnen, fälschlicherweise in die Zählung kommen. Die Ursachen für solche Messfehler sind im Allgemeinen sehr vielfältig. Fehler können durch die Erhebungsbeauftragten entstehen, wenn diese Personen nicht auffinden oder Personen angeben, die nicht an der Adresse wohnen.

Weiter können Fehler dadurch entstehen, dass sich Personen der Erhebung entziehen oder doppelt erfasst werden.

Im Allgemeinen ist es sehr schwierig, Mess- und Erhebungsfehler zu quantifizieren. Typischerweise werden zu solchen Fragen unabhängige Kontrollzählungen durchgeführt (sog. „Post Enumeration Surveys“). Dies liefern dann Hinweise auf mögliche systematische Messfehler. Weiter kann die Qualität der Befragung durch Erhebungen zu dem Verlauf der Befragung (sog. Metadaten) untersucht und verbessert werden. Beim Zensus liegt für Bremerhaven eine Dokumentation der örtlichen Erhebungsstelle vor (Schreiben des Statistischen Landesamt Bremen vom 25.4.2014). Hier wurden Schulung und Ablauf dokumentiert. Daraus ergibt sich im Wesentlichen ein reibungsloser Ablauf der Erhebung. Das Thema möglicher Erhebungsfehler ist dort aber nicht thematisiert.

In einer Befragung von Erhebungsstellen durch die Stadt München, siehe dazu Statistisches Amt München (2013), wurden Probleme bei den Erhebungsstellen bundesweit untersucht. Hier wird als wichtiges Problem das „Vorgehen der Interviewer nach Namensliste“ genannt. Den Erhebungsbeauftragten standen Namenslisten der gemeldeten Personen bei den Adressen zur Verfügung. Aus der Survey Forschung ist bekannt, dass die Nutzung solcher Listen zu Fehlern führen, siehe Eckmann und Kreuter (2011). Der Fehler wird als „Confirmation Bias“ bezeichnet und beschreibt die Tendenz der Interviewer, eine Liste zu bestätigen. Bei der Befragung im Rahmen des Zensus bedeutet dies, dass Personen, die nicht auf der Liste stehen, tendenziell nicht aufgefunden werden und Personen, die auf der Liste stehen und nicht mehr an der Adresse wohnen, fälschlicherweise gezählt werden.

Weiter wird in der Befragung der Erhebungsstellen auch dokumentiert, dass das Problem, dass bestimmte Personen nicht befragt wurden (Kinder, Ausländer) in seltenen Fällen auftritt. Daher ist es möglich, dass insbesondere nicht gemeldete Personen nicht erfasst wurden. Weiter könnten sich Personen der Zensusbefragung entziehen, indem sie bzw. die Bewohner der Adresse angeben, hier nicht mehr zu wohnen. Auch wenn es nach meinem Wissen keine Veröffentlichung zum Zensus 2011 gibt, die sich explizit mit dem Thema Mess- und Erhebungsfehler auseinandersetzt, so finden sich z.B. in der Veröffentlichung von Hirner und Siglmayer (2013) dazu einige Aspekte. Beispielsweise gab es bei unplausiblen Ergebnissen zu Anschriften weitere Nachprüfungen. Allerdings ist insgesamt aus den Veröffentlichungen nicht genau zu entnehmen, wie Datenbereinigungen und Korrekturen bezüglich der Erhebungsdaten erfolgten.

Weiter ist zu beachten, dass Messfehler bei den Adressen, bei denen eine Vollerhebung durchgeführt wurde (sensible Sonderanschriften) auch zu einer Verzerrung der Ergebnisse führen können. Hinweise darauf ergeben sich auch aus der Befragung des Statistischen Amtes der Stadt München.

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.

Insgesamt lässt sich aber kein Nachweis für eine Vergrößerung des Fehlers in der Einwohnerzahl erbringen, aber es gibt zusätzliche Indizien, die für eine Verzerrung der Ergebnisse sprechen:

- Die Analysen von Christensen et al. (2014) zeigen, dass es einen nicht unerheblichen Unterschied zwischen Gemeinden unter und über 10000

Einwohnern bezüglich der Korrektur der Einwohnerzahlen gibt. Die Gemeinden unter 10000 Einwohnern zeigen deutlich Unterschiede in der prozentualen Korrektur der Einwohnerzahlen. In der Analyse wird deutlich, dass es sich um einen Methodeneffekt handeln könnte, da der Effekt auch bei einem eingeschränkten Bereich von Städten zwischen 8000 und 12000 Einwohnern erhalten bleibt. Aus der Befragung der Erhebungsstellen ergeben sich Gesichtspunkte, die für Probleme bei der vollständigen Erfassung sprechen.

- Der von dem Kläger angegebene erhebliche Unterschied in der Anzahl der Ausländer(innen) ist sehr allgemein zu beobachten. Nach dem Zensus leben in Deutschland deutlich wenige 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).

Zu beachten ist, dass bei dem Vorgehen der Regressionsschätzung, systematische Fehler bei der Erhebung der Einwohnerzahl zu einer Verzerrung der Schätzung der Einwohnerzahl führen. Das träfe auch zu, wenn die Daten aus dem Melderegister korrekt wären. Das ist eine Folge der Schätzstrategie, die Registerdaten in der Hochrechnung als Hilfsgröße zu verwenden und die in der Befragung erhobenen Werte als korrekte relevante Zielgröße zu betrachten. Da die Regressionsschätzung auf Gemeindeebene durchgeführt wurde, wirken sich Erhebungsfehler auf der Gemeindeebene direkt auf die Schätzung aus.

Im Rahmen des Zensus wurde eine Wiederholungsbefragung (5% der gezogenen Haushalte) durchgeführt. Aus den Ergebnissen dieser Stichprobe könnten sicher einige Erkenntnisse zu dem Messfehler gewonnen werden. Eine Dokumentation und Auswertung dieser Wiederholungsstichprobe wäre sehr hilfreich und würde teilweise eine quantitative Abschätzung von Messfehlern ermöglichen. Mir liegt die Veröffentlichung einer solchen Analyse nicht vor.

## **Prozessfehler**

Als Prozessfehler sind Fehler, die bei der *Dateneingabe* und Verarbeitung auftreten. Hier ist die Zusammenarbeit zwischen den Erhebungsstellen und den Statistischen Ämtern relevant. Da es sich um komplexe Vorgänge handelt, können hier Probleme auftreten. Allerdings wurde sowohl von der Erhebungsstelle der Stadt Bremerhaven

als auch von Grohnbach und Loth (2013) von einer weitestgehend reibungslosen Zusammenarbeit der Erhebungsstellen mit den Statistischen Landesämtern auch bezüglich der Nutzung der entsprechenden EDV berichtet. In der Befragung des Statistischen Amtes der Stadt München wird allerdings von einigen Problemen mit der Zensus- Software berichtet. Bei der Dateneingabe und der Übertragung ist zu beachten, dass es sich bei den für die Berechnung der Einwohnerzahl relevanten Eingaben primär um Zählungen handelt, deren Verarbeitung nicht so schwierig ist wie bei anderen komplexeren Fragbögen. Es ist daher davon auszugehen, dass die Prozessfehler einen geringen Einfluss auf das Ergebnis der Einwohnerzählung haben.

## **Zusammenfassung**

Insgesamt ist bei der Fehlerabschätzung der Einwohnerzahl beim Zensus 2011 durch das Statistische Bundesamt nur der reine „Sampling error“ angegeben. Design und Berechnung hierzu sind klar nachvollziehbar, ausreichend validiert und korrekt und nach wissenschaftlichen Standards durchgeführt worden.

Eine Betrachtung anderer Fehlerquellen ergibt eine mögliche systematische Verzerrung und eine Vergrößerung des Standardfehlers. Zentral ist dabei aus meiner Sicht der Messfehler in der Erhebung.

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. Es ist aber davon auszugehen, dass der Gesamtfehler bei der Schätzung der Einwohnerzahl beim Zensus höher ist als der jeweils ausgewiesene relative Standardfehler.

---

## Literaturverzeichnis

- Berg, Andreas und Bihler, Wolf (2011): Das Stichprobendesign zur Haushaltsstichprobe des Zensus 2011. *Wirtschaft und Statistik* April 2011: 317–328.
- Berg, Andreas und Bihler, Wolf (3/2014): Der Auswahlplan für die Ziehung der Neuzugänge der Haushaltsstichprobe des Zensus 2011. *Wirtschaft und Statistik* 3/ 2014: 151-154.
- Berg, Andreas und Bihler, Wolf (4/2014): Das Hochrechnungsverfahren zur Ermittlung der Einwohnerzahl im Zensus 2011. *Wirtschaft und Statistik* 4/ 2014: 229-236.
- Christensen, Björn, Christensen, Sören, Hoppe, Tim und Spandel, Michael (2014). Everthing counts! – Warum kleine Gemeinden die Gewinner der Zensuserhebung 2011 sind. Arbeitspapier, 8.9. 2014 unter <http://arxiv.org/abs/1409.1360>.
- Eckmann, Stefanie und Kreuter, Frauke (2011): Confirmation Bias in Housing unit listing. *Public Opinion Quaterly* 75 Nr.1: 1-12.
- Fuchs, Marek (2009): (Optimal) governance of research support by „Survey Methodology“. German Council for Social and Economic Data (RatSWD). Working paper Nr. 98, Juli 2009.
- Gronbach, Markus und Loth, Anett (2013): Zensus 2011 – Die Zusammenarbeit mit den kommunalen Erhebungsstellen in Niedersachsen. *Statistische Monatshefte Niedersachsen* 5: 248–251.
- Hirner, Stephanie und Stiglmayr, Susanne (2013): Der Referenzdatenbestand im Zensus 2011. *Wirtschaft und Statistik* 1/2013: 30–41.
- Münnich, Ralf und Gabler, Siegfried (2012): Stichprobenoptimierung und Schätzung im Zensus 2011. *Statistik und Wissenschaft* Band 21. Wiesbaden: Statistisches Bundesamt.
- Münnich, Ralf; Gabler, Siegfried; Ganninger, Matthias; Burgard, Jan Pablo und Kolb, Jan-Philipp (2012): Das Stichprobendesign desregistergestützten Zensus 2011. *Methoden – Daten – Analysen* 5(1): 37–61.
- Münnich, Ralf; Gabler, Siegfried; Ganninger, Matthias; Burgard, Jan Pablo und Kolb, Jan-Philipp (2013): Validierungsprojekt zum deutschen Zensus 2011. Abschlussbericht. Trier: Universität Trier. Fachbereich IV.
- Statistisches Amt der Stadt München (2013). Ergebnisse der Befragung zu den Erhebungsstellen im Zensus 2011.
- Ziprik, Katja (2012): Qualitätsaspekte des Anschriften- und Gebäuderegisters im Zensus 2011. *Wirtschaft und Statistik* 11: 947–956.