

Qualitätskontrolle bei Erhebungen durch Rohdatenanalyse am Beispiel des Mikrozensus

Nauenburg, Ricarda

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Nauenburg, Ricarda: Qualitätskontrolle bei Erhebungen durch Rohdatenanalyse am Beispiel des Mikrozensus.

In: Rehberg, Karl-Siebert (Ed.) ; Deutsche Gesellschaft für Soziologie (DGS) (Ed.): *Die Natur der Gesellschaft: Verhandlungen des 33. Kongresses der Deutschen Gesellschaft für Soziologie in Kassel 2006. Teilbd. 1 u. 2.* Frankfurt

am Main : Campus Verl., 2008. - ISBN 978-3-593-38440-5, 1950-1955.. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-152283>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Qualitätskontrolle bei Erhebungen durch Rohdatenanalyse am Beispiel des Mikrozensus

Ricarda Nauenburg

Erhebungen der amtlichen Statistik sind für ihre hohe Datenqualität bekannt. Die amtliche Statistik kommt damit ihrem Auftrag nach, statistische Informationen bereitzustellen und zu verbreiten, die objektiv, unabhängig und qualitativ hochwertig sind. In Zeiten knapper öffentlicher Kassen sind jedoch auch hier neue Ideen gefragt, um diesem Anspruch weiter gerecht werden zu können.

Eine hohe Datenqualität ist besonders für den Mikrozensus – der größten Erhebung zu den Lebens- und Arbeitsbedingungen der Bevölkerung in Deutschland wichtig, denn auf seiner Grundlage werden weitreichende politische Entscheidungen getroffen. Auch die wissenschaftliche Nutzung des Mikrozensus hat sich seit der Einrichtung der Forschungsdatenzentren der amtlichen Statistik erheblich intensiviert (vgl. Zühlke/Christians 2006). Allerdings sind die gegenwärtig eingesetzten Methoden zur Qualitätssicherung sowohl zeit- und personalintensiv als auch verbesserungsfähig. So lag es nahe, ein neues Verfahren zur Identifikation fälschender Interviewer, das bereits im Sozio-oekonomischen Panel (SOEP) erfolgreich getestet wurde (Schäfer u.a. 2005), auf seine Eignung für den Mikrozensus zu untersuchen. Es entstand ein Kooperationsprojekt zwischen dem Fraunhofer Institut für Rechnerarchitektur und Softwaretechnik FIRST.IDA, dem Statistischen Landesamt Berlin und dem Deutschen Institut für Wirtschaftsforschung (DIW) Berlin (vgl. auch Schäfer u.a. 2006), über dessen Ergebnisse hier zu berichten ist. Ein weiteres noch anzumerkendes und nicht zu unterschätzendes Novum des Projekts ist auch, dass die amtliche Statistik bereit war, externen Wissenschaftlern Zugriff auf Rohdaten einzuräumen mit der Aussicht auf möglicherweise kritische Ergebnisse.

Gefälschte Interviews bzw. fälschende Interviewer sind in der Branche ein heikles Thema und werden von den Umfrageinstituten lediglich intern behandelt. Fest steht jedoch, dass Fälschungen spätestens dann offenbar werden, wenn auch nach ihnen gesucht wird. Zu finden ist dabei natürlich die ganze Bandbreite vom unbeabsichtigten Fehleintrag über das fragwürdige Nachtragen vergessener Filter bis zur böswilligen Fälschung ganzer Interviews. Paneluntersuchungen erschweren das Fälschen, weil in der nächsten Welle die Fälschung offenbar wird. An auf diese Weise konventionell identifizierten Fälschern im Sozio-oekonomischen Panel wurde die neue Methode zuerst getestet. Sie konnten bereits in den Daten der ersten Welle ge-

funden bzw. bestätigt werden. Demnach kann das Verfahren auch bei Querschnittsuntersuchungen bereits eindeutige Hinweise geben.

Das neue Verfahren beruht auf der Auswertung der Rohdaten einer Erhebung. Es kommen zwei Methoden zum Einsatz, die – parallel angewendet – sich gegenseitig bestätigen und damit die Ergebnisse robuster machen. Beide Methoden berechnen aus allen Interviews eines Interviewers jeweils einen Kennwert. Danach wird die Wahrscheinlichkeit dieses Kennwerts in Bezug auf die vorliegende Stichprobe ermittelt. Anhand der Wahrscheinlichkeiten erfolgt ein Ranking der Interviewer. Interviewer, deren Kennwerte die geringste Wahrscheinlichkeit für ein zufälliges Zustandekommen aufweisen, sind potentiell verdächtig, Interviews gefälscht zu haben.

Beide Methoden nutzen zur Berechnung und Bewertung der Kennwerte empirische Gesetzmäßigkeiten und Tatsachen aus. Die erste Methode wendet das Benford-Gesetz an. Laut »Benfords Law« tritt die Ziffer Eins als erste Ziffer aller Zahlen vieler Datensätze mit höchster Wahrscheinlichkeit und die größte Ziffer mit niedrigster Wahrscheinlichkeit auf. Die Wahrscheinlichkeiten folgen einer logarithmischen Form. Daraus ergibt sich zum Beispiel im Dezimalzahlensystem für die Ziffer Eins eine Häufigkeit von 30,1 Prozent, für die Ziffer Zwei eine Häufigkeit von 17,6 Prozent und für die Ziffer Neun eine Häufigkeit von nur noch 4,6 Prozent. Voraussetzung ist jedoch, dass die zu testenden Werte metrisch, positiv und potentiell unbegrenzt sind. Empirisch feststellbare Abweichungen von der theoretischen Verteilung nach Benford können dann mögliche Fälschungen indizieren. Zur Beurteilung der Abweichungen wird der Chi-Quadrat-Test eingesetzt. Benfords Law ist ein empirisches Gesetz, das von Frank Benford 1938 zuerst veröffentlicht wurde. Theodore P. Hill konnte es 1995 theoretisch begründen. Benfords Law wird auch bei der Bilanzprüfung angewendet.

Die zweite zum Einsatz kommende Methode – die »Variability Method« – wurde speziell für die SOEP-Untersuchung entwickelt. Sie stützt sich auf die Annahme, dass die von Interviewern gefälschten Fragebögen im Vergleich eine unterdurchschnittliche Varianz der Antworten aufweisen. Diese Annahme rechtfertigt sich durch empirische Beobachtungen: Fälschende Interviewer kennen den Fragebogen und füllen ihn deshalb vollständig aus, machen keine Fehler und vermeiden vor allem extreme Antworten. Sie wollen damit eine genauere Plausibilitätskontrolle mit eventueller Nachfrage bei den Befragten unbedingt verhindern. Gefälschte Interviews sind daher »unauffällig«. Eine statistische Größe, die das abbilden kann, ist die Varianz bzw. Variabilität. Für die Berechnung der Testgröße werden für alle Variablen aller Fragebögen eines Interviewers die Varianzen ermittelt und addiert. Dann werden aus der Menge aller Fragebögen viele Male gleich große Fragebogencluster zufällig gezogen und für jedes die Variabilität ermittelt. Daraus ergibt sich eine Vergleichskurve, anhand derer die Wahrscheinlichkeit für ein zufälliges Zustande-

kommen des Interviewertestwertes bestimmt wird. Je kleiner diese Wahrscheinlichkeit, desto stärker der Verdachtsmoment gegen den betreffenden Interviewer. Die folgende gezielte Überprüfung kann jedoch auch ergeben, dass die geringe Variabilität durch eine sehr homogene Befragtengruppe (z.B. in einem Altenheim) entstanden ist. Solche Clustereffekte müssen natürlich im weiteren Vorgehen ausgeschlossen werden. Die Variabilitätsmethode ist anders als Benfords Law nicht skalenabhängig, auch ordinale oder nominale Variablen können analysiert werden. Denn die Annahme, dass Interviewer ihre Fälschungen in der (optischen) Mitte platzieren, gilt auch für diese Skalenniveaus. Außerdem geht es hier nicht um die inhaltliche Interpretation des Kennwertes, sondern um abstrakte Verteilungsmuster, die erkannt werden sollen. Allerdings müssen die Variablen mindestens drei Kategorien in einer lückenlosen Skala aufweisen und vor der Analyse normiert werden, damit nicht einzelne Variablen die Variabilitätskennwerte dominieren.

Beide Methoden sollten auf den Daten des Berliner Mikrozensus 2004 angewendet werden. Für den Mikrozensus 2004 wurden in Berlin rund 18.000 Haushalte mit circa 34.000 Personen befragt und die Erhebung von 170 Interviewern durchgeführt. Der Mikrozensus enthält jedoch im Gegensatz zum Sozio-oekonomischen Panel fast keine metrischen Variablen mit unbeschränktem Wertebereich – die Voraussetzung für die Anwendung von Benfords Law. Deshalb musste auf die Anwendung von Benfords Law erst einmal verzichtet werden. Ein zweiter Punkt ist die Unterstellung einer homogenen Befragtenpopulation. Die Datenerhebung beim Berliner Mikrozensus 2004 erfolgte auf verschiedenen Wegen. Prioritär waren Face-to-face-Interviews durch Interviewer. Diese benutzten aber nicht nur Papierfragebögen, sondern je nach Verfügbarkeit auch Laptops mit einer Befragungssoftware für die Interviews. Die Befragungssoftware stellt eine saubere Filterführung sicher und führt schon während des Interviews erste Plausibilisierungen durch. Wurden Zielfamilien wiederholt nicht angetroffen, erhielten sie die Möglichkeit, einen ausgefüllten Papierfragebogen an das Statistische Landesamt Berlin zu schicken oder das Interview mit einem Mitarbeiter telefonisch durchzuführen. Außerdem gab es noch ein zusätzliches Befragungsmodul – die Arbeitskräftestichprobe der Europäischen Union –, das nur ein Teil der Befragten auszufüllen hatte. Mit Hilfe der zusätzlichen Information konnten Antworten im Hauptteil des Fragebogens plausibilisiert werden. Der Erhebungsprozess des Mikrozensus legt deshalb von vornherein nahe, dass die Befragtenpopulationen der verschiedenen Erhebungsmodi nicht gleich sind und ihre Homogenität zumindest überprüft werden muss. Die Varianzunterschiede in den verschiedenen Samples könnten sonst die ohnehin schwachen Interviewereffekte überdecken.

Die Variability Method wurde deswegen zuerst auf die Überprüfung der Varianzhomogenität zwischen den verschiedenen Erhebungsmodi angewendet, mit dem Ergebnis, dass die Varianzen tatsächlich nicht homogen sind. Ebenfalls ent-

standen Varianzunterschiede je nachdem, ob Befragte in der EU-Stichprobe waren oder nicht. Die Variability Method müsste so innerhalb der Befragungsmodi angewendet werden. Prekärer aber ist, dass die Interviews des Berliner Mikrozensus 2004 den Interviewern gar nicht eindeutig zuzuordnen waren. Zum Beispiel teilten sich mehrere Interviewer einen Laptop. Oder ein Interview, das vom Statistischen Landesamt aus telefonisch erfasst wurde, ist dem Interviewer, der es ursprünglich durchführen sollte, »gutgeschrieben« worden. Die näherungsweise Zuordnungen mit einer durchschnittlichen Fehlerquote von 30 Prozent sind für die Variabilitätsmethode nicht akzeptabel. Die Überprüfung der Interviewer musste deswegen in diesem Test erst einmal ausfallen. Äußerst interessant ist jedoch, was im Homogenitätstest der Befragungsmodi gewissermaßen als unerwartetes Nebenprodukt angefallen ist. Bei der variablenweisen Durchführung sind nämlich diverse Unterschiede in der Varianz von Variablen offenbar geworden, die auf Erhebungseffekte, wenn nicht sogar -fehler hindeuten. Eine gezielte Suche nach Erklärungen ergab folgende Gründe für die Varianzunterschiede:

- abweichende Interpretation von Fragen bei geführten und Selbstausfüller-Interviews;
- unvollständige Codierungsschemen;
- inkonsistente Codierungsschemen;
- bei Selbstausfüllern Schwierigkeiten bei der Ausfüllung von Filtern.

Diese Mängel bei der Fragebogenkonstruktion, der Fragenformulierung bzw. der Erhebungsdurchführung führen je nach Erhebungsmodus zu charakteristischen Missverständnissen und daraus folgenden Erhebungsfehlern durch die Interviewer und im Fall der Selbstausfüller auch durch die Interviewten. Mikrozensus-Interviewer haben sich zum Beispiel daran gewöhnt, mit der Ziffer Acht ein »Nein« in binären Fragen zu vercoden. Bei der Frage an Migranten, wie viele ihrer Kinder im Ausland leben, tragen sie für die Antwort »keine« häufig auch die Ziffer Acht ein. Eine notorische Ungenauigkeit im Mikrozensus, die gemeinsame Vercodung von »keine Antwort« und »trifft nicht zu« in *einem* Wert bei manchen Fragen ist durch die Variability Method ebenfalls erkannt worden. Die bei den Laptop-Interviews eingesetzte Erhebungssoftware verhindert zwar Interviewabbrüche und führt einige Plausibilisierungen schon während der Erhebung durch, aber die Tests mit der Variability Method haben ergeben, dass auch sie nicht einwandfrei funktioniert: die automatische Plausibilisierung zum Beispiel des Schuleintrittsalters auf frühestens sechs Jahre bzw. des frühesten Geburtsjahres auf 1910 muss als zu streng betrachtet werden.

Ob die Testmethode auch beim Mikrozensus helfen kann, fälschende Interviewer zu identifizieren, wird sich frühestens mit den Daten des Jahres 2006 überprüfen lassen, denn erst dann ist die Zuordnung der Interviewer zu den Interviews

eindeutig erfasst. Aber auch mit den bis jetzt vorliegenden Ergebnissen können am Erhebungsdesign des Mikrozensus erhebliche Verbesserungen vorgenommen werden, die die Datenqualität spürbar erhöhen. Zum Beispiel ist ein computergestütztes Interview mit dynamischer Filterführung dem normalen Face-to-face-Interview vorzuziehen, da Filterfehler vermieden und Filterfragen komplett ausgefüllt werden. Selbstausfüller-Interviews sind dagegen so weit wie möglich zu vermeiden, weil sie am fehleranfälligsten sind. Der Mikrozensus wird dezentral in den Statistischen Ämtern der Länder erfasst. Eine Aufgabe für die Landesämter ist deshalb vor allem, in Abstimmung mit dem Bundesamt die durch die Methode aufgedeckten Missstände zu beseitigen, die Erhebungsdokumentation des Mikrozensus allgemein zu verbessern und dafür in allen Bundesländern einheitliche Standards einzuführen. Diese jetzt schon möglichen Korrekturen sollten nicht unterschätzt werden, sie werden die Datenqualität des Mikrozensus in der Summe deutlich mehr erhöhen als die eventuelle Überführung einzelner fälschender Interviewer.

Bei der Anwendung des neuen Verfahrens ist zu beachten, dass es keine Plausibilitätskontrolle ersetzen kann. Es kann zum Beispiel keine falschen Einzelwerte identifizieren. Seine Aussagen beruhen auf Wahrscheinlichkeiten, so dass auch eine gewisse Irrtumswahrscheinlichkeit in Kauf genommen werden muss. Deswegen sollten beide Methoden – Benford und die Variability Method – wenn möglich zusammen angewandt werden. Allerdings kann das Verfahren qualitätsrelevante Dateneigenschaften erkennen, die eine übliche – automatische oder »händische« – Plausibilisierung nicht erfasst. Mit Hilfe der ermittelten Kennwerte ist es auch möglich, die Interviewerkontrolle bei Erhebungen erheblich effizienter durchzuführen: Das neue Verfahren liefert gezielte Hinweise, welche Interviewer kontrolliert werden sollten.

Das Statistische Landesamt Berlin hat die Absicht, aus dieser Methode ein anwenderfreundliches, universelles Software-Tool entwickeln zu lassen und für die Mikrozensuserhebung einzusetzen.

Literatur

- Benford, Frank (1938), »The Law of Anomalous Numbers«, *Proceedings of the American Philosophical Society*, Bd. 78, Nr. 4, S. 551–572.
- Hill, Theodore P. (1995), »Base-Invariance Implies Benford's Law«, *Proceedings of the American Mathematical Society*, Bd. 123, S. 887–895.
- Schäfer, Christin/Schräpler, Jörg-Peter/Müller, Klaus-Robert/Wagner, Gert G. (2005), »Automatic Identification of Faked and Fraudulent Interviews in the German SOEP«, *Schmollers Jahrbuch – Zeitschrift für Wirtschafts- und Sozialwissenschaften*, Jg. 125, H. 1, S. 183–193.

-
- Schäfer, Christin/Bömermann, Hartmut/Nauenburg, Ricarda u.a. (2006), »Qualitätssicherung des Faktors Interviewer für den Mikrozensus«, in: Forschungsdatenzentrum der Statistischen Landesämter (Hg.), *Amtliche Mikrodaten für die Sozial- und Wirtschaftswissenschaften*, S. 7–21.
- Zühlke, Sylvia/Christians, Helga (2006), »Datenangebot und Datenzugang im Forschungsdatenzentrum der Statistischen Landesämter«, *Statistische Analysen und Studien Nordrhein-Westfalen*, Bd. 29, S. 3–11.