Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

Universität Bamberg

Methoden und Potenziale des Zensus 2011

Dokumentation zur Tagung am 26. und 27. Juli 2012
Bibliothek des Staatlichen Bauamts, Bamberg

Weitere Informationen unter www.statistik.bayern.de

Statistik-Tage Bamberg-Fürth 2012

Organisation:
Bayerisches Landesamt für Statistik und Datenverarbeitung
Lehrstuhl für Statistik und Ökonometrie der Otto-Friedrich-Universität Bamberg

Ansprechpartnerin:
Dipl.-Pol. Daniela Lamprecht
Tel.: (0911) 98208-273
E-Mail: statistiktage@lfstad.bayern.de
Internet: www.statistik.bayern.de

Inhalt

Einführung

Dr. Michael Fürnrohr:
„Überblick über den Zensus“

Abstract:

Der Vortrag bildet den Einstieg in die Veranstaltung und gibt einen Überblick über das Gesamtprojekt des registergestützten Zensus 2011. Zunächst wird die Projektstruktur vorgestellt und erläutert, aus welchen einzelnen Datenquellen sich die später zu veröffentlichenden Zensusdaten zusammensetzen.

Wie an der Bezeichnung „registergestützter Zensus“ bereits erkennbar ist, werden verschiedene Register, aber auch primärstatistische Erhebungen zur Gewinnung der benötigten Zensusdaten eingesetzt. Ein solches Kombinationsmodell macht die Anwendung von Techniken der Datenfusion und von Record-Linkage-Verfahren erforderlich, die auch den thematischen Schwerpunkt des ersten Blocks der Veranstaltung bilden.

Grundlage des Zensus bildet das Anschriften- und Gebäuderegister, das aus Daten der Melderegister, der Bundesagentur für Arbeit und der Vermessungsämter erstellt wurde. Die demographischen Grunddaten des Zensus liefern die Melderegister, die zu zwei Zeitpunkten von den Kommunen geliefert, abgeglichen und in Einzelfällen primärstatistisch korrigiert wurden. In der primärstatistisch als Vollerhebung durchgeführten Gebäude- und Wohnungszählung wurden alle Eigentümer von Gebäuden und Wohnungen postalisch befragt. Für diese Befragung wurde ein Gebäude- und Wohnungseigentümerregister aus den Daten der Grundsteuerstellen erstellt und für die Erhebung genutzt.

Die zweite große primärstatistische Erhebung des Zensus war die Haushaltsstichprobe. Hier wurden zehn Prozent der Bevölkerung von Interviewern befragt. Die Konzeption des Stichprobendesigns, das Hochrechnungsverfahrens sowie die Umsetzung von Korrekturen im Rahmen der Haushaltegenerierung sind die Themen des zweiten Teils im Rahmen der Statistik-Tage Bamberg-Fürth zum Zensus.

All diese Daten fließen in die Haushaltegenerierung ein, die die Daten zusammenführt, korrigiert und letztendlich den Zensuseinzeldatenbestand erstellt, der fachlich und räumlich hoch differenzierte Auswertungen ermöglicht. Veröffentlicht werden die Ergebnisse mit Hilfe einer Auswertungsdatenbank, durch die Kommunen, Wissenschaft sowie auch die Öffentlichkeit die Ergebnisse des Zensus nutzen können. Über das hierfür erforderliche Geheimhaltungsverfahren sowie über den Aufbau und die Funktionalitäten der Auswertungsdatenbank wird im dritten Teil der Veranstaltung informiert.

Zur Person:

Dr. Michael Fürnrohr ist Leiter der Abteilung „Bevölkerung, Haushalte, Zensus, Erwerbstätigkeit, Finanzen, Rechtspflege, Schulen“ im Bayerischen Landesamt für Statistik und Datenverarbeitung. In dieser Funktion ist er auch Projektleiter „Zensus 2011“ für Bayern und Mitglied der Projektleitung der Statistischen Ämter von Bund und Ländern. Neben dem Zensus engagiert er sich insbesondere bei den Themen „Demographischer Wandel“ und „Bevölkerung mit Migrationshintergrund“.

Vortragsfolien:

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

zensus 2011

Überblick über den Zensus 2011

Dr. Michael Fürnrohr
Bayerisches Landesamt für Statistik und Datenverarbeitung

Gliederung

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

- Überblick
- Anschriften- und Gebäuderegister
- Melderegister
- Gebäude- und Wohnungszählung
- Haushaltsstichprobe
- Sonderbereiche
- Haushaltegenerierung
- Auswertungsdatenbank

Überblick I

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

Daten und Fakten zum Zensus 2011

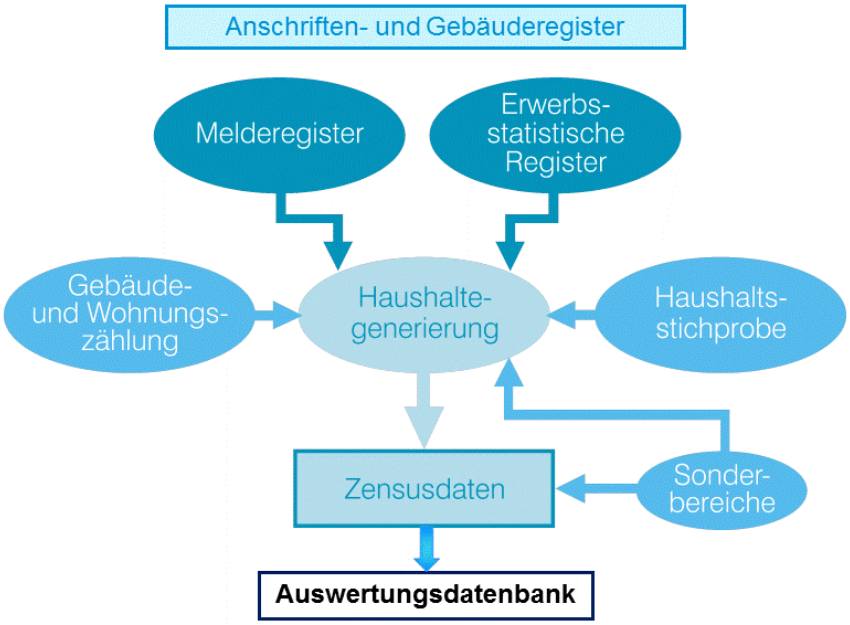
- Letzte Volkszählungen: alte Bundesländer 1987; neue Bundesländer 1981
- Erstmals in Deutschland „registergestützt“ statt primärstatistischer Vollerhebung
- Bundesweite Kosten: 720 Mio. Euro
- Rechtsgrundlagen: EU-Verordnung, Zensusvorbereitungsgesetz, Zensusgesetz
- Vorbereitung seit Mitte 2005

2/13

Überblick II

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.



3/13

Anschriften- und Gebäuderegister (AGR) I

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

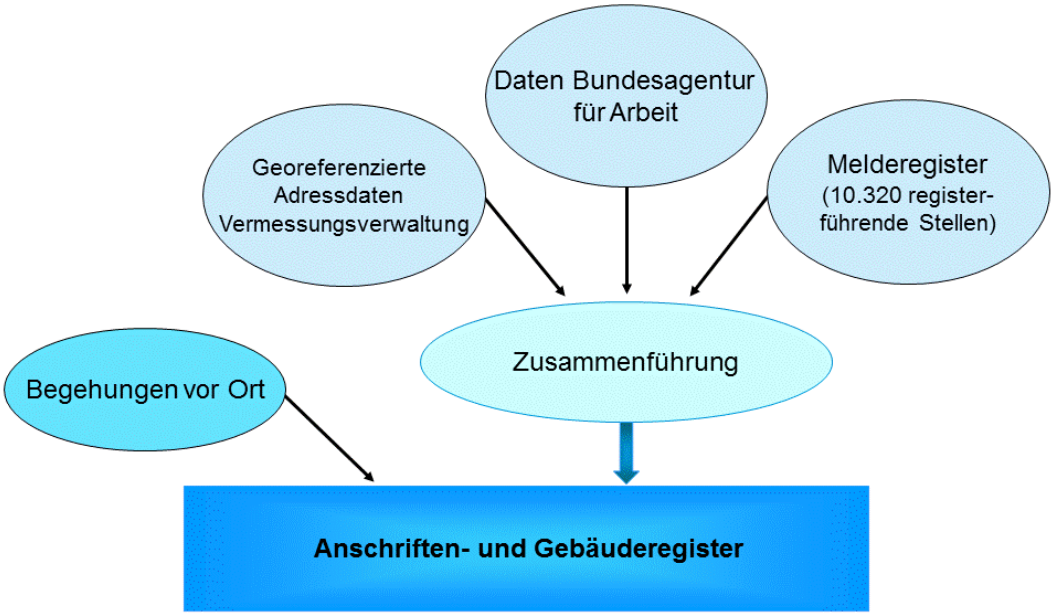
- Ziel: Verzeichnis aller Gebäude und Wohnungen in Deutschland
- Verbindung verschiedener Datenquellen, die über die Anschrift zusammengeführt werden müssen
- Ermittlung aller existierenden Gebäude mit Wohnraum einschließlich aller bewohnten Unterkünften
- Zweck: Datenbasis aller Erhebungsteile des Zensus 2011

4/13

Anschriften- und Gebäuderegister (AGR) II

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.



5/13


Melderegister
Bayerisches Landesamt für Statistik und Datenverarbeitung
Mit Bayern rechnen.
Melderegisterlieferung zum 1. November 2010
Melderegisterlieferung zum 9. Mai 2011
Melderegisterlieferung zum 9. August 2011
Zusammenführung
Melderegisterdaten
Mehrfachfallprüfung
Zusammenführung mit Erwerbsstatistik
6/13


GWZ:
Gebäude- und Wohnungseigentümerregister
Bayerisches Landesamt für Statistik und Datenverarbeitung
Mit Bayern rechnen.
Grundsteuerdaten
Daten der Ver- und Entsorger
automatisiertes Liegenschaftsbuch
Zusammenführung
Gebäude- und Wohnungseigentümerregister
(Ermittlung der Auskunftspflichtigen zum AGR-Bestand)
Primärstatistische Vollerhebung aller Gebäude und Wohnungen
7/13

Gebäude und Wohnungszählung (GWZ)

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

- Erhebungsart: Vollerhebung mit Auskunftspflicht
- Erhebungseinheiten: ca. 19,2 Mio. Wohngebäude/Eigentumswohnungen
- Befragte: ca. 19,0 Mio. Eigentümer von Wohngebäuden/Eigentumswohnungen
- Befragungsart: Postalischer Versand von Fragebogen
- Ziel: Datengewinnung über Gebäude und Wohnungen für wohnungspolitische und raumplanerische Entscheidungen

8/13

Haushaltsstichprobe

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

- Erhebungsart: Stichprobe mit Auskunftspflicht in Gemeinden über 10.000 Einwohnern
- Erhebungseinheiten: ca. 2,0 Mio. Anschriften; ca. 7,5 Mio. existente Personen
- Stichprobenverfahren entwickelt von Prof. Dr. Münnich und PD Dr. Gabler
- Befragungsart: Interview
- Ziele:
  1. Korrektur der Über- und Untererfassungen der Melderegister
  2. Gewinnung zusätzlicher Informationen (z.B. Bildung und Religion)

9/13

Sonderbereiche

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

- Unterteilung der Sonderbereiche in sensible und nicht-sensible Bereiche
- Erhebungsart: Vollerhebung mit Auskunftspflicht
- Erhebungseinheiten: ca. 25.800 nicht-sensible Sonderbereiche
  ca. 35.000 sensible Sonderbereiche
- Befragungsart: Interview
- Ziel: Feststellung der Über- und Untererfassungen der Melderegister

10/13

Haushaltegenerierung

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

- Ziele: 1. Zensustypischer Datensatz
  2. Basisdaten für die Bevölkerungsfortschreibung
  3. Ermittlung von Daten zur Zahl und Struktur von Haushalten
- Synthese verschiedener Erhebungsteile des Zensus 2011
- Ermöglicht fachlich und regional tiefgegliederte, erhebungsteilübergreifende Kombinationsauswertungen

11/13

Auswertungsdatenbank

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

- „Herzstück“ der Veröffentlichung im Internet sowie der Veröffentlichung insgesamt
- Breites Spektrum an Auswertungsmöglichkeiten für unterschiedliche Nutzergruppen
- Öffentlicher Zugriff über www.zensus2011.de
- Veröffentlichung von Ergebnissen in Form von „statischen“ und „dynamischen“ Inhalten
- Geheimhaltungsverfahren „SAFE“
- Forschungsdatenzentrum

12/13

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

Vielen Dank
für Ihre Aufmerksamkeit!

zensus2011

13/13

Vortragsblock I: Matching- und Linkageverfahren

Prof. Susanne Rässler:
„Statistische Matching-Verfahren“

Abstract:

Der Vortrag gibt einen Überblick über gängige statistische Matching-Verfahren. Diese lassen sich den Missing-Data-Techniken zuordnen und als Ergänzungs- bzw. Imputationsverfahren verstehen. Im Gegensatz zum Record-Linkage werden nicht identische Einheiten (Personen, Haushalte, Firmen) in verschiedenen Datensätzen anhand von identifizierenden Schlüsseln (wie Sozialversicherungsnummer, Name und Adresse) gesucht, sondern in bestimmten Kovariablen (etwa Geschlecht, Alter, Bildung) ähnliche Einheiten. Von diesen ähnlichen Einheiten werden dann bestimmte Merkmale, die der Einheit im anderen Datensatz fehlen, übertragen. Der Vortrag strukturiert unterschiedliche Ausfallmuster, definiert den Begriff der Datenfusion und gibt einen Einblick in die Voraussetzungen und Annahmen von Imputationsverfahren. Abschließend werden die Vor- und Nachteile von einfachen („single“) und mehrfachen („multiple“) Imputationsverfahren gegenüber gestellt.

Zur Person:

Prof. Susanne Rässler ist Inhaberin des Lehrstuhls für Statistik und Ökonometrie an der Otto-Friedrich-Universität Bamberg und Mitglied der Zensus-Kommission, dem wissenschaftlichen Beirat des Großprojekts. Zuvor war sie Leiterin des Kompetenzzentrums Empirische Methoden am Institut für Arbeitsmarkt- und Berufsforschung und des Bereichs Produkt- und Programmanalyse der Bundesagentur für Arbeit sowie Professorin für Computational Statistics an der Frankfurt School of Finance & Management in Frankfurt a.Main. Ihre Forschungsschwerpunkte liegen u.a. im Bereich der Missing-Data-, Imputations- und Datenfusionstechniken sowie der Gütemessung bei statistischen Matching-Verfahren.

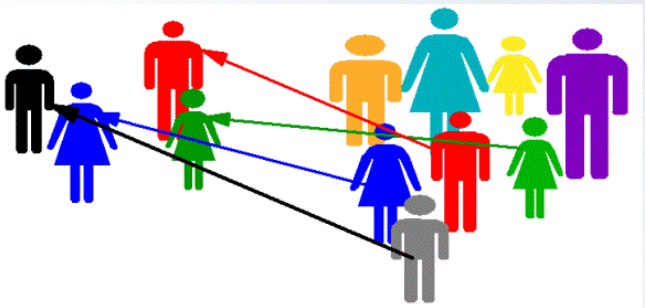
Vortragsfolien:

Statistische Matching-Verfahren

Erste Statistik-Tage 2012 Bamberg | Fürth
26./27. Juli 2012 in Bamberg

Prof. Dr. Susanne Rässler
Lehrstuhl für Statistik und Ökonometrie
in den Sozial- und Wirtschaftswissenschaften
Otto-Friedrich-Universität Bamberg

Statistische Matching-Verfahren = suche einen statistischen Zwilling



Agenda

- Einführung: Rekord Linkage vs. Datenfusion
- Babylonische Sprachverwirrung
- Missing Data: Ausfallmuster, -mechanismen und Techniken
- Statistisches Matching: Definition, Anwendung und Verfahren
- Einfache und mehrfache Imputationsverfahren
- Zusammenfassung

Agenda

- Einführung: Rekord Linkage vs. Datenfusion
- Babylonische Sprachverwirrung
- Missing Data: Ausfallmuster, -mechanismen und Techniken
- Statistisches Matching: Definition, Anwendung und Verfahren
- Einfache und mehrfache Imputationsverfahren
- Zusammenfassung

Record Linkage

- Verknüpfung von Datensätzen aus verschiedenen Quellen ABER mit denselben Objekten (Haushalten, Personen, Firmen)
- Zusammenführung über identifizierende Schlüssel (Sozialversicherungsnummer, Name und Adresse, …)

Datenfusion

- Verknüpfung von Datensätzen aus verschiedenen Quellen mit unterschiedlichen Objekten (Haushalten, Personen, Firmen)
- Zusammenführung über identische/ähnliche Ausprägungen (Geschlecht, Alter, Bildung, Familienstand, Bundesland,…)

Beispiel: Datenfusion

Fernsehpanel

Nr.	Sex	Alter	...	Fernseh-verhalten
1	...	...	...	...
...	...	...	...	...
i	1	36	...	...
...	...	...	...	...
n(R)	...	...	...	...

Verbraucherpanel

Nr.	Sex	Alter	...	Verbraucher-verhalten
1	...	...	...	...
...	...	...	...	...
j	1	38	...	...
...	...	...	...	...
n(D)	...	...	...	...

Fusionsstichprobe

Nr.	Sex	Alter	...	Fernsehverh.	Verbaucherverh.
1	...	...	...	...	...
...	...	...	...	...	...
i	1	36	...	...	...
...	...	...	...	...	...
n(R)	...	...	...	...	...

➔ **Traditionelle Verfahren fusionieren i.a. „nächste Nachbarn“, d.h. verwenden statistische Matching-Verfahren**

Agenda

- Einführung: Rekord Linkage vs. Datenfusion
- **Babylonische Sprachverwirrung**
- Missing Data: Ausfallmuster, -mechanismen und Techniken
- Statistisches Matching: Definition, Anwendung und Verfahren
- Einfache und mehrfache Imputationsverfahren
- Zusammenfassung

Babylonische Sprachverwirrung I

Datenfusion

Propensity Score Matching

Statistisches Matching

Fusion

Single Imputation

Multiple Imputation

Babylonische Sprachverwirrung II

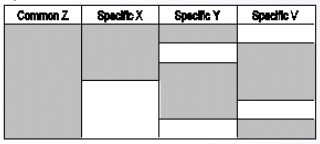
- **USA und Kanada: „Statistical Matching“ meint meistens Datenfusion (Data Fusion) mit dessen Identifikationsproblem**
  - ➔ Siehe Rässler, S. (2002). Statistical Matching: A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches. Lecture Notes in Statistics 168, Springer, New York.
- **Europa: Statistisches Matching meint das Auffinden von „statistischen Zwillingen“**
  - ➔ Siehe Bacher, J. (2002). Statistisches Matching: Anwendungs-möglichkeiten, Verfahren und ihre praktische Umsetzung in SPSS, ZA-Informationen, 51, S. 38-66.
- **Imputation: Ergänzung von fehlenden Werten einmal (single) oder mehrfach (multiple)**
  - ➔ Siehe Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. Wiley, New York.

Agenda

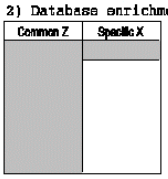
- Einführung: Rekord Linkage vs. Datenfusion
- Babylonische Sprachverwirrung
- **Missing Data: Ausfallmuster, -mechanismen und Techniken**
- Statistisches Matching: Definition, Anwendung und Verfahren
- Einfache und mehrfache Imputationsverfahren
- Zusammenfassung

Missing Data: Ausfallmuster

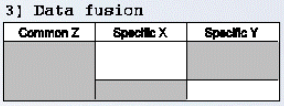
1) General situation of variables missing in



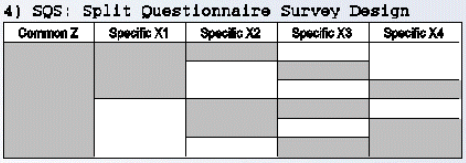
2) Database enrichment



3) Data fusion



4) SQS: Split Questionnaire Survey Design



- **Datenfusion: die spezifischen Variablen X und Y werden nicht gemeinsam beobachtet; es liegen keine gemeinsamen Einheiten vor**
- **Statistische Matching-Verfahren: Die Ergänzung von Merkmalen erfolgt auf Basis nächster Nachbarn, d.h. „statistischer Zwillinge“**

Missing Data: Ausfallmechanismen

Lfd. Nr.	Geschlecht	Alter	Bildung	Gesundheit	Personen-netto-EK	...
1	Weiblich	40-45	Hoch	Gut	?	...
2	Männlich	30-35	Mittel	Schlecht	4500-5000	...
3	Weiblich	>60	?	Mittel	4000-4500	...
4	Männlich	20-25	Hoch	?	?	...
5	Männlich	20-25	Gering	?	1500-2000	...
6	Weiblich	30-35	Gering	Gut	1500-2000	...
...	...	...	...	...	...	...

Löschung der Fälle

Lfd. Nr.	Geschlecht	Alter	Bildung	Gesundheit	Personen-netto-EK	...
2	Männlich	30-35	Mittel	Schlecht	4500-5000	...
6	Weiblich	30-35	Gering	Gut	1500-2000	...
...	...	...	...	...	...	...

- **Missing Completely at Random (MCAR): rein zufälliger Datenausfall**
- **Missing at Random (MAR): bedingt zufälliger Datenausfall**
- **Not Missing at Random (NMAR): systematischer, verzerrender Datenausfall**
- **Missing by Design: Nicht erfragte Merkmale (Datenfusion, SQS)**

Missing Data: Techniken

- **Verfahren, die nur die verfügbaren (AC) oder die vollständigen (CC) Informationen verwenden: Problem MCAR Annahme und hoher Datenverlust**
- **Gewichtung, i.allg. bei Teilnahmeverweigerung also Totalausfall des Interviews**
- **Likelihood-basierte Parameterschätzungen, z.B. Expectation-Maximization Algorithmus von Dempster, Laird und Rubin (1977)**
- **Einfache Ergänzung / Single Imputation (Vor. MAR) und Korrektur der Varianzschätzung (!) für statistisch valide Inferenz**
- **Mehrfache Ergänzung / Multiple Imputation (Vor. MAR) nach Rubin (1978, 1987, ...) mit Standardschätzung auf mehreren (m) Datensätzen und Kombination der Ergebnisse nach Rubin‘s Combining Rules**

Agenda

- Einführung: Rekord Linkage vs. Datenfusion
- Babylonische Sprachverwirrung
- Missing Data: Ausfallmuster, -mechanismen und Techniken
- **Statistisches Matching: Definition, Anwendung und Verfahren**
- Einfache und mehrfache Imputationsverfahren
- Zusammenfassung

Statistisches Matching: Definition & Anwendung

- **Suche für jede Person i aus B_1 in B_2 ein oder mehrere Fälle i*, die sich von der Person i in den Variablen X_i nicht oder nur geringfügig unterscheiden, also z.B. eine Person, die gleich alt ist, dieselbe Schulbildung hat und dasselbe Geschlecht hat wie Fall i. Ergänze die interessierenden Merkmale von i* bei i**
- **Anwendungsgebiete nach Bacher (2002):**

➔ **Datenfusion:** Zwei Datensätze sollen über eine Menge gemeinsamer Merkmale fusioniert werden

➔ **Bestimmung einer Kontrollgruppe:** Zu einer Untersuchungsgruppe soll zur Effektschätzung eine Kontrollgruppe aus anderen Daten gezogen werden, die sich hinsichtlich einer Menge an Kovariablen nicht unterscheidet

➔ **Item Nonresponse:** Ergänzung fehlender Information in einem Datensatz, z.B. fehlt bei einigen Personen die Einkommensangabe

Statistisches Matching: Verfahren

- **Auswahl von geeigneten Variablen:** Alter, Geschlecht, Familienstand, …, im Quadrat? Logs?
- **Auswahl eines Suchverfahrens:** Zufallsanordnung, mit oder ohne Zurücklegen, …
- **Auswahl eines Verfahrens zur Berechnung der Ähnlichkeit:** Propensity Score Matching vs. Distanzmaße (z.B. Mahalanobis Distanz Matching, Minkowski q-Metrik, …)
- **Ergänzung:** einfach oder gar mehrfach?

➔ **Achtung: Das kommt auf die Fragestellung an!!!!!**

Datenfusion: Bedingte Unabhängigkeit

Originäre Beziehung von Kauf- und Fernsehverhalten

Kauf von Knoblauchpillen	Kinder		Ältere Leute		Summe
	Werbung: ja	Werbung: nein	Werbung: ja	Werbung: nein	
Ja	5	0	50	10	65
Nein	95	100	50	90	335
Summe	100	100	100	100	400

Verbraucherpanel

Kauf von Knoblauchpillen	Kinder	Ältere Leute	Summe
Ja	5	60	65
Nein	195	140	335
Summe	200	200	400

Fernsehpanel

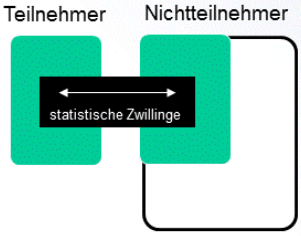
Werbung gesehen	Kinder	Ältere Leute	Summe
Ja	100	100	200
Nein	100	100	200
Summe	200	200	400

Durchschnittliche künstliche Fusionsstichprobe aus 1000 Simulationen

Kauf von Knoblauchpillen	Kinder		Ältere Leute		Summe
	Werbung: ja	Werbung: nein	Werbung: ja	Werbung: nein	
Ja	2.49	2.51	29.92	30.03	64.95
Nein	97.51	97.49	70.08	69.97	335.05
Summe	100.00	100.00	100.00	100.00	400.00

- ➔ **Distanzmaße zum Matching geeignet**
- ➔ **Propensity Score Matching NICHT (!) für Datenfusion geeignet**
- ➔ **Multiple Imputationsverfahren generell geeignet**
- ➔ **Aber: Nach der Fusion sind die spezifischen Variablen bedingt unabhängig gegeben die gemeinsamen Variablen**

Bestimmung einer Kontrollgruppe (zur Schätzung eines Behandlungseffekts)



Beispiel:
Haben sich die Beschäftigungschancen der Maßnahmeteilnehmer durch die Förderung verbessert?

Vergleichsmaßstab:
Beschäftigungschancen von statistischen Zwillingen ohne Förderung

- ➔ **Distanzmaße zum Matching geeignet (aber Dimensionalitätsproblem!)**
- ➔ **Propensity Score Matching sehr gut geeignet**
- ➔ **Erweiterungen durch parametrische Imputationsverfahren können sehr sinnvoll sein**

Agenda

- Einführung: Rekord Linkage vs. Datenfusion
- Babylonische Sprachverwirrung
- Missing Data: Ausfallmuster, -mechanismen und Techniken
- Statistisches Matching: Definition, Anwendung und Verfahren
- **Einfache und mehrfache Imputationsverfahren**
- Zusammenfassung

Item Nonresponse: Single Imputation

Ersetze einen fehlenden Wert durch einen "Schätzer“, z.B.

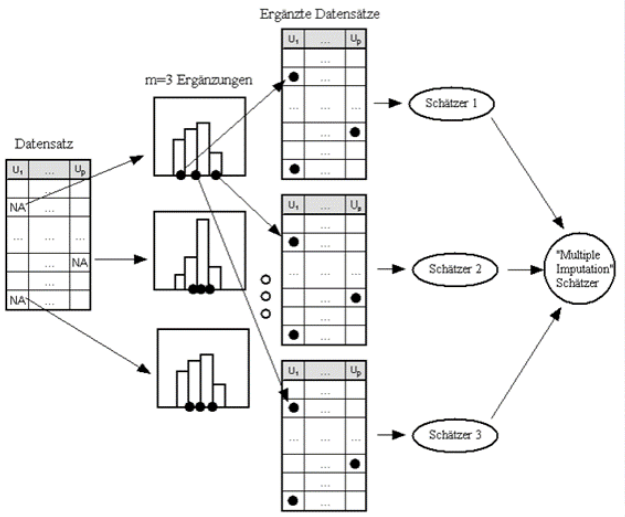
- Mittelwerte
- Bedingte Mittelwerte
- Hot Deck Verfahren
- Alle statistischen Matching-Verfahren
- Regressionsergänzung
- Regressionsergänzung mit Zufallsfehler

➔ Annahme: MAR

➔ Liefert unverzerrte Punktschätzung

➔ Diese Verfahren tendieren alle dazu, Varianzen und Korrelationen der Teststatistiken zu “zerstören”, d.h. erzeugen zu kleine p-Werte, zu signifikante Ergebnisse SOFERN keine Korrektur vorgenommen wird…

Item Nonresponse: Multiple Imputation



➔ Basis: Generiere Zufallszüge für die fehlenden Werte = parametrische MI

➔ Alle statistischen Matching-Verfahren über Bayesian Bootstrap einbindbar

➔ Annahme: MAR

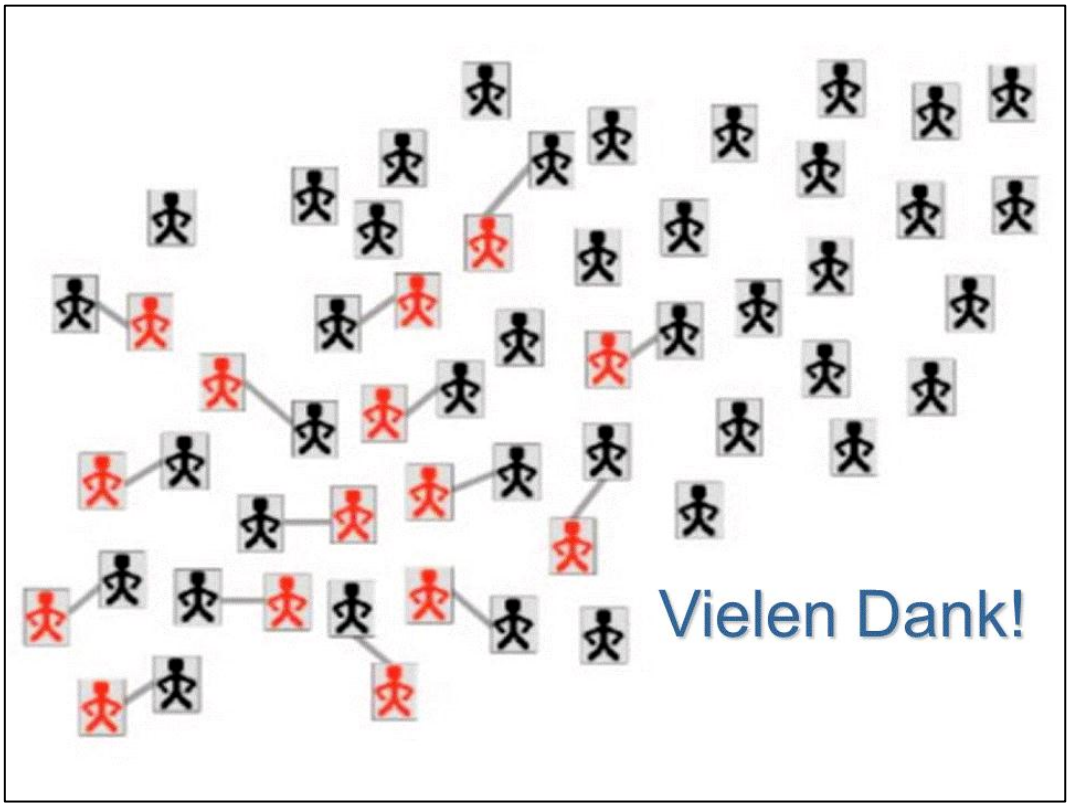
➔ Analysen mit Standardsoftware statistisch valide

Agenda

- Einführung: Rekord Linkage vs. Datenfusion
- Babylonische Sprachverwirrung
- Missing Data: Ausfallmuster, -mechanismen und Techniken
- Statistisches Matching: Definition, Anwendung und Verfahren
- Einfache und mehrfache Imputationsverfahren
- **Zusammenfassung**

Zusammenfassung

- **Bitte zunächst genau die Begrifflichkeiten klären!**
- **Datenfusion:** Statistische Zwillingssuche über Distanzmaße (NICHT PS-Matching) funktioniert nur unter der CIA, besser noch sind parametrische Imputationsverfahren mit informativer a priori!
- **Kontrollgruppen:** Gold-Standard ist Propensity Score Matching für Kontrollgruppen zur Analyse von Behandlungseffekten
- **Zur Ergänzung fehlender Werte ganz allgemein:** Gold-Standard sind multiple Ergänzungsverfahren, sofern Interesse an valider statistischer Inferenz besteht
- **Über den sog. „approximate Bayesian Bootstrap“ können die vorgestellten statistischen Matchingverfahren eingebunden werden**


Vielen Dank!

Prof. Rainer Schnell:

„Statistische Verfahren des Record-Linkage“

Abstract:

Durch die Verknüpfung bereits vorhandener Forschungsdaten aus verschiedenen Quellen (Record-Linkage) kann deren Analysepotential bei relativ geringem Ressourceneinsatz beträchtlich gesteigert werden. Allerdings erfordert Record-Linkage häufig den Einsatz besonderer Verfahren, da die meisten Datenbestände nicht mit eineindeutigen Identifikationsnummern versehen sind, sondern nur über Identifikatoren wie Namen und Adressangaben verknüpft werden können. Für eine erfolgreiche Anwendung von Record-Linkage-Verfahren sind Kenntnisse ihrer theoretischen Grundlagen unverzichtbar. Der Vortrag stellt einführend die Theorie von Record-Linkage-Verfahren in der Praxis vor. Dargestellt werden deterministisches, distanzbasiertes und probabilistisches Record-Linkage, Stringähnlichkeits funktionen, Blocking-Verfahren und Schwellenwertbestimmung. Abschließend werden Möglichkeiten erwähnt, wie auch verschlüsselte Identifikatoren fehlertolerant abgeglichen werden können.

Zur Person:

Prof. Dr. Rainer Schnell ist Inhaber des Lehrstuhls für empirische Sozialforschung an der Universität Duisburg-Essen. Seine Forschungsschwerpunkte liegen vor allem im Bereich des Entwurfs komplexer Stichproben, Analysen zu den Ursachen, der Vermeidung und Korrektur von Nonresponse sowie der Entwicklung von Record-Linkage-Verfahren. Er ist Herausgeber der Zeitschrift „Survey Research Methods“ und Verfasser der Lehrbücher „Graphisch gestützte Datenanalyse“ (1994), „Nonresponse in Bevölkerungsumfragen“ (1997), „Methoden der empirischen Sozialforschung“ (9. Auflage 2012) und „Survey-Interviews: Methoden standardisierter Befragungen“ (2012).

Vortragsfolien:

Der Inhalt des Vortrags wird voraussichtlich im Juni 2013 veröffentlicht. Der Arbeitstitel der Monographie lautet Schnell, R./Bachteler,T.: Record-Linkage.

Dipl.-Math. Marco Reisch:

„Record-Linkage im Zensus 2011: Der maschinelle Namensabgleich der Haushaltegenerierung“

Abstract:

Die Europäische Union hat für den Zeitraum 2010/2011 einen gemeinschaftsweiten Zensus angeordnet. In Deutschland wird dieser Zensus nach dem Zensusgesetz 2011 zum Stichtag 9. Mai 2011 erstmalig als registergestütztes Verfahren durchgeführt. Dabei werden in weiten Teilen Daten aus Verwaltungsregistern verwendet, welche mit den Ergebnissen gesonderter Befragungen ergänzt werden. Die besondere Herausforderung besteht anschließend in der Verknüpfung der einzelnen Datenquellen, um daraus Informationen zu Wohnhaushalten gewinnen zu können – der Haushaltegenerierung.

Der maschinelle Namensabgleich als Bestandteil der Haushaltegenerierung dient der Zuordnung von Melderegisterdaten und Wohnungsdaten aus dem Erhebungsteil der Gebäude- und Wohnungszählung. Dieses Verfahren ist notwendig, da in den Verwaltungsregistern keine Informationen darüber enthalten sind, in welcher Wohnung eine Person an einer Anschrift lebt. Um dies herauszufinden, werden die in der Gebäude- und Wohnungszählung erfragten bis zu zwei Wohnungsnutzer pro Wohnung einer Anschrift mit den im Melderegister gemeldeten Personen der Anschrift zusammengeführt.

Der Vortrag gibt einen Einblick in die einzelnen Verfahrensschritte des maschinellen Namensabgleichs und vermittelt ein Bild, wie mit diversen Hürden – zum Beispiel Schwächen der Beleglesung – umgegangen wird.

Zur Person:

Marco Reisch ist Referent im Sachgebiet „Zensus – Register, Haushaltegenerierung, Auswertung, Querschnittsaufgaben“ des Bayerischen Landesamts für Statistik und Datenverarbeitung. Zu seinem Aufgabengebiet im Rahmen des Zensus-Teilprojekts Haushaltegenerierung zählt unter anderem die Konzeption und Weiterentwicklung des maschinellen Namensabgleichs, der bundesweiten Zusammenführung von Melderegister- und Wohnungsdaten. Vor seiner Tätigkeit im Projekt „Zensus 2011“ studierte er Diplom-Mathematik an der Katholischen Universität Eichstätt-Ingolstadt.

Vortragsfolien:

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

RECORD-LINKAGE IM zensus2011

Marco Reisch
Bayerisches Landesamt für Statistik und Datenverarbeitung

Gliederung

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

- **Überblick: Record-Linkage im Zensus 2011**
- **Die Haushaltegenerierung**
- **Der maschinelle Namensabgleich**

Übersicht: Record-Linkage im Zensus 2011

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

- **Erstellung des Anschriften- und Gebäuderegisters**
  → Zusammenführung **unterschiedlicher** Quellen, um ein Register aller Anschriften Deutschlands mit Wohnraum zu gewinnen
- **Zusammenführung unterschiedlicher Melderegisterbestände**
  → Konsolidierter Melderegisterbestand zum Stichtag des Zensus 2011
- **Identifikation von Personendubletten für die Mehrfachfallprüfung**
  → Bereinigung des konsolidierten Melderegisterbestands um Personen mit doppeltem Hauptwohnsitz bzw. alleinigem Nebenwohnsitz
- **Anbindung erwerbsstatistischer Daten an den Melderegisterbestand**
  → Erweiterung des Merkmalskranzes der Melderegisterpersonen, um Auswertung übergreifender Merkmalskombinationen zu ermöglichen
- **Erstellung des Gebäude- und Wohnungseigentümerregisters**
  → Register aller Auskunftspflichtiger zum AGR-Bestand für die Durchführung der Gebäude- und Wohnungszählung
- **Abgleich von elektronischen Erhebungslisten der Haushaltebefragung mit dem konsolidierten Melderegister**
  → abschließende Existenzfeststellung für Personensätze im konsolidierten Melderegister

Gliederung

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

- **Übersicht: Record-Linkage im Zensus 2011**
- **Die Haushaltegenerierung**
- **Der maschinelle Namensabgleich**

Die Haushaltegenerierung - Das Modell

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

4/18

Gliederung

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

- Übersicht: Record-Linkage im Zensus 2011
- Die Haushaltegenerierung
- Der maschinelle Namensabgleich

5/18

Konzept der Zusammenführung

Verfahrensschritte des Datenabgleichs

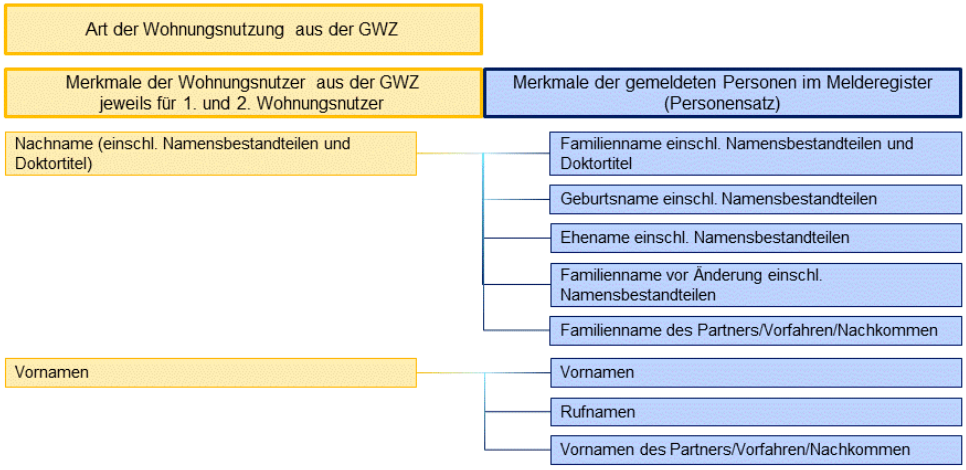
- **Bestimmung zu identifizierender Merkmale**
- **Vergleich der Datensatzpaare**
- **Klassifikation der Datensatzpaare in identisch/nicht identisch**

  Komponenten einer festen Bewertungsregel für die Merkmale m_j

  - **Vergleichsfunktionen $f_j(m_j)$**
  - **Bewertungsfunktion $\lambda(f_j(m_j))$**
  - **Entscheidungsfunktion $\delta(\lambda)$**

Konzept der Zusammenführung

Bestimmung zu identifizierender Merkmale



Konzept der Zusammenführung

Vergleich der Datensatzpaare: vorbereitende Maßnahmen

- **Ersetzen von Sonderzeichen durch Blanks**
  Ausnahme: Wildcard-Symbol der Beleglesung ‚*'
- **Ersetzen von überzähligen Blanks**
  Beispiel: doppelte Blanks, Blanks am Anfang oder Ende von Namen
- **Umsetzen von Umlauten**
  Beispiel: ä → ae
- **Umsetzen von Zeichen/Zeichenkombinationen**
  Beispiel: ß → ss, ph → f, th → t
- **Entfernen von Akzenten**
  Beispiel: é → e, ç → c
- **Umsetzen aller Klein- in Großbuchstaben**

8/18

Der maschinelle Namensabgleich

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

Konzept der Zusammenführung

Vergleich der Datensatzpaare: Problemstellung

- **Problem: Namen häufig nicht vollständig identisch**
  - **Fehlerhafte Angaben in der GWZ**
  - **Fehler bei der Datenerfassung (Beleglesung)**
    Falsch bzw. nicht erkannte Zeichen
  - **Fehler in den Melderegistern**
- **Ziel: Algorithmus, der auch „Ähnlichkeiten" von Namen aufdeckt**
- **Anforderungen an den Algorithmus**
  - **Maximierung der Trefferquote**
    (= Minimierung des manuellen Aufwands)
  - **Minimierung von Falschzuordnungen**

9/18

Konzept der Zusammenführung

Vergleich der Datensatzpaare: Das Ähnlichkeitsmodul

- Vergleich innerhalb eines Suchraums

  GWZ: BIERMANN

  MR: BIRMANN

  - Suchraum begrenzt
- Bewertung des Abgleichs

$$\text{Bewertungsquotient} = \frac{\text{Summe der Laengen der gemeinsamen Teilstrings}}{\text{Laenge des laengeren Namens}}$$

$$\text{Bewertungsquotient} = \frac{2+5}{8} = 0{,}875$$

Der maschinelle Namensabgleich

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

Konzept der Zusammenführung

Vergleich der Datensatzpaare: Das Ähnlichkeitsmodul

- Beispiele:

Gesuchter String	Vergleichsstring	Länge der gemein-samen Teilstrings	Länge des längeren Namens	Bewertungs-quotient
Franziska	Francisca	4+2	9	0,67
Marie	Maria	4	5	0,8
Jeanette	Jeannette	4+4	9	0,89
Gretchen	Grete	4	8	0,5
Erwin	Ervin	2+2	5	0,8
Drakomena	Draomina	3+2+2	9	0,78
Rieki	Rilki	2+2	5	0,8
Atanassioni	Atanasiou	6+2	11	0,73
Bandisch	Baudisch	2+5	8	0,88
Lieskovsky	Lieszkovszky	4+4+2	12	0,83

Konzept der Zusammenführung

Klassifikation der Datensatzpaare

- **Vergleich zweier Datenbestände:**

  paarweiser Vergleich zwischen jedem Datensatz des einen Datenbestandes mit jedem Datensatz des anderen Datenbestandes

  → wenig praktikabel

- **Prinzip der sukzessiven Massenreduktion:**
  - **ersten Satz des einen Datenbestandes mit jedem Datensatz des anderen Datenbestandes vergleichen**
  - **Klassifikation in identisch/nicht identisch**
  - **Fortfahren mit nächstem Satz des einen Datenbestandes mit den verbleibenden Sätzen des anderen Datenbestandes**

Der maschinelle Namensabgleich

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

Konzept der Zusammenführung

Klassifikation der Datensatzpaare

- **Methodisches Problem des Prinzips der sukzessiven Massenreduktion**

GWZ		
Lfd. Nr.	Familienname	Vorname
1	Müller	Petra
2	Müller	Peter

Melderegister		
Lfd. Nr.	Familienname	Vorname
1	Miller	Hans-Peter
2	Müller	Peter

- **Lösung: Verwendung mehrerer hierarchisch abgestufter Bewertungsregeln**

Konzept der Zusammenführung

Klassifikation der Datensatzpaare: Hierarchisches Stufenmodell

- Zunächst Vergleich zweier Datenbestände mit sehr restriktiver Bewertungsregel nach dem Prinzip der sukzessiven Massenreduktion
- Im weiteren werden die Bewertungsregeln zunehmend „weicher“
- Fortfahren mit nächstem Satz des einen Datenbestandes mit den verbleibenden Sätzen des anderen Datenbestandes
  - Bei voller Übereinstimmung zweier Namen Bewertung von 1
  - Bei Namensähnlichkeiten Bewertung mittels Ähnlichkeitsmodul, Ähnlichkeitsmaß im Intervall [0, 1]

  Schwellwert für Klassifikation in identisch situationsbedingt (einfache/doppelte Ähnlichkeit, Ähnlichkeitsbewertung von Vornamen/Nachnamen)

14/18

Der maschinelle Namensabgleich

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

Konzept der Zusammenführung

Klassifikation der Datensatzpaare: Hierarchisches Stufenmodell

- Gesamtablauf gliedert sich in drei Stufen
  - 1. Stufe: Namensabgleich im engeren Sinne (46 Unterstufen)

    Abgleich diverser Kombinationsmöglichkeiten der Wohnungsnutzer-namen der GWZ mit den persönlichen Namensangaben der Melderegisterpersonen der selben Anschrift
  - 2. Stufe: Namensabgleich über Verzeigerungen (11 Unterstufen)

    Abgleich diverser Kombinationsmöglichkeiten der Wohnungsnutzer-namen der GWZ mit den Namensangaben zum Partnern/Vorfahren/ Nachkommen der Melderegisterpersonen der selben Anschrift
  - 3. Stufe: Unechter Namensabgleich (4 Unterstufen)

    Abgleich des Nachnamens eines Wohnungsnutzers der GWZ mit den Familiennamen der Melderegisterpersonen der selben Anschrift
    - Einschränkungen: GWZ-Vorname leer & GWZ-Nachname für keine weitere Wohnung angegeben

15/18

Konzept der Zusammenführung

Klassifikation der Datensatzpaare: Hierarchisches Stufenmodell

- **Unterstufen der Stufen 1 und 2 zum Teil verstärkt durch Synonym-Suche**
  - **Ausschließlich Unterstufen ohne Ähnlichkeits- oder Teilvornamen-Suche**
  - **betrifft 11 Unterstufen der Stufe 1 → 57 ‚Unterstufen'**
  - **betrifft 2 Unterstufen der Stufe 2 → 13 ‚Unterstufen'**
- **Unterstufen der Stufe 1 zum Teil verstärkt durch Umlaut-Suche**
  - **Keine Kombination mit Synonym-Suche**
  - **betrifft alle 46 Unterstufen der Stufe 1 → 103 ‚Unterstufen'**

16/18

Der maschinelle Namensabgleich

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

Konzept der Zusammenführung - Testergebnisse

- **Zensustest**
  - **Auswertung der maschinellen Verknüpfung von 1683 paarigen Fällen**
  - **Richtigzuordnungen bei 99,3 % (Rest keine Zuordnung)**
- **Simulation eines Beleglese-Szenarios**
  - **Beleglesung von rund 18900 handschriftliche Wohnungsnutzerangaben**
  - **Betrachtung auch nichtpaariger Fälle**
  - **Richtigzuordnungen bei 88 %, Falschzuordnungen bei 0,0019%**
- **Testrechnungen mit unplausibilisierten Rohdaten**
  - **rund 49,5 Mio. Wohnungsnutzerangaben bei etwa 40,5 Mio. Wohnungsdatensätzen**
  - **Erhebungsformen: Beleg, IDEV, Core-Meldungen**
  - **Trefferquote bei 81,6 %**

17/18

Vielen Dank
für Ihre Aufmerksamkeit!

Marco Reisch
Tel.: 089 / 2119 3649
E-Mail: Marco.Reisch@lfstad.bayern.de

zensus 2011

18/18

Abschließendes Beispiel: Doppelte Ähnlichkeit

Hinreichende Ähnlichkeit der Familien- und Vornamen (Stufe 1)

GWZ		
Lfd. Nr.	Familienname	Vorname
1	ZIMMERMANN	JOHANNES

Melderegister		
Lfd. Nr.	Familienname	Vorname
1	ZIMMERER	HANS
2	SEMMERMANN	JOHANN
3	ZIMMER	HANNES

- **Schwellwerte für Abgleich mit doppelter Ähnlichkeit**
  - **Schwellwert Familiennamenbewertung: 0,6**
  - **Schwellwert Vornamenbewertung: 0,5**
  - **Schwellwert Gesamtbewertung: 0,6**

Abschließendes Beispiel: Doppelte Ähnlichkeit

Hinreichende Ähnlichkeit der Familien- und Vornamen (Stufe 1)

- **Vergleichsfunktionen:**
  - Eine mögliche Zusammenführung soll nur bei Datensatzpaaren erfolgen, bei denen die Übereinstimmung beider Merkmale mindestens die geforderten Schwellwerte beträgt.
  - Vergleichsfunktion Familiennamenbewertung

$$f_1(F) = \begin{cases} p(F), & \text{falls } p(F) \geq 0{,}6 \\ 0, & \text{sonst} \end{cases}$$

  - Vergleichsfunktion Vornamenbewertung

$$f_2(V) = \begin{cases} p(V), & \text{falls } p(V) \geq 0{,}5 \\ 0, & \text{sonst} \end{cases}$$

Der maschinelle Namensabgleich

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

Abschließendes Beispiel: Doppelte Ähnlichkeit

Hinreichende Ähnlichkeit der Familien- und Vornamen (Stufe 1)

- **Bewertungsfunktion:**
  - Datensatzpaare erhalten eine Gesamtbewertung > 0, wenn das arithmetische Mittel der einzelnen Namensähnlichkeiten mindestens dem geforderten Schwellwert für die Gesamtbewertung entspricht.

$$\lambda(f_1(F), f_2(V)) = \begin{cases} 0{,}5(f_1(F) + f_2(V)), & \text{falls } 0{,}5(f_1(F) + f_2(V)) \geq 0{,}6 \\ & \text{und } f_1(F), f_2(V) > 0 \\ 0, & \text{sonst} \end{cases}$$

Abschließendes Beispiel: Doppelte Ähnlichkeit

Hinreichende Ähnlichkeit der Familien- und Vornamen (Stufe 1)

- **Entscheidungsfunktion:**
  - **Das Datensatzpaar mit der höchsten Gesamtbewertung der Übereinstimmung (> 0) wird als identisch klassifiziert.**

$$\delta(\lambda) = \begin{cases} \text{identisch}, & \text{falls } \lambda > 0 \text{ und } \lambda = \max_i \lambda_i \\ \text{Nicht identisch}, & \text{sonst} \end{cases}$$

Der maschinelle Namensabgleich

Bayerisches Landesamt für Statistik und Datenverarbeitung | Mit Bayern rechnen.

Abschließendes Beispiel: Doppelte Ähnlichkeit

Hinreichende Ähnlichkeit der Familien- und Vornamen (Stufe 1)

- **Ähnlichkeitmodul liefert folgende Ähnlichkeitsmessungen:**

Paar	GWZ		Melderegister		Ähnlichkeit	
	Familienname	Vorname	Familienname	Vorname	p(F)	p(V)
(a, b_1)	ZIMMERMANN	JOHANNES	ZIMMERER	HANS	0,6	0,375
(a, b_2)	ZIMMERMANN	JOHANNES	SEMMERMANN	JOHANN	0,8	0,75
(a, b_3)	ZIMMERMANN	JOHANNES	ZIMMER	HANNES	0,6	0,75

- **Vergleich, Bewertung und Entscheidung:**

Paar	Vergleich		Bewertung	Entscheidung
	f1(F)	f1(V)	λ(f1 (F), f2 (V))	δ(λ)
(a, b_1)	0,6	0	0	nicht identisch
(a, b_2)	0,8	0,75	0,775	identisch
(a, b_3)	0,6	0,75	0,675	nicht identisch

→ **Johannes Zimmermann wird mit Johann Semmermann zusammengeführt.**

Vortragsblock II: Stichprobenverfahren

PD Dr. Siegfried Gabler:
„Das Stichprobendesign des Zensus 2011“

Abstract:

2011 wurde in Deutschland ein registergestützter Zensus durchgeführt. Das neue Zensusmodell beinhaltet u.a. eine statistische Korrektur von Melderegisterdaten um Karteileichen und Fehlbestände sowie eine Erhebung zusätzlicher, nicht aus Registern verfügbarer Merkmale. Beides wird durch eine Haushaltebefragung auf Stichprobenbasis realisiert. Damit wurde ein Paradigmenwechsel in der deutschen amtlichen Statistik herbeigeführt, der die Erforschung neuer statistischer Methoden erforderlich machte. Das Statistische Bundesamt vergab einen Forschungsauftrag an das Forscherteam um Ralf Münnich (Universität Trier) und Siegfried Gabler (GESIS Mannheim), Stichprobendesign und Schätzmethodik für die Haushaltsstichprobe zu untersuchen – das Zensus-Stichprobenforschungsprojekt. Aufbauend auf Ergebnissen des Zensustests wurde in diesem Projekt ein Stichprobendesign entwickelt, das unter gewissen Nebenbedingungen optimal ist und im Folgenden dargestellt wird. Die vollständigen Ergebnisse und Empfehlungen des Stichprobenforschungsprojekts zum deutschen Zensus 2011 werden im Band 21 der Reihe Statistik und Wissenschaft des Statistischen Bundesamtes abgehandelt.

Zur Person:

PD Dr. Siegfried Gabler ist wissenschaftlicher Mitarbeiter der Abteilung „Survey Design und Methodology“ der gesis, Leibniz-Institut für Sozialwissenschaften, in Mannheim. Er ist dort zuständig für die Entwicklung von Telefonstichproben für Deutschland, Stichproben- und Schätzverfahren, insbesondere deren entscheidungstheoretische Begründung, sowie Gewichtung und Designeffekte. Des Weiteren ist er Teil des Forscherteams, das im Auftrag des Statistischen Bundesamts das Stichprobendesign der Zensus-Haushaltsstichprobe entwickelt hat. Als Privatdozent lehrt er an der Fakultät für Rechtswissenschaft und Volkswirtschaftslehre der Universität Mannheim. Er ist Herausgeber und (Co-)Autor mehrerer Bücher und vieler Zeitschriftenaufsätze in internationalen Journals. Unter anderem ist er Mitglied im Sampling Expert Panel des European Social Survey.

Vortragsfolien:

gesis Leibniz-Institut für Sozialwissenschaften

Das Stichprobendesign des Zensus 2011

Statistiktage Bamberg, 2012

Siegfried Gabler

GESIS - Leibniz-Institut für Sozialwissenschaften

Bamberg, 26. Juli 2012

gesis Leibniz-Institut für Sozialwissenschaften

Team

Zensus 2011: Projekt Team
v.l.n.r: Jan-Philipp Kolb, Jan Pablo Burgard, Ralf Münnich, Siegfried Gabler, Matthias Ganninger

gesis Leibniz-Institut für Sozialwissenschaften

Definition

Der Zensus 2011 ist in Deutschland eine registergestützte, durch eine Stichprobe und eine Vollerhebung in Gemeinschaftsunterkünften ergänzte Bevölkerungszählung, die - mit einer Gebäude- und Wohnungszählung kombiniert - zum Stichtag 9. Mai 2011 stattfindet.

Quelle: Zensus_2011_Methodentext_16.pdf

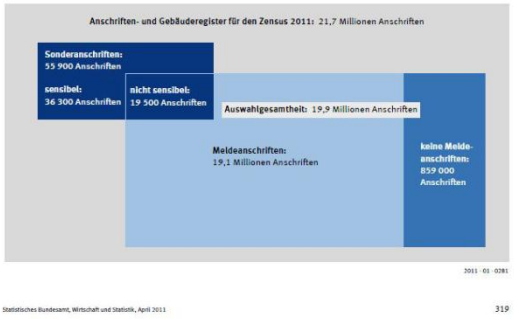
Paradigmenwechsel

Von der klassischen Volkszählung ... zum registergestützten Zensus

Es wird eine ergänzende Stichprobe erhoben.Diese dient zum einen einer Abschätzung der Zahl der Karteileichen (KAL) und Fehlbestände (FEB) in den Melderegistern und mithin der Ermittlung der Amtlichen Einwohnerzahl**[Ziel 1]**. Zum anderen sollen in den Registern nicht erfasste Informationen erhoben werden **[Ziel 2]**. Zur Erarbeitung eines geeigneten Stichprobendesigns hat das Bundesministerium des Inneren in Zusammenarbeit mit dem Statistischen Bundesamt (DESTATIS) ein Forschungsprojekt in Auftrag gegeben. Das Forschungsprojekt sollte vor allem Antworten auf die Frage geben, welches Stichprobendesign unter den gegebenen Restriktionen empfohlen werden kann sowie die Frage beantwortet werden, welche Schätzstrategien verfolgt werden sollen.

Anschriften- und Gebäuderegister als Auswahlgrundlage

Im Zensus 2011 werden komplette Anschriften aus dem Anschriften- und Gebäuderegister (AGR) gezogen.



Statistisches Bundesamt, Wirtschaft und Statistik, April 2011 319

Begründung für Schichtung

In einer ausgewählten Anschrift wird die Anzahl an Personen mit einer bestimmten interessierenden Eigenschaft ermittelt. Bei dieser Eigenschaft handelt es sich im Fall von Ziel 1 um die Existenz einer Person in der Anschrift (Ziel 1 Variable) und bei Ziel 2 um die konkrete Ausprägung einer interessierenden Variablen (Ziel 2 Variable).
Im AGR ist die Anschriftengrösse (ADG), also die Anzahl der an der Anschrift registrierten Personen, enthalten. Es liegt nahe, diese Information durch eine Schichtung schon in das Auswahlverfahren einfliessen zu lassen, da für etliche interessierende Merkmale, insbesondere Anzahl der Karteileichen und Fehlbestände, ein Zusammenhang mit der Anschriftengrösse vermutet wird. Hierzu wird die Variable Anschriftengrösse in Klassen eingeteilt und die so entstandene Anschriftengrössenklasse (ADK) als Schichtungsvariable verwendet.

Schichtungsvarianten

Bezeichner	Beschreibung
ADK1	6 Schichten mit Schichtgrenzen:
	1, 2, 3, 4-6, 7-10, 11+ registrierte Personen in der Anschrift
ADK2	4 mit registrierten Personen gleich stark besetzte Schichten pro SMP
ADK3	8 mit registrierten Personen gleich stark besetzte Schichten pro SMP

Tabelle: Schichtungsvarianten

Vorgaben des Auftraggebers (DESTATIS)

Neben der Festlegung der Schichten ist eine wesentliche Aufgabe des Forschungsprojekts, eine geeignete Aufteilung des Stichprobenumfangs auf die Schichten zu finden. Wünschenswert ist die Einhaltung von Präzisionsanforderungen an die Schätzungen sowie das Nicht-Überschreiten eines vorab festgelegten Gesamtstichprobenumfangs bezogen auf in Deutschland registrierte Personen.

Stichprobenbasiseinheiten (SMP)

Hierarchische Struktur von Zusammenfassungen regionaler Einheiten, dass die einzelnen Ebenen der Präzisionsanforderungen widerspruchsfrei und eindeutig berücksichtigt werden. Diese Struktur dient als Basis, um eine regionale Aufteilung des Gesamtstichprobenumfangs zu ermöglichen. Eine anschliessende Schichtung zur Erhöhung der Präzision der Schätzungen bleibt davon unberührt.
Eine Stichprobenbasiseinheit ist als regionale Einheit definiert, aus der eine Stichprobe gezogen wird, wobei der Stichprobenumfang noch festzulegen ist.
Die SMPs werden in vier Typen eingeteilt.

Einteilung in Sampling Points

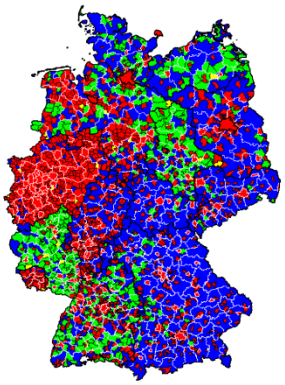
Typ 0 (SDT): Stadtteile ab 200.000 Einwohnern aus Städten mit mindestens 400.000 Einwohnern;

Typ 1 (GEM): Gemeinden und Städte mit mindestens 10.000 Einwohnern, sofern sie nicht zum Typ 0 gehören;

Typ 2 (VBG): Kleine Gemeinden (unter 10.000 Einwohnern) innerhalb eines Gemeindeverbands beziehungsweise einer Verbandsgemeinde werden zusammengefasst, sofern sie in der Summe mindestens 10.000 Einwohner betragen;

Typ 3 (KRS): Zusammenfassung aller Gemeinden eines Kreises, die bis dahin noch keinem Typ zugeordnet wurden.

- Berücksichtigung der Präzisionsanforderungen zu Ziel 1 und 2



Die Einfärbung wurde mit Hilfe der Klassifikation der SMPs gemacht (SDT: gelb; GEM: rot; VBG: grün; KRS: blau) diese sind durch feine schwarze Linien abgegrenzt. Die Grenzen der Stadtteile sind synthetisch erzeugt worden.

Man erkennt, dass zwischen den Bundesländern sehr unterschiedliche Verteilungen der Typen von Stichprobenbasiseinheiten vorliegen. Nordrhein-Westfalen, Rheinland-Pfalz sowie Bayern.

Auswahlsätze von Anschriften

SMP	SDT	GEM	VBG	KRS	Summe
Baden-Württemberg	2	244	126	35	407
Bayern	8	216	30	71	325
Berlin	12	0	0	0	12
Brandenburg	0	71	5	14	90
Bremen	3	1	0	0	4
Hamburg	7	0	0	0	7
Hessen	3	168	0	21	192
Mecklenburg-Vorpommern	0	24	30	12	66
Niedersachsen	2	205	68	34	309
Nordrhein-Westfalen	12	339	0	17	368
Rheinland-Pfalz	0	46	122	20	188
Saarland	0	40	0	5	45
Sachsen	4	69	13	22	108
Sachsen-Anhalt	0	60	27	11	98
Schleswig-Holstein	0	53	52	11	116
Thüringen	0	33	6	17	56
Deutschland	53	1569	479	290	2391

Letztendlich hat sich die tatsächlich verwendete Anzahl an SMPs noch um 26 reduziert. Die tatsächlich verwendeten Zahlen sind in (Berg und Bihler 2011, S. 321) zu finden.

Präzisionsanforderungen im Zensus 2011

Ziel 1: Schätzungen nur bei Gemeinden ab 10.000 EW

- Gemeinden ab 10.000 EW: $\mathrm{RRMSE}(\widehat{\tau}_{Z,<\mathrm{area}>}) \leq 0{,}5\%$
- Stadtteile von Grossstädten: $\mathrm{RRMSE}(\widehat{\tau}_{Z,<\mathrm{area}>}) \leq 0{,}5\%$

Ziel 2: Betrachtet wird bei $\frac{\tau_{Y,<\mathrm{area}>}}{\tau_{Z,<\mathrm{area}>}} \approx p$ mit $p \geq \frac{1}{15}$) :

- Gemeinden ab 10.000 EW: $\mathrm{RRMSE}(\widehat{\tau}_{Y,<\mathrm{area}>}) \leq \frac{1}{p}$
- Stadtteile von Grossstädten: $\mathrm{RRMSE}(\widehat{\tau}_{Y,<\mathrm{area}>}) \leq \frac{1}{p}$
- Kreise: $\mathrm{RRMSE}(\widehat{\tau}_{Y,<\mathrm{area}>}) \leq \frac{1}{p}$
- VBG in RLP: $\mathrm{RRMSE}(\widehat{\tau}_{Y,<\mathrm{area}>}) \leq \frac{1}{p}$

Ziel:	1	2	2	2	2	2	2
p (in %):	100	80	50	30	20	10	6,7
$\mathrm{RRMSE}_{\max}$ (in %):	0,5	1,25	2	3,33	5	10	15

Anforderungen Übersicht

1. Erfüllung aller gestellten Präzisionsanforderungen,
2. Verwendung eines möglichst geringen Auswahlsatzes,
3. Begrenzung der Gesamtkosten sowie
4. keine zu grosse Variation der Auswahlwahrscheinlichkeiten.

Beispiel

- Gewichte 1,2,3,4 sollen auf das Intervall [1,3] gestutzt werden, so dass die Summe der Gewichte identisch bleibt.
- 4 auf 3 stutzen und die restlichen Zahlen mit 7/6 multiplizieren.
- Ergebnis:
- 7/6, 7/3, 7/2, 3
- Dieses Vorgehen liefert also kein akzeptables Ergebnis.

Lösung durch Gabler/Ganninger/Münnich-Algorithmus

Ausgangspunkt: Optimale Aufteilung

- Es gelte $n_h = n\frac{N_h \cdot S_h}{\sum_{h=1}^{H} N_h \cdot S_h} = n\frac{d_h}{\sum_{h=1}^{H} d_h}$
- Es kann hierbei jedoch vorkommen, dass die n_h
  - grösser sind als M_h und/oder
  - kleiner sind als ein geforderter Mindestumfang m_h sowie
- $\sum_{h=1}^{H} n_h N p_h$ soll $Np\theta$ sein

Lösung durch Gabler/Ganninger/Münnich-Algorithmus

Zusammenfassend ist daher folgendes Minimierungsproblem zu lösen: Minimiere als Funktion von $n_{<g>h}$

$$\sum_{g=1}^{G}\sum_{h=1}^{H} N_{<g>h}^2 \cdot \frac{S_{<g>h,R}^2}{n_{<g>h}}$$

unter den Nebenbedingungen

$$0 < m_{<g>h} \leq n_{<g>h} \leq M_{<g>h}$$

$$\sum_{g=1}^{G}\sum_{h=1}^{H} n_{<g>h,A} \cdot \frac{\tau_{<g>h,R}}{N_{<g>h,A}} = \tau_R \cdot \theta \quad .$$

Ergebnis des optimalen Algorithmus

Das so gestellte Problem einer nicht-linearen Optimierung unter Nebenbedingungen kann durch einen einfachen Algorithmus gelöst werden. Dabei macht sich der Algorithmus die Tatsache zunutze, dass die Schichten exakt drei Klassen zugeordnet werden können. In Schichten, die der ersten Klasse U_1 angehören, wird der Stichprobenumfang exakt auf die untere Schranke m_h gesetzt, in Schichten der zweiten Klasse U_2 wird n_h exakt auf die obere Schranke M_h gesetzt und in der dritten Klasse U_3 wird der verbleibende Stichprobenumfang $n - \sum_{h\in U_1} m_h - \sum_{h\in U_2} M_h$ optimal im Sinne von Neyman-Tschuprov aufgeteilt. Das Problem besteht somit darin, diejenige Zusammensetzung der Klassen zu finden, für die insgesamt eine bestimmte Zielfunktion minimiert wird. Der Algorithmus löst dieses Problem dadurch, dass zunächst zwei geordnete Reihen gebildet werden, in denen die Schichten entsprechend ihrer Ausprägung auf $N_h \cdot S_{h,Y}$ in aufsteigender, beziehungsweise absteigender Reihenfolge angeordnet sind. Anschliessend werden die Kombinationen dieser Ordnungen Schritt für Schritt abgearbeitet. Die erste Lösung, bei der alle Elemente aus

Aktuelle Festlegungen der Box-Constraints

Für $p_h := \frac{m_h}{N_h} \leqslant \frac{n_h}{N_h} \leqslant \frac{M_h}{N_h} =: P_h$ gelten folgende Festlegungen:

	SMP-Typ									
	0		1		2 (RLP)		2 ($\overline{\text{RLP}}$)		3	
GemGK	p_h	P_h	p_h	P_h	p_h	P_h	p_h	P_h	p_h	P_h
I	—	—	—	—	—	—	—	—	0,05	0,05
II	—	—	0,05	0,50	0,05	0,50	0,05	0,05	0,05	0,05
III	—	—	0,04	0,40	0,04	0,40	0,05	0,05	0,05	0,05
IV	0,02	0,40	0,02	0,40	0,02	0,40	0,05	0,05	0,05	0,05

- Festlegung der GemGK-Grenzen
  I: 0 bis unter 10.000 EW / II: 10.000 bis unter 30.000 EW
  III: 30.000 bis unter 100.000 EW / IV: ab 100.000 EW
- Alle SMPs werden nach Anschriftengrössenklassen in 8 gleich grosse Schichten eingeteilt.
- max / min der Design-Gewichte: 25

Danke für die Aufmerksamkeit!

Dipl.-Stat. Josef Schäfer:

„Was wird beim Zensus hochgerechnet?“

Abstract:

Wesentlicher Bestandteil des Zensus 2011 bildet eine Haushaltsstichprobe mit zufällig ausgewählten Anschriften als Auswahlgrundlage. Ziele dieser Stichprobenerhebung sind zum einen die Feststellung von Über- und Untererfassungen der Melderegister zur Ermittlung der amtlichen Einwohnerzahl in Gemeinden mit mindestens 10.000 Einwohnern (Ziel 1) zum anderen die Erhebung von Merkmalen, die nicht in den für den Zensus 2011 genutzten Verwaltungsregistern enthalten sind, wie z. B. Angaben zur Migration, zur Religion, zur Bildung und zur Erwerbstätigkeit (Ziel 2).

Unterschiedliche Teilaufgaben bei der Erstellung der Zensusergebnisse erfordern differenzierte Hochrechnungen aus der Haushaltsstichprobe. Neben der Ermittlung der amtlichen Einwohnerzahl und der Auswertung weiterer Merkmale liefern diese Hochrechnungen wichtige Eckzahlen für Verfahren zur Korrektur von Melderegisterauszählungen und für die Haushaltegenerierung. Daneben werden auch die Ergebnisse der Wiederholungsbefragung als deskriptive Kontrolle zur Bewertung der Qualität der Zensusergebnisse nach § 17 ZensG 2011 hochgerechnet.

Die einzelnen Hochrechnungen erfolgen als verallgemeinerte Regressionsschätzung (GREG) mit Melderegisterangaben (Zahl der an der Anschrift gemeldeten Personen insgesamt sowie untergliedert nach demographischen Merkmalen) als potenziellen Bezugsmerkmalen auf Anschriftenebene.

Zur Person:

Josef Schäfer ist derzeit einer von zwei Projektleitern für den Zensus 2011 bei IT.NRW in Düsseldorf, dem für Nordrhein-Westfalen zuständigen Statistischen Landesamt. Neben unterschiedlichen anderen Aufgaben, zuletzt als Leiter des Referats für interfachliche Erhebungen und mathematisch-wissenschaftliche Aufgaben, hat er dort bereits an der Volkszählung 1987 (als wissenschaftlicher Mitarbeiter) und am Zensustest 2001 (in leitender Funktion) mitgewirkt. An der Universität Dortmund hat Herr Schäfer den Abschluss eines Diplom-Statistikers erworben.

Vortragsfolien:

zensus 2011

Information und Technik Nordrhein-Westfalen
Geschäftsbereich Statistik

Methoden und Potenziale des Zensus 2011
Was wird beim Zensus 2011 hochgerechnet?

Statistik-Tage Bamberg-Fürth 2012

zensus 2011

Information und Technik Nordrhein-Westfalen
Geschäftsbereich Statistik

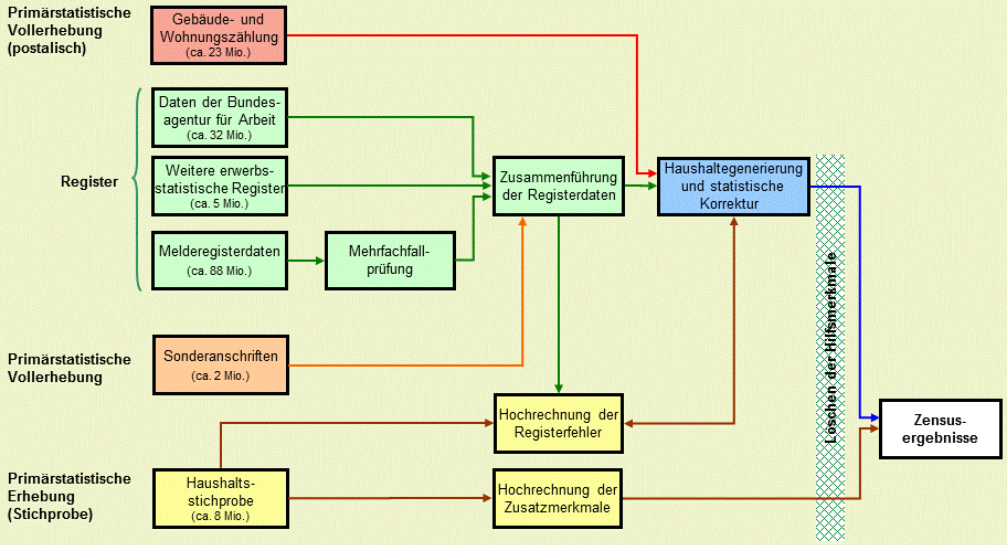
Gliederung

1. Modell des registergestützten Zensus 2011
2. Vorliegende Daten
3. Ergebniserstellung
4. Unterstützung der Haushaltegenerierung
5. Deskriptive Kontrolle
6. Verfahren

zensus 2011

Information und Technik Nordrhein-Westfalen
Geschäftsbereich Statistik

1. Modell des Registergestützten Zensus 2011



zensus 2011

Information und Technik Nordrhein-Westfalen
Geschäftsbereich Statistik

2. Vorliegende Daten (1)

Flächendeckend

- Melderegister zum 09.05.2011
  bereinigt um
  - kurzfristig nicht registrierte Zu- und Fortzüge
  - Mehrfachfälle (Hauptwohnungsdubletten)
  - ausschließliche Nebenwohnsitze
- Gebäude- und Wohnungszählung
- Erhebungen in Sonderbereichen

zensus 2011

Information und Technik Nordrhein-Westfalen
Geschäftsbereich Statistik

2. Vorliegende Daten (2)

Zusätzlich in Gemeinden unter 10.000 Einwohner

Für systematisch ausgewählte Anschriften
- Befragung zur Klärung von Unstimmigkeiten

Für zufällig ausgewählte Anschriften
- Haushaltsstichprobe (Auswahlsatz 5 %)

zensus 2011

Information und Technik Nordrhein-Westfalen
Geschäftsbereich Statistik

2. Vorliegende Daten (3)

Zusätzlich in Gemeinden ab 10.000 Einwohner

Für zufällig ausgewählte Anschriften
- Ergebnisse der Haushaltsstichprobe (Auswahlsatz zwischen 2 % und 50 % je nach Gemeindegröße und Anschriftengröße; vgl. Vortrag von Herrn Dr. Gabler)
- Wiederholungsbefragung bei 5 % der Anschriften der Haushaltsstichprobe

zensus 2011

Information und Technik Nordrhein-Westfalen
Geschäftsbereich Statistik

3. Ergebniserstellung (1)

Gemeinden unter 10.000 Einwohner

Amtliche Einwohnerzahl

- Auszählung der bereinigten Melderegister, korrigiert um die Ergebnisse der Befragung zur Klärung von Unstimmigkeiten

Weitere Merkmale

- Hochrechnung der Haushaltstichprobe zur Vervollständigung der Kreisergebnisse

zensus 2011

Information und Technik Nordrhein-Westfalen
Geschäftsbereich Statistik

3. Ergebniserstellung (2)

Gemeinden ab 10.000 Einwohner (1)

Amtliche Einwohnerzahl

- Korrektur der bereinigten Melderegister um hochgerechnete Über- und Untererfassungen, dazu erforderlich
  - Auszählung der Melderegister
  - Hochrechnung der Haushaltsstichprobe
  - Hochrechnung der „paarigen Fälle“

Über- und Untererfassungen ergeben sich aus Differenzbildungen.

zensus 2011

Information und Technik Nordrhein-Westfalen
Geschäftsbereich Statistik

3. Ergebniserstellung (3)

Gemeinden ab 10.000 Einwohner (2)

Die Ergebnisse weiterer Merkmale zum ersten Veröffentlichungstermin (vor der Haushaltegenerierung) basieren auf

- Melderegisterauszählungen
- hochgerechneten Randverteilungen von Über- und Untererfassungen der Melderegister nach demografischen Merkmalen
- „Korrektur“ von mehrdimensional gegliederten Tabellen durch „Herunterbrechen“ spezifischer Über- und Untererfassungen aus den Randverteilungen mit Hilfe von loglinearen Modellen. Die Umsetzung erfolgt durch eine Anpassung der Hochrechnungsfaktoren.

zensus 2011

Information und Technik Nordrhein-Westfalen
Geschäftsbereich Statistik

3. Ergebniserstellung (4)

Gemeinden ab 10.000 Einwohner (3)

Die Ergebnisse weiterer Merkmale zum zweiten Veröffentlichungstermin (nach der Haushaltegenerierung) basieren auf

- Melderegisterauszählungen
- hochgerechneten Randverteilungen von Über- und Untererfassungen der Melderegister differenziert nach
  - demografischen Merkmalen
  - Vorhandensein von Angaben in erwerbsstatistischen Registern
  - „Ranking“ aus der Haushaltegenerierung
  - Haushaltstypen
  - vollständig oder nur teilweise fehlerhaft registrierten Haushalten
- einem an die Randverteilungen angepassten einzelfallbezogenen Korrekturverfahren (vgl. Vortrag von Frau Hofmeister)

zensus 2011

Information und Technik Nordrhein-Westfalen
Geschäftsbereich Statistik

4. Unterstützung der Haushaltegenerierung

Schätzung des Anteils schwer erkennbarer Haushaltskonstellationen, z. B. nichteheliche Lebensgemeinschaften (auf Gemeinde- bzw. Kreisebene) zur Unterstützung der Haushaltegenerierung.

zensus 2011

Information und Technik Nordrhein-Westfalen
Geschäftsbereich Statistik

5. Deskriptive Kontrolle

Gemeinden unter 10.000 Einwohner

- Auswertung einer Unterstichprobe der Haushaltsstichprobe (0,3 % der Bevölkerung) zur Überprüfung des Verfahrens für „kleine“ Gemeinden
- Ebenen: Bundesland, grobe Gemeindegrößenklassen

Gemeinden ab 10.000 Einwohner

- Auswertung der Wiederholungsbefragung (erneute Befragung von 5 % der Haushaltsstichprobe)
- Ebene: Bundesland, grobe Gemeindegrößenklassen

6. Verfahren (1)

Verallgemeinerte Regressionsschätzung (GREG) (1)

Nutzung von Melderegisterauszählungen als Bezugsmerkmale

- Zahl der Personen insgesamt sowie gegliedert nach
  - Geschlecht & Staatsangehörigkeit
  - Familienstand
  - Altersklassen
  - Vorhandensein von Einträgen in Erwerbsregistern

6. Verfahren (2)

Verallgemeinerte Regressionsschätzung (GREG) (2)

Formel (1):

$$\hat{t}_{y,d,GREG} = \sum_{i \in s_d} w_i y_i + \sum_{j=1}^{J} \hat{\beta}_j \left(\sum_{i \in U_d} x_{ji} - \sum_{i \in s_d} w_i x_{ji} \right)$$

$$\hat{\beta} = \left(\sum_{i \in s} w_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \sum_{i \in s} w_i \mathbf{x}_i y_i$$

zensus 2011

Information und Technik Nordrhein-Westfalen
Geschäftsbereich Statistik

6. Verfahren (3)

Verallgemeinerte Regressionsschätzung (GREG) (3)

Formel (2):

y_i : Zielvariable y an der i-ten Anschrift

w_i : Designgewicht der i-ten Anschrift

x_{ji} : Bezugsvariable j an der i-ten Anschrift, Vektorschreibweise $x_i = \left(x_{1i}, \ldots, x_{Ji} \right)'$,

U_d : Menge der Anschriften der Zielgesamtheit in der Gemeinde (Domain) d

s_d : Menge der Stichprobenanschriften in der Gemeinde (Domain) d

$\hat{\beta}_j$: *j-te* Komponente des Vektors der geschätzten Regressionskoeffizienten $\hat{\beta}$

zensus 2011

Information und Technik Nordrhein-Westfalen
Geschäftsbereich Statistik

Vielen Dank für Ihre Aufmerksamkeit!

zensus 2011
Wissen, was morgen zählt

Josef Schäfer
Landesbetrieb Information und Technik
Nordrhein-Westfalen (IT.NRW)
– Geschäftsbereich Statistik –

Dipl.-Geogr. Katrin Hofmeister:

„Das Korrekturverfahren im Rahmen der Haushaltegenerierung beim Zensus 2011“

Abstract:

Beim Zensus 2011 wird in Gemeinden mit 10.000 oder mehr Einwohnern eine Haushaltsstichprobe durchgeführt. Zweck dieser Stichprobe ist neben der Erhebung von nicht in Registern verfügbaren Daten primär die gemeindeweise Gewinnung von demographischen und haushaltsstatistischen Informationen zu Über- und Untererfassungen (Karteileichen und Fehlbestände) in den Melderegistern. Mit diesen Informationen sollen die potenziellen Fehler einer unkontrollierten Registerauszählung vermieden werden. Um einen qualitativ hochwertigen, fachlich und regional flexibel auswertbaren Zensuseinzeldatensatz zu erhalten, muss eine Bereinigung der Karteileichen und Fehlbestände auf der Basis der Einzeldaten vorgenommen werden.

Zu diesem Zwecke war es erforderlich, ein Verfahren zu entwickeln, welches die gemeindeweise aggregierten Vorgaben aus der Haushaltsstichprobe möglichst genau umsetzt. Dabei ist jedoch zu berücksichtigen, dass eine solche Korrektur der Einzeldaten nur statistisch erfolgen kann, d.h. nicht die buchhalterisch betrachtete „Richtigkeit“ des Einzelfalls ist relevant und auch realisierbar, sondern die strukturelle Qualität der Zensusergebnisse. Der Vortrag soll einen Überblick über das im Rahmen der Haushaltegenerierung eingesetzte Korrekturverfahren sowie dessen Stärken aber auch Grenzen geben.

Zur Person:

Katrin Hofmeister ist Referentin und stellvertretende Leiterin im Sachgebiet „Zensus – Register, Haushaltegenerierung, Auswertung, Querschnittsaufgaben“ des Bayerischen Landesamts für Statistik und Datenverarbeitung und dort zuständig für das Zensus-Teilprojekt Haushaltegenerierung. Insbesondere war sie an der Entwicklung des Korrekturverfahrens zur Bereinigung von Karteileichen und Fehlbeständen in den Registerdaten maßgeblich beteiligt und wird dessen Umsetzung im Rahmen der Haushaltegenerierung für den gesamtdeutschen Datenbestand begleiten. Vor ihrer Tätigkeit im Projekt „Zensus 2011“ studierte sie Diplom-Geographie an der Julius-Maximilians-Universität Würzburg.

Vortragsfolien:

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

Das Korrekturverfahren (KV) im Rahmen der Haushaltegenerierung im Zensus 2011

Katrin Hofmeister
Bayerisches Landesamt für Statistik und Datenverarbeitung

Statistik-Tage Bamberg-Fürth 2012

zensus 2011

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

Gliederung

- **Grundlagen und Ziele**
- **Problemstellung**
- **Modell der HHGen incl. Korrekturverfahren**
- **Exkurs Ranking**
- **Modell des Korrekturverfahrens**
- **Bewertung Korrekturverfahren**

Grundlagen und Ziele

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

- Korrekturstichprobe in Gemeinden mit 10 000 oder mehr Einwohnern gemäß § 7 ZensG 2011
- Stichprobenergebnis legt die Höhe der Korrektur fest (Anzahl Karteileichen, Anzahl Fehlbestände - fachlich gegliedert) *
- Problem: Zensuseinzeldatensatz für flexible Auswertungen

→ Notwendigkeit eines statistischen Korrekturverfahrens zur Umsetzung der Summenergebnisse aus der Stichprobe im Personendatenbestand der HHGen

→ Das Korrekturverfahren verändert nicht die amtliche Einwohnerzahl!

Grundlagen und Ziele

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

Korrektur der Melderegisterdaten anhand der Ergebnisse der Stichprobe!

Stichprobe

Einzeldaten

Hochgerechnete Ergebnisse

Korrektur der Stichprobenanschriften

Statistische Korrektur der Nicht-Stichprobenanschriften

Die Korrektur bewirkt die Herstellung vollzähliger Zensuseinzeldatensätze, die flexible Auswertungen ermöglichen!

Problemstellung

Bayerisches Landesamt für Statistik und Datenverarbeitung | Mit Bayern rechnen.

- Karteileichen und Fehlbestände weisen eine signifikant andere demografische und haushaltsstatistische Struktur auf als die Grundgesamtheit der Bevölkerung → Rein zufällige Korrektur würde Verzerrung der demografischen und haushaltsstatistischen Struktur bewirken. *

  → Grundlage für die statistische Korrektur an **Nicht-Stichprobenanschriften** sind die demografischen und haushaltsstatistischen Anpassungsrahmen aus der Stichprobe pro Gemeinde

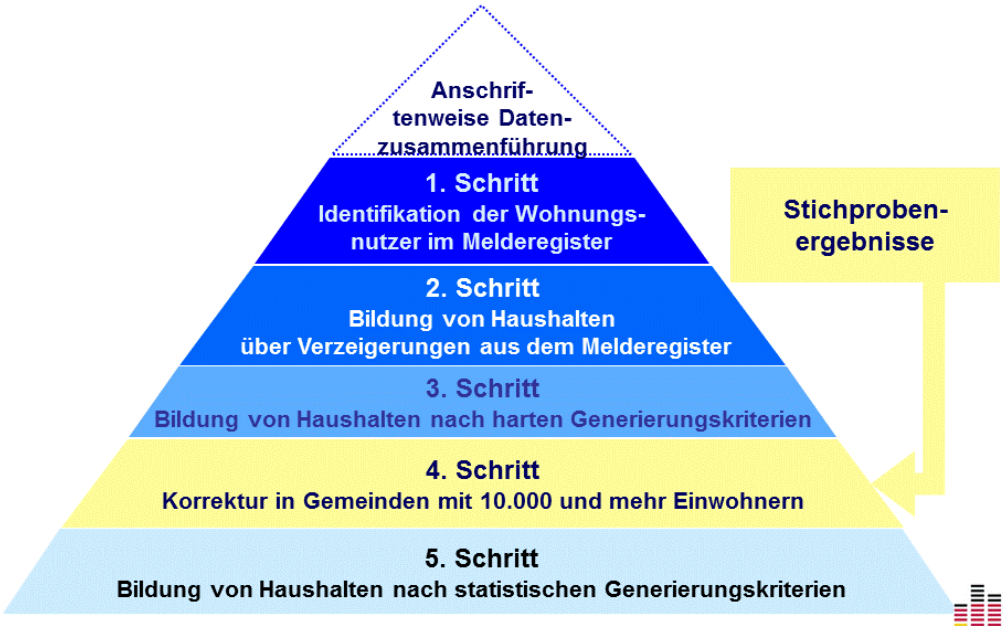
- Karteileichen und Fehlbestände weisen eine unterschiedliche demografische und haushaltsstatistische Struktur auf und treten nur selten in einem Haushaltszusammenhang auf

  → Notwendigkeit einer zweigleisigen Korrektur: Über- und Untererfassungen werden in einzelnen Korrekturmodulen bereinigt.

→ Notwendigkeit eines wesentlich komplexeren Verfahrens als das bloße Löschen und Hinzufügen von Datensätzen

Modell der HHGen incl. Korrekturverfahren

Bayerisches Landesamt für Statistik und Datenverarbeitung | Mit Bayern rechnen.



5/23

EXKURS RANKING

Bayerisches Landesamt für Statistik und Datenverarbeitung

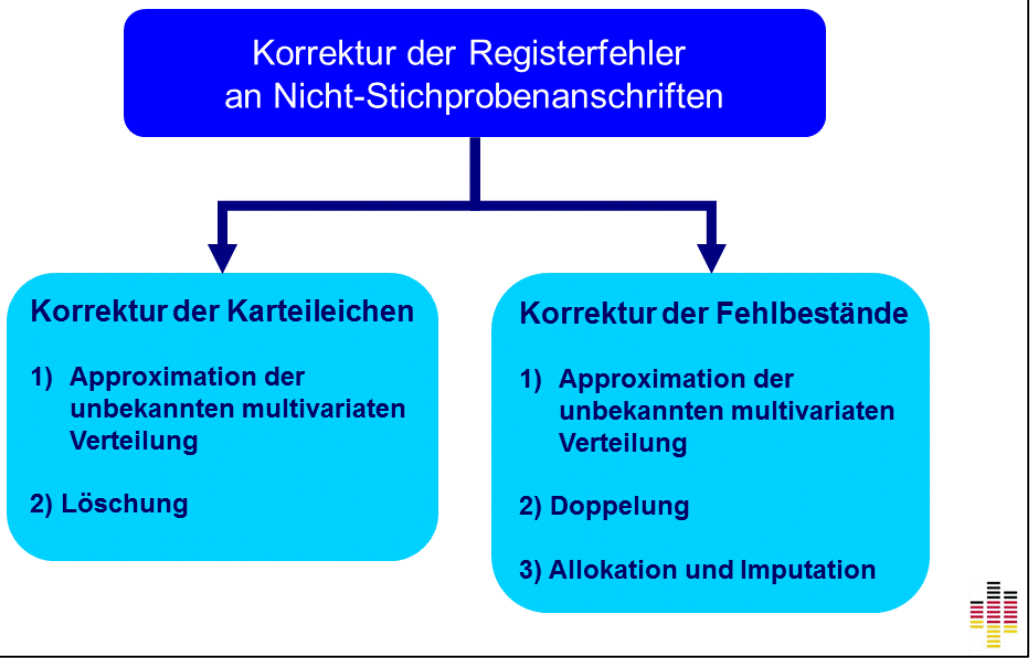
Mit Bayern rechnen.

- Ranking = Verknüpfungsstatus einer Person nach Schritt 3
- Vier Ausprägungen: Wohnungsnutzer, in Schritt 3 verknüpft, unverknüpfter Deutscher, unverknüpfter Ausländer
- Karteileichenrate der verknüpften Personen nach Schritt 3: 1 %
  Karteileichenrate der unverknüpften Personen nach Schritt 3: 17 %
  Karteileichenrate der unverknüpften Ausländer nach Schritt 3: 33 %
- Konsequenz: Verwendung des Rankings im Korrekturverfahren!

Modell Korrekturverfahren

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.



Karteileichen-bereinigung

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

Zweiteiliges Verfahren:

1) Approximation der Verteilung der zu löschenden Datensätze unter demografischen Randbedingungen (sowie Rankinginformation)
   - Alter
   - Geschlecht
   - Staatsangehörigkeit
   - Familienstand
   - Erwerbstätigkeit
2) Iteratives Verfahren zur Löschung der Karteileichen unter haushaltsstatistischen Randbedingungen
   - Drei- oder Mehrpersonenhaushalte
   - Zweipersonenhaushalte
   - Einpersonenhaushalte

Fehlbestands-imputation

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

Dreiteiliges Verfahren:

1) Approximation der Verteilung der zu importierenden Datensätze unter demografischen Randbedingungen
   - Alter
   - Geschlecht
   - Staatsangehörigkeit
   - Familienstand
2) Doppelung auf Grundlage der approximierten Verteilung unter haushaltsstatistischen Randbedingungen
   - Drei- oder Mehrpersonenhaushalte
   - Zweipersonenhaushalte
   - Einpersonenhaushalte
3) Allokation und Imputation der duplizierten Personen/ Haushalte

Problemstellung Approximation

Bayerisches Landesamt für Statistik und Datenverarbeitung | Mit Bayern rechnen.

- Multivariate demografische Verteilung der Karteileichen bzw. Fehlbestände nicht bekannt → unbekanntes theoretisches Ergebnis *
- Aufgabenstellung: Konzeption eines Verfahrens zur Annäherung ans Optimum → kein wirkliches Kriterium → Nicht nachprüfbar, wie nah Verfahren an Realität liegt
- Merkmalsausprägungen werden zufällig gezogen anhand des sog. Anpassungsfaktors: Gibt an, um wie viel häufiger (oder auch seltener) als in der Grundgesamtheit vorhanden eine bestimmte Merkmalsausprägung ausgewählt werden soll:
  - $a = \frac{kl / KL}{gg / GG}$
  - Alle Ausprägungen bilden zusammen die sog. Klasse

Konzept der Approximation

Bayerisches Landesamt für Statistik und Datenverarbeitung | Mit Bayern rechnen.

- Beispiel zur Berechnung der Anpassungsfaktoren anhand des Merkmals Familienstand

Familien-stand	Karteileichen	Häufigkeit der Ausprägung in den Karteileichen	Grundgesamtheit	Häufigkeit der Ausprägung in der Grundgesamtheit	Anpassungsfaktor a
	1	2	3	4	5
Ledig	600	600/1000 = 0,6	5 000	5000/10000=0,5	1,20
Verheiratet	300	300/ 1000 = 0,3	4 000	4000/10000=0,4	0,75
Verwitwet	10	10/1000=0,01	200	200/10000=0,02	0,50
Geschieden	90	90/1000=0,09	800	800/10000=0,08	1,13
Insgesamt	1 000		10 000		3,58

- Prüfung auf Zulässigkeit der Klasse
- Neuberechnung der Anpassungsfaktoren

Beispiel Karteileichenlöschung

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

- Beispiel: Löschung der Karteileichen
- Ausgangslage: Hochgerechnete Randverteilung aus der Stichprobe

Merkmal	Unkorrigiertes Melderegister	Karteileichen (aus der Stichprobe)	Korrigiertes Melderegister
Personen insgesamt	10 000	500	9 500
Männlich	5 000	300	4 700
Weiblich	5 000	200	4 800
< 25 Jahre	2 000	100	1 900
25 bis unter 50 Jahren	4 000	300	3 700
≥ 50 Jahre	4 000	100	3 900
Einpersonenhaushalte	5 000	250	4 750
Zwei- oder Mehrpersonenhaushalte	2 000	100	1 900

Beispiel Fehlbestands-imputation

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

- Beispiel: Imputation der Fehlbestände
- Ausgangslage: Hochgerechnete Randverteilung aus der Stichprobe

Merkmal	Unkorrigiertes Melderegister	Fehlbestände (aus der Stichprobe)	Korrigiertes Melderegister
Personen insgesamt	9 500	550	10 050
Männlich	4 700	350	5 050
Weiblich	4 800	200	4 000
< 25 Jahre	1900	50	1 950
25 bis unter 50 Jahren	3 700	400	4 100
≥ 50 Jahre	3 900	100	4 000
Einpersonenhaushalte	4 750	100	4 850
Zwei- oder Mehrpersonenhaushalte	1 900	50	1 950

Bewertung Korrekturverfahren

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

- Das Korrekturverfahren wurde anhand der Zensustestdaten überprüft
- Regionale Verzerrung
- Keine Abweichungen in der Gesamtanzahl der Personen
- Abweichungen in den von der Stichprobe vorgegebenen demografischen Strukturen: Sehr selten
- Abweichungen in den von der Stichprobe vorgegebenen Haushaltsstrukturen: Häufig

Bayerisches Landesamt für Statistik und Datenverarbeitung

Mit Bayern rechnen.

Vielen Dank
für Ihre Aufmerksamkeit!

Bei Fragen:
Katrin Hofmeister
Bayerisches Landesamt für Statistik und Datenverarbeitung

E-Mail: Katrin.Hofmeister@lfstad.bayern.de
Tel: 089/ 2119 - 3646

zensus 2011

Vortragsblock III: Datenzugang

Dr. Jörg Höhne:
„Statistische Geheimhaltung des Zensus 2011“

Abstract:

Der vertrauliche Umgang mit den erhobenen Einzelangaben ist eine grundlegende Voraussetzung für die Akzeptanz des Zensus in der Bevölkerung. Bei der Bereitstellung von Zensusergebnissen kommen deshalb statistische Geheimhaltungsverfahren zum Einsatz, die einen Rückschluss auf die Angaben einzelner Personen verhindern. Mit dem Zensus 2011 findet ein methodischer Wechsel von bisher bei Volkszählungen verwendeten Zellsperrverfahren zu einem pre-tabularen Geheimhaltungsverfahren statt. Bei pre-tabularen Verfahren werden die Mikrodaten anonymisiert und aus dem anonymisierten Mikrodatenbestand alle Auswertungstabellen erzeugt. Der Vortrag erläutert die Entscheidung der amtlichen Statistik zum Wechsel vom Zellsperrverfahren zur pre-tabularen Geheimhaltung. Für die Ergebnisse aus Registern wird das Verfahren SAFE, eine Variante der Mikroaggregation, verwendet. Im Vortrag wird das Verfahren SAFE beschrieben und einige Ergebnisse von Tests mit historischen Zensusdaten werden präsentiert. Diese ermöglichen erste Aussagen zur erwarteten Qualität der Ergebnisse des Zensus 2011 nach der Anonymisierung.

Zur Person:

Dr. Jörg Höhne ist Referatsleiter im Amt für Statistik Berlin-Brandenburg. Er arbeitet seit 2002 an verschiedenen Forschungsprojekten zur Anonymisierung von Einzeldaten mit. Diese werden durchgeführt, um Einzeldaten im Rahmen der Forschungsdatenzentren der Statistischen Ämter bereitzustellen. Er studierte Statistik und Wirtschaftsmathematik in Berlin und Moskau und promovierte 2009 an der Universität Tübingen mit einer Arbeit über „Verfahren zur Anonymisierung von Einzeldaten“.

Vortragsfolien:

Amt für Statistik Berlin-Brandenburg

Methoden und Potenziale des Zensus 2011

Statistische Geheimhaltung

Statistik-Tage Bamberg-Fürth 2012

26./27.7.2012

Dr. Jörg Höhne IT- Internet / IT - Wahlen

2

Gliederung

- Wozu statistische Geheimhaltung?
- Bisherige Vorgehensweise
- Verschiedene Datenquellen / verschiedene Risiken?
- Verfahren für Registerdaten
- Verfahren für Stichprobendaten

Wozu statistische Geheimhaltung?

Statistische Geheimhaltung sichert die Vertraulichkeit der Informationen über die einzelnen statistische Einheiten (Personen, Unternehmen u.a.).

Ziel ist nicht das Verschweigen von unliebsamen Informationen über die Gesellschaft sondern der Schutz des Einzelnen vor Schäden durch Datenmissbrauch.

Statistische Geheimhaltung wird außerdem durchgeführt weil:

- vertraulicher Umgang mit den Informationen über den Einzelnen ist die grundlegende Voraussetzung für:
  - die Bereitschaft zur Herausgabe wahrheitsgemäßer Informationen
  - für die Sicherung einer hohen Qualität der statistischen Daten
- Bundesstatistikgesetz, Bundesdatenschutzgesetz, ...

Bisherige Vorgehensweise

Statistische Geheimhaltung wurde früher durch Verfahren der Informationsreduktion gesichert.

Kritische Informationen wurden nicht veröffentlicht.

- „Zellsperrungen“ – Tabellenfelder aus denen Informationen über den Einzelnen gewonnen werden könnten werden gesperrt „/“. Ggf. werden zusätzliche Tabellenfelder mit entfernt um Rückrechnungen zu verhindern.
- Zusammenfassung von Informationen
- Sperrung der Tabellen insgesamt

Je größer der Umfang an Tabellen umso schwieriger ist es eine Rückrechenbarkeit sicher zu verhindern.

Verschiedene Quellen – verschiedene Risiken?

Im Zensus 2011 werden Informationen aus Registern / Vollerhebungen und Stichproben genutzt.

- Register / Vollerhebungen
  - Umfang der enthaltenen Informationen ist begrenzt
  - Informationen sind aber für alle vorhanden (Teilnahmekenntnis)

  Anonymisierung auf Ebene der Mikrodaten (Verfahren SAFE)

- Stichprobeninformationen
  - Umfang der erhobenen Informationen ist höher
  - (teilweise) fehlende Teilnahmekenntnis
  - höhere Qualität, da aktuelle Selbstauskunft

  Anonymisierung bei der Auswertung, da Hochrechnungen erforderlich

Grundidee der Mikrodatenanonymisierung

Anonyme Mikrodaten:

- Mikrodaten sind dann anonym (ausreichend geschützt), wenn „sie dem Befragten oder Betroffenen nicht zuzuordnen sind.“ (siehe BStatG §16 (1) Punkt 4)

- Die Verhinderung einer eindeutigen Zuordnung kann durch die Veränderung der originalen Merkmale oder durch Mehrdeutigkeit im Datenbestand erreicht werden.

- Die Schutzwirkung bei SAFE-Mikrodaten basiert primär auf der Mehrdeutigkeit im Datenbestand.

7

Vor- und Nachteile der Mikrodatenanonymisierung

Vorteile:

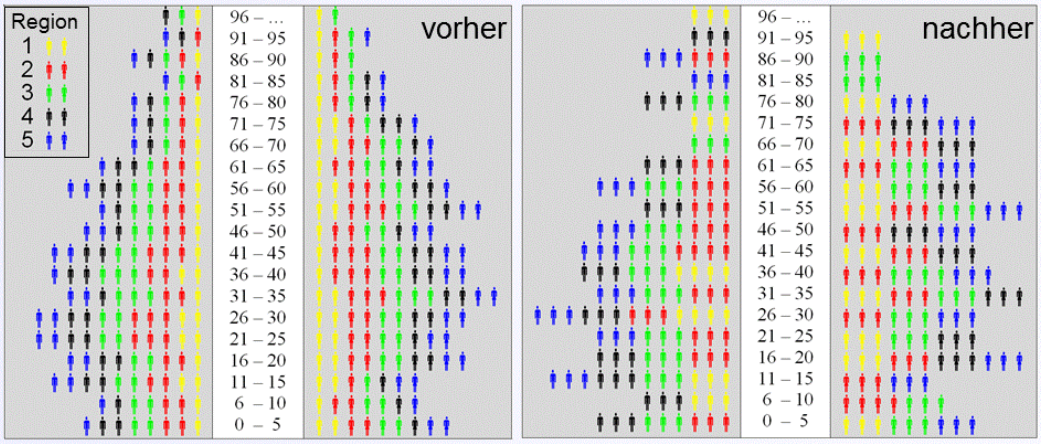
- Die Lösung der Geheimhaltung ist eine einmalige Aufgabe.
- Alle Auswertungen erfolgen aus den anonymen Daten und sind untereinander immer konsistent.
- Eine Sperrung von Feldern (Primär- und Sekundärsperrungen) ist nicht nötig.

Nachteile:

- Die Interpretation von mit anonymisierten Mikrodaten erstellten Tabellen muss durch den Nutzer neu "gelernt" werden.
- Die Sperrungen in Auswertungen werden durch Unsicherheiten ersetzt.
- Das nachträgliche Erweitern der möglichen Auswertungen um vorher nicht betrachtete Merkmale ist nur schwer möglich. (stochastische Effekte)
- Der einmalige Rechenaufwand kann relativ hoch sein.

8

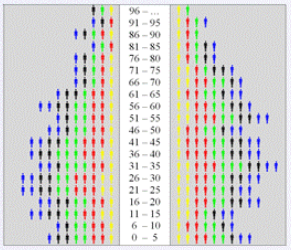
Grundidee des Verfahrens SAFE (1)



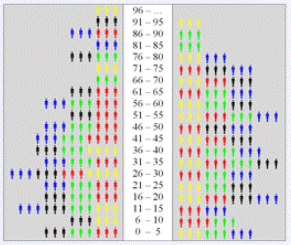
- Innerhalb des anonymisierten Datenbestandes sind alle statistischen Objekte mindestens dreimal mit identischen Merkmalsausprägungen vorhanden.
- Es gibt keine Geheimhaltungsprobleme mehr, da die Merkmalsausprägungen nicht mehr eindeutig einzelnen Objekten zugeordnet werden können.

9

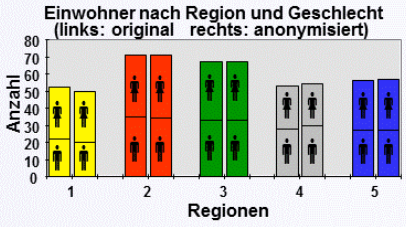
Grundidee des Verfahrens SAFE (2)

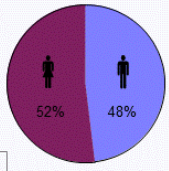


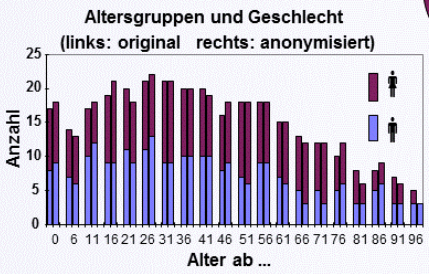
vorher



nachher







- Die so erhaltene Basisdatei ist einerseits als Gesamtinformation teilweise stark verfälscht.
- Bei den Auswertungen wird jedoch eine höchstmögliche Ähnlichkeit zur originalen Basisdatei angestrebt. Das gilt für alle potentiell sinnvoll möglichen Auswertungen des Datenbestandes.

10

Mathematisches Modell von SAFE (vereinfacht) (1)

$$ZF: \quad \min\left(\max_i\left(|e_i| - g_i\right)\right)$$

$$Ax = t + e$$

$$\text{mit}: \quad x_j = 0,3,4,\ldots$$

mit:

x - Vektor der Häufigkeiten möglicher Sätze im Datenbestand
- einmalige Originalsätze ($x_j^o=1$)
- völlig identische Originalsätze zusammengefasst ($x_j^o>1$)
- künstliche Sätze ($x_j^o=0$)

t - Vektor der Häufigkeiten aller Tabellenfelder in den zu kontrollierenden Tabellen bei Auswertung des Originaldatenbestandes

A - Zuordnungsmatrix
($a_{ij}=1$, wenn Objekt im Tabellenfeld j gezählt wird, sonst $a_{ij}=0$)

e - Vektor aller Fehler bei der Tabellierung e_i ist Fehler im Tabellenfeld t_i

g - Vektor eventueller Gewichtungen der Tabellenfelder

11

Mathematisches Modell von SAFE (vereinfacht) (2)

$$Ax^{\circ} = t + e^{\circ}$$

mit :

$$x_j^{\circ} = 0,1,2,3,4,\ldots$$

$$e_j^{\circ} = 0$$

- Originallösung x^{o} enthält zwar keine Fehler, aber die potentiellen Geheimhaltungsfälle durch (x_j=1 und x_j=2)
- Die Dimension des Problems (Anzahl Objekte * Anzahl Tabellenfelder) verhindert eine direkte Lösung mit klassischer Optimierungssoftware. Deshalb wird ein Näherungsverfahren verwendet.

12

Arbeitsweise des Verfahrens SAFE

Arbeitsschritte:

1. - Sortierung der Originaldatei und Zusammenfassung von Objekten mit identischen Merkmalskombinationen
2. - Vorgabe eines zulässigen Anfangswertes des Maximalfehlers
3. - schrittweises Bearbeiten der Datei und Verändern der Häufigkeit von originalen Sätzen zu (0 oder 3,4 usw.) bei gleichzeitigem Versuch der Minimierung der Randsummenfehler unter der Restriktion des zulässigen Maximalfehlers
4. - Wenn erforderlich (Stagnation), erfolgt eine Erhöhung des zulässigen Maximalfehlers und die Wiederholung von Schritt 3

SAFE ist ein iteratives Verfahren.

SAFE – Tests mit der Volkszählung 1987

Beispiel:

Registermerkmale für Personen

Datensätze (Personen):	63 202 834	
Merkmale:	20	(30 mit Aggregationsstufen)
Kontrolltabellen:	416	
mit Tabellenfeldern:	11 485 249	
Tabellenfelder mit Geheimhaltungsfällen:	2 134 034	

SAFE–Tests mit der Volkszählung 1987

Mikrodatenfile für das Personenregister

Datensätze (Personen): 63 202 834
kontrollierte Tabellen: 416
Tabellenfelder: 11 485 249
Maximale Abweichung: 10 insgesamt (3 eindim.)

Tabellenfelder nach Größe von ... - bis ...	Anzahl an Tabellenfeldern	maximale Abweichung insgesamt	maximale Abweichung eindim.	mittlere Abweichung
1 - 9	4 471 429	6	2	1.65
10 - 49	2 822 413	7	2	2.39
50 - 99	1 008 565	8	2	2.67
100 - 149	505 857	9	2	2.95
150 - 199	318 245	9	2	3.19
200 - 999	1 254 910	10	2	3.41
1 000 - 9 999	854 690	10	2	3.41
10 000 - 99 999	215 132	10	3	3.28
100 000 - 999 999	31 177	10	3	3.13
1 000 000 und mehr	2 831	9	3	3.08

SAFE–Tests mit der Volkszählung 1987

Mikrodatenfile Personenregister

Abweichung im Tabellenfeld	Anzahl an Tabellenfeldern nach der Größe von … bis …							
	1 - 9	10 - 49	50 - 99	100 - 149	150 - 199	200 - 999	1 000 - 9 999	10 000 und mehr
0	521 383	357 877	113 036	51 228	30 265	110 616	77 217	23 772
1	1 842 827	637 243	219 070	98 896	57 794	214 492	148 823	45 021
2	1 224 414	550 070	176 839	91 929	54 089	200 664	136 501	41 396
3	539 738	526 629	150 607	72 071	47 656	179 562	119 681	35 595
4	248 322	424 206	159 125	61 408	36 894	153 140	102 549	30 200
5	89 584	235 239	119 751	62 601	31 166	123 561	85 691	25 240
6	5 161	88 668	55 700	45 068	30 230	109 787	72 499	21 143
7	-	2 481	14 358	18 576	20 577	91 814	57 427	14 858
8	-	-	79	4 068	8 310	51 827	35 858	8 069
9	-	-	-	12	1 264	19 343	17 482	3 778
10	-	-	-	-	-	104	962	68
>10	-	-	-	-	-	-	-	-
Σ	4 471 429	2 822 413	1 008 565	505 857	318 245	1 254 910	854 690	249 140

SAFE–Tests mit der Volkszählung 1987

Mikrodatenfile Personenregister

Abweichung im Tabellenfeld	Anteil an Tabellenfeldern nach der Größe von … bis …							
	1 - 9	10 - 49	50 - 99	100 - 149	150 - 199	200 - 999	1 000 - 9 999	10 000 und mehr
=0	11.7%	12.7%	11.2%	10.1%	9.5%	8.8%	9.0%	9.5%
≤1	52.9%	35.3%	32.9%	29.7%	27.7%	25.9%	26.4%	27.6%
≤2	80.3%	54.7%	50.5%	47.9%	44.7%	41.9%	42.4%	44.2%
≤3	92.3%	73.4%	65.4%	62.1%	59.6%	56.2%	56.4%	58.5%
≤4	97.9%	88.4%	81.2%	74.2%	71.2%	68.4%	68.4%	70.6%
≤5	99.9%	96.8%	93.0%	86.6%	81.0%	78.3%	78.4%	80.8%
≤6	100.0%	99.9%	98.6%	95.5%	90.5%	87.0%	86.9%	89.3%
≤7	100.0%	100.0%	100.0%	99.2%	97.0%	94.3%	93.6%	95.2%
≤8	100.0%	100.0%	100.0%	100.0%	99.6%	98.5%	97.8%	98.5%
≤9	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	99.9%	100.0%
≤10	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Interpretation der Ergebnisse

Besonderheiten von mit SAFE anonymisierten Daten:

1. Veröffentlichte Tabellenfelder sind mit einem gewissen Fehler überlagert.
2. Die Größe des Fehlers ist bekannt, und kann berücksichtigt werden.
3. Der Fehler wird relativ um so größer, je kleiner die Zellwerte sind.
4. In Tabellen fehlende Merkmalskombinationen können trotzdem existieren.
5. In einer Reihe (Zeile/Spalte) einmalige Kombinationen (Randsummenprobleme) sind nicht unbedingt in der originalen Gesamtheit einmalig.
6. Strukturen in den Daten werden gut abgebildet, und nicht wegen ggf. erforderlicher „Komplementärsperrungen“ verschleiert.

Grundidee der Stichprobenanonymisierung (1)

- Geheimhaltungsproblem:
  Teilnahmekenntnis ist unter bestimmten Voraussetzungen vorhanden.

- Stichprobenergebnisse werden mit Hilfe von Hochrechnungsfaktoren aus den erhobenen Eigenschaften der Stichprobeneinheiten hochgerechnet.

  Hochrechnungsfaktor = 1 / Auswahlsatz

Aber

- Hochrechnungsfaktoren werden skaliert damit sie Eigenschaften für den Gesamtbestand möglichst gut widerspiegeln.

  Hochrechnungsfaktor ≠ 1 / Auswahlsatz

  (verschiedene Einheiten → verschiedene Hochrechnungsfaktoren)

- Bei fein untergliederten Tabellen und genauen Hochrechnungen lässt sich die originale Zusammensetzung aus den einzelnen Einheiten unter bestimmten Konstellationen reproduzieren.

Grundidee der Stichprobenanonymisierung (2)

Beispiel: Kleine Gemeinde ca. 200 EW, Auswahlsatz 10%
(20 Personen in der Stichprobe)

Tabelle 1:

	männlich	weiblich	ins-gesamt
Deutsche	83	94	177
Ausländer	9	11	20
Insgesamt	92	105	197

Tabelle 2:

	erwerbs-tätig	erwerbs-los	ins-gesamt
Deutsche	155	22	177
Ausländer	9	11	20
Insgesamt	164	33	197

Wer von den beiden Ausländern in der Stichprobe ist erwerbslos?

Stichprobenergebnisse sind nicht automatisch sicher!!!

Grundidee der Stichprobenanonymisierung (3)

Deshalb Rundung veröffentlichter Ergebnisse (z.B. auf 10).

Tabelle 1:

	männlich	weiblich	ins-gesamt
Deutsche	80	90	180
Ausländer	10	10	20
Insgesamt	90	100	200

Tabelle 2:

	erwerbs-tätig	erwerbs-los	ins-gesamt
Deutsche	160	20	180
Ausländer	10	10	20
Insgesamt	160	30	200

Rundung ist auch wegen der statistischen Unsicherheit sinnvoll.
Aber: Rundung allein ist nicht immer ausreichend.

Tabelle 1:

	männlich	weiblich	ins-gesamt
Deutsche	90	90	180
Ausländer	0	10	10
Insgesamt	90	100	190

Tabelle 2:

	erwerbs-tätig	erwerbs-los	ins-gesamt
Deutsche	160	20	180
Ausländer	0	10	10
Insgesamt	160	30	190

Grundidee der Stichprobenanonymisierung (4)

Deshalb:

Rundung veröffentlichter Ergebnisse in Kombination mit Mindestfallzahlregel (z.B. Rundung auf 10, Mindestfallzahl 5).

Tabelle 1:

	männlich	weiblich	insgesamt
Deutsche	90	90	180
Ausländer	0	0	0
Insgesamt	90	100	190

Tabelle 2:

	erwerbstätig	erwerbslos	insgesamt
Deutsche	160	0	180
Ausländer	0	0	0
Insgesamt	160	0	190

Keine klassische Unterscheidung der Tabellenwerte zwischen nicht existent „-“ und geheim „/“ zu halten. Diese Unterscheidung ist als sichere Aussage für die Gesamtheit aus Stichproben nicht möglich.

Stichprobentabellen enthalten nur statistisch belastbare Tabellenwerte.

Vielen Dank für Ihre Aufmerksamkeit!

Dipl.-Volksw. Barbara Sinner-Bartels:
„Die Auswertungsdatenbank Zensus 2011“

Abstract:

Die Ergebnisse des Zensus 2011 bieten ein vielfältiges Analysepotential für die unterschiedlichsten Nutzergruppen. Für Entscheidungsträger aus Politik und Verwaltung stehen zunächst die aktuellen Einwohnerzahlen im Vordergrund. Für die Wissenschaft sind zudem die detaillierten Informationen über die Bevölkerung und den Gebäude- und Wohnungsbestand von besonderem Interesse.

Das breite Spektrum der Nutzergruppen erfordert ein vielschichtiges Datenangebot. Neben klassischen Printveröffentlichungen wird es eine Zensus-Auswertungsdatenbank geben, welche frei im Internet zugänglich sein wird. Sie umfasst zum einen vorgefertigte Tabellen und Grafiken. Darüber hinaus besteht aber auch die Möglichkeit, Tabellen selbst zusammenzustellen. Aus dem umfangreichen Themenkatalog des Zensus können Merkmale individuell und flexibel kombiniert, mit Grafiken visualisiert, heruntergeladen werden. Für wissenschaftliche Einrichtungen besteht die Möglichkeit, über die Forschungsdatenzentren komplexe Auswertungen auf Basis von Mikrodaten des Zensus durchzuführen.

Es ist selbstverständlich sichergestellt, dass die Regeln der statistischen Geheimhaltung berücksichtigt sind und keine Angaben über einzelne Personen an die Öffentlichkeit gelangen können.

Zur Person:

Barbara Sinner-Bartels leitet im Statistischen Landesamt Baden-Württemberg neben der Abteilung „Bevölkerung und Kultur“ mit den Bereichen Bevölkerungsstatistiken, Bildungsstatistiken, Beschäftigung und Arbeitsmarkt sowie Mikrozensus und Wahlen die Projektgruppe Zensus. Im Verbund der Statistischen Ämter des Bundes und der Länder hat Baden-Württemberg eine besondere fachliche Verantwortung für die Konzeption der Auswertungsdatenbank Zensus 2011 übernommen. Frau Sinner-Bartels hat an der Universität Tübingen den Abschluss einer Diplom-Volkswirtin erworben.

Vortragsfolien:

Statistik-Tage 2012
vom 26. bis 27. Juli 2012 in Bamberg
„Methoden und Potenziale des Zensus 2011“

Die Auswertungsdatenbank Zensus 2011

Barbara Sinner-Bartels
Statistisches Landesamt Baden-Württemberg

Bamberg, 27. Juli 2012

Baden-Württemberg
STATISTISCHES LANDESAMT

Agenda

1. Ziele und Rahmenbedingungen der Auswertungsdatenbank (ADB)
2. Struktur der Auswertungsdatenbank
3. Zugang der Wissenschaft zu den Zensusdaten

Baden-Württemberg
STATISTISCHES LANDESAMT

1. Ziele und Rahmenbedingungen der ADB

Die Auswertung muss den individuellen Ansprüchen der verschiedenen Nutzergruppen gerecht werden

Folgende Ansprüche an die Auswertung des Zensus 2011 konnten identifiziert werden:

- Flexibilität für die Erstellung individueller Auswertungen
- möglichst freie Verfügbarkeit für Nutzergruppen
- Etablierung eines bundesweit einheitlichen Standards
- Nachvollziehbarkeit der Zensusergebnisse
- umfassendes Auswertungsprogramm (soweit methodisch vertretbar)

1. Ziele und Rahmenbedingungen der ADB

Methodische Schranken sind erkannt, berücksichtigt und werden bei Bedarf transparent gemacht

Zensus-Modell mit „Multiple Source Mixed Mode“-Design

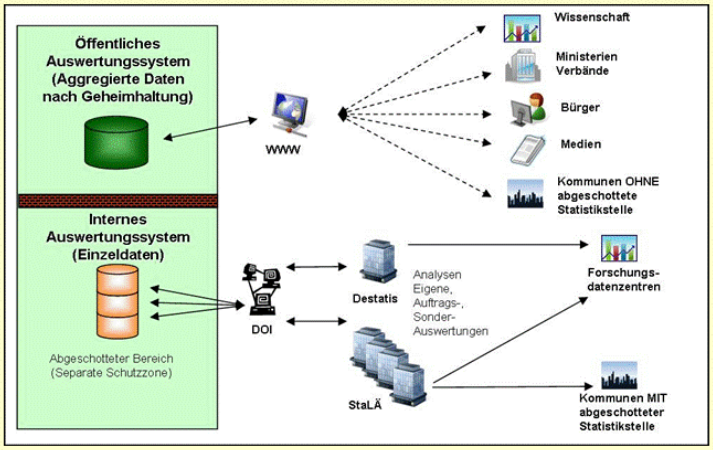
- Unterschiedliche Datenquellen für Ergebnisse zu den beiden Veröffentlichungsterminen (erste Ergebnisse/Ergebnisse nach Haushaltegenerierung)
- Neuland bei Geheimhaltung (Einsatz von SAFE)
- „Baukasten“ bei erwerbsstatistischen Merkmalen
- Vorgehensweise bei Ergebnissen mit zu geringer Fallzahl, sogenannte „unsichere“ oder „unzuverlässige“ Ergebnisse

→ fachlicher Abstimmungsbedarf

→ Erklärungsbedarf gegenüber Datennutzern

2. Struktur der Auswertungsdatenbank

Die Nutzergruppen der Auswertungsdatenbank erhalten Zugriff auf spezifische Datenbestände

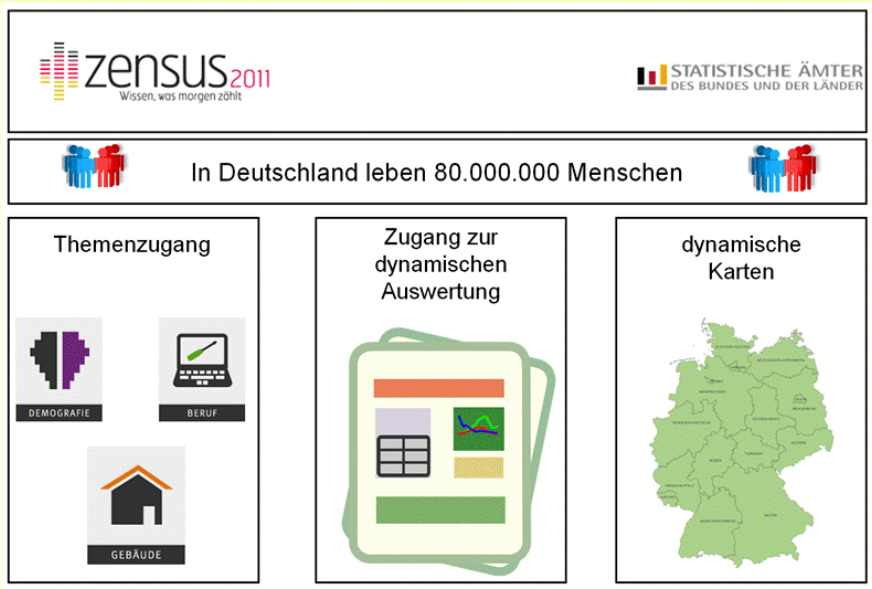


2. Struktur der ADB: das öffentliche System

Anders als bei der Volkszählung 1987 liegt der Fokus auf der Ergebnisveröffentlichung im Internet

- Zentrales Veröffentlichungsmedium des Zensus 2011bildet die **Auswertungsdatenbank**
- Freie Verfügbarkeit im WWW
- Anforderungen spezifischer Nutzergruppen (z.B. der Wissenschaft) wurden aufgegriffen und in der Konzeption berücksichtigt
  - → vorgefertigte Tabellen und Grafiken (statischer Bereich)
  - → individuelle Gestaltung von Tabellen und Grafiken (dynamischer Bereich)

2. Struktur der ADB: das öffentliche System



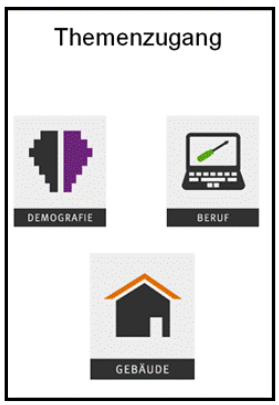
2. Struktur der ADB: das öffentliche System

Fokus:

Nutzer mit spezifischem Themeninteresse

Aufbau:

thematische Gliederung des Zensus 2011 → bis auf Einzelmerkmalsebene



Vorteil:

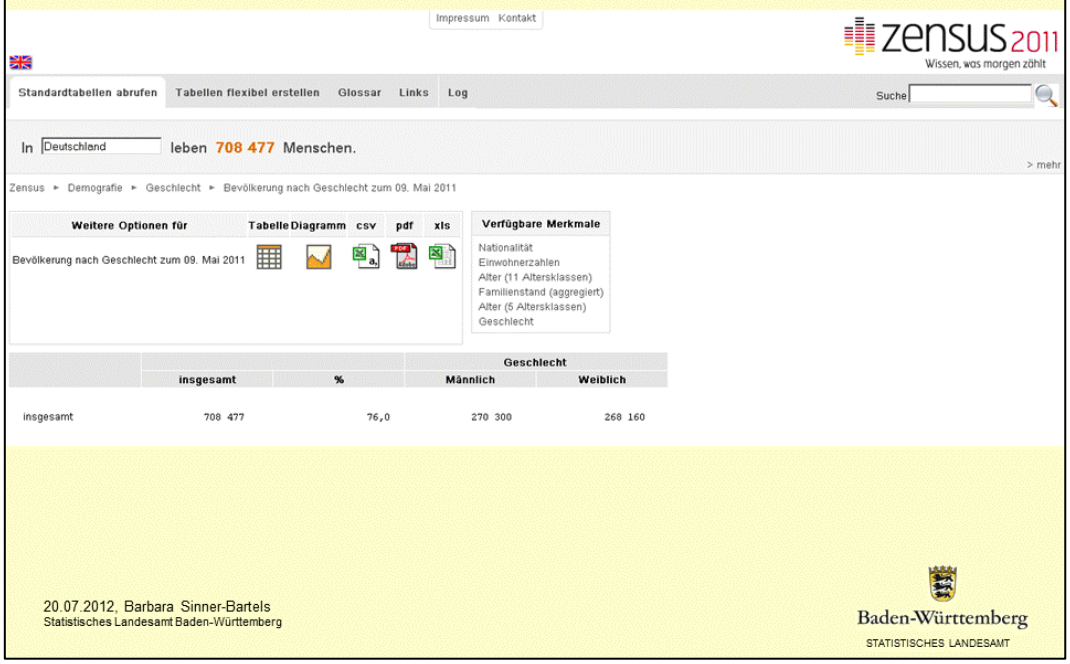
Schnellzugriff auf themenspezifische Ergebnistabellen & Grafiken

2. Struktur der ADB: das öffentliche System

20.07.2012, Barbara Sinner-Bartels
Statistisches Landesamt Baden-Württemberg

Baden-Württemberg
STATISTISCHES LANDESAMT

2. Struktur der ADB: das öffentliche System



20.07.2012, Barbara Sinner-Bartels
Statistisches Landesamt Baden-Württemberg

Baden-Württemberg
STATISTISCHES LANDESAMT

2. Struktur der ADB: das öffentliche System

Fokus:

Nutzer mit speziellem Informationsbedarf und/oder Statistikkenntnissen

Aufbau:

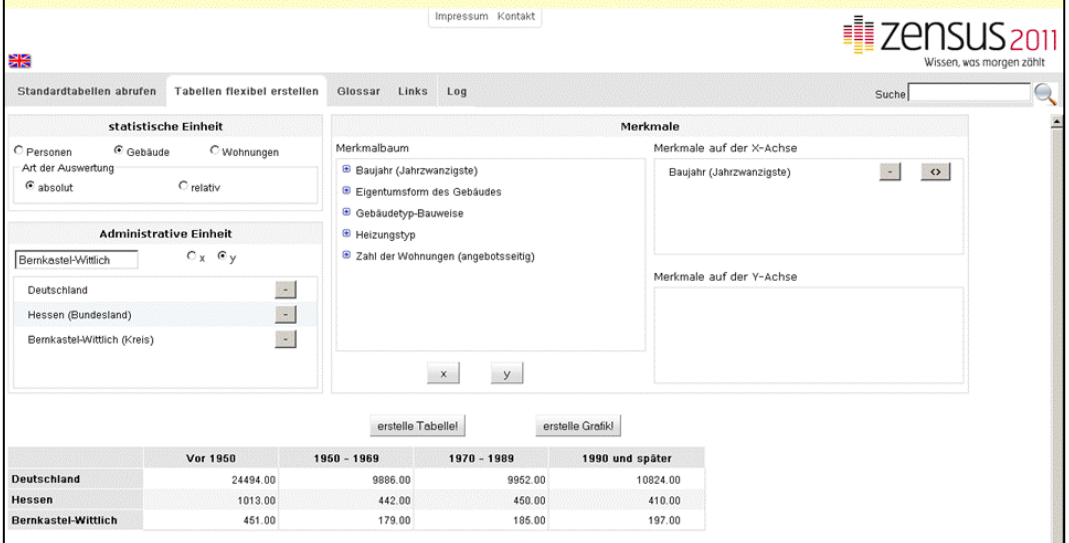
individuelle Aufbereitung und Gestaltung von Ergebnistabellen und Grafiken

Vorteil:

Kreuzkombinationen mit bis zu 5 fachlichen Merkmalen möglich

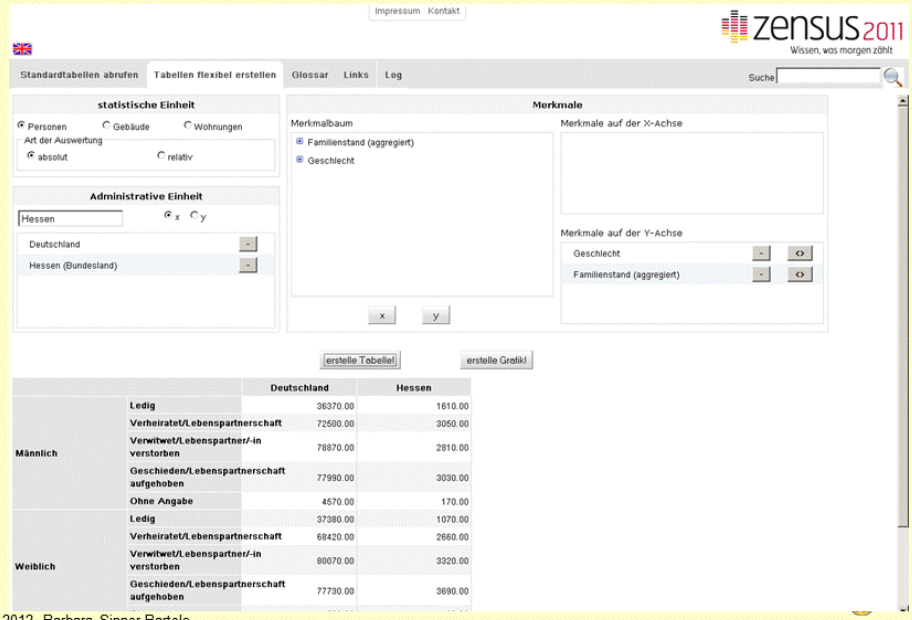
Zugang zur dynamischen Auswertung

2. Struktur der ADB: das öffentliche System



	Vor 1950	1950 - 1969	1970 - 1989	1990 und später
Deutschland	24494.00	9886.00	9952.00	10824.00
Hessen	1013.00	442.00	450.00	410.00
Bernkastel-Wittlich	451.00	179.00	185.00	197.00

2. Struktur der ADB: das öffentliche System



2. Struktur der ADB: das öffentliche System

Fokus:

Nutzer mit spezifischem Themeninteresse

Aufbau:

dynamische indikatorengestützte Kartendarstellungen zum zweiten Veröffentlichungstermin (VÖT2)

Vorteil:

Mehr als 230 vordefinierte Indikatoren stehen zur Auswahl

Alle Themenbereiche des Zensus werden über Indikatoren visualisiert

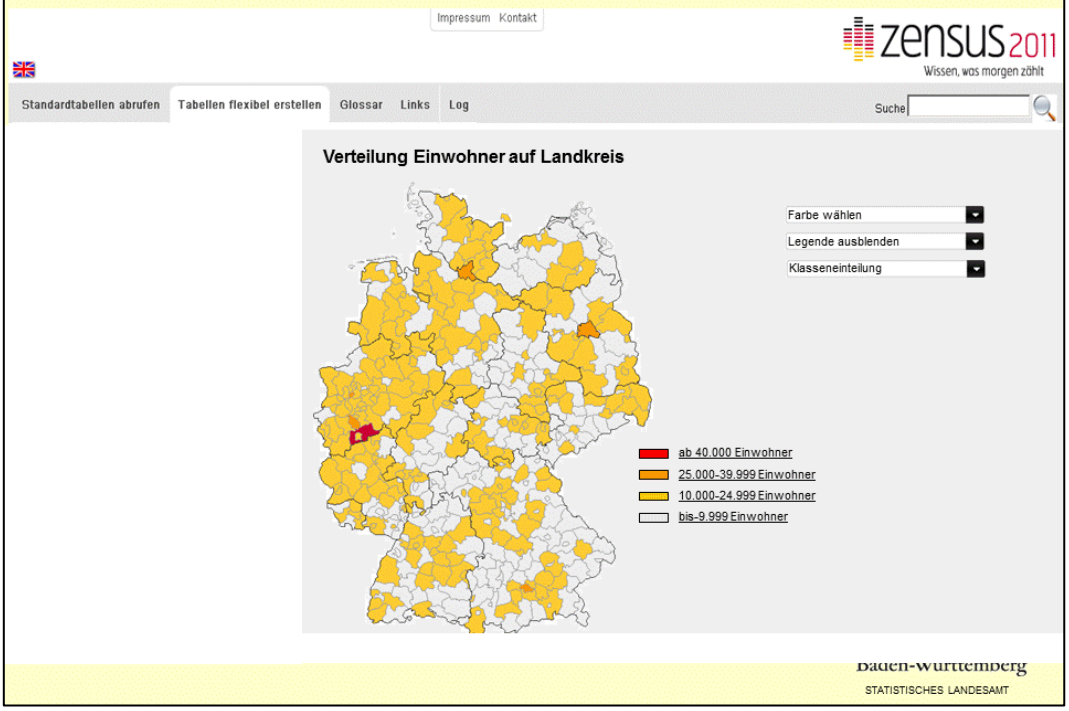
Bis zu drei fachliche Merkmalsdimensionen werden über Indikatoren abgebildet

Nahezu flächendeckende Bereitstellung der Indikatoren für alle regionalen Auswertungseinheiten

dynamische Karten

20.07.2012, Barbara Sinner-Bartels
Statistisches Landesamt Baden-Württemberg
Baden-Württemberg
STATISTISCHES LANDESAMT

2. Struktur der ADB: das öffentliche System



3. Zugang der Wissenschaft zu Zensusdaten

Der Informationsbedarf der Wissenschaft kann über verschiedene Kanäle gedeckt werden

Zugangsweg	Datenverfügbarkeit
Auswertungs-datenbank	Dynamischer Auswertungsbereich mit mehr als 500 vordefinierten Datenquadern Beauftragung von Sonderauswertungen für individuelle Forschungsfragen bzw. Fragestellungen
Forschungs-datenzentren	Diverse Zugangswege → On-Site / Off-Site Erweiterung des Gesamtproduktportfolios um Zensusdaten unterschiedlicher Granularität → SUF / PUF / Gastwissenschaftler / kontrollierte Datenfernverarbeitung

20.07.2012, Barbara Sinner-Bartels
Statistisches Landesamt Baden-Württemberg

Baden-Württemberg
STATISTISCHES LANDESAMT

3. Datenzugang über Forschungsdatenzentren

Die amtliche Statistik stellt den Forschungsdaten-zentren umfangreiches Datenmaterial zur Verfügung

Der bereitgestellte Datenumfang **erstreckt sich** über:

→ alle Erhebungs- und Auswertungsmerkmale (inkl. Wohnstatus),

→ Lösch- und Imputationskennzeichen im Rahmen des integrierten Korrekturverfahrens auf Personenebene,

→ Hochrechnungsfaktoren.

Nicht enthalten sind:

→ Hilfsmerkmale gemäß ZensusG 2011,

→ Imputationskennzeichen für Gebäude, Wohnungen und Personen im Rahmen des Aufbereitungsprozesses,

→ Imputationskennzeichen auf Merkmalsebene,

→ Existenzkennzeichen sowie Ergebniskennzeichen der Abgleiche verschiedener Erfassungsquellen.

3. Datenzugang über Forschungsdatenzentren

Für die Datenbereitstellung müssen gesetzliche Vorgaben beachtet werden

Geokoordinaten können der Wissenschaft aufgrund der gesetzlichen Rahmenbedingungen derzeit nicht zur Verfügung gestellt werden.

Forschungsvorhaben, für die weitere Merkmale benötigt werden, können ggf. in gemeinsamen Forschungsprojekte der Statistischen Ämter und der Wissenschaft realisiert werden.

Vielen Dank für Ihr Interesse!

www.zensus2011.de
www.statistik-bw.de/zensus

20.07.2012, Barbara Sinner-Bartels
Statistisches Landesamt Baden-Württemberg

Baden-Württemberg
STATISTISCHES LANDESAMT

Vortragsblock IV: Erwartungen der Wissenschaft

Neben der wissenschaftlichen Statistik, die sich mit der Evaluation und Weiterentwicklung der Methodik befasst, sind die generierten Individual- und kleinräumigen Aggregatdaten des bundesweiten Zensus vor allem für die empirische Sozialforschung von großem analytischem Interesse. Die Demographie hat ihren Fokus insbesondere auf den Bevölkerungszahlen und -strukturen, die Geographie in erster Linie auf kleinräumigen sozialstrukturellen Aspekten. Für die Migrations- und Integrationsforschung bietet sich durch die Erfassung verschiedener Merkmale aus dem Bereich Migrationshintergrund und Religionszugehörigkeit eine wertvolle Datenquelle und auch die Arbeitsmarkt- und Berufsforschung erhält durch die Verknüpfung von Stichproben-, Prozess- und Registerdaten eine reichhaltige Auswertungsbasis. Mit Frau Prof. Engelhardt-Wölfler, Prof. Jürgen Rauh, Prof. Peter Schimany und Prof. Uwe Blien sprechen vier Experten u.a. über die Ansprüche an den Zensus aus Sicht ihres Forschungsbereichs, die Potenziale des Zensus in seiner jetzigen Form für ihre Disziplin, ihre Erwartungen an Datenzugangs- und Auswertungsmöglichkeiten und ihre Änderungsvorschläge für die Zukunft.

Prof. Henriette-Engelhardt-Wölfler:
„Der Zensus aus der Sicht der Demographie“

Zur Person:

Prof. Henriette Engelhardt-Wölfler ist Inhaberin der Professur für Bevölkerungswissenschaft an der Otto-Friedrich-Universität Bamberg. Nach ihrem Studium der Soziologie und Statistik an der Universität Mannheim wurde sie an der Universität Bern 1998 promoviert und 2005 habilitiert. Sie arbeitete als wissenschaftliche Mitarbeiterin an der Universität Bern (1992-1998), dem Max-Plack-Institut für Bildungsforschung in Berlin (1998-2000), dem Max-Planck-Institut für demographische Forschung in Rostock (2002-2002) und dem Vienna Institute of Demography (2002-2006). Darüber hinaus war sie Gastwissenschaftlerin an der Duke University, North Carolina (2000), dem International Institute for Advanced Systems Analysis (IIASA) in Laxenburg, Österreich (2002) und an der Eidgenössischen Technischen Hochschule Zürich (2004). Ihre Forschungsinteressen liegen im Bereich der Sozial- und Familiendemographie sowie der demographischen Alterung und Kausalanalyse.

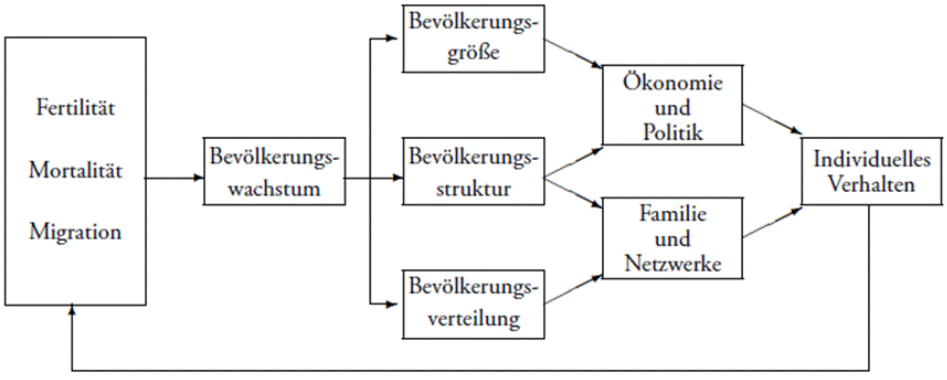
Vortragsfolien:

Otto-Friedrich-Universität Bamberg

Der Zensus aus der Sicht der Demographie

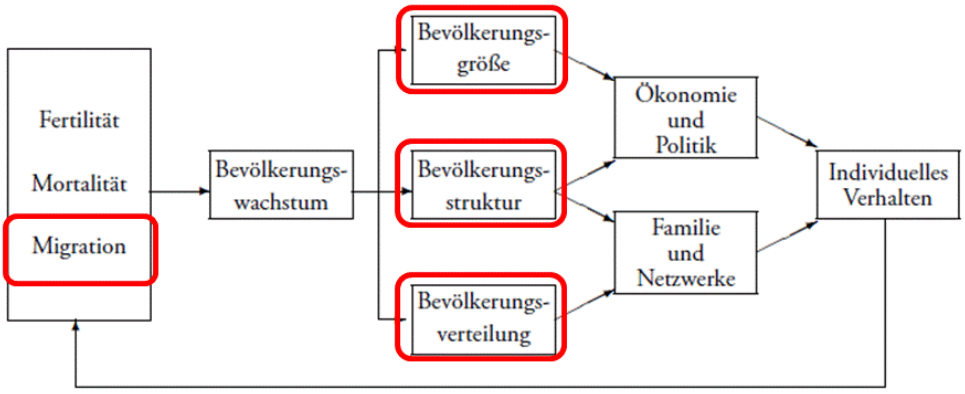
Prof. Dr. Henriette Engelhardt-Wölfler

Kerngeschäft der Demographie



Quelle: Engelhardt (2011)

Kerngeschäft der Demographie und Zensusdaten



Quelle: Engelhardt (2011)

Festlegung eines Gesamtbestandes

„Ausgangspunkt jeder demographischen Analyse und Messung ist eine Festlegung eines Gesamtbestandes zu irgendeinem Zeitpunkt, differenziert nach den wichtigsten Strukturmerkmalen wie Alter, Geschlecht, Familienstand oder Erwerbsbeteiligung."
(Dinkel 1989: 15)

Definition des Zählmasse: Empfehlung der UN (1980):

- "modified de facto population" – alle Personen, die zu einem Stichtag auf dem Gebiet gefunden werden, abzüglich ausländischer Militärpersonen und Diplomaten mit Familien, und zuzüglich einheimischer Militärpersonen und Diplomaten mit Familien, die sich im Ausland befinden, sowie zuzüglich einheimischer Seeleute, die sich gerade auf See befinden.

→ Kombination von Aufenthalts- und Residenzprinzip

→ Wie ist der Gesamtbestand im Zensus 2011 definiert?

Bevölkerungsgröße

Bevölkerungsgröße als Ausgangsbasis:

- Bevölkerungsfortschreibung:
  - Alters- und geschlechtsspezifische Bevölkerungsgröße
  - Aber auch nach Nationalität, Konfession, Bildung, etc.
- Bevölkerungsprognosen und -projektionen:
  - Alters- und geschlechtsspezifische Bevölkerungsgröße
  - Differentielle Prognosen nach Regionen, Nationalität, Konfession, Bildung, etc.
- Fertilitäts-, Mortalitäts- und Migrationsraten:
  - Bevölkerung nach Alter, Geschlecht, Nationalität, Bildung, Region, etc. zur Berechnung differentieller Raten : Nenner

Fertilität

- Verlässliche Daten zu Geburten aus den Registern der Einwohnermeldeämter
- Verfügbare Informationen:
  - Geburtsdatum, Geschlecht, Alter der Mutter und Staatsangehörigkeit (seit 1950)
  - Bis 2008 Information zur Rangfolge nur für eheliche Geburten
- Demographische Fertilitätsindikatoren:
  - Gesamtfertilitätsrate (TFR = Summe der alters- und paritätsspezifischen Periodenfertilitätsraten)
  - Durchschnittliche Kinderzahl
  - Anteil Kinderlosigkeit
  - Durchschnittliches Alter der Mutter bei Erstgeburt

Fertilität

Konsequenz:

- Keine Information zur Fertilität von Männern
- Keine Analyse der Kohortenfertilität möglich
- Keine paritätsspezifischen Analysen vor 2008 möglich

Analysemöglichkeiten der Zensusdaten:

- Korrektur der aktuellen Fertilitätsziffern (Fertilität wird sich vermutlich erhöhen, aufgrund Dezimierung der Bezugspopulation)
- Gesamtfertilitätsrate unterschätzt die tatsächliche Anzahl der Geburten, aufgrund des paritätsspezifischen Aufschiebens der Geburten; Bereinigung um Tempoeffekte möglich bei Nutzung der paritätsspezifischen Information über das Alter bei Geburt (auch für Projektionen)
- Differentielle Fertilität: Familienstand, Bildung, Migrationshintergrund

Mortalität

- Zensusdaten liefern Bevölkerungsgröße zur Berechnung von differentiellen Mortalitätsraten
- Befund aus der Bevölkerungsstatistik des Statischen Bundesamtes: Menschen mit Migrationshintergrund leben deutlich länger als die einheimische Bevölkerung
- „Healthy-Migrant“-Effekt: Migranten sind zum Zeitpunkt der Einwanderung gesünder als die einheimische Bevölkerung und haben in Folge eine höhere Lebenserwartung
- Statistisches Artefakt?

Mortalität

- Mangelnde Registrierung von Fortzügen und im Ausland verstorbenen Ausländern
  - Anzahl der Ausländer in der Statistik wird zu hoch ausgewiesen
  - Anzahl der Sterbefälle wird zu gering ausgewiesen

→ Ausmaß der Unterschätzung der Mortalität/Langlebigkeit?

→ Annäherung der Mortalität/Langlebigkeit mit Aufenthaltsdauer?

→ Differentielle Mortalität/Langlebigkeit nach Herkunftsland?

Migration

- Zensus ermöglicht Identifizierung von Personen mit Migrationshintergrund
- Ermöglicht differentielle Analyse der
  - Bevölkerungsstruktur
  - Bevölkerungsverteilung
  - Fertilität
  - Mortalität
- Relevant für differentielle Prognosen (anstelle Mikrozensusdaten)

→ Mehr zu Migration von Prof. Schimany

Aggregat- vs. Individualanalyse

- Aggregatanalysen erlauben nicht die Übertragung der Ergebnisse auf individuelle Ebene
  → Gefahr des ökologischen Fehlschlusses
- Bsp. Zusammenhang von Fertilität und Erwerbstätigkeit
  Individualebene: negativ korreliert,
  Aggregatanalyse: positiv korreliert
- Bsp. Zusammenhang von Fertilität und Scheidung
  Individualebene: negativ korreliert
  Aggregatebene: positiv korreliert
- Zusammenhang von Fertilität und Heirat
  Individualebene: negativ korreliert
  Aggregatebene: positiv korreliert

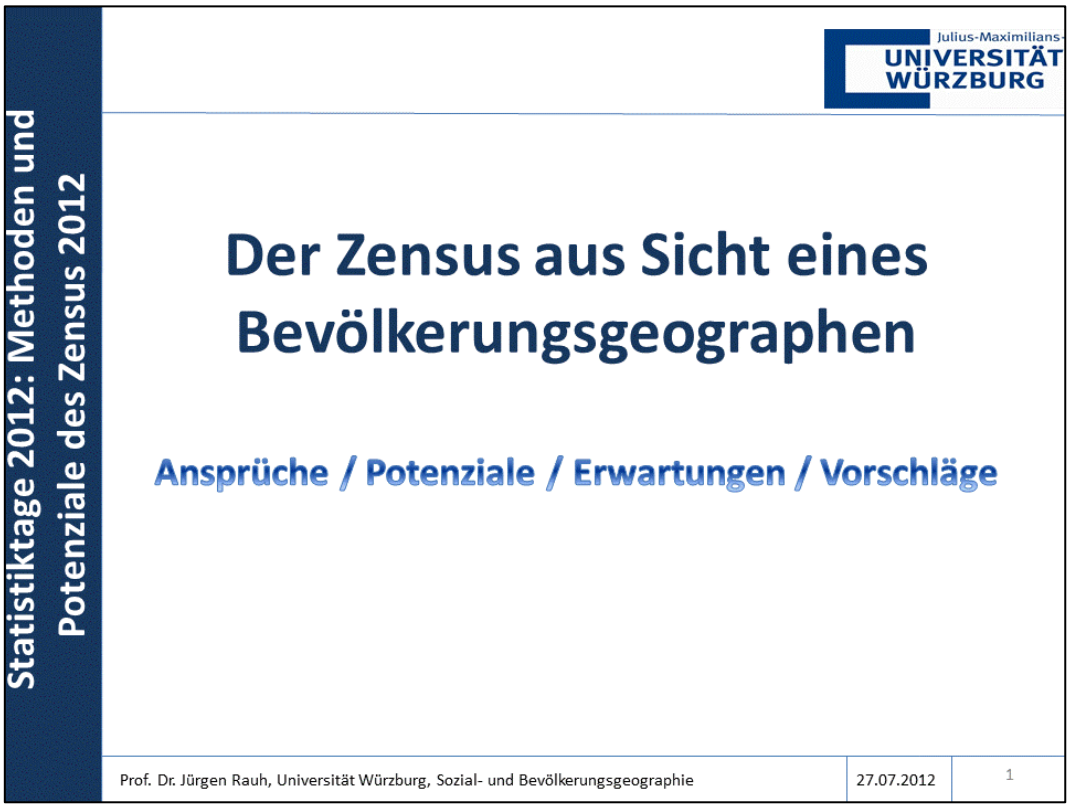
→ Analyse von Individualdaten ist erforderlich!

Prof. Jürgen Rauh:
„Der Zensus aus Sicht eines Bevölkerungsgeographen“

Zur Person:

Prof. Jürgen Rauh ist Inhaber der Professur für Sozialgeographie mit Schwerpunkt Bevölkerungsgeographie und regionalwissenschaftliche Methodenlehre an der Julius-Maximilians-Universität Würzburg. Nach seinem Studium der Geographie, Statistik und Ökonometrie an der Universität Regensburg, war er zunächst Gesellschafter eines Regionalplanungsbüros und im Anschluss bei der Stadt Regensburg beschäftigt, bevor er während seiner Zeit als wissenschaftlicher Mitarbeiter und Dozent an der Universität Regensburg 1991 promoviert und 1998 habilitiert wurde. Im Anschluss übernahm er Lehrstuhlvertretungen an der Technischen Universität München, bevor ihn sein Weg in die Region Mainfranken führte, zu der er aktuell auch in mehreren Forschungsprojekten Analysen und Konzepte erstellt. Seine allgemeinen Forschungsinteressen liegen im Bereich der Sozial- und Bevölkerungsgeographie, der Regionalforschung und der geographischen Handelsforschung sowie der regionalwissenschaftlichen Methodenlehre und Geoinformatik.

Vortragsfolien:



Der Zensus aus Sicht eines Bevölkerungsgeographen

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Ansprüche, Potenziale, Erwartungen, Vorschläge:

5 Facetten

1. Fachliche Analysepotenziale

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

1. Fachliche Analysepotenziale der demographischen und sozioökonomischen Strukturdaten, Wohnungs- und Gebäudedaten, Pendlerbeziehungen

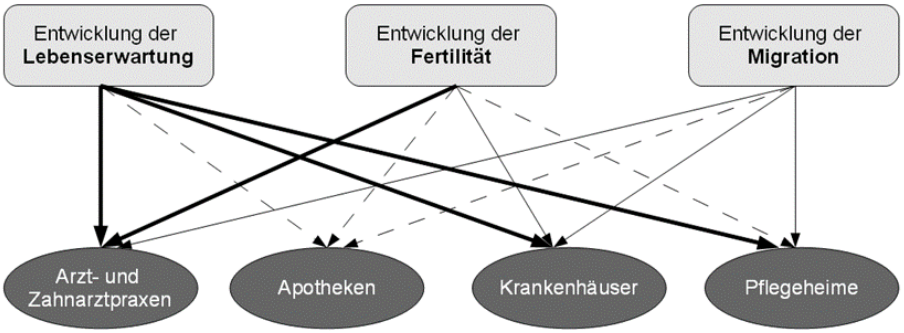
- große Potenziale v.a. aufgrund zusammengeführter Registerdaten
- Forschungsfragen mit kleinräumigen Informationsbedarf (z.B. Grundlagendaten für Forschungen zu Quartiersbezogenen Themen, Quartiersmanagement, demographischen Wandel, sozialräumliche Differenzierungen, Business Improvement Districts, kleinräumige Wohnsituation/ -markt (auch Gemeinschaftsunterkünfte), Gebäudenutzung, Leerstände)
- Forschungsfragen zu Migrationen (International, regional (?))
- Forschungsfragen zur geographischen Arbeitsmarktforschung (z.B. Telearbeit, Wohn-Arbeitsortbeziehungen, Bildungs-/Erwerbschancen nach regionaler Herkunft)
- eigene themenbezogene Regionalisierungen

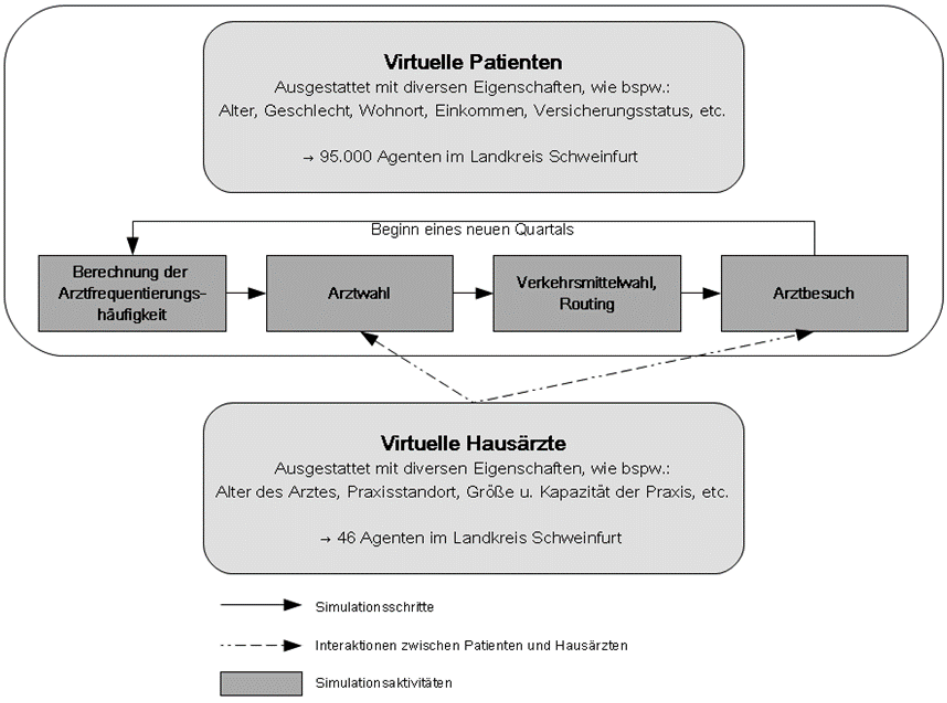
1. Fachliche Analysepotenziale

Julius-Maximilians-UNIVERSITÄT WÜRZBURG

Beispiel Forschungsprojekt „Demographischer Wandel und medizinische Versorgung im ländlichen Raum: Simulation mit einem Multiagentenmodell“

Auswirkungen des demographischen Wandels auf die medizinische Infrastruktur

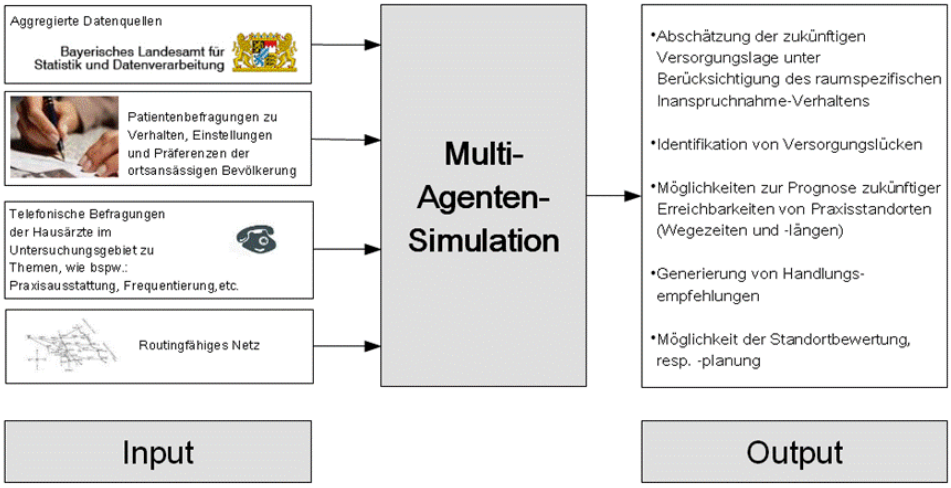




Statistiktage 2012: Methoden und Potenziale des Zensus 2012

1. Fachliche Analysepotenziale

Julius-Maximilians-UNIVERSITÄT WÜRZBURG



Statistiktage 2012: Methoden und Potenziale des Zensus 2012

2. Methodische Potenziale

Julius-Maximilians-UNIVERSITÄT WÜRZBURG

2. Zensus besitzt große Bedeutung für humangeographische Methodenforschung

- Zensusdaten als Referenz- und Hochrechnungsrahmen für Primärerhebungen in der Bevölkerungs-, Sozial- und Arbeitsmarktgeographie
- regionalvergleichende Analysen von Register- und Befragungsdaten
  ➔ Diskrepanzen zwischen beiden Quellen

Statistiktage 2012: Methoden und Potenziale des Zensus 2012

3. Flexible Geographien

Julius-Maximilians-UNIVERSITÄT WÜRZBURG

3. Flexible Geographien: Daten für gesellschaftlich, ökonomisch und ökologisch relevante Raumeinheiten

- je nach Fragestellung: **flexible Geographien**

vgl. Stellungnahme der Zensuskommission (2009):

- für jede Fragestellung ist eine bestimmte „Körnigkeit" der Daten oder ein spezifischer räumlicher Maßstab angemessen
- eine zu grobe oder zu feine Aufschlüsselung kann zur Folge haben kann, dass mit den Daten keine sinnvolle Analyse mehr möglich ist.
- möglichst feinkörnig erheben und die Feinkörnigkeit dauerhaft beibehalten (d.h. Baublockseite als Pflichtmerkmal), um für weitere Aggregationen so **flexibel** wie möglich zu bleiben.

Statistiktage 2012: Methoden und Potenziale des Zensus 2012

3. Flexible Geographien

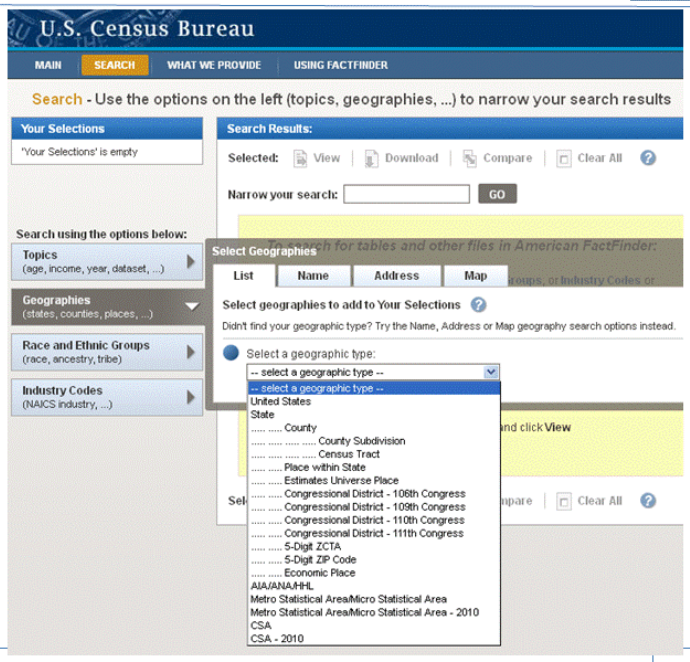
Julius-Maximilians-UNIVERSITÄT WÜRZBURG

- Online-Auswertungsdatenbank gut (incl. Kartendarstellung bis auf Gemeindeebene)!
- **Geocodierung** der Zensusdaten, so dass sie auf die jeweils geeignetste regionale Konfiguration für die jeweilige Forschungs- und Planungsaufgabe aggregiert werden können (faktische Anonymisierung)
- Problem Kommunen ohne abgeschottete Statistikstelle: Kleinräumigkeit der Daten?
  ➔ Jedes Datum sollte die Geocodierung erhalten, die rechtlich zulässig ist und methodisch sinnvoll ist (ideal: Blockseite)
- Flexibilität für die Erstellung individueller Auswertungen

Statistiktage 2012: Methoden und Potenziale des Zensus 2012

3. Flexible Geographien

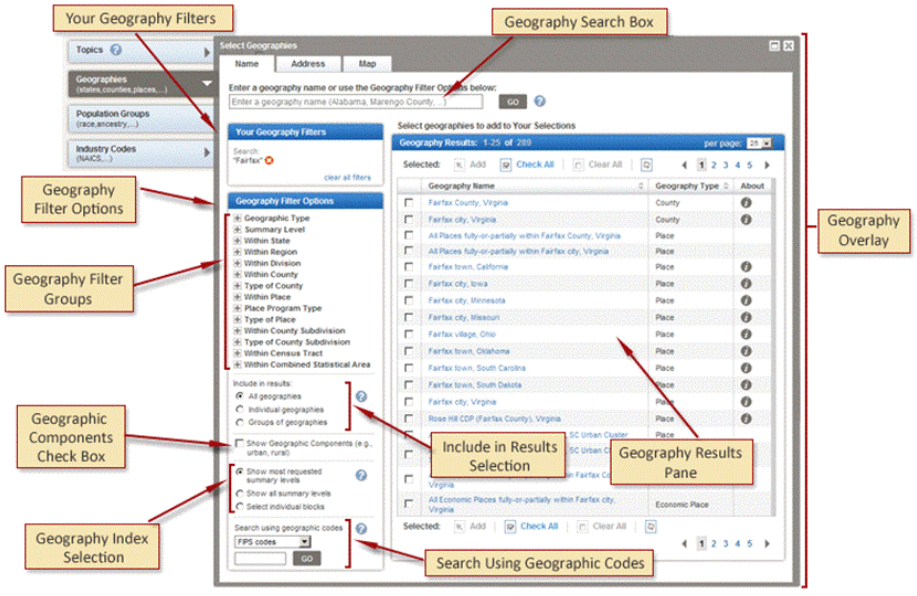
Julius-Maximilians-UNIVERSITÄT WÜRZBURG



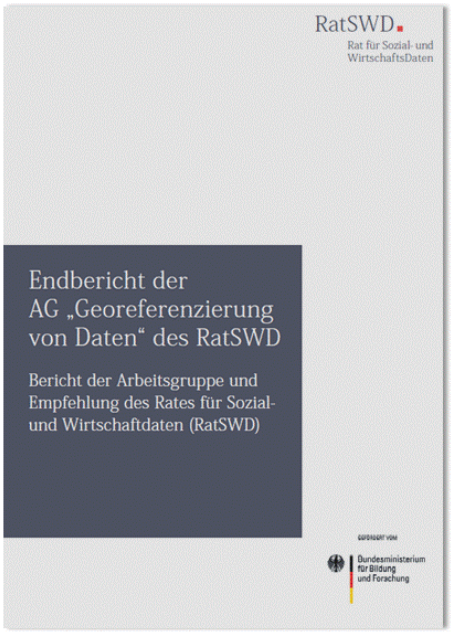
Statistiktage 2012: Methoden und Potenziale des Zensus 2012

3. Flexible Geographien

Julius-Maximilians-UNIVERSITÄT WÜRZBURG



3. Flexible Geographien



Georeferenzierte Daten

„Die AG beklagt das Fehlen von flexibel auswertbaren, kleinräumigen Daten aus der amtlichen Statistik.“

Forderung nach „Einführung kleinräumiger, nicht administrativer Bezugseinheiten (z. B. Gitterzellen) in der amtlichen Statistik“

Statistiktage 2012: Methoden und Potenziale des Zensus 2012

3. Flexible Geographien

Zuordnung georeferenzierter Adressen zu Gitterzellen

Quelle: Maack/Rösel (2010): Das GeoStat-Projekt. Wohin bewegt sich die Statistik in Europa? Vortrag bei Statistische Woche 2010, München

4. Mobile Gesellschaft

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

4. Bevölkerung und Raum stellen keinen fixen Konnex dar

Die Bevölkerung interagiert zwischen einer Reihe an Orten und ist nicht nur an einem Wohnort fixiert

➔ Zensus weist Bevölkerung nur Wohnort und überwiegenden Arbeitsort zu

➔ gebraucht wird aber: tatsächliche Bevölkerung mit einer Differenzierung nach permanenten und nicht-permanenten Bevölkerungsgeographien

➔ andere Zensusfragen wünschenswert
(z.B. Unterscheidung nach Tag-/Nachtbevölkerung, Interaktionsbeziehungen …)

Statistiktage 2012: Methoden und Potenziale des Zensus 2012

5. Erhebung der Migration

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

5. Wachsende internationale Migration und ihre große Diversität (permanent, semi-permanent, temporär, Transmigration, Remigration, …) erfordert differenziertere Erfassung sowie internationale Abstimmung von Migrationsdaten!

- Grund: Fokus des Zensus liegt auf der Erfassung der Immigranten und nicht der Emigranten

➔ keine Informationen zur Diaspora, Netzwerken, Transmigration …

➔ differenziertere Erfassung und Synchronisation von nationalen Zensen und Migrationsstatistiken ➔ Quell-/Zielmatrizen

Julius-Maximilians-
UNIVERSITÄT WÜRZBURG

6. Fazit

vgl. Stellungnahme der Zensuskommission (2009):

- ein Zensus kann nur eine begrenzte Anzahl von Erhebungsmerkmalen erfassen
- Neben systematischen Grundinformationen über Bevölkerung, Wohnungs- und Arbeitsmarkt sollten Informationen zu folgenden Herausforderungen bereitgestellt werden:
  - Migration und Integration,
  - Berufliche Mobilität und Verkehrsverhalten,
  - Energieversorgung und Umweltverhalten
- In zukünftigen Zensus: Insbesondere Informationen zu Transmigration, Multilokalität, hauptsächlich gesprochene Sprache im Haushalt, Pendlerbeziehungen, Energiequelle/Heizung
- Rechtliche Voraussetzung schaffen, damit alle Zensusdaten in der kleinräumig möglichsten Georeferenzierung bereitgestellt werden können
- Verweis auf Mikrozensus als Alternative aus geographischer Perspektive nicht zufriedenstellend: Mikrozensus aufgrund seiner noch geringeren Stichprobengröße für viele kleinräumige Forschungen ungeeignet

Statistiktage 2012: Methoden und Potenziale des Zensus 2012

Julius-Maximilians-
UNIVERSITÄT WÜRZBURG

Vielen Dank!

Prof. Peter Schimany:

„Der Zensus aus Sicht der Migrations- und Integrationsforschung“

Zur Person:

Prof. Peter Schimany ist Referatsleiter im Bundesamt für Migration und Flüchtlinge in Nürnberg und apl. Professor für Soziologie an der Universität Passau. Zuvor war er als Wissenschaftler am Institut für Demographie der Österreichischen Akademie der Wissenschaften in Wien tätig. Er promovierte an der Universität Erlangen-Nürnberg und habilitierte sich an der Universität Passau. Seine Forschungsinteressen liegen im Bereich der Demographie und der Migrationsforschung sowie der Wissenschaftssoziologie und der Zeitgeschichte. Unter anderem ist er Mitherausgeber der Bände „Integration von Zuwanderern. Erfahrungen, Konzepte, Perspektiven“ (Bielefeld: transcript Verlag 2010) und „Viele Welten des Alterns. Alternde Migranten im alternden Deutschland“ (Wiesbaden: VS Verlag 2012) sowie Autor von „Die Alterung der Gesellschaft. Ursachen und Folgen des demographischen Umbruchs“ (Frankfurt/New York: Campus Verlag 2003) und „Migration und demographischer Wandel“ (Nürnberg: BAMF-Forschungsbericht 2007).

Vortragsmanuskript:

0. Vorbemerkungen

Aus Sicht der Migrations- und Integrationsforschung kommt im Zensus den Erhebungsmerk-malen „Migrationshintergrund“ und „Religionszugehörigkeit“ besondere Bedeutung zu. Zum einen nimmt die Bevölkerung mit Migrationshintergrund weiter zu. Dem Mikrozensus 2010 zufolge haben 15,7 Mio. Menschen einen Migrationshintergrund, was etwa einem Fünftel (19,3%) an der Gesamtbevölkerung entspricht (Stat. Bundesamt 2011) . Zum anderen ist mit dem Migrationsgeschehen auch eine zunehmende Vielfalt in kultureller und religiöser Hin-sicht verbunden. Darüber hinaus sind die Zensusdaten für die Arbeit der Forschungsgruppe des Bundesamtes für Migration und Flüchtlinge in mehrfacher Hinsicht von Interesse. Nachfolgend möchte ich daher zuerst die Bedeutung der Zensusdaten für die Migrations- und Integrationsforschung im Bundesamt ansprechen. Anschließend werde ich auf die beiden Erhebungsmerkmale „Migrationshintergrund“ und „Religionszugehörigkeit“ eingehen. Hierbei werde ich zunächst das Konzept „Personen mit Migrationshintergrund“ im Mikrozensus behandeln. Danach gehe ich auf die Erhebung des „Migrationshintergrundes“ und der „Religionszugehörigkeit“ im Zensus ein. Abrunden werde ich mein Statement mit einem Fazit.

1. Bedeutung der Zensusdaten für die Migrations- und Integrationsforschung

Die Migrations- und Integrationsforschung umfasst thematisch und institutionell ein weites Feld (Schimany/Schock 2012). Ein zentraler Bestandteil der Forschungslandschaft ist seit 2005 die Gruppe für Migrations- und Integrationsforschung im Bundesamt für Migration und Flüchtlinge (BAMF). Für die Arbeit der Forschungsgruppe bzw. des Bundesamtes sind die Zensusdaten in mehrfacher Hinsicht von Interesse:

- Datenbasis für die Migrations- und Integrationsforschung

Erstens können die Zensusdaten generell als Datengrundlage für die Migrationsforschung gemäß § 75 Nr. 4 AufenthG dienen, wonach das Bundesamt wissenschaftliche Forschungen über Migrationsfragen zur Gewinnung analytischer Aussagen für die Steuerung der Zuwanderung betreiben soll.

Zweitens können die Zensusdaten als Auswahlgrundlage für empirische Erhebungen dienen (z.B. „Repräsentativbefragung ausgewählter Migrantengruppen (RAM)“).

Drittens können die Zensusdaten im Rahmen von Sekundäranalysen für die Integrationsberichterstattung herangezogen werden. Die Vergleichbarkeit von strukturellen Merkmalen von Personen ohne und mit Migrationshintergrund ermöglichen Analysen zum Integrationsstand einzelner Bevölkerungsgruppen.

- Planungsgrundlage für den Integrationsbereich

  Die Zensusdaten enthalten räumlich tiefer gegliederte Informationen zu den in Deutschland lebenden Personen mit Migrationshintergrund. Diese Daten können als Grundlage zur Durchführung, Steuerung und Förderung von Integrationsmaßnahmen wie Integrationskurse, Migrationsberatungen und Integrationsprojekte genutzt werden. Im Idealfall können die Teilnehmerpotenziale von Integrationsmaßnahmen genauer bestimmt und einzelne Aktivitäten mit Blick auf bestimmte Zielgruppen genauer konzipiert werden.

- Beurteilung der Datenqualität des Ausländerzentralregisters (AZR)

  Im AZR werden Ausländer erfasst, die sich nicht nur vorübergehend in Deutschland aufhalten, sondern länger als drei Monate hier leben. Mit den Zensusdaten wird die Bevölkerungsfortschreibung auf eine neue Ausgangsbasis gestellt. Dies betrifft auch Zahl und Struktur der ausländischen Bevölkerung. Insofern ergeben sich aus den Zensus-Ergebnissen auch Orientierungswerte zur Beurteilung der Datenqualität des AZR. Eine Korrektur der Melderegisterdaten über den gemäß § 90b AufenthG vorgesehenen Datenabgleich zwischen Meldebehörden und Ausländerbehörden dürfte auch eine Verbesserung der Datenqualität im AZR zur Folge haben. Denn letztlich sind die Daten des AZR nur so gut wie die gelieferten Daten der Ausländerbehörden.

- Datenbasis für die kulturelle Vielfalt der Bevölkerung

  Die Zensusdaten können auch als Informationsgrundlage für die kulturelle und religiöse Vielfalt der Bevölkerung dienen. Kenntnisse hierzu sind für das Bundesamt auch aus institutionellen Gründen wichtig. Zum einen ist hier die Geschäftsstelle der Deutschen Islam Konferenz eingerichtet und zum anderen ist das Bundesamt mit dem Aufnahmeverfahren für jüdische Zuwanderer (gemäß § 23 Abs. 2 AufenthG) betraut. Anfragen beziehen sich daher häufig auch auf das Merkmal „Religionszugehörigkeit“.

2. Konzept und Definition „Personen mit Migrationshintergrund“ im Mikrozensus

Die amtlichen Bevölkerungsstatistiken unterscheiden in der Regel nur zwischen Deutschen und Ausländern. Aufgrund der Vielfalt des Migrationsgeschehens und der Reform des Staatsangehörigkeitsrechts im Jahr 2000 lassen sich Stand und Entwicklung von Personen mit Migrationshintergrund, zu denen neben Ausländern und ihren

Nachkommen auch (Spät-) Aussiedler und Eingebürgerte zählen, nur noch unzureichend abbilden. Die amtliche Bevölkerungsstatistik hat auf diese Defizite reagiert und das Konzept „Bevölkerung mit Migrationshintergrund“ eingeführt.

Seit 2005 ist im Mikrozensus eine tiefergehende Identifizierung von Personen mit Migrationshintergrund möglich. Danach liegt ein Migrationshintergrund bei folgenden Personengruppen vor (Stat. Bundesamt 2011:6):

a) Zugewanderten seit dem 01.01.1950,
b) Ausländerinnen und Ausländern,
c) Eingebürgerten und
d) Kindern mit mindestens einem im Ausland geborenen und zugewanderten, ausländischen oder eingebürgerten Elternteil.

Zur Bevölkerung mit Migrationshintergrund zählen demnach Personen, die selbst zugewandert oder Nachkommen von Zuwanderern sind. Mit dem Konzept „Personen mit Migrationshintergrund“ im Mikrozensus können somit die beiden Dimensionen „Herkunft“ und „Generationenstatus“ differenziert erfasst werden (Gresch/Kristen 2011).

Der Mikrozensus ist derzeit die einzige amtliche und repräsentative Datenquelle zur Bevölkerung mit Migrationshintergrund. Allerdings gibt es Ansätze, den Migrationshintergrund auch in anderen amtlichen Statistiken abzubilden wie der Schulstatistik sowie der Kinder- und Jugendhilfestatistik. Zudem wird der Migrationshintergrund in der Kommunalstatistik und im Rahmen der empirischen Sozialforschung wie der PISA-Untersuchung und der Berufsbildungsforschung erfasst. Schließlich wird der Migrationshintergrund zukünftig auch in Arbeitsmarktstatistiken erhoben (Fritz/Gericke 2012).

Die einzelnen Definitionen des Migrationshintergrundes lehnen sich eng an die Abgrenzung des Statistischen Bundesamtes an. Die Erfassung von Personen mit Migrationshintergrund erfolgt aber auf Grundlage weniger tief gehender Definitionen als im Mikrozensus. Zu erwarten wäre daher gewesen, dass der Zensus die Definition des Mikrozensus übernimmt, um eine einheitliche und differenzierte Erfassung von „Personen mit Migrationshintergrund“ zu gewährleisten.

3. Erhebung des „Migrationshintergrundes“ im Zensus

Welche einzelnen Merkmale im Zensus erhoben werden, wurde anhand des „Zensusdurchführungsgesetzes“ festgelegt. Den Mindestrahmen bildete das in der EU-Zensus-Verordnung vorgegebene Erhebungsprogramm, wodurch eine EU-weite Vergleichbarkeit der Volkszählungsergebnisse gewährleistet werden soll. Darüber hinaus konnten vom nationalen Gesetzgeber zusätzliche Erhebungsmerkmale bestimmt werden. So wurde der „Migrationshintergrund“ durch weitere Angaben differenzierter als ursprünglich erfasst.

Im Rahmen des Zensus erfolgt die Bestimmung des Migrationshintergrundes über die Haushaltsstichprobe und über das Melderegister. Sowohl in der Haushaltsstichprobe als auch im Melderegister werden die Personen mit Migrationshintergrund abweichend zum Mikrozensus definiert.

Werden Haushaltsstichprobe und Melderegister mit dem Mikrozensus verglichen, dann zeigen sich – wie Kreuzmair (2012) in einer Übersicht darlegt – mehrere Abweichungen:

- Im Mikrozensus wird das Merkmal „alle nach 1949 auf das heutige Gebiet der Bundesrepublik Deutschland Zugewanderten“ verwendet, während in der Haushaltsstichprobe die Zuwanderung erst nach 1955 erfasst wird (Frage 14). Mit dieser Jahresabgrenzung werden Ausländer, die vor 1955 migriert sind, unterschätzt. Warum der Gesetzgeber dieses Jahr gewählt hat, ist nicht kommentiert. In der Beschlussempfehlung des Bundestag-Innenausschusses (2009) wird das Jahr 1955 ohne weitere Erläuterung genannt, obwohl der Bundesrat (2009) mit Blick auf vergleichende Auswertungen von Zensus und Mikrozensus dafür plädiert hat, dass als Zuwanderung im Sinne der statistischen Erfassung das Datum „1. Januar 1950“ gelten soll.

- Zweitens können im Zensus Deutsche mit eigener Migrationserfahrung wie Spätaussiedler und Eingebürgerte über die Haushaltsstichprobe nicht und über das Melderegister kaum differenziert werden, sondern nur als Gruppe insgesamt ausgewiesen werden.

- Einschränkungen ergeben sich im Zensus auch bei Deutschen ohne eigene Migrationserfahrung. So sind in der Haushaltsstichprobe Kinder von in Deutschland geborenen Ausländern und Eingebürgerten nicht erfassbar. Es wird zwar erhoben, ob die Eltern zugezogen sind oder nicht, nicht jedoch, ob sie Ausländer oder Eingebürgerte sind (Fragen 17-22).

Vorteile ergeben sich im Zensus aber hinsichtlich der regionalen Auswertungsmöglichkeiten des Migrationshintergrundes. So können über die Haushaltsstichprobe Gemeinden mit 10.000 Einwohnern ausgewertet werden und über das Melderegister sogar alle beliebig kleinen regionalen Einheiten, während im Mikrozensus aufgrund der geringeren Stichprobengröße (1% vs. 10% der Bevölkerung) mehrere Kreise zusammengefasst werden müssen.

4. Erhebung der „Religionszugehörigkeit“ im Zensus

Neben einer differenzierteren Erfassung des „Migrationshintergrundes“ wurde vom Gesetzgeber auch entschieden, das Merkmal „Religionszugehörigkeit“ ins Erhebungsprogramm aufzunehmen. Hierfür hatten u.a. die christlichen Kirchen, der Bundesrat und die Wissenschaft votiert.

Die Frage nach der Zugehörigkeit zu einer öffentlich-rechtlichen Religionsgemeinschaft wurde zu einer Zensus-Pflichtfrage (Frage 7 der Haushaltsstichprobe). Die Filterantwort „keine“ führt zudem zu der freiwillig zu beantwortenden Frage nach dem Religionsbekenntnis (Frage 8).

Bei den bisherigen Volkszählungen in der Bundesrepublik Deutschland wurde aus mehreren Gründen immer das Merkmal der Religionszugehörigkeit erhoben:

- Ergebnisse zur Religionszugehörigkeit stellen zentrale Informationen für Gesetzgebung, Verwaltung und Planung von Bund und Ländern dar, da sie im Zusammenhang mit anderen Erhebungsmerkmalen Informationen über den Einfluss der Religionszugehörigkeit auf demographische, wirtschaftliche und soziale Tatbestände erlauben (Bundesrat 2009).

- Von Seiten der Evangelischen Kirche wurde eingehend betont, dass die Bedeutung der religiösen Bindung für gesellschaftliche Prozesse und das Zusammenleben der Menschen in einer offenen Gesellschaft in letzter Zeit verstärkt diskutiert wird. Der Zensus bietet daher die Möglichkeit, „auch im Hinblick auf die religiöse Zusammensetzung der Bevölkerung wieder zu einem belastbaren Zahlenmaterial zu kommen“ (Deutscher Bundestag Drucksache 16/12219, 2009:14).

Indem aber die Beantwortung der Fragen nach dem Religionsbekenntnis teilweise freiwillig ist, sind die Informationen zur religiösen Zusammensetzung der Bevölkerung eingeschränkt. Ursache hierfür ist das Grundgesetz, wonach niemand verpflichtet ist, seinen religiösen Glauben mitzuteilen.

Mit der Frage nach der Religionsgesellschaft werden nur die Personen sicher erfasst, die einer öffentlich-rechtlichen Religionsgesellschaft angehören. Hierzu zählen christliche Kirchen und jüdische Gemeinden, nicht jedoch Angehörige einer islamischen Religionsgesellschaft (Sunniten, Schiiten und Aleviten) sowie Buddhisten und Hindus. Diese werden nur erfasst, wenn sie die freiwillige Nachfrage zum Religionsbekenntnis beantworten.

Auch in Zukunft sind daher Studien wie „Muslimisches Leben in Deutschland“ (Haug et al. 2009) notwendig, um Zahl und Anteil der Muslime begründet zu schätzen. Folgt man dieser Studie, dann leben derzeit zwischen 3,8 und 4,3 Mio. Muslime in Deutschland. Dies entspricht einem Anteil an der Bevölkerung von 4,6% bis 5,2%. Aus der Studie „Islamisches Gemeindeleben in Deutschland“ geht zudem hervor, dass die Gemeindelandschaft deutlich sunnitisch dominiert ist (Halm et al. 2012).

5. Fazit

Der Zensus kann grundsätzlich als ein zentrales Instrument zur Verbesserung der Erkenntnislage im Migrations- und Integrationsbereich angesehen werden. Gleichwohl lassen sich aufgrund des begrenzten Erhebungsprogramms und der erwähnten Einschränkungen bei den Merkmalen „Migrationshintergrund“ und „Religionszugehörigkeit“ nicht alle Erkenntnisdefizite beheben. Vor diesem Hintergrund wird der Mikrozensus seine zentrale Bedeutung als amtliche Datenquelle für die Migrations- und Integrationsforschung behalten. Darüber hinaus werden Studien der empirischen Sozialforschung weiterhin notwendig sein, um vertiefende Erkenntnisse zu gewinnen. Dies gilt nicht zuletzt für den Zusammenhang von Religion bzw. Islam und Integration bzw. Arbeitsmarktintegration, der vielfach überbewertet wird (Stichs/Müssig).

Prof. Uwe Blien:

„Der Zensus aus Sicht der Arbeitsmarkt- und Berufsforschung“

Zur Person:

Prof. Uwe Blien ist Leiter des Forschungsbereichs Regionale Arbeitsmärkte am Institut für Arbeitsmarkt- und Berufsforschung (IAB) in Nürnberg, für das er seit 1990 tätig ist. Zuvor war er wissenschaftlicher Mitarbeiter an den Universitäten Erlangen-Nürnberg und Regensburg sowie am Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim. 1999 erlangte er die Habilitation und erhielt die Venia Legendi für Volkswirtschaftslehre an der Universität Kaiserslautern, wo er 2007 zum apl. Professor berufen wurde. Seit 2008 ist er Inhaber des Lehrstuhls für Arbeitsmarkt- und Regionalforschung an der Otto-Friedrich-Universität Bamberg und seit 2010 Vorsitzender der „Gesellschaft für Regionalforschung“ (GfR), der deutschsprachigen Sektion der European Regional Science Association (ERSA). Seine Forschungsinteressen betreffen u. a. regionale Arbeitsmärkte, die Anwendung von Wirtschafts- und Arbeitsmarktpolitik, die ökonometrische Anwendung von Mehrebenenmodellen, die Arbeitsmarktwirkungen des Strukturwandels und die Entwicklung sozialer Normen und Institutionen.

Vortragsfolien:

Zensus eine einmalige Datenquelle für die Arbeitsmarkt- und Berufsforschung

- Umfassende Datenquelle
- Differenzierung in großer Tiefe möglich

Arbeitsmarkt- und Berufsforschung beschäftigt sich mit dem ökonomischen Erfolg und Mißerfolg des größten Teils Bevölkerung. Dementsprechend wichtig ist die Verfügbarkeit aussagefähiger Datenquellen.

Nachteile des Zensus

- Keine Totalerhebung
- Kein Einkommen

Nutzen des Zensus für die Berufsforschung

- Auswertung der Beschäftigungsstatistik bereits möglich, jedoch nur für die sozialversicherungspflichtig Beschäftigten
- Mit Zensus Inklusion der Selbständigen, geringfügig Beschäftigten und Beamten
- Realistische Analyse des Bestands in vielen Berufen (z. B. zum Thema „Fachkräftebedarf“)
- Wegen der Vielzahl der Berufe ist eine differenzierte Datenquelle notwendig! (Selbst der Mikrozensus versagt)

4

Nutzen des Zensus für die Berufsforschung (II)

- Der Übergang in die Selbständigkeit ist für viele Berufe eine realistische Möglichkeit, jedoch mit der Beschäftigungs-statistik nicht zu erfassen.
- Dadurch kann ein wichtiger Zu- und Abstrom zum / vom Arbeitsmarkt nicht ohne Zensus erfasst werden.

5

Zensus wichtig für die (Forschung über) Arbeitsmarktpolitik

- Die Zusammenordnung der Einzeldaten in Haushalte erlaubt eine Abschätzung der Bedeutung von Unterstützungs-leistungen.

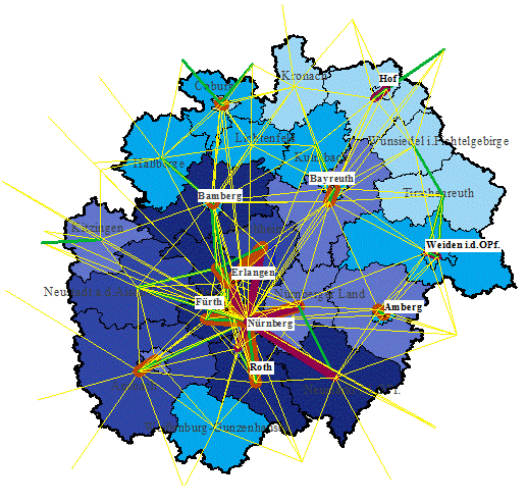
Nutzen des Zensus für die regionale Arbeitsmarktforschung/ Regionalforschung

- Auswertung der Beschäftigungsstatistik bereits möglich, jedoch nur für die sozialversicherungspflichtig Beschäftigten
- Mit Zensus Inklusion der Selbständigen, geringfügig Beschäftigten und Beamten
- Auswertbar auf der Ebene von „Blocks“
- Ermöglicht die Analyse der Konzentration wirtschaftlicher Aktivität (Agglomerationseffekte, Dispersion etc.)

Zensus wichtig für die (Forschung über) (Berufs-)Pendlerverflechtungen

- Die Beschäftigungsstatistik enthält Angaben zum Arbeitsort und zum Wohnort für sozialversicherungspflichtig Beschäftigte.
- Im Zensus sind diese Angaben auch für andere Erwerbstätige verfügbar.
- Daraus können dann Pendlerverflechtungen konstruiert werden.
- Die gewonnene Information ist für die Verkehrsplanung wichtig.
- Sie ist aber auch zur Analyse der Arbeitsmarktverflechtungen und von Entlastungen bedeutsam.

8



Pendelverflechtungen in der Metropolregion Nürnberg 30 Juni 2010

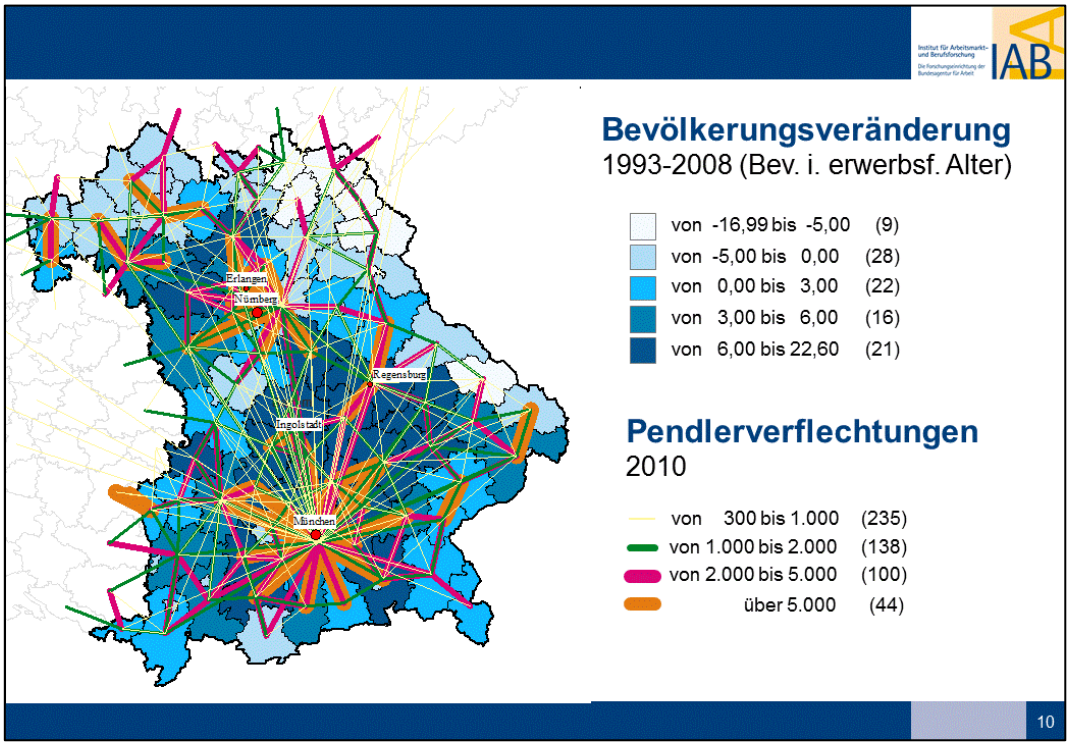
- von 200 bis 2.000
- von 2.000 bis 4.000
- von 4.000 bis 10.000
- ab 10.000

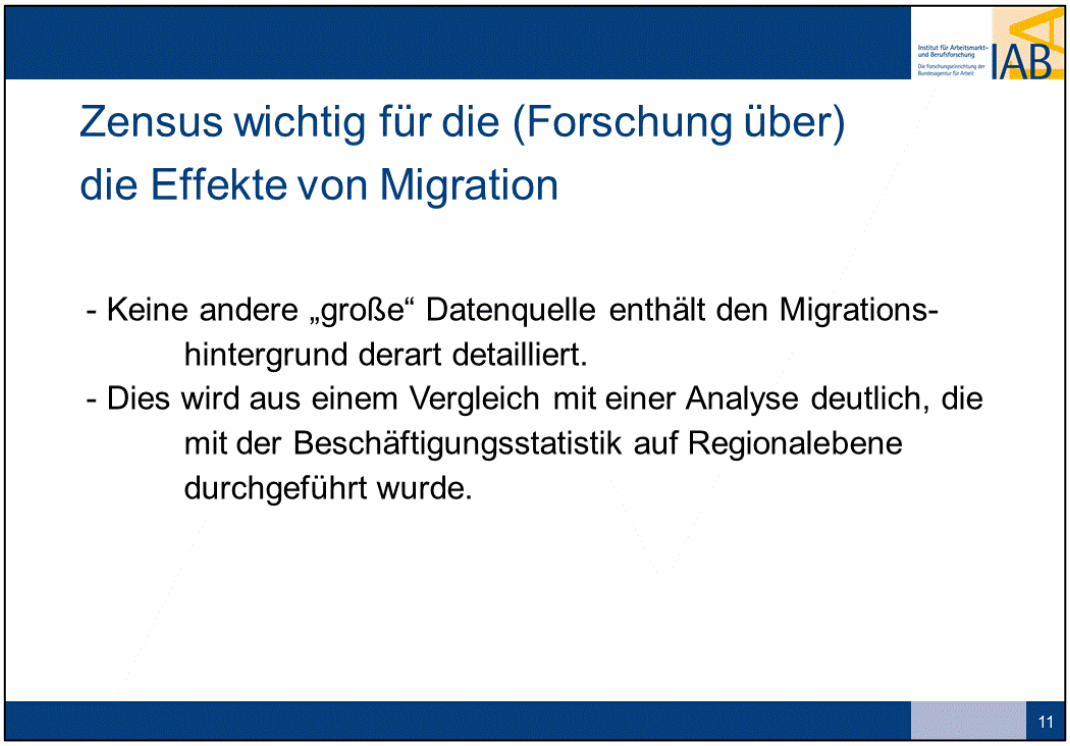
Bevölkerungsveränderung
1993 - 2010

von -14,13 bis -5,00	(6)
von -5,01 bis 0,00	(10)
von 0,01 bis 3,00	(6)
von 3,01 bis 6,00	(5)
von 6,01 bis 9,60	(6)

Min: -14,13% (Wunsiedel i.Fichtelgebirge)
Max: 9,6% (Fürth)

9


Institut für Arbeitsmarkt- und Berufsforschung
IAB
Erlangen
Nürnberg
Regensburg
Ingolstadt
München
Bevölkerungsveränderung
1993-2008 (Bev. i. erwerbsf. Alter)
von -16,99 bis -5,00 (9)
von -5,00 bis 0,00 (28)
von 0,00 bis 3,00 (22)
von 3,00 bis 6,00 (16)
von 6,00 bis 22,60 (21)
Pendlerverflechtungen
2010
von 300 bis 1.000 (235)
von 1.000 bis 2.000 (138)
von 2.000 bis 5.000 (100)
über 5.000 (44)
10


Institut für Arbeitsmarkt- und Berufsforschung
IAB
Zensus wichtig für die (Forschung über) die Effekte von Migration
- Keine andere „große“ Datenquelle enthält den Migrations-hintergrund derart detailliert.
- Dies wird aus einem Vergleich mit einer Analyse deutlich, die mit der Beschäftigungsstatistik auf Regionalebene durchgeführt wurde.
11

Effekte der kulturellen Diversität auf die Beschäftigung und die Löhne von Deutschen (Modell mit räumlicher Autokorrelation)

Beobachtungseinheiten: Kreise Westdeutschlands	**Löhne**	**Beschäftigung**
Anteil der hochqualifizierten Ausländer	0.0849*** (.025)	0.1988*** (.069)
Diversitätsindex der hochqualifizierten Ausländer	0.0082 (.005)	0.0141 (.014)
Anteil der nicht formal qualifizierten Ausländer	- 0.0446*** (.013)	- 0.1588*** (.039)
Diversitätsindex der nicht formal qualifizierten Ausländer	0.0187*** (.006)	0.0271 (.017)
Andere Kontrollvariablen	ja	ja
Fixe Effekte	ja	ja

12

Vorteil des Zensus bei einer Analyse des beschriebenen Typs

- „Kulturelle Diversität“ kann über den Migrationshintergrund gemessen werden (und nicht über die Staatsbürgerschaft wie mit der Beschäftigungsstatistik).
- Dies erlaubt eine bessere Abschätzung der Migrationseffekte auf Beschäftigung und Einkommen.

(keine Verwendung von Einzeldaten notwendig, da Regionalanalyse)

13

Fazit: Der Zensus ist für viele Analysen unbedingt notwendig

Ausländer, Berufe und Regionen sind differenziert zu erfassen und dies ist nur mit dem Zensus möglich