

DAS DEFECT-PROJEKT: SAMPLING-ERRORS UND NONSAMPLING- ERRORS IN KOMPLEXEN BEVÖLKERUNGSSTICHPROBEN

RAINER SCHNELL UND FRAUKE KREUTER

Einleitung

Die in der empirischen Sozialforschung verwendeten Datensätze beruhen fast ausnahmslos auf mehrstufigen Klumpenauswahlen. Soziale Prozesse (wie z.B. Alterssegregation) und die Tätigkeit der Interviewer führen zu relativer Homogenität der ausgewählten Klumpen. Beide Mechanismen bedingen die Unterschätzung der Standardfehler und damit die Verfälschung aller Konfidenzintervalle und aller Signifikanztests aller Untersuchungen, die auf Bevölkerungsstichproben beruhen. In der Regel werden diese Effekte in der Forschungspraxis einfach ignoriert. In der methodischen Forschung entsteht hingegen das Problem, daß die Effekte, die durch die Interviewer bedingt werden, nicht von den Effekten, die durch die Sampling-Points bedingt werden, getrennt werden können. Um diese Effekte in einer realistischen Bevölkerungsstichprobe der gesamten Bundesrepublik trennen zu können, wurde durch unsere Arbeitsgruppe das DEFECT-Projekt durchgeführt¹.

Das Hauptziel des Projekts ist die empirische Schätzung des Effekts, der durch die räumliche Klumpung der Untersuchungseinheiten hervorgerufen wird. Solche Effekte werden in der statistischen Literatur als „Design-Effect“ bezeichnet (Kish 1995). Die Studie erlaubt für die Bundesrepublik erstmalig die Berechnung dieser Effekte. Weiterhin gibt es nur sehr wenige Studien, die den Effekt der Durchführung einer Befragung durch verschiedene Institute untersuchen. Auch dies sollte durch das Design des DEFECT-Projekts ermöglicht werden. Ebenso sollten mögliche Effekte durch ver-

¹ Das DEFECT-Projekt wurde mit Mitteln der Deutschen Forschungsgemeinschaft (Projektnummer SCHN 586/2-1) finanziert.

schiedene Erhebungsmodi („face-to-face“, postalische und telefonische) Befragung untersucht werden. Zusätzlich sollte das Projekt den ersten systematischen Vergleich einer bundesweiten Quoten-Studie mit bundesweiten Zufallsstichproben seit mehr als vierzig Jahren ermöglichen. Schließlich lag für die Bundesrepublik bislang kein Datensatz vor, der die vollständigen Kontaktprotokolle auch für die Nonrespondenten einer Befragung enthielt. Damit waren bislang weder Anwendungen potentieller Korrekturtechniken noch Untersuchungen über das Interviewerkontaktverhalten mit BRD-Daten möglich.

Mit den Daten des DEFECT-Projekts liegt nun ein Datensatz vor, der die Untersuchung der genannten Sampling-Errors und Nonsampling-Errors für bundesweite Befragungen kommerzieller Datenerhebungsinstitute erlaubt.

Design

Das methodische Ziel der empirischen Abschätzung der Standardfehler komplexer Bevölkerungstichproben erforderte die Implementierung eines speziellen Stichprobendesigns, so genannter „interpenetrierter Stichproben“ (Bailar 1983). Bei interpenetrierten Stichproben werden innerhalb der selben Sampling-Points unabhängig arbeitende Interviewer eingesetzt. In der vorliegenden Studie haben in jedem Sampling-Point zwei Institute unabhängig voneinander „face-to-face“-Befragungen bei vorgegebenen Haushalten durchgeführt. Jeder Interviewer hat dabei in genau einem Sampling-Point gearbeitet².

Um Vergleiche zwischen diesen Randomstichproben und einer Quoten-Auswahl zu ermöglichen, hat ein drittes Institut in den gleichen Sampling-Points mit Quoten-Auswahl eine weitere „face-to-face“-Befragung durchgeführt³. Zum Vergleich der Erhebungsmodi wurde in den gleichen Sampling-Points eine CATI-Befragung durch ein viertes Institut durchgeführt. Schließlich wurde von den Mitarbeitern des DEFECT-Projekts eine postalische Befragung in den gleichen Sampling-Points erhoben.

Die Vorbereitungen zur Datenerhebung und die eigentliche Datenerhebung der DEFECT-Hauptstudie erfolgten zwischen August 1999 und April 2000. Alle Erhebungen fanden in den gleichen Sampling-Points, zum gleichen Zeitpunkt mit einem

² Allerdings muß erwähnt werden, daß es nicht möglich war in allen Sampling-Points pro Institut nur einen Interviewer einzusetzen.

³ Ursprünglich sollte die Quoten-Erhebung durch das Institut für Demoskopie Allensbach durchgeführt werden. Leider zog das Institut im Juni 1999 sein Angebot zur Zusammenarbeit zurück.

identischen Fragebogen statt und wurden von verschiedenen voneinander unabhängigen Instituten durchgeführt⁴. Beteiligt waren ACADEMIC DATA Gesellschaft für Umfragen, Methodenberatung und Analysen mbH (Essen), foerster & thelen Marktforschung Feldservice GmbH (Bochum), infas Institut für angewandte Sozialwissenschaft GmbH (Bonn) und INRA Deutschland Gesellschaft für Markt- und Sozialforschung mbH (Möln).

Ziehung der Zufallsstichproben

Die Stichprobe für die Random-Erhebungen wurde nach dem ADM-Design (ADM 1999) gezogen und umfaßt 160 Sampling-Points⁵. Die 160 Sampling-Points wurden aus dem ADM-Mastersample basierend auf der Datei des Bundeswahlleiters gezogen. Die Ziehung erfolgte nach Absprache mit der Arbeitsgemeinschaft ADM-Stichproben durch die Firma BIK (Hamburg)⁶. Bei der Ziehung wurden nur Stichprobennetze verwendet, die noch nie an Institute vergeben worden waren, so daß das Risiko einer Mehrfachbefragung in den Sampling-Points minimiert werden konnte.

Der zweite Schritt der Stichprobenziehung bestand aus der Auswahl der Zielhaushalte. Diese erfolgte mit Hilfe eines so genannten „Address-Random“-Verfahrens (Behrens/Löffler 1999: 81). Die Adressensammlung in den 160 Sampling-Points erfolgte unter anderem aus datenschutztechnischen Gründen durch insgesamt acht Mitarbeiterinnen des Projekts an der Universität Konstanz⁷. Dabei wurden in jedem Sampling-Point 110 Adressen notiert, wobei jede dritte Klingel ausgehend von einem Startpunkt aufgelistet wurde. Der Startpunkt wurde entweder den von BIK erstellten Begehungsunterlagen der Sampling-Points entnommen oder in den wenigen sehr kleinen Gemeinden von den Begeherinnen ausgewählt. Zur Auswahl standen dazu zentral gelegene öffentliche Gebäude, in der Regel das Rathaus oder andere öffentliche

⁴ Allen beteiligten Instituten sei an dieser Stelle für ihr außergewöhnlich großes Engagement im Rahmen des DEFECT-Projekts gedankt.

⁵ Die Zahl von 160 Sampling-Points entspricht in ihrer Größenordnung derer des ALLBUS 1994 und 1996 sowie der des Deming-Plans der früheren DIVO-Stichproben (Schnell 1997a: 58-59).

⁶ Für die Genehmigung der Durchführung einer Sonderstichprobe aus dem ADM-Mastersample bedanken wir uns herzlich bei der Arbeitsgemeinschaft der ADM-Stichproben und besonders bei deren Sprecher Christian von der Heyde.

⁷ Wäre die Stichprobe der Zielhaushalte durch eines der beteiligten Institute erfolgt, hätten die Adressen nicht an die Forschungsgruppe oder unabhängig arbeitende Institute weitergegeben werden können.

Gebäude, die als Wahllokal des Stimmbezirks dienen. Die Haushaltsadressen und Namen der Zielhaushalte wurden mit Hilfe eines Handheld-Computers (Psion 6) bereits bei der Begehung maschinell erfaßt und fast täglich über Modem an die Projektleitung übermittelt. Um die Begehung einheitlich durchzuführen, waren die Begeherinnen angehalten, bei jeder aufkommenden Unklarheit über Handy Kontakt zur Projektleitung aufzunehmen.

Während der Begehung wurden alle Straßen, in denen der Random-Walk erfolgte, von den Projektmitarbeiterinnen entweder gefilmt oder fotografiert. Weiterhin wurde von den Begeherinnen für jede Adresse ein Beobachtungsbogen über Sicherheitseinrichtungen (Alarmanlagen etc.), Gebäudeklassifikation und Umgebungsmerkmale (Wald, Graffiti, etc.) ausgefüllt. Die Begehung der Sampling-Points fand im Zeitraum vom 16.08.1999 bis zum 29.09.1999 statt.

Die gesammelten Adressen wurden manuell und maschinell mit Hilfe von AWK-Makros (Schnell 1997b) um offensichtliche Fehler bereinigt. Anschließend wurden die Adressen ebenfalls maschinell mit AWK-Makros zufällig auf fünf unabhängige Stichproben verteilt. Drei dieser Stichproben wurden an die Institute verteilt, mit einer weiteren wurde die postalische Befragung durchgeführt. Die fünfte Stichprobe diente als Reserve, falls die gelieferten Adressen den Instituten nicht ausreichen sollten, um die vereinbarten acht Interviews pro Sampling-Point zu realisieren.

Als Zielpersonen kamen alle Personen in Frage, die zur deutschsprachigen Wohnbevölkerung ab 18 Jahren gehörten und in Privathaushalten lebten. Die Auswahl der Zielperson innerhalb der Haushalte erfolgte in den „face-to-face“-Erhebungen über einen „Schwedenschlüssel“, in der telefonischen Befragung wurde diejenige Person ausgewählt, welche als letztes Geburtstag hatte. Auch in der postalischen Befragung waren die Personen im Haushalt angehalten, den Fragebogen an diejenige Person über 18 Jahren weiterzugeben, welche als letztes Geburtstag hatte⁸.

Ziehung der Quoten-Stichprobe

Ausgehend von den aus dem ADM-Mastersample gezogenen 160 Sampling-Points wurde für die Quoten-Erhebung eine Parallelstichprobe gezogen. Dabei wurde für alle Sampling-Points, die in Gemeinden mit mehr als 10.000 Einwohnern lagen, der

⁸ Da bislang für die Bundesrepublik anscheinend keine Ergebnisse einer bundesweiten postalischen Zufallsstichprobe veröffentlicht wurden, lagen keine Erfahrungswerte über die Praktikabilität dieses Vorgehens vor.

Stimmbezirk in derselben Gemeinde mit einer um eins verminderten Stimmbezirksnummer gezogen. Für alle kleineren Sampling-Points wurde die nächstliegende Nachbargemeinde mit gleicher Ortsgröße und gleichem Regierungsbezirk ausgewählt.

Die Auswahl der Zielpersonen innerhalb der Sampling-Points erfolgte nach einer Kombination der Variablen Geschlecht, Altersgruppe und Berufstätigkeit sowie einer Kombination der Variablen Berufsgruppe und Gemeindegrößenklasse. Als Altersgruppen wurden die Gruppen 18-29, 30-44, 45-59 sowie 60 und mehr Jahre verwendet. Als berufstätig galten Erwerbstätige mit mindestens einer Stunde bezahlter Arbeit einschließlich Studenten und Auszubildende.

Die Berufsgruppen waren

- Landarbeiter, Bauern und mithelfende Familienangehörige,
- Arbeiter,
- Angestellte und Beamte,
- Selbständige und mithelfende Familienangehörige und
- Sonstige.

Als Ortsgrößenklassen wurden die Kategorien bis 5.000 Einwohner, 5.000-20.000, 20.000-100.000, 100.000-500.000 sowie 500.000 und mehr verwendet.

Zum Einsatz kamen zwei Quotentabellen. In der ersten Quotentabelle wurden die Variablen Geschlecht*Altersgruppe* Berufstätigkeit kombiniert. Diese Tabelle wurde für jedes Bundesland aus den Daten des Mikrozensus 1995 berechnet. In der zweiten Quotentabelle wurden die Variablen Berufsgruppe*Gemeindegrößenklasse kombiniert. Die Daten dieser Tabelle wurden vom Institut über alle Bundesländer aus den Daten einer großen ADM-Stichprobe berechnet. Von der DEFECT-Projektgruppe wurden ausgehend von der ersten Quotentabelle für jeden Sampling-Point individuelle Soll-Vorgaben berechnet. Ausgehend von der zweiten Quotentabelle wurden dem Institut zusätzlich die Soll-Vorgaben dieser Tabelle pro Bundesland vorgeschrieben⁹.

⁹ Hierzu wurde unter Verwendung der von Sudman (1976: 198) empfohlenen Rundung ein iteratives Optimierungsprogramm geschrieben, das die Abweichungen zwischen den kumulierten Soll-Vorgaben pro Point und den Verteilungen der beiden Quotentabellen minimiert. Interessanterweise sind solche Details der Vorgehensweise bei Quotenstichproben in der gesamten Literatur nicht dokumentiert.

Realisierung der Stichproben

Pro Stichprobe wurde für jeden der 160 Sampling-Points in einem ersten Schritt 16 Adressen vergeben. Mit diesen sollten für jede Stichprobe acht Interviews pro Sampling-Point erreicht werden, also insgesamt $n=1.280$ realisierte Fälle pro Stichprobe. Um den neutralen Ausfällen¹⁰ gerecht zu werden und die Chance auf eine für die Analysen ausreichend große Fallzahl pro Sampling-Point zu erhöhen, wurden zunächst pro Sampling-Point vier Ersatzadressen zur Verfügung gestellt und bei den „face-to-face“-Erhebungen im Rahmen einer Nachbearbeitung später weitere Adressen vergeben. Der sukzessive Einsatz von Adressen war notwendig, da sonst aus dem gegebenen Adressen-Pool lediglich die gut erreichbaren und kooperativen Personen befragt worden wären, bis das vorgegebene Maximum von acht Interviews realisiert worden wäre.

Es zeigte sich rasch, daß in der ursprünglich vorgesehenen Erhebungszeit nur ein Teil der beabsichtigten Interviews realisiert werden konnte. In allen „face-to-face“-Erhebungen einschließlich der Quoten-Erhebung waren Nachbearbeitungsphasen notwendig¹¹. Für die beiden Random-„face-to-face“-Erhebungen wurden nach Beendigung der Nachbearbeitung telefonische Nonresponse-Studien mit drei- bis vierwöchiger Feldzeit durchgeführt¹².

Eine ähnliche telefonische Nachbefragung fand für die Nonrespondenten der postalischen Erhebung statt. Da der größte Teil der Nonrespondenten der CATI-Erhebung bereits während der Feldzeit der Hauptstudie erneut kontaktiert wurde, gab es keine gesonderte Nonresponse-Studie für die CATI-Erhebung¹³. Insgesamt ergeben sich damit die mittlerweile üblichen ausgedehnten Feldzeiten (vgl. Tabelle 1).

¹⁰ Straße oder Hausnummer gibt es nicht, Wohnung nicht bewohnt, Haushalt unbekannt, Haushalt hat neue Adresse, keiner über 18, keiner spricht deutsch.

¹¹ Die Nachbearbeitung in der Quoten-Studie wurde dadurch notwendig, daß 23 Interviewer irrtümlich in mehr als einem Sampling-Point eingesetzt wurden. Dadurch wurden 17 Points teilweise, 16 Point vollständig bearbeitet und 10 Points vollkommen neu ausgewählt und bearbeitet.

¹² Aus Datenschutzgründen mußte die Nonresponse-Studie jeweils durch das gleiche Institut durchgeführt werden, das auch die Haupterhebung durchführte: Ansonsten hätten personenbezogene Daten zwischen den Instituten ausgetauscht werden müssen, was leider nicht möglich ist.

¹³ Aus den üblichen Gründen wurden explizite Verweigerer in keiner Nonresponse-Studie erneut kontaktiert.

Tabelle 1: Feldzeiten der Erhebungen

	Haupterhebung	Nachbearbeitungsphase	Nonresponse-Studie
„face-to-face“ I	25.10.99 – 29.02.00	Ab erster KW 2000	28.02.00 – 30.03.00
„face-to-face“ II	15.10.99 – 28.02.00	28.01.00 – 23.02.00	02.03.00 – 19.03.00
CATI-Erhebung	19.10.99 – 14.12.00		
Postalische Erhebung	25.10.99 – 22.12.99 (99% des Rücklaufs)		13.01.00 – 19.01.00 03.02.00 – 04.02.00 02.03.00 – 18.03.00
Quoten-Erhebung	21.10.99 – 31.12.99	09.03.00 – 07.04.00	

Unter den genannten Rahmenbedingungen gelang es den beiden Random-„face-to-face“-Instituten die Ergebnisse der Tabelle 2 zu erzielen¹⁴.

Für die telefonische Befragung wurden ebenfalls 20 Adressen pro Sampling-Point vergeben. Die insgesamt 3.200 Adressen wurden manuell in einer Telefon-CD-ROM gesucht¹⁵. Insgesamt konnten 2.012 Telefonnummern den Adressen zugeordnet werden (62.9%).

Um Vergleiche zwischen nicht gelisteten und gelisteten Telefonnummern durchführen zu können und um die notwendige Fallzahl von mindestens 8 Interviews pro Sampling-Point zu realisieren, wurden zusätzlich insgesamt 4.465 „Randomized Last Digit“-Nummern generiert und dem Institut übergeben¹⁶. In den Sampling-Points wurden die RLD-Nummern erst dann bearbeitet, wenn die Nummern aus den durch die Begehung gesammelten Adressen weitgehend abgearbeitet waren. Somit wurden in einigen Sampling-Points keine RLD-Nummern verwendet. Bei jeder bearbeiteten Telefonnummer wurden bis zu 12 Kontaktversuche unternommen¹⁷.

¹⁴ Hier muß wie bei allen hier berichteten Ergebnissen darauf hingewiesen werden, daß es sich um vorläufige Ergebnisse handelt. Die Datenaufbereitung der Kontaktprotokolle erwies sich als überaus aufwendig, so daß mit endgültigen Ergebnissen erst im Frühjahr 2001 gerechnet werden kann.

¹⁵ Von den 3.200 Adressen enthielten 149 keinen Haushaltsnamen, diesen konnte keine Telefonnummer zugeordnet werden.

¹⁶ Die Stichprobenziehung erfolgte durch AWK-Macros (Schnell 1997b).

¹⁷ Aufgrund eines Programm-Fehlers des Herstellers der CATI-Software (Ci3) wurden nur die Kontaktversuche 1-4 und der jeweilige letzte Kontaktversuch protokolliert.

Tabelle 2: Ausschöpfungsstatistik der beiden Random-“face-to-face“-Erhebungen

	N	%	N	%
Delieferte Adressen ¹⁸	4889	100,0	3868	100,0
Nicht verwendete Adressen	1231	25,2	445	11,5
Brutto-Stichprobe	3658	100,0	3423	100,0
Neutraler Ausfall	232	6,4	230	6,7
Bereinigte Brutto-Stichprobe	3426	100,0	3193	100,0
Haushalt nicht erreichbar	582	17,0	735	23,0
Zielperson nicht erreichbar oder befragbar	88	2,6	79	2,5
Termin vereinbart	-	-	11	0,3
Kontaktperson verweigert	1161	33,9	537	16,8
Zielperson verweigert	241	7,0	423	13,2
Interview abgebrochen	2	0,1	7	0,2
Interview ungültig	4	0,1	71	2,2
Status unklar	3	0,1	4	0,1
Interview erfolgreich durchgeführt	1345	39,3	1326	41,5

Zwischen dem 19.10.1999 und dem 14.12.1999 wurden insgesamt 6439 Telefonnummern verwendet, von diesen waren 68,9% RLD-Nummern¹⁹. Nach Abzug aller neutralen Ausfälle stand ein bereinigtes Brutto von 4.592 Telefonnummern zur Verfügung. Mit diesen Telefonnummern wurden 1.350 Interviews realisiert (vgl. Tabelle 3).

Die niedrigere Ausschöpfung in der CATI-Erhebung gegenüber den anderen Random-Erhebungen kann vermutlich auf die deutlich kürzere Feldzeit der CATI-Erhebung zurückgeführt werden. Die effektive Feldzeit der CATI-Erhebung dauerte lediglich fünf Wochen. In den ersten beiden Wochen der Feldzeit wurden ausschließlich Telefonnummern der Adressenstichprobe bearbeitet. In den letzten drei Wochen der Feldzeit wurden die verbliebenen Nummern aus der Adressenstichprobe und die RLD-Nummern bearbeitet. Vom 03.11.99 bis 18.11.99 wurden keine Interviews durchgeführt, da in diesem Zeitraum die Zahl der zur Verfügung stehenden Point-Adressen nur noch so gering war, daß keine systematische Bearbeitung möglich war. In

¹⁸ Die unterschiedliche Zahl von Adressen erklärt sich durch die unterschiedlichen Strategien der Institute während der Nachbearbeitungsphase.

¹⁹ Ein weiterer Programm-Fehler innerhalb von Ci3 führte zum Verlust von zwei Kontaktprotokollen, so daß sich für zwei Interviews nicht mehr nachvollziehen läßt, ob es sich um RLD-Nummern handelte oder um eine gelieferte Adresse.

diesem Zeitraum wurde die zum weiteren Einsatz vorgesehenen RLD-Stichprobe erstellt²⁰.

Tabelle 3: Ausschöpfungsstatistik der telefonischen Befragung

	N	%
Gelieferte Adressen	6477	100,0
Nicht verwendete Adressen	38	0,6
Brutto-Stichprobe	6439	100,0
Neutraler Ausfall	1847	28,7
Bereinigte Brutto-Stichprobe	4592	100,0
Haushalt nicht erreichbar	788	17,2
Anschluß besetzt	92	2,0
Anrufbeantworter	326	7,0
Zielperson nicht erreichbar	111	2,4
Zielperson nicht befragbar	360	7,8
Rückruf vereinbart (außerhalb der Feldzeit)	137	3,0
Kontaktperson verweigert	902	19,6
Zielperson verweigert	451	9,8
Interview abgebrochen	42	0,9
Technischer Abbruch	33	0,7
Interview erfolgreich durchgeführt	1350	29,4

Auch der postalischen Befragung wurden für jeden Sampling-Point 16 Adressen zugewiesen. Davon waren 51 Adressen lediglich Beschreibungen der selektierten Wohnung und konnten nicht angeschrieben werden. Weitere 49 Adressen enthielten ebenfalls keinen Namen aber eine ansonsten vollständige Anschrift. Während alle anderen Briefe an „Familie <<Vorname>> <<Nachname>>“ adressiert waren, wurden in diesen 49 Fällen die Briefe „An die Bewohner“ adressiert, sofern aus den Begehungsunterlagen hervorging, daß es sich um ein Einfamilienhaus handelte. Insgesamt konnten 95% aller versandten Briefe zugestellt werden und gelten somit als kontaktierte Haushalte.

²⁰ Mit der Erstellung der RLD-Nummern konnte erst nach Abschluß der Bereinigung der Adressenstichprobe begonnen werden. Für sämtliche generierten RLD-Nummern wurde dann geprüft, ob eine dieser Nummern bereits in der Adressenstichprobe enthalten war. Natürlich konnte diese Prüfung nur für diejenigen RLD-Nummern vorgenommen werden, die in den Telefonverzeichnissen eingetragen waren. Um zu gewährleisten, daß die „face-to-face“-Interviewer nicht auf Befragte treffen, die bereits im CATI-RLD-Teil befragt wurden, mußten die RLD-Nummern verzögert eingesetzt werden.

Zur Durchführung der postalischen Befragung wurde sorgfältig auf alle prinzipiellen Details von Dillmanns „Total Design Method“ geachtet. Der Fragebogen wurde vor dem Versand durch einen Brief, dem bei 50% der Adressen ein Hochglanzfaltblatt der Fakultät für Verwaltungswissenschaft der Universität Konstanz beigelegt war, angekündigt. Eine Woche danach wurde der Fragebogen verschickt. Eine Woche nach dem Versand des Fragebogens wurde eine Mahnpostkarte verschickt, drei Wochen nach dem Versand des Fragebogens wurde ein Mahnbrief zusammen mit einem neuen Fragebogen verschickt.

Der insgesamt 20 Seiten lange Fragebogen mit 71 Fragen wurde im A4 Format broschiert gedruckt, bei 50% der Adressen wurde stets eine handschriftliche Unterschrift des Projektleiters verwendet. Auf dem Deckblatt des Fragebogens war eine kostenfreie Telefonnummer für Nachfragen vermerkt, zudem enthielten die Anschreiben die Telefonnummern der Projektleitung und der Polizeidirektion Konstanz, die über das Projekt vorab informiert worden war. Alle diese Telefonnummern wurden im Feldverlauf auch einige Male in Anspruch genommen. Insgesamt konnte durch diese Maßnahmen das Ergebnis in Tabelle 4 erzielt werden.

Tabelle 4: Endstatus der schriftlichen Befragung am 16.03.00

	N	%
Gesamtzahl der angeschriebenen Personen	2509	100,0
Adresse ist Ferienhaus	2	0,1
Empfänger unbekannt	85	3,4
Unbekannt verzogen	37	1,5
Adresse ungenügend	2	0,1
Firma erloschen	1	0,0
Empfänger verstorben	14	0,6
Empfänger verzogen – Nachsendeauftrag liegt nicht vor	4	0,2
Bereinigte Bruttostichprobe	2364	100,0
Personen, von denen keine Rückmeldung einging	1095	46,3
Personen, die explizit verweigert hatten	81	3,4
Gesundheitliche Gründe	18	0,8
Abwesend	1	0,0
Sonstiges	8	0,3
Personen, die einen ausgefüllten Fragebogen ohne Rücklaufnummer zurückschickten	9	0,4
Personen, die einen ausgefüllten Fragebogen mit Rücklaufnummer zurückschickten	1152	48,7

Erwartungsgemäß verlief die Feldarbeit der Quotenerhebung im Vergleich zu den anderen Erhebungen relativ unproblematisch. Trotzdem zeigte sich während der Feldarbeit, daß die Quotenvorgaben etwas gelockert werden mußten²¹. In der gesamten Feldzeit wurden 1.276 Interviews entlang der Quotenvorgaben realisiert. Lediglich bei den Berufsgruppen kam es zu nennenswerten Abweichungen von der Sollvorgabe. Dabei wurden bei den „Arbeitern“ 112% der Sollvorgabe befragt, bei den „Selbständigen“ 111% und bei „Angestellten und Beamten“ 95% der ursprünglich vorgegeben Sollvorgabe. Die größte Abweichung ergab sich bei der anscheinend zu ungenau definierten Berufsgruppe „Sonstige“ bei der nur 76% der Sollvorgabe befragt wurden.

Nonrespondenten

Über die Nonrespondenten der Random-Erhebungen liegen in diesem Projekt Informationen aus drei Quellen vor. Zunächst wurden während der Begehung von den Projektmitarbeiterinnen Gebäude- und Umgebungsmerkmale für alle Adressen gesammelt. Weiterhin waren die Interviewer dazu verpflichtet, für jede Adresse ein detailliertes Kontaktprotokoll zu führen. Dieses Protokoll enthält neben Datum, Uhrzeit und Ergebnis des Kontaktes auch den Modus des Kontaktversuchs und ist auch für die Nonrespondenten Bestandteil des Datensatzes („Nonresponse-Records“, sogenannte Brutto-Stichprobe). Schließlich wurden alle Adressen der bereinigten Bruttostichproben, bei denen in der eigentlichen Feldphase oder der Nachbearbeitungsphase kein Interview realisiert werden konnte, noch einmal telefonisch kontaktiert²². Nach Möglichkeit wurde hier ein vollständiges Interview realisiert. War dies nicht möglich, wurde versucht, zwei Fragen zur zentralen abhängigen Variable des Projekts („basic questions“, Kersten/Betlehem 1984) zu stellen.

Interviewerschulung und Einsatz

Die Interviewer wurden von den Instituten, die die „face-to-face“-Studien durchführten, schriftlich geschult. Die Schulung der CATI-Interviewerinnen erfolgte mündlich im Telefonstudio.

²¹ Anfang Dezember wurde mit dem Institut vereinbart, daß die Quotenvorgaben über eine Gemeindengrößenklasse nach oben oder unten variiert werden konnten. Weiterhin wurde den Interviewern erlaubt, etwas über die Pointgrenzen hinauszugehen. Hierbei wurde allerdings ausgeschlossen, daß die Interviewer in den der Random-Erhebungen zugewiesenen Straßen Interviews durchführten.

²² Ausgeschlossen waren selbstverständlich harte Verweigerer und nicht befragbare Personen.

Die im Jahr 1999 eingeführte Regelung zur Verhinderung von Scheinselbständigkeit erhöhte die Gefahr einer Mehrfach­tätigkeit der Interviewer. Für das oben formulierte Projektziel mußte deshalb explizit ausgeschlossen werden, daß die Interviewer für ein weiteres am Projekt beteiligtes Institut Interviews für diese Projekt durchführen. Aus diesem Grund tauschten die Institute Listen der Geburtsdaten und Namens­kürzel der Interviewer aus, um so für jeden Sampling-Point festzustellen, ob einer der Interviewer eventuell bei einem anderen Institut eingesetzt wurde oder werden sollte. Zudem ließen sich die Institute von den Interviewern versichern, daß diese bei der vorliegenden Studie ausschließlich für ein Institut arbeiten.

Erhebungsinstrument

Thema der Erhebung ist Kriminalitätsfurcht. Hier soll die Erhebung dazu führen, die bisher unklaren Meßverfahren zu verbessern und klären, ob einige in der Literatur dokumentierte Anomalien auf Methodeneffekte zurückzuführen sind. Ein weiterer Grund für die Wahl des Themas bestand in der vermuteten ubiquitären Betroffenheit der Population. Zudem erlaubt das Thema eine ausgewogene Mischung von Fakten- und Einstellungsfragen, ohne dabei auf politische Einstellungen zurückgreifen zu müssen.

Die Entwicklung des Erhebungsinstruments begann im Juli 1998, acht Monate vor Genehmigung des Projektes und 15 Monate vor Beginn der Feldzeit. Begonnen wurde mit einigen experimentellen Studien mit Studierenden zur Überprüfung der bisher in der kriminologischen Forschung verwendeten Instrumente. Daran anschließend wurden in der verschiedenen Städten insgesamt 15 qualitative Intensiv-Interviews mit einer Zufallsstichprobe aus der allgemeinen Bevölkerung durchgeführt.

Ausgehend von den experimentell und qualitativ gewonnen Ergebnissen wurde im Juli 1999 eine erste Fassung des Fragebogens erarbeitet. Diese Fassung wurde in insgesamt 32 Schritten überarbeitet. Zwischen den einzelnen Schritten wurden die Überarbeitungen mit den Instituten diskutiert. Nach jeder größeren Veränderung wurde der Fragebogen mit kognitiven Pre-Test Techniken an kleinen Stichproben (5 bis 20 Personen) bestehend aus Studierenden und nichtwissenschaftlichen Angestellten (Hausmeister, Techniker, etc.) überprüft. In den letzten Schritten wurden auch Personen außerhalb der Universität hinzugezogen. Abschließend wurde der Fragebogen von einem der Institute in zwei CATI-Stichproben (n=59 und n=205) aus der allgemeinen Bevölkerung auf seine Tauglichkeit für Nicht-Selbstaufüller getestet.

Der Fragebogen umfaßt mit seinen 71 Fragen bei Berücksichtigung aller Filter insgesamt 135 Items. Hierzu gehören Fragen zur Kriminalitätsfurcht, subjektiven Viktimisierungswahrscheinlichkeit und tatsächlichen Viktimisierungserfahrung für mehrere Delikte, Sicherheitsmaßnahmen am Haus und im Alltag, Gesundheitszustand und Wohnsituation der Befragten sowie einige Fragen der Standarddemographie.

Der Fragebogen der „face-to-face“-Erhebungen enthielt für die Interviewer noch einige Fragen zur Klassifikation des Wohngebäudes und des Wohngebietes, zur Lage der Wohnung, zu den sichtbaren Sicherheitsvorkehrungen sowie Fragen zur Interviewsituation. Der Interviewerfragebogen enthält darüber hinaus noch Fragen zum Schulabschluß, Alter und Geschlecht des Interviewers.

Ergänzende Informationen

Derzeit werden für alle 320 Sampling-Points (160 Random-Points, 160 Quoten-Points) Daten der kleinräumigen Polizeilichen Kriminalstatistik (PKS) gesammelt. Zusätzlich werden dem Datensatz Daten über die Bevölkerungszahl pro Quadratkilometer, den Ausländeranteil und die Arbeitslosenquote in den Sampling-Points hinzugefügt. Schließlich werden die Begehungsvideos digitalisiert und durch zwei unabhängige Beurteiler noch einmal in Hinsicht auf Siedlungsstruktur und „signs of incivility“ klassifiziert. Wir beabsichtigen, den Datensatz gegen Ende des Jahres 2001 bereinigt und aufbereitet dem Zentralarchiv in Köln zu übergeben. Ein vollständiger Projektbericht wird zum gleichen Zeitpunkt vorliegen.

Korrespondenzadresse

*Prof. Dr. Rainer Schnell
Universität Konstanz
Fachbereich für Politik- und Verwaltungswissenschaft
Universitätsstraße 10 - Postfach D92
78457 Konstanz
Rainer.Schnell@uni-konstanz.de
<http://www.uni-konstanz.de/FuF/Verwiss/Schnell/defect/>*

Literatur

Bailar, B., 1983: Interpenetrating Subsamples. S. 197-201 in: Johnson, N. L./Kotz, S. (eds.), *Encyclopedia of Statistical Sciences*. New York, Vol. 4.

Behrens, K./Löffler, U., 1999: Aufbau des ADM-Stichproben-Systems. S. 69-91 in: ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. AG.MA Arbeitsgemeinschaft Media-Analyse e.V. (eds.), *Stichprobenverfahren in der Umfrageforschung*. Opladen: Leske+Budrich.

Kersten, H. M. P./Betlehem, J. G., 1984: Exploring and reducing the nonresponse bias by asking the basic question. *Statistical Journal of the United Nations, UNECE*, 2: 369-380.

Kish, L., 1995: Methods for Design-Effects. *Journal of Official Statistics*, 11, 1: 55-77.

Schnell, R., 1997a: *Nonresponse in Bevölkerungsumfragen*, Opladen: Leske+Budrich.

Schnell, R., 1997b: Praktische Ziehung von Zufallsstichproben für Telefon-Surveys. *ZA-Information* 40: 45-59.

Sudman, S., 1976: *Applied Sampling*, New York: Wiley.

STAND DER BERUFSVERCODUNG

ALFONS J. GEIS UND JÜRGEN H.P. HOFFMEYER-ZLOTNIK

Der vorliegende Beitrag beschäftigt sich mit dem Thema Berufsvercodung. Anlaß für diese Veröffentlichung sind die in letzter Zeit vermehrt aufgetretenen Fragen sowie Mißverständnisse hinsichtlich des Aufwands und der Qualität der Vercodung von Berufsangaben. Im folgenden werden wir auf einige neuralgische Punkte in der Vercodungspraxis hinweisen. Fragen zur Vergleichbarkeit von Klassifikationen, zur Prestigeenergie oder zum Computereinsatz werden nur kurz behandelt; diese sollen in nachfolgenden Beiträgen diskutiert werden.

Die Vercodung von offenen Berufs- und Branchenangaben gehört seit über 20 Jahren zum Standardangebot von ZUMA; dies gilt sowohl für die konventionelle Vorgehensweise durch geschulte Fachkräfte als auch für die computerunterstützte Vercodung.

Die Berufsvercodung besteht - vereinfacht ausgedrückt - darin, die Berufsangabe des Befragten zu lesen und festzustellen, welche der mehreren hundert Kategorien ausschließlich zutrifft, und abschließend die Ziffer der zutreffenden Kategorie, Code genannt, zu notieren. Ist die Angabe ohne weitere Informationen präzise genug, um die gültige Berufskategorie zu wählen, ist der Codiervorgang für beide Vorgehensweisen, die coder- und die computergestützte, denkbar einfach und unproblematisch. Denn diese eindeutigen Angaben liegen mittlerweile in Form eines bei ZUMA empirisch gewachsenen Wörterbuchs vor: das Programm oder der Codierer durchsucht diese Liste nach dem angegebenen Begriff und vergibt den zugehörigen Code (vgl. Tabelle 1).

Das Berufe-Wörterbuch für vier Klassifikationen, zwei vom Statistischen Bundesamt und zwei vom International Labour Office in Genf, umfaßt mittlerweile circa 4300 Stichworte. Es wird kontinuierlich gepflegt, d.h. seine Validität wird kontrolliert, indem Eintragungen präzisiert, korrigiert und ergänzt werden. Auf der Basis dieses Wörterbuchs läßt sich etwa die Hälfte der Angaben aus Repräsentativumfragen automatisch codieren. Die Sprachkompetenz des Codierers erhöht diesen Wert um etwa 15% bei der rein manuellen Bearbeitung.

Warum nicht mehr Fälle anhand derartiger Auflistungen von Berufsangaben gelöst werden können, bzw. worin die Schwierigkeiten der Berufsvercodung bestehen, soll im folgenden dargestellt werden.

Tabelle 1: Wörterbuchauszug (ISCO-68)

599	FLUGHOSTESS
041	FLUGINGENIEUR
041	FLUGKAPITÄN
041	FLUGLEHRER
359	FLUGLEITER
359	FLUGLOTSE
844	FLUGMOTORENSCHLOSSER
359	FLUGSICHERUNGSKONTROLLDIENSTLEITER
359	FLUGVERKEHRSLOTSE
873	FLUGZEUGBAUER
024	FLUGZEUGBAUINGENIEUR
873	FLUGZEUGBAUMEISTER
855	FLUGZEUGELEKTRIKER
842	FLUGZEUGFEINMECHANIKER

1. Grundlagen

1.1 Abfrage des Berufs in empirischen Erhebungen

In Umfragen sind geschlossene Fragen immer vorzuziehen. Sie sind kontrollierbarer, kostengünstiger und zeitsparender als andere Erhebungsformen. Ist der Untersuchungsgegenstand aber nicht ausreichend bekannt - explorative Phase - oder sind die bekannten Antwortalternativen zu umfangreich oder zu komplex, dann kommt eine Feldvercodung nicht in Frage. In solchen Fällen wird man offen abfragen, die Antwort notieren oder aufnehmen und später codieren. Dies trifft auf die Berufe zu: Die Codierung nach der „International Standard Classification of Occupations“ des International Labour Office in Genf von 1990 (ISCO-88) hat zwar nur ca. 390 Kategorien, die aber bestimmten Bedingungen genügen müssen, und die Klassifikation der Berufe (KldB) des Statistischen Bundesamtes hat über 2000 Kategorien, die teilweise auch weitere Informationen (wie z.B. Branche) erfordern.

Die Frageformulierung muß auf das anzuwendende Klassifikationsschema abgestimmt sein. „Geben Sie bitte Ihren Beruf an“ ist nicht ausreichend, denn während man für die StaBu-Codierung nach der KldB 92 die Berufsbezeichnung oder Berufsbenennung braucht, ist für die Verschlüsselung nach der ISCO die Beschreibung der

konkreten beruflichen Tätigkeit, unabhängig von Bezeichnung oder Titel, maßgeblich. Folgende dreistufige Frageformulierung hat sich bewährt:

„Welche berufliche Tätigkeit üben Sie zur Zeit aus?“, „Bitte beschreiben Sie mir diese Tätigkeit genau.“, „Hat dieser Beruf noch einen besonderen Namen?“

Die erste Frage lenkt den Fokus auf die Tätigkeit, unabhängig von der Bezeichnung oder dem Titel, die oft wenig über das aussagen, was die Person wirklich tut; die Information darüber soll durch die zweite Frage ergänzt werden. Und erst als letztes wird gefragt, was bei unspezifischer Fragestellung meist zuerst genannt wird, nämlich die Berufsbezeichnung. Die enthält zwar zu etwa 50 Prozent die für die Codierung ausreichende Informationen, die andere Hälfte benötigt aber Zusatzinformationen. Die Erfahrung hat gezeigt, daß die Informationsmenge und -genauigkeit größer sind, wenn diese drei Fragen getrennt nacheinander gestellt und ihre Antworten jeweils separat notiert werden.

Bei der Interviewerschulung muß darauf hingewiesen werden, daß Titel oder berufliche Stellung nicht ausreichen, sondern das angegeben werden soll, was jemand konkret tut. Der Interviewer sollte im Zweifelsfall nachfragen. Folgende Angaben sind z.B. nicht zu verarbeiten: Angestellter, Beamter, Eisenbahner, Postbeamter, Technischer Angestellter, Abteilungsleiter, Selbständiger usw. Sie alle würden den Code 004 für nicht näher zu bestimmende Berufsangaben erhalten müssen, denn aus diesen Angaben ist nicht ersichtlich, welche Tätigkeit ausgeübt wird. Die Spezifizierung der Angabe muß der des Kategorienschemas entsprechen und darf nicht zu allgemein, wenn nicht gar nichtssagend sein: statt Postbeamter z.B. muß es heißen Briefträger, Schalterbeamter, Kraftfahrer, Fernmeldetechniker; statt Technischer Angestellter z.B. Technischer Zeichner, Materialprüfer, Chemotechniker; statt Angestellter z.B. wissenschaftlicher Mitarbeiter im Finanzministerium, Textilverkäuferin, Angestellter in der Personalverwaltung; statt nur Geschäftsführer, Abteilungsleiter, Filialleiter u.a.m. muß zusätzlich auch die Angabe der Berufsstellung, der Branche, des Produktes, die Beschreibung der Abteilung u.ä. angegeben werden. Bei vielen Berufen wie z.B. Kaufmann, Versicherungskaufmann, Gärtner, Fuhrunternehmer ist es für die Einordnung wichtig zu wissen, ob die befragte Person selbständig oder angestellt, ob allein oder mit Angestellten arbeitet oder wieviele Mitarbeiter der Betrieb überhaupt hat, weil diese Merkmale jeweils zu anderen Kategorien führen.

All diese Informationen müssen im Fragebogen strukturiert erhoben werden, um eine unproblematische Berufsvercodung nach den gängigen Klassifikationen zu ermöglichen.

Für die Berufsvercodung stehen zwei Klassifikationssysteme zur Verfügung, die des Statistischen Bundesamtes - KldB genannt - („Klassifizierung der Berufe - Systematisches und alphabetisches Verzeichnis der Berufsbenennungen“ - Fassungen von 1975 und 1992) und die Internationale von 1968 und 1988. Im folgenden sollen beide kurz vorgestellt werden.

1.2 Berufsklassifikationen

International: ISCO

Das International Labour Office in Genf hat 1968 eine „International Standard Classification of Occupations“ vorgelegt (ISCO-68), die den Anspruch der internationalen Gültigkeit hat und eine systematische Klassifikation für alle Beschäftigungen der gesamten zivilen Erwerbsbevölkerung darstellen will. In den Folgejahren ist die ISCO-68 Ausgangsdatei für Prestige- und Statusskalen geworden. Zwanzig Jahre später wurde die „ISCO-68“ überarbeitet und umstrukturiert. Das Ergebnis ist die „ISCO-88“, die die „ISCO-68“ ablöst, aber nicht unbedingt ersetzt, da sie einem anderen Strukturkonzept folgt.

ISCO-68

Die Struktur der Klassifikation von ISCO-68 hat vier Ebenen, die, aufeinander aufbauend, sukzessive eine feinere Untergliederung bieten (Tabelle 2).

Tabelle 2: Klassifikationsstruktur der ISCO-68

Ebene	Kategorienschema	Anzahl der Kategorien
1	major groups	8
2	minor groups	83
3	unit groups	284
4	occupational categories	1506

(ILO 1968: 1)

Die „major groups“ orientieren sich an den Beschäftigungssektoren (Primärer Sektor: Code „6“; Sekundärer Sektor: Codes „7-9“; Tertiärer Sektor: Codes „2-4“), ohne jedoch diese klar voneinander abzugrenzen, da der Bereich der leitenden Tätigkeiten (Codes „0-2“) nicht gemäß den Sektoren klassifiziert wird. Die „minor groups“ stellen eine grobe Zusammenfassung von Tätigkeitengruppen dar. Die „unit groups“ gründen auf der beruflichen Tätigkeit: Es werden vergleichbare Tätigkeiten auf verwandten Gebieten zusammengefasst. Diese Ebene der Kategorisierung ist für den praktischen

Gebrauch gedacht. Mit der vierten Ebene, der Ebene der „Occupations“, wird ein Typ von Arbeit identifiziert.

In der Praxis finden der dreistellige Code auf der Ebene der „unit groups“ und der fünfstelligen Code auf der Ebene der „Occupations“ Anwendung. Der in den Umfragen der Survey-Forscher zu erhebende Code sollte daher so detailliert erfaßt werden, so daß er dazu ausreicht, den fünfstelligen Code der „Occupations“ zu generieren. Dieser klassifiziert sehr präzise berufliche Tätigkeiten, ohne hierbei auf die Arbeitsplatzebene hinunter zu gehen.

Die Klassifikation einzelner Berufsangaben geschieht also hierarchisch, bei der Oberkategorie des Beschäftigungssektors beginnend, über eine immer detailliertere Klassifikation, bis zur endgültigen Klassifikation der exakten Tätigkeit aufschlüsselnd. Hierbei wird ein Dezimalsystem verwendet, das von der sektoralen Gliederung nach Beschäftigtensektoren in zwei Schritten zur fach- oder branchenbezogenen Gliederung übergeht. Diese wird schließlich in der vierten und fünften Stelle durch eine Klassifikation der Tätigkeit präzisiert. Diese vier Schritte werden von den Autoren der ISCO an folgendem Beispiel verdeutlicht (ILO 1968: 1):

Eine „Schreibkraft“ erhält den Code 3-21.40. Dieser Code setzt sich über die Ebenen wie folgt zusammen: Der Code „3“ in der „major group“ bedeutet „*Clerical and related worker*“; fügt man auf der zweiten Ebene, der „minor group“ den Code „2“ präzisierend ein, so wird aus dem „Büroarbeiter“ im Code „3-2“ ein „*Stenographers, typists and card- and tape-punching machine operators*“, in der „unit group“ im Code „3-21“ ein „*Stenographers, typists and teletypists*“ und auf der vierten Ebene in der Kategorie „occupations“ über den Code „3-21.40“ eine Person, deren Hauptaufgabe das Maschineschreiben ist.

In der Mehrzahl der Fälle dürfte es unproblematisch sein, die „Occupations“ der ISCO mit nationalen Berufskategorien soweit in Übereinstimmung zu bringen, daß die jeweiligen Arbeitsbereiche in den meisten Elementen, die ausgeübte Tätigkeit betreffend, übereinstimmen. In manchen Fällen hat ein und derselbe Beruf in unterschiedlichen Kulturen allerdings eine unterschiedliche Bedeutung. In solchen Fällen sollte dieses durch einen nationalen Zusatzcode kenntlich gemacht werden. Solche nationalen Zusatzcodes haben die Funktion, eine Kategorie zu präzisieren.

ISCO-88

Durch die Zunahme der beruflichen Spezialisierung und das technologiebedingte Entstehen neuer Berufe wurde eine Revision der ISCO-68 notwendig. Die revidierte Fassung der ISCO, bekannt als ISCO-88, wurde 1990 vorgelegt. Die Klassifikationen

von 1968 und 1988 unterscheiden sich durch grundlegende methodische und strukturelle Neuerungen (siehe: Statistisches Bundesamt 1992: 12).

Die ISCO von 1968 war als Instrument für unterschiedliche Zwecke angelegt: Sie war nicht allein als Instrument für die amtliche Statistik konzipiert, sondern stellte zusätzlich ein Instrument dar, das für berufskundliche Fragestellungen, für Fragestellungen zur sozialen Mobilität und für eine international vergleichende Forschung einsetzbar sein sollte. Daher wurde besonderer Wert auf eine detaillierte Beschreibung der untersten Gliederungseinheit gelegt, deren Grundlage die Tätigkeitsbeschreibung ist. Die ISCO von 1988 ist primär für Anwendungszwecke der amtlichen Statistik konzipiert. Nicht mehr die Differenzierung steht im Vordergrund, sondern eine strukturierte Reduktion. Damit erhält bei diesem Kategoriensystem die oberste Hierarchieebene, die erste Stelle, eine zentrale strukturierende Rolle.

Die ISCO von 1988 hat als neue Bestimmungsfaktoren zur Konkretisierung der Gleichheit bzw. der Verwandtschaft von beruflichen Tätigkeiten die Merkmale „skill level“ und „skill specialisation“ eingeführt:

„Skill level“ stellt hierbei im Sinne eines Anforderungsniveaus ein berufs- und arbeitsplatzbezogenes Merkmal dar. Das „skill level“-Konzept findet nur Anwendung auf die Bestimmung der „major groups“. Es unterscheidet in Anlehnung an die „International Standard Classification of Education (ISCED)“ nach vier relativ groben Kategorien von Bildungsqualifikation. Hierbei ist allein wichtig, daß ein bestimmtes Kenntnis- und Fertigungsniveau vorhanden ist:

The first ISCO skill level: „comprising primary education which generally begins at the age of 5, 6 or 7 and lasts about five years“.

The second ISCO skill level: „comprising first and second stages of secondary education. The first stage begins at the age of 11 or 12 and lasts about three years, while the second stage begins at the age of 14 or 15 and also lasts about three years. A period of on-the-job-training and experience may be necessary, sometimes formalised in apprenticeships. This period may supplement the formal training or replace it partly or, in some cases, wholly.“

The third ISCO skill level: „comprising education which begins at the age of 17 or 18, lasts about four years, and leads to an award not equivalent to a first university degree.“

The fourth ISCO skill level: „comprising education which also begins at the age of 17 or 18, lasts about three, four or more years, and leads to a university or postgraduate university degree, or the equivalent.“ (ILO 1990: 2-3)

„Skill specialisation“ ist ein berufsfachliches Kriterium. Es hebt einerseits auf die Art der ausgeübten Tätigkeit innerhalb einer gegebenen Qualifikationsebene ab und versucht andererseits Merkmale wie verarbeitetes Material, Berufsmilieu, Art der produzierten Güter und Dienstleistungen, Besonderheiten der Fertigungsverfahren etc. mit einzubeziehen.

Im Gegensatz zur ISCO-68 geht die ISCO-88 nicht auf die Ebene der beruflichen Tätigkeit („occupations“) hinunter. Auch wenn in der Regel mehr als eine berufliche Tätigkeit einer Berufsgruppe („unit group“) zuzuordnen sind, so erscheint die Ebene der „unit groups“ den Autoren der ISCO-88 (ILO 1990: 4) die sinnvollere und aussagekräftigere Ebene, denn: auf unterschiedlichen nationalen Ebenen müssen die Definitionen von Tätigkeiten voneinander abweichen „depend on the size of the economy and the level of economic development, the level and type of technology, work organisation and historical circumstances“ (ILO 1990: 4).

Wegen der grundlegenden Unterschiede in der Struktur läßt sich die Ebene der „major groups“ der ISCO-88 nicht mit denen der ISCO-68 vergleichen. Kontinuität zum Zwecke des Vergleichs wurde dennoch angestrebt: Ein Vergleich sei auf der Ebene der „unit groups“ gegeben, d.h. die 284 dreistelligen „unit groups“ der ISCO-68 sind mit den 390 vierstelligen „unit groups“ der ISCO-88 vergleichbar. Sie sind allerdings nicht automatisch von einem Code in den anderen übersetzbar.

National: KldB-75 und -92

Die derzeit gültige „Klassifizierung der Berufe“ (KldB) in der Bundesrepublik Deutschland von 1992 stellt eine Weiterentwicklung der von 1975 dar. Die Gliederung ist auch in dieser Klassifikation eine hierarchische über unterschiedliche Ebenen. Die Obereinheit stellen die „Berufsbereiche oder -abschnitte“ dar. Es folgt die Ebene der „Berufsgruppen“ (Zweisteller), von denen die 1992 revidierte Fassung 88 umfaßt. Die dritte Ebene stellen die dreistelligen „Berufsordnungen“ dar, die als statistische Basiseinheiten deklariert sind (StaBu 1992: 7). Ein Problem hierbei ist jedoch die Möglichkeit der De-Anonymisierung bei zu gering besetzten Kategorien, so daß eine Mindestbesetzung von 5.000 Erwerbspersonen pro Basis-Einheit im (Mikro-)Zensus erforderlich sein muß, um eine Kategorie als solche auszuweisen. Die vierte Ebene stellen die „Berufsklassen“ dar, die sich in einem Berufsbenennungskatalog niederschlagen. Dieser umfaßt in der 1992er Version 29.500 Benennungen. Die Aktualisierung des Berufsbenennungskatalogs mit Streichungen überholter und Aufnahme 10.300 neuer Berufsbenennungen erweiterte die Ausgabe von 1992 gegenüber der von 1975 von ehemals insgesamt 22.000 (1975) auf 29.500 (1992). In diesen Zahlen zeigt sich die

Dynamik des Wandels, die bei einer Auflistung von Berufsbenennungen, wie sie beim Instrument der KldB geschieht, nicht unberücksichtigt bleiben darf.

Betrachtet man den generellen Unterschied zwischen der Logik der ISCO und der KldB, so besteht dieser darin, daß bei der ISCO auf der tiefsten Ebene Definitionen von Berufen stehen, während bei der KldB die unterste Ebene eine Auflistung von Berufsbezeichnungen darstellt. Berufsbezeichnungen sind Moden und Schwankungen unterworfen, bieten aber den großen Vorteil, daß bei einer ständigen Überarbeitung der Liste der Benennungen eine hohe Trefferquote bei der Zuordnung von Befragtenantwort zu vorhandenem Listeneintrag gegeben ist, denn es wird der Eintrag, nicht eine dahinter zu vermutende Definition betrachtet. Im Gegensatz hierzu stellt ein Katalog voller Definitionen von Berufen eine härtere Anforderung an eine Codierung. Definitionen von Berufen bieten aber, im Gegensatz zu Bezeichnungen, einen geringeren Interpretationsspielraum, auch über die Zeit.

2. Methode der Inhaltsanalyse

Die Inhaltsanalyse ist eine Methode der empirischen Sozialforschung, mit der Kommunikationsinhalte systematisch und nachvollziehbar entsprechend einer vorgegebenen Forschungsfrage untersucht werden. Die Vorgehensweise der Inhaltsanalyse besteht - kurz gesagt - darin, daß Kommunikationsinhalte nach festen Regeln vorgegebenen Kategorien zugeordnet werden. Ausführliche methodische Grundlagen und Hinweise finden sich in einschlägigen Methodenbüchern wie denen von Früh (1998), Friedrichs (1973), Merten (1995), u.a.m. Die Zuordnung zu bestimmten Kategorien wird Codierung genannt und ist vergleichbar der Einstufung von Antwortvorgaben in einer Umfrage. Aber nicht immer sind Vorgaben im Fragebogen selbst oder eine Feldvercodung (Einordnen/Codieren der Antwort an Ort und Stelle des Interviews) möglich. Mit Berufsangaben, die nach der Internationalen Standardklassifikation der Berufe (ISCO-88) mit 390 Kategorien (=Vorgaben) codiert werden, wären Interviewer und Befragte überfordert.

Stellt schon die Menge der Kategorien/Vorgaben ein Problem dar, so scheitert eine Feldvercodung erst recht an der Komplexität. Den Inhalt des über 400-seitigen Buches der ISCO-Klassifikationen mit Beschreibungen und Definitionen präsent zu haben, ist kaum möglich. Deshalb wird die Frage nach der Berufstätigkeit offen gestellt; die Antwort wird schriftlich festgehalten und erst später durch geschulte Codierkräfte codiert.

An die Inhaltsanalyse werden die gleichen Forderungen gestellt wie an jede wissenschaftliche Methode: sie muß wissenschaftlich *relevant* sein, *systematisch* betrieben werden und *intersubjektiv* nachvollziehbar sein. Die erste Forderung kann als gegeben vorausgesetzt werden: Ein Indiz dafür ist nicht zuletzt die ungebrochene Nachfrage der Dienstleistung „Berufsvercodung“. Die Intersubjektivität wird u.a. durch gute Dokumentation des Codiervorgangs erreicht, so daß letztlich die Systematik als Hauptforderung bleibt, und zwar einerseits an das theoretische Konstrukt, das Kategorienschema, und andererseits an den Codiervorgang.

Kategorien müssen folgende Kriterien erfüllen: sie müssen *relevant* im Sinne der Fragestellung, *eindimensional* (nur eine logische Ebene), *vollständig* (damit die Hypothese falsifizierbar ist) und *trennscharf* (im Sinne eines systematischen Vorgehens) sein. Die Vorgaben „Akademiker“ und „selbständig“ überschneiden sich, sie sind weder trennscharf (ein Anwalt ist Akademiker und kann gleichzeitig selbständig sein) noch eindimensional (mit der schulisch-beruflichen Ausbildung und der beruflichen Stellung werden zwei unterschiedliche logische Ebenen angesprochen).

Die Kategorien für die Berufsvercodung sind vorgegeben; die Frage nach der wissenschaftlichen Systematik wird insoweit positiv beantwortet, als es gelingt, die Angaben der Befragten nach den vorliegenden Kategorien zu codieren, ihnen also einen Code zuzuweisen, der den Inhalt der entsprechenden Kategorie repräsentiert. Dazu dienen die allgemeinen Hinweise in den Einleitungskapiteln der Klassifikationen, zusammenfassende Kategorien-Namen, die Definitionen, die Beispiele und die Auflistung der Ausnahmen.

Die Operationalisierung ist der Übersetzungsprozeß der abstrakten Idee in eine konkrete Erscheinung, das Bindeglied zwischen der Hypothese und den Daten, die Steuerung der Datenerhebung im Sinne des Untersuchungszieles. Dem Codierer werden zu dieser Steuerung Handlungsanweisungen in Form von allgemeinen Codierregeln, Kategorienbegriffen, Definitionen, Indikatoren, Beispielen und Codezuweisungen an die Hand gegeben. Alle Handlungsanleitungen finden sich im sogenannten Codebuch. Der *konventionelle* Codiervorgang sieht so aus, daß die geschulte Kraft sich den Text (der Berufsangabe) ansieht und überlegt, welche der über 390 Kategorien zutrifft. Wenn die Interviewer und Befragte gut gearbeitet haben, ist ein großer Teil sofort einzuordnen; da genügt manchmal ein einziges Stichwort, das klarstellt, welche Kategorie zutrifft. Das trifft für etwa gut die Hälfte der Angaben in einer Repräsentativbefragung zu. Der andere Teil erfordert genauere Überlegungen, Nachschlagen im ISCO-Buch, Prüfen weiterer Angaben des Befragten, manchmal sogar Telefonate und weitere Recherchen. Zu diesem Zweck wurde bei ZUMA ein Regelwerk zusammengestellt, das

Konventionen, Beispiele und Entscheidungshilfen enthält, wie sie im ISCO-Buch nicht zu finden sind. Darüber hinaus stehen eine Reihe weiterer Hilfsmittel/Veröffentlichungen zur Verfügung, wie z.B. „Beruf aktuell“ (Bundesanstalt für Arbeit, jährliche Neuauflagen), „Bildungssystem in der Bundesrepublik Deutschland“ (MPI Berlin für Bildungsforschung, 1994), DDR-Ausbildungsberufe (IAB Bundesanstalt für Arbeit 1990), „Bildungssystem der BRD“ (Mohr 1991).

3. Codiervorgang

Die Klassifikation der Berufe (KldB), herausgegeben vom Statistischen Bundesamt (1992) kennt keine Definitionen einzelner Kategorien, sondern führt in den Einleitungskapiteln der Veröffentlichung nur einige allgemeingültige Grundsätze auf. Im übrigen werden alle Bezeichnungen aufgelistet, die zu einer jeweiligen Kategorie gehören, was meist ergänzt wird von einer Ausnahmeliste, überschrieben mit „Nicht einzubeziehen“. Abstrakte Überlegungen und logische Schlüsse müssen weniger angestellt werden. Im folgenden wird deshalb fast ausschließlich auf die ISCO-Klassifikationen eingegangen. Nur die allgemeinen Regeln gelten auch für die Codierung nach der KldB.

3.1 Coderbasiert

Ein Großteil der Angaben wird immer noch durch Codierkräfte bearbeitet. Man spricht dann von konventioneller oder manueller oder coderbasierter (im Gegensatz zur computerunterstützter) Berufsvercodung.

Allgemeine Regeln und Grundsätze

Die im folgenden aufgeführten Regeln sind einerseits aus den offiziellen Klassifikationen des Statistischen Bundesamtes (1971, 1992) und des Internationalen Arbeitsamtes - ILO - (1990) abgeleitet; andererseits haben darüber hinaus hier die aus der Codier Erfahrung getroffenen Vereinbarungen und Festlegungen ihren Niederschlag gefunden.

- Die ISCO-Klassifikationen sind hierarchisch aufgebaut, die Codes sind nach dem Dezimalsystem vergeben, wobei die erste Ziffer die allgemeinste Information wiedergibt, z.B. 2 für Wissenschaftler, und die dritte (ISCO-68) bzw. vierte (ISCO-88) die spezifischste, z.B. 2441 für Wirtschaftswissenschaftler. Es wird so spezifisch codiert, wie es die Informationen erlauben, d.h. möglichst keine MD-Codes (missing-data-Codes für fehlende Werte). Dazu sind alle Informationen (wie z.B. Branche, Stellung im Beruf, Dienststelle, Firmengröße u.ä.) des Fragebogens heranzuziehen.

- Maßgeblich für die Codierung sind in erster Linie die ausführlichen Definitionen, wie sie vom Statistischen Bundesamt (1971) und dem Internationalen Arbeitsamt (1990) veröffentlicht wurden.
- Sind mehrere verschiedene Berufe mit gleicher Genauigkeit angegeben, dann wird der erste verschlüsselt. Beispiel: „Landwirt und Maurer“ = Landwirt 611 (ISCO 68), 6130 (ISCO 88).
- Sind mehrere Berufsbezeichnungen für den gleichen Beruf angegeben, dann wird die konkretere verschlüsselt. Beispiel: „kaufmännischer Angestellter, Buchhalter“ = Buchhalter 331 (ISCO 68), 3433 (ISCO 88).
- Im ISCO-88 wird nach Hand-/Maschinenarbeit unterschieden; wenn dies nicht möglich ist, wird die Tätigkeit gewählt, die meistens zutrifft (Erfahrung), z.B. Näherin in Fabrik 8263, Weberin in Fabrik 8262. Heute stellt die reine Handarbeit eher die Ausnahme dar.
- „Produktionsleiter“ werden im ISCO-88 je nach Branche unterschiedlich codiert.
- „Poliere“ werden entsprechend dem Beruf codiert.
- Selbständige: bei Angabe des Berufs, z.B. Maler, Schreiner usw., und einer Betriebsgröße bis zu 9 Personen wird der Beruf codiert; bei Angabe „Betriebsleiter eines Handwerksbetriebs“ wird auch bei einer Betriebsgröße von weniger als 9 Personen Betriebsleiter 211/1310 codiert, wenn der spezifische Beruf nicht bekannt ist.

Nicht vollständig codierbare Angaben

In Repräsentativumfragen können gewöhnlich 75 bis 85% der Angaben vollständig codiert werden, d.h. es liegen in diesen Fällen genügend Informationen vor, um nach der ISCO bis zur letzten Ebene spezifisch einordnen zu können. Die restlichen 15-25% der Fälle lassen sich folgendermaßen beschreiben:

Allgemeine, spezifische, sonstige Angaben

Als allgemeine Angaben werden die Antworten bezeichnet, die zwar eine Tätigkeit beschreiben, aber nicht genau genug, um sie bis zur letzten Code-Stelle einem Beruf zuzuordnen zu können. Um dies mit einem Beispiel zu verdeutlichen: mit „Lehrer“ kann eine Lehrkraft für die Vorschule, die Grundschule, das Gymnasium oder auch für die Volkshochschule u.a. gemeint sein. Vorausgesetzt, es ist festzustellen, daß es sich um eine wissenschaftliche Lehrkraft handelt, kann in der ISCO-88 eine Festlegung nur für die erste und zweite Stelle des Codes getroffen werden: 2300. Ist nicht auszuschließen, daß es auch eine nicht wissenschaftliche Lehrkraft ist (z.B. Fahrlehrer mit Code 3340), ist auch hier nur der Code 0004 für „Nicht codierbar“ zulässig.

Die Kategorien, die mit "sonstige" beginnen oder den Zusatz "anderweitig nicht klassifiziert" o.ä. haben, sind den Angaben vorbehalten, die zwar einen Beruf genau beschreiben, für den es aber in der ISCO keine eigene Kategorie gibt.

Zusammenfassend läßt sich sagen, das Attribut „allgemein“ trifft dann zu, wenn eine Kategorie vorhanden ist, aber die genaue Information fehlt, und um eine „sonstige“ Kategorie handelt es sich, wenn die Information vorliegt, aber eine passende Kategorie fehlt.

Unzureichende Angaben - unvollständige Codes

Aufgrund fehlender Informationen können ISCO-Codes oft nicht bis in die letzte Stelle vergeben werden, da präzise Angaben zur Bestimmung von Berufsfeld und -gattung, oft sogar der Untergruppe fehlen. Wenn keine exakte Einordnung erfolgen kann, wird "unvollständig" codiert:

Beispiele für unvollständige Codes in der ISCO-68 sind:

Hüttenarbeiter = 72- / nicht bekannt, ob Ofenmänner (= 721), (Metall-)Walzer (= 722), Schmelzer (= 723) oder Gießler (= 724)

Chemiefacharbeiter = 74- / nicht bekannt, ob Mischer (= 741), Wärmebehandler (= 742), Filtrierer (= 743), Destillierer (= 744) oder Raffinierer (= 745)

Keramiker = 89- / nicht bekannt, ob Töpfer (= 892), Keramdreher (= 892), Kerambrenner (= 893), Kerammale (= 895) oder Tonmassemischer (= 899)

Jurist = 12- / nicht bekannt, ob Anwalt (= 121), Richter (= 122), Notar (= 129), oder Syndikus (= 129)

In der ISCO-68 dient ein Blank (Leerstelle:-) auf der dritten Stelle dazu, in der Angabe nicht weiter spezifizierte oder spezifizierbare Berufe (allgemeine Angaben) zu kennzeichnen.

a. Kann nur die Berufshauptgruppe sicher verschlüsselt werden, dann bedeuten:

19- nicht weiter spezifizierte Wissenschaftler usw.

39- nicht weiter spezifizierte Bürokräfte

49- nicht weiter spezifizierte Handelsberufe

59- nicht weiter spezifizierte Dienstleistungsberufe

99- nicht weiter spezifizierte gütererzeugende Tätigkeiten usw.

b. Kann auch die (zweistellige) Berufsuntergruppe verschlüsselt werden, aber noch nicht die Berufsgattung, dann wird die dritte Stelle der Untergruppe durch ein Blank ersetzt; z.B.: Spinnereiarbeiter = 75-, denn dieser Beruf gehört zur Gruppe 75 (Spinner, Weber, Stricker usw.), kann aber dort nicht genauer lokalisiert werden.

c. Wenn die angegebene Berufsbezeichnung mehrere Tätigkeitsuntergruppen umfaßt, ist nur der Code 99- möglich. Nicht weiter spezifizierte Arbeiter (z.B. Fabrikarbeiter) werden unter 99- verschlüsselt. Desgleichen alle Arbeiter, die nur ihren Arbeitsplatz angeben; z.B.: Arbeiter in einer Spinnerei = 99-, aber Spinnereiarbeiter = 75-

In der *ISCO-88*, die nach dem Dezimalsystem gegliedert ist, zeigt eine Null auf einer bestimmten Dezimalstelle an, daß es auf dieser Ebene keine weitere Differenzierung gibt. Allerdings ist nicht jeder ISCO-88-Code, der an letzter Stelle eine Null führt, ein unvollständiger Code. Die Kategorie 2460 (Geistlicher, Seelsorger) existiert z.B. nur als Berufsuntergruppe und ist daher nicht weiter untergliedert.

Völlig unzureichend ist eine Angabe dann, wenn die gegebenen Informationen nicht einmal ausreichen, auch nur die Berufshauptgruppe zu bestimmen. "Angestellte" oder "Beamtin" beschreibt das Arbeitsverhältnis oder die Zugehörigkeit zu einem System der Rentenversicherung, nicht aber die Tätigkeit. Angestellte können z.B. Lagerarbeiter, Verkäufer, Bankkaufleute, wissenschaftliche Mitarbeiter usw. sein. Ein Müllarbeiter in der Kommune kann bei gleicher Tätigkeitsbeschreibung Arbeiter, Angestellter oder Beamter sein. Ähnlich vage ist der Informationsgehalt, wenn anstelle von Berufen Branchen angegeben werden. In all diesen Fällen wird die 0004 (nicht codierbar) vergeben; sie stellen in repräsentativen Studien 2 bis 6% der Angaben dar.

Weitere Beispiele für nicht codierbare Angaben (Code 0004):

- *Außendienstmitarbeiter*: (ohne Angaben zu Branche und Berufsausbildung), selbst mit kaufmännischer Lehre könnte es z.B. ein Handelsvertreter, Inkassoangestellter oder Verkäufer mit jeweils unterschiedlichen Codes sein.
- *Abteilungsleiter*: auch mit Kennz. 52 (Angestellte, die schwierige Aufgaben ...), Schulabschl. mittlerer Reife und gewerblicher Lehre nicht codierbar.
- *Angestellte* im öffentlichen Dienst: wenn eine kfm. Lehre vorliegt, ist mit 39-/4000 codierbar.
- *Eisenbahner*: auch mit Kennziffer nicht codierbar; mit Kennz. 40 (Beamte im einfachen Dienst) ergeben sich u.a. folgende Möglichkeiten, die jeweils einen anderen Code fordern: Weichenwärter, Ladedienst, Schaffner oder Stellwerkwärter: Mit Kennziffer 62 (gelernte und Facharbeiter) ist mit 99-/7000 codierbar.
- *Technische Angestellte*: Der Tätigkeitsbereich kann vielfältig sein. In der ehemaligen DDR wurden z.B. Putzfrauen auch als „Technische Angestellte“ bezeichnet.

Tabelle 3: Zusatzcodes

ISCO-68	ISCO-88	Angabe	Erläuterung
000	0000	TNZ - trifft nicht zu	Die Frage nach dem Beruf wurde dem Befragten nicht gestellt, es wird keine Angabe erwartet.
000	0001	Soldat	Angehörige der Streitkräfte werden nach der ISCO nur durch den Code 0110 erfaßt; besonders für die Generierung von Prestige- oder Status-Werte ist die Unterscheidung wenigstens zwischen Mannschafts- und höheren Dienstgraden nützlich.
002	0002	Offizier	vgl. Soldat
003	0003	arbeitssuchend	Je nachdem, wie die Filter in einem Fragebogen gesetzt werden, können auch Arbeitssuchende im Sinne von Berufsanfängern betroffen sein.
004	0004	nicht codierbar	Die Informationen sind so unzureichend, daß auch nicht mit Allgemein- oder Sonderkategorien codiert werden kann.
005	0005	in der Ausbildung	Betrifft nur die schulische, nicht berufliche Ausbildung. Eine Lehre wird wie Berufstätigkeit codiert.
006	0006	Hausfrau /-mann	Im Selbstverständnis vieler Befragten zählt die Arbeit im Haus als Berufstätigkeit.
007	0007	Rentner	Hier handelt es sich um einen Filterfehler oder eine unzureichende Angabe des Befragten, der z.B. seinen früheren Beruf angeben sollte.
008	0008	weiß nicht	Eine mögliche Antwort, wenn z.B. nach dem Beruf der Partnerin oder der Eltern gefragt wurde.
009	0009	keine Angabe	Es liegt ein Fehler vor, sei es, daß der Befragte wirklich keine Antwort gegeben hat (verweigert) oder der Interviewer die Frage nicht stellte bzw. die Antwort nicht notierte.
13-	2500	Erzieher	Ohne nähere Angabe (könnte u.a. Sozialarbeiter, Kindergärtnerin sein). - Konvention in ISCO-68, zusätzlicher Sondercode in ISCO-88.
399	4300	Bürohilfskräfte	Spezifizierung innerhalb der Büroberufe – vgl. KldB-Codes 7802, 7840 ff
310	4400	Verwaltungsbeamte	Gehobener Dienst Post (ab Inspektor Kennz. 42). - Konvention in ISCO-68, zusätzlicher Sondercode in ISCO-88.
310	4500	Verwaltungsbeamte	Gehobener Dienst Bahn (ab Inspektor Kennz. 42). - Konvention in ISCO-68, zusätzlicher Sondercode in ISCO-88.
99-	7510	Betriebsschlosser	Ohne nähere Angabe - Konvention in ISCO-68, zusätzlicher Sondercode in ISCO-88.
99-	7510	Schlosser	Ohne nähere Angabe - Konvention in ISCO-68, zusätzlicher Sondercode in ISCO-88.
	7520	Elektroniker	Ohne nähere Angabe - Konvention in ISCO-68, zusätzlicher Sondercode in ISCO-88.
700	7900	Aufsichtskräfte in der Produktion	Ohne nähere Angabe - Konvention in ISCO-68, zusätzlicher Sondercode in ISCO-88.
700	7900	Meister, Vormänner in der Produktion	Nur im Sinne von Aufsichtsperson (u.U. auch Vorarbeiter, der Kolonnen beaufsichtigt), sonst entsprechende Tätigkeit - Konvention in ISCO-68, zusätzlicher Sondercode in ISCO-88.

Zusatzcodes

Nicht für alle Angaben der Befragten lassen sich Entsprechungen in der Klassifikation finden, sei es, daß die Angabe zu allgemein ist und ISCO keine Sammelkategorie vorsieht oder daß es sich gar nicht um Berufsangaben i.e.S. handelt, sondern nur um Angaben, die den beruflichen Status beschreiben. Deshalb wurden zusätzliche Codes eingeführt. Es handelt sich hierbei nicht um ISCO-Codes, sondern um Ergänzungen, die sich aus der ZUMA-Codierpraxis ergeben haben; sie dienen in erster Linie dazu, die fehlenden Werte zu spezifizieren (Tabelle 3).

Verfügbare Hilfsmittel

Berufewörterbuch

Das Wörterbuch, auch Diktionär genannt, ist eine alphabetische Liste, die alle Angaben enthält, die nur durch Nennung des Begriffs codierbar sind und keine weiteren Informationen erfordern, also als harte Indikatoren gelten können. Durch die Sprachkompetenz des/der Codierers/in ist das Wörterbuch in der coderbasierten Berufsvercodung jedoch umfänglicher zu nutzen: Während der Computer z.B. nicht erkennt, daß eine Textangabe im Verzeichnis keine Entsprechung hat, weil ein (kleiner) Schreibfehler vorliegt, kann der/die Codierer/in erkennen, daß es sich um eine nach dem Wörterbuch codierbare Angabe handelt. Ähnliches gilt für unterschiedliche Schreibweisen, Abkürzungen, ausführliche Beschreibungen der Tätigkeit usw.

Wörterbücher gibt es für ISCO-68, ISCO-88, StaBu-75 und StaBu-92, die zusammengefaßt und mit Kommentaren, Anmerkungen, Querverweisen, Beispielen, Einschränkungen usw. versehen sind. Sie enthalten die bei ZUMA kumulierte und codifizierbare Erfahrung in der Berufsvercodung.

Konventionen und Einzelfallregelungen (Fallsammlung)

Es gibt gewisse problematische Angaben, die immer wieder zur Entscheidung anstehen. Um die Daten auch in nicht eindeutigen Fällen konsistent zu halten wurde eine Sammlung mit Angaben und deren Codierentscheidungen angelegt (Tabelle 4).

Buchveröffentlichungen der Berufsklassifikationen

- International Labour Office (Ed.): International Standard Classification of Occupations: ISCO-88, Geneva 1990.
- Classification of Occupations des Internationalen Arbeitsamtes, Genf 1968, deutsche Ausgabe 1968, Verlag W. Kohlhammer, Stuttgart und Mainz 1971.
- Statistisches Bundesamt Wiesbaden (Hrsg.): Klassifizierung der Berufe. Systematisches und alphabetisches Verzeichnis der Berufsbenennungen. Ausgabe 1975. Stuttgart und Mainz 1975.

- Statistisches Bundesamt (Hrsg.): Klassifizierung der Berufe, systematisches und alphabetisches Verzeichnis der Berufsbezeichnungen, Ausgabe 1992, Verlag Metzler Poeschel, Stuttgart 1992.

Zusatzinformationen

Alle Informationen, die über die reine Berufsangabe hinausgehen, werden Zusatzinformationen genannt. Sie dienen dazu, den Beruf genauer zu spezifizieren. Meist werden sie im Zuge der statistischen Angaben in einem Fragebogen erfaßt, sie können aber auch aus anderen Quellen stammen. Im folgenden werden die wichtigsten aufgeführt.

Tabelle 4: Fallsammlung (Auszug)

ISCO 68	ISCO 88	Text
19-	2000	Wissenschaftliche Mitarbeiter
133	2300	Lehrer ohne nähere Angabe (wissenschaftliche Lehrkräfte, Lehrkräfte an Schulen)
13-	3300	nicht wissensch. Lehrkräfte, Nachhilfelehrer (meist Nebenerwerb)
49-	3000	Kaufleute (Handel)
39-	4000	Büroangestellte: diese Angabe ist nur als Büroberuf spezifizierbar (Hauptgruppe 3), weitere Angaben fehlen. Ebenso Bürofachkräfte, Sachbearbeiter o.n.A.
59-	5000	Dienstleistungsberufe
62-	6100	Pflanzenbau-/Tierwirtschaftsberufe (nicht zu unterscheiden, ob Pflanzenbau oder Tierwirtschaft) - meist Genossenschaftsbauern, LPG-Arbeiter
75-		z.B. Spinnereiarbeiter, denn dieser Beruf gehört zur Gruppe 75 (Spinner, Weber, Stricker usw.), kann aber dort nicht lokalisiert werden.
99-	7000	Arbeiter in einer Spinnerei, weil kein Facharbeiter; aber Spinnereiarbeiter (i.S.v. Facharbeiter) ist mit 75- zu codieren Handwerk./gütererzeugende Berufe Kennziffer 62 bis 64 (bei 64 nicht auf-sichts-führend, 61 unter Vorbehalt)
95-		z.B. Bauhilfsarbeiter, weil der Gruppe der Bauarbeiter zuzuordnen
99-	8000	Maschinenbediener o.n.A. (auch in Produktion)

- Kontext: Wenn z.B. Kinder den Beruf ihrer Eltern angeben, hat es eine andere Qualität als die Beschreibung des eigenen Berufs (die Angaben der Kinder sind meist konkreter auf den tatsächlichen Beruf bezogen, während man bei der Selbstangabe damit rechnen muß, daß nur der Ausbildungsberuf angegeben wurde, unabhängig davon, welche Tätigkeit aktuell ausgeübt wird). Zu wissen, wann und wo ein Beruf ausgeübt wurde, ob z.B. vor/nach dem Krieg, in der BRD oder in der DDR, kann eine Entscheidungshilfe darstellen, ebenso die Tatsache, daß es sich vielleicht um den Wunschberuf von Jugendlichen handelt (der sich einerseits an gängigen Mustern orientiert, andererseits auch Fantasievorstellungen enthalten kann). Diese und ähnliche In-

formationen finden sich im Fragebogen oder in der Projektbeschreibung. Sie beschreiben also den Kontext, in dem die Angabe interpretiert werden muß.

- Stellung im Beruf: Sie beschreibt nicht, wie oft fälschlich angenommen wird, den Beruf, sondern nur das Vertragsverhältnis und die Position des Berufs. Es gibt unterschiedlich differenzierte Vorgaben. Die ausführlichere Liste, wie sie in der Standarddemographie und den „Demographischen Standards“ für Deutschland (StaBu 1999) enthalten ist und z.B. im ALLBUS verwendet wird, gibt Hinweise auf die Hierarchie im Berufsleben, die Ebene der Beamtenlaufbahn, das Ausmaß der Entscheidungautonomie, das Ausmaß der Eigenverantwortlichkeit des Handelnden, ob der Befragte selbständig ist, wieviele Mitarbeiter er hat, wie groß der landwirtschaftliche Betrieb ist, u.a.m. Für viele Berufskategorien sind dies Merkmale, die darüber entscheiden, welche der möglichen Spezifikationen zutrifft. Ein selbständiger Heizungsinstallateur, der alleine arbeitet, wird z.B. anders codiert als jemand, der 10 Mitarbeiter hat.

- Branche/Rechtsform (z.B. öffentlicher Dienst): Die Wirtschaft und damit die Berufswelt wird zu ihrer Beschreibung nach Branchen/Wirtschaftszweigen gegliedert, wie z.B. Landwirtschaft, Bau, Chemie, Handel, Gesundheitswesen usw. Die Information darüber kann eine zunächst nicht codierbare Angabe zumindest teilweise präzisieren. Ein „Holzarbeiter“ in der Forstwirtschaft erhält einen anderen Code als einer in der Möbelherstellung; und ohne nähere Angaben kann der Holzarbeiter nur in der Berufshauptgruppe (Einsteller) codiert werden. In der KldB (Klassifikation der Berufe) des Statistischen Bundesamtes finden die meisten Spezifizierungen über die Branchenangabe statt. Für viele Codierentscheidungen spielt es eine Rolle, ob die Person in der Privatwirtschaft oder im Öffentlichen Dienst beschäftigt ist.

- Betriebsgröße/Anzahl der Mitarbeiter/Untergebenen: Je nach Größe des Betriebs muß unterschiedlich codiert werden, besonders in den ISCO-88-Kategorien. Außerdem läßt die Größe des Betriebs Rückschlüsse auf die Größe des Verantwortungsbereichs zu, der maßgeblich eine Reihe von Kategorien definiert, besonders die der ISCO-Hauptgruppen 1 und 2.

- Schulbildung: Bestimmte berufliche Positionen und Tätigkeiten setzen bestimmte Bildungs- und Ausbildungsqualifikationen voraus (als „skill level“ in der ISCO-88 beschrieben). Läßt die Berufsangabe mehrere Codiermöglichkeiten zu, so hat es eine gewisse Plausibilität, den zu der Schulbildung passenden Beruf zu bevorzugen. Aber Vorsicht: diese Schlußfolgerung ist nicht grundsätzlich zulässig, denn je nach Arbeitsmarktlage ist Überqualifizierung keine Seltenheit. Eine gute Übersicht zur „Grundstruktur des Bildungswesen in der Bundesrepublik Deutschland“ findet sich bei Mohr (1991).

– Berufsabschluß: Auch hier müssen Berufsausbildung und ausgeübter Beruf nicht unbedingt übereinstimmen, dennoch erhält man mit dieser Angabe eine weitere Beschreibung der befragten Person.

- Einkommen: Auch das Einkommen kann (muß aber nicht) ein guter Indikator für die Position in der Berufshierarchie sein. Wenn „Promotor“ als berufliche Tätigkeit mit einem monatlichen Verdienst von unter 2000 DM angegeben wird, dann dürfte es sich wahrscheinlich um Werbetätigkeit auf der Straße oder im Kaufhaus, aber nicht um die Gestaltung und Organisation des Vertriebs handeln.

Tabelle 5 zeigt, in welchem Umfang die Zusatzinformationen im Durchschnitt bei nicht eindeutigen Angaben für die Codierentscheidung genutzt werden.

Tabelle 5: Nutzung der Zusatzinformationen

Stellung im Beruf	33%
Branche	28%
Ausbildung	18%
Betriebsgröße	9%
Sonstige Informationen	12%

Arbeitsschritte

In den Veröffentlichungen finden sich so gut wie keine Hinweise, wie die Codierung der Berufsangaben im einzelnen durchzuführen ist. Wenn ein Codierer mit der Angabe des Befragten zunächst nichts anfangen kann, weil es ein seltener Beruf ist oder dem Codierer die einschlägigen Erfahrungen fehlen, stellen die bloßen Auflistungen und Beschreibungen der Kategorien eine geringe Hilfe dar. Im folgenden finden sich einige Hinweise, wie auch relativ ungeübte Codierkräfte eine Codierung durchführen können.

Die aufgeführten Stichworte kennzeichnen die einzelnen Stufen, bis endgültig ein Code vergeben werden kann, angefangen bei dem einfachsten Fall, daß sich die Angabe im Wörterbuch findet und damit der Fall abgeschlossen ist, bis hin zu dem Ergebnis, daß nach allen möglichen Recherchen und Bemühungen eine Codierung nicht möglich ist und nur der Code für fehlende Werte vergeben werden kann. Die Hinweise beziehen sich fast ausschließlich auf die ISCO-Codierung. Die Codierung nach der Klassifikation der Berufe (KldB92) erfordert in viel geringerem Ausmaß Regeln, da sie nach Listen erfolgt, die (theoretisch) jede Berufsbezeichnung enthalten.

- Meist dürfte ohnehin automatisch codiert worden sein, so daß nur noch die Angaben zu codieren sind, die nicht im Wörterbuch aufgeführt sind: die sogenannten Leftover-Fälle. Aber auch dann lohnt sich ein Blick in das *Berufewörterbuch*. Sind le-

diglich Schreibfehler oder abweichende Schreibweisen der Grund für die nicht automatische Codierung, kann der Code direkt aus dem Wörterbuch übernommen werden. Ansonsten muß man sich bei ähnlichen, aber nicht identischen Begriffen zuerst vergewissern, ob sich hinter der abweichenden Bezeichnung auch eine andere Tätigkeit verbirgt. „Bürohilfskraft“ ist eine definierte Tätigkeitsbeschreibung, was für „Hilfe im Büro“ nicht zutrifft.

- Ist die Berufsbezeichnung nicht bekannt, und es fehlt jeder Ansatzpunkt zur Suche nach der passenden Kategorie, so kann das *alphabetische Verzeichnis der Berufsbezeichnungen* des Statistischen Bundesamtes (1992) als erste Orientierung dienen, um sich über das Berufsumfeld zu informieren. Es enthält knapp 30.000 Berufsbezeichnungen mit Codes nach der KldB92. Ein Umsteigeschlüssel von dem dreistelligen Code der KldB 92 zur ISCO-88 liefert zwar nicht immer das richtige Ergebnis, wohl aber einen oder mehrere Vorschläge einer möglichen Verortung oder Angabe in der ISCO-88. Über den angegebenen Code findet man darüber hinaus die Liste aller Berufe, die mit diesem StaBu-Code zusammengefaßt werden, so daß eine Ähnlichkeit der Tätigkeiten unterstellt werden kann. So kann zwar nicht der angegebene, aber vielleicht ein ähnlicher Beruf in der ISCO-Klassifikation gefunden werden
- Als nächstes sind die *Codierregeln, Konventionen und Fallsammlung* zu Rate zu ziehen. Diese Sammlung enthält die Angaben, für die zwar eine Regel aufgestellt werden kann, die sich aber nicht als Wörterbucheinträge dokumentieren lassen, weil z.B. weitere Bedingungen abgefragt werden müssen, und sei es auch die, ob weitere Informationen vorliegen. Zu den Codierregeln werden auch die Ausführungen in den Einleitungskapiteln der jeweiligen Klassifikationen über die Grundprinzipien gerechnet (StaBu 1971 und 1992, ILO 1990). Es empfiehlt sich, sich diese von Zeit zu Zeit immer wieder mal zu vergegenwärtigen.
- Ausführliche Klassifikation/Codebuch: Hiermit sind die *Veröffentlichungen* vom StaBu (1992) und ILO (1990) gemeint, in denen jede Berufskategorie dargestellt wird. Diese Beschreibungen und Definitionen müssen dann nachgelesen werden, wenn es es sich um selten genannte Berufe handelt, für die sich noch keine Codier-Routinen herausgebildet haben, oder deren Bedeutungsfeld unklar ist.
- Im Laufe der Jahre wurden beim ZUMA mehrere Hunderttausend Berufsvercodungen durchgeführt. Diese Codierentscheidungen wurden zusammen mit den Texten und den Randbedingungen (Zusatzinformationen) in einer *Datenbank* zusammengetragen. Das Projekt ist noch nicht abgeschlossen; mittlerweile stehen über 100.000 Codierungen als Beispiele zur Verfügung.
- Die Codierregeln und Definitionen der Standardklassifikationen liefern u.U. den Hinweis, daß mehr Informationen als nur die Berufangabe gebraucht werden (vgl.

Abschnitt 4.1.3 - *Zusatzinformationen*), daß eine Kategorie nur dann vergeben werden darf, wenn bestimmte Bedingungen erfüllt sind: z.B. Größe des Betriebs oder eine bestimmte Branche. In der Regel erhält man diese Informationen aus den Umfragedaten, als eigens erhobene Variable oder erschließbar aus der Untersuchungsanlage (Population, Ort u.ä.).

Die Berufsangabe „Altenpfleger“ kann z.B. vier verschiedenen Kategorien der ISCO-68 zugewiesen werden, je nachdem, ob die Tätigkeit im fürsorglichen Bereich (dann Code 193), in nicht privaten Heimen (Code 599), in Privathaushalten (Code 540) oder in der Krankenpflege (Code 071) ausgeübt wird. Beispiel Selbständige: Für fast alle Selbständigen in handwerklichen Berufen kommt es darauf an, wie groß der Betrieb ist, weil man davon ausgeht, daß der Besitzer in einem kleinen Betrieb die handwerkliche Tätigkeit selbst mit ausübt, in einem Großbetrieb aber mit Geschäftsführungsaufgaben beschäftigt ist, was jeweils eine unterschiedliche Tätigkeit darstellt: ISCO codiert Tätigkeiten, nicht aber Berufsbezeichnungen!

- Wenn trotz Hinzuziehen aller vorhandenen Informationen und Angaben sich keine eindeutige Entscheidung abzeichnet, dann kann dennoch die eine oder andere Alternative aufgrund von *Erfahrungswerten* als die „wahrscheinlichere“ gelten, weil man beispielsweise aus anderen Studien (mit umfangreicheren Informationen) weiß, daß für 95% und mehr derartiger Angaben immer eine bestimmte Tätigkeit zutrifft. Mit der Angabe „in der Gastronomie beschäftigt“ wird z.B. in den allermeisten Fällen die Tätigkeit einer Bedienung beschrieben.
- Führen keine der aufgeführten Vorgehensweisen oder Maßnahmen zum Erfolg, kann man sich manchmal weiterhelfen, indem bei den entsprechenden Firmen nachgefragt, im Lexikon oder im Branchenverzeichnis nachgeschlagen oder bei Fachleuten Auskunft eingeholt wird u.ä.m.
- Als letztes bleibt die Einstufung als *nicht codierbare Angabe*. Die vorliegenden Informationen und Erfahrungen reichen nicht aus, die Angabe auch nur unvollständig (nur Ein- oder Zweisteller) zu codieren. Der Berufecode ist in diesem Fall immer 004 bzw. 0004.

Empirisch kommen die oben genannten Maßnahmen in folgendem Umfang zur Anwendung (siehe Tabelle 6).

3.2 Computerunterstützte Berufsvercodung

Voraussetzungen

Computerunterstützte Inhaltsanalyse (cui) stellt im Prinzip dieselbe Methode wie die konventionelle Codierung dar. Allerdings wird ein Teil der Codierarbeit, nämlich die

Codezuweisung, mit Hilfe des Computers erledigt. Diese Hilfe bleibt jedoch auf die Untersuchungsgegenstände beschränkt, die sich als regelmäßige physikalisch-materielle Vorgänge beschreiben lassen. Bedeutung oder Sinn sind für den Computer nicht zu verarbeiten.

Tabelle 6: Bedeutung der einzelnen Maßnahmen (in Prozent)

Automatisch codiert bzw. im Wörterbuch	50%
nicht eigens im Wörterbuch, aber erschließbar	15%
Regeln, Konventionen, Fallsammlung (schließt meist die Berücksichtigung der Zusatzinformationen ein)	20%
Veröffentlichung des StaBu und ILO	3%
Zusatzinformationen werden herangezogen	14%
Erfahrung	2%
Nicht codierbar	4%

Die Maßnahmen schließen sich gegenseitig nicht aus, so daß die Prozentwerte über 100 ergeben.

3.2 Computerunterstützte Berufsvercodung

Voraussetzungen

Computerunterstützte Inhaltsanalyse (cui) stellt im Prinzip dieselbe Methode wie die konventionelle Codierung dar. Allerdings wird ein Teil der Codierarbeit, nämlich die Codezuweisung, mit Hilfe des Computers erledigt. Diese Hilfe bleibt jedoch auf die Untersuchungsgegenstände beschränkt, die sich als regelmäßige physikalisch-materielle Vorgänge beschreiben lassen. Bedeutung oder Sinn sind für den Computer nicht zu verarbeiten.

Die Fälle, in denen weitere Informationen und Überlegungen herangezogen werden müssen, sind für die cui zunächst nicht geeignet, es sei denn, man wollte für jede denkbare sprachliche Variante der Befragten einen Algorithmus schreiben, der den Entscheidungsprozeß des Codierens nachvollzieht - dies widerspricht jedem Rentabilitätsgedanken. Anders verhält es sich dagegen bei Angaben, in denen Stichworte/Begriffe vorkommen, deren Nennung allein ausreicht, um eindeutig die Kategorie bestimmen zu können (harter Indikator). Hier bot es sich an, eine Liste der eindeutigen Nennungen anzulegen, um die Codierung zu erleichtern. Dies wurde beim ZUMA bereits 1976 durch Klingemann und Schönbach begonnen (Klingemann/Schönbach 1984) und wird seither fortgesetzt. Ohne sich tiefgehend mit der Systematik auseinandersetzen zu müssen, braucht der Codierer nur noch das Stichwort ausfindig zu machen und den dazugehörenden Code zu übernehmen - genaugenommen eine intellektuell wenig anspruchs-

volle Arbeit. Und an diesem Punkt setzt die cui an, sie kann viel genauer und schneller nachschauen, ob ein Wort in der Liste (Wörterbuch oder Diktionär) steht und dann ggf. den entsprechenden Code übernehmen.

Es sind die Vorgänge, die für einen Menschen ermüdend und damit fehleranfällig sind, die nach dem Prinzip „immer wenn ..., dann ...“ ablaufen. Da jedoch Sprachverständnis und die Fähigkeit zu logischen Schlußfolgerungen oder Analogiebildungen nicht vorausgesetzt werden können, muß jeder nur denkbare Fall aufgezählt werden. Die Kategorien werden ausschließlich durch Listendefinitionen oder Beispielsammlungen definiert (vgl. Tabelle 7 und 8).

Das Wörterbuch/Diktionär ist das wichtigste Element der cui, es stellt die Operationalisierung schlechthin dar. Neben jedem Wörterbucheintrag steht der Code, mit dem dieses Wort in die Analyse eingeht. Die cui ist nur dann zur Codierung geeignet, wenn es gelingt, die zu erhebenden Merkmale mit Kategorien zu erfassen, die sich ausschließlich durch Listen definieren lassen. Gründe, Argumente oder Definitionen lassen sich mit der cui nicht codieren, wohl aber z.B. Themenstrukturen.

Bis auf das Pünktchen auf dem i muß der Wörterbucheintrag mit dem Wort in der Berufsangabe übereinstimmen, um als codierbar erkannt zu werden. So könnte es sich lohnen, auch falsche, aber häufig vorkommende Schreibweisen in dem Wörterbuch zu berücksichtigen. Man spricht von harten Indikatoren, d.h. von Begriffen, deren Auftreten allein ausreicht, um sie eindeutig einer Kategorie zuzuweisen zu können, unabhängig von Kontext, Kontext-Informationen oder sonstigen Bedeutungen des Wortes. Nur insoweit dies zutrifft, ist das Kategorienschema - hier das Wörterbuch - valide.

Wörterbuch

In den Anfängen der cui galt es, möglichst sparsam mit Speicherplatz und Rechenzeit umzugehen. Deshalb wurden nur die Berufsbezeichnungen in das Wörterbuch übernommen, die auch tatsächlich genannt und damit gebraucht wurden.

Heute spielt Speicherplatz und Rechenzeit keine Rolle mehr. Daher könnte man versucht sein, alphabetische Verzeichnisse, wie sie in den Anhängen der Dokumentationen zu den Berufsklassifikationen zu finden sind, ohne weitere Prüfung in das Wörterbuch zu übernehmen.

Aber: Alle Einträge müssen vor der Verwendung im Wörterbuch zumindest auf Validität geprüft werden. Darüber hinaus gibt es auch Gründe, gültige Einträge nicht wahllos in das Wörterbuch aufzunehmen, besonders wenn der nicht genutzte Anteil für die einzelne Studie beispielsweise über 90% beträgt, bzw. nur 10% davon genutzt werden.

Tabelle 7: „Definition“ der Kategorie 2148 (Kartographen und Vermessungsingenieure) in dem ISCO-88-Wörterbuch

2148 GEOMETER
2148 KREISVERMESSUNGSAHMANN
2148 LANDVERMESSER
2148 MARKSCHEIDER
2148 OBERMARKSCHEIDER
2148 STADTVERMESSUNGSAHMANN
2148 VERMESSUNGSAHMANN
2148 VERMESSUNGSDIRIGENT
2148 VERMESSUNGSINGENIEUR
2148 VERMESSUNGSOBERAHMANN
2148 VERMESSUNGSRAT
2148- REGIERUNGSVERMESSUNGS DIREKTOR
2148- REGIERUNGSVERMESSUNGSREFERENDAR
2148- STADTVERMESSUNGSOBERINSPEKTOR
2148- VERMESSUNGS DIPLOM INGENIEUR
2148- VERMESSUNGS DIREKTOR
2148- VERMESSUNGSINSPEKTOR
2148- VERMESSUNGSOBERINSPEKTOR
2148+ INGENIEUR FUER VERMESSUNG
2148+ INGENIEUR FUER VERMESSUNGSTECHNIK

Tabelle 8: „Definition“ der Kategorie 7212 (Schweißer und Brennschneider) in dem ISCO-88-Wörterbuch

7212 AUTOGENSCHWEISSER
7212 BLEILOETER
7212 KOLBENLOETER
7212 LEHRSCHEISSER
7212 SCHMELZSCHWEISSER
7212 SCHWEISSER
7212 SCHWEISSERIN
7212 SCHWEISSERLEHRLING
7212 SCHWEISSERMEISTER
7212 SCHWEISSERVORARBEITER
7212 SCHWEISSMEISTER
7212 VORSCHWEISSER
7212- AUTOMATENSCHWEISSER
7212- BETRIEBSSCHWEISSER
7212- BUNTSCHWEISSER
7212- EISENLOETER
7212- ELEKTROSCHWEISSER
7212- HOCHDRUCKSCHWEISSER
7212- KESSELSCHWEISSER
7212- LOETER
7212- METALLSCHWEISSER
7212- PUNKTSCHWEISSER
7212- ROHRSCHEISSER
7212- THERMITSCHWEISSER

Bei der Verwendung in mobilen Geräten, z.B. bei computerunterstützten persönlichen Interviews (CAPI), ist es fraglich, ob der hohe Speicherplatzbedarf von mehreren Zehntausend ungenutzten Antwortvorgaben für eine einzige Frage zu rechtfertigen ist. Die damit verbundenen längeren Suchzeiten verzögern den Interviewablauf.

Ein Wörterbuch muß gewartet werden, das bedeutet, von Zeit zu Zeit müssen die Einträge auf ihre Gültigkeit überprüft werden. Eine Veränderung in der Berufswelt oder in der Bezeichnung von Tätigkeiten macht neue Kategorienzuordnungen notwendig. Während z.B. „Manager“ und „Direktor“ vor 20 Jahren noch auf eine gehobene Position hindeuteten, werden diese Bezeichnungen heute inflationär gebraucht. Sortier- und Suchfunktionen erleichtern zwar die Arbeit, dennoch ist eine Datei mit 3.000 Zeilen einfacher zu handhaben als eine mit 30.000.

Auch außerhalb der Codierung mit Computern ist ein ausgedrucktes Wörterbuch eine wertvolle Hilfe bei der konventionellen Codierung (bei kleiner Fallzahl oder nicht maschinenlesbaren Berufsangaben), und eine Liste nur mit empirisch genutzten Angaben ist voll ausreichend, besonders weil durch Analogieschlüsse und Vergleiche der Nutzungsgrad viel höher sein wird als durch den Computer.

Fehler

Die wortbasierte automatische Codierung, die in der Berufsvercodung fast ausschließlich verwendet wird, setzt voraus, daß Mehrdeutigkeiten des einzelnen Wortes ausgeschlossen werden müssen. Dies ist weitgehend dadurch gewährleistet, daß die gestellte Frage den Rahmen für die Bandbreite der Antworten vorgibt. Ist der Kontext weiter gefaßt und werden die Antworten in Form eines ausführlichen Berichts gegeben, steigt die Fehlerrate der automatischen Codierung. Die Befragtenangabe „... und ich sage den Reisenden, wann die Züge abfahren.“ führt zu dem Code 432/3415, weil es einen Wörterbucheintrag „REISENDE“ im Sinne von Handlungsreisender gibt.

Während der letzten 15 Jahre wurden die Korrekturen der automatisch vergebenen Codes stichprobenweise ausgezählt: insgesamt waren von den 120.000 automatischen Codes im nachhinein 2.500 (2,1%) abgeändert worden. Die Veränderungen bezogen sich zum größten Teil auf die unterste Gliederungsebene oder betreffen Codealternativen, die im üblichen Entscheidungsspielraum liegen.

Effizienz

Wenn schon nach dem coderbasierten Verfahren bei der Berufsvercodung eine maximale Quote von ca. 80% vollständiger Codes erreicht wird, so kann die automatische Codierung auf keinen Fall einen höheren Codieranteil ergeben.

Die Erfahrungswerte auf der Basis von etwa 500.000 automatischen Codierungen nach ISCO liegen in den letzten 10 Jahren zwischen 45% und 55%. Angesichts der teilweise vielfältigen Überlegungen, die einer endgültigen Codierentscheidung vorausgehen, und der zu prüfenden Randbedingungen darf diese Erfolgsquote nicht unterbewertet werden. Eine deutliche Steigerung der Effizienz der automatischen Berufsvercodung versprechen wir uns von der Einbeziehung von Randbedingungen, wie sie in den sogenannten Zusatzinformationen zur Verfügung stehen.

Ein Programm müßte alle nicht mit dem Wörterbuch codierten Angaben darauf hin überprüfen, ob es sich um eine Beschreibung handelt, die durch sonstige zur Verfügung stehende Variablen näherer spezifiziert und codiert werden kann. Dazu muß für jede Berufsangabe festgelegt werden, bei welchen Variablenwerten (oder Kombinationen davon) welcher Code vergeben werden soll, etwa im Sinne eines Entscheidungsbaumes. Das bedeutet eine sehr umfangreiche Programmierarbeit, die erfordert, daß die Datenformate und Variablen standardisiert sind wie z.B. in der "Standard-Demographie". Entsprechende Bemühungen werden derzeit von Burke Infratest Sozialforschung und ZUMA unternommen. Hierüber soll in einer der nächsten Ausgaben der ZUMA-Nachrichten berichtet werden.

Literatur

- Bundesanstalt für Arbeit (Hrsg.), 2000: Beruf Aktuell, Ausgabe 2000/2001. Nürnberg.
- Bundesanstalt für Arbeit (Hrsg.), 1990: DDR-Ausbildungsberufe. Vergleichbare und verwandte Berufe in der Bundesrepublik Deutschland, 5 Bände. Nürnberg.
- International Labour Organisation (Ed.), 1968: International Standard Classification of Occupations, Revised Edition 1968. Genf.
- International Labour Organisation (Ed.), 1990: ISCO-88: International Standard Classification of Occupations. Genf.
- Friedrichs, J., 1973: Methoden empirischer Sozialforschung. Reinbek bei Hamburg: Rowohlt.
- Früh, W., 1998: Inhaltsanalyse. Theorie und Praxis. Konstanz: UVK Medien.
- Klingemann, H.-D., 1984: Computerunterstützte Inhaltsanalyse in der empirischen Sozialforschung. Frankfurt/Main: Campus Verlag GmbH.
- Merten, K., 1995: Inhaltsanalyse. Einführung in Theorie, Methode und Praxis. Opladen: Westdeutscher Verlag GmbH.
- Mohr, B., 1991: Bildung und Wissenschaft in Deutschland West. Köln: Bundesanzeiger.
- Molle, F., 1975: Wörterbuch der Berufs- und Berufstätigkeitsbezeichnungen. Grenzland-Verlag-Wolfenbüttel.

MPI Berlin für Bildungsforschung (Hrsg.), 1994: Bildungssystem in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick. Reinbek.

Schönbach, K. 1977: Berufsverschlüsselung. ZUMA-Nachrichten 1: 18 - 20.

Statistisches Bundesamt (Hrsg.), 1971: Systematische Verzeichnisse: Internationale Standardklassifikation der Berufe, Deutsche Ausgabe 1968. Stuttgart und Mainz: Verlag W. Kohlhammer.

Statistisches Bundesamt (Hrsg.), 1975: Klassifizierung der Berufe: Systematisches und alphabetisches Verzeichnis der Berufsbenennungen. Stuttgart: Metzler-Poeschel.

Statistisches Bundesamt (Hrsg.), 1992: Personensystematik. Klassifizierung der Berufe: Systematisches und alphabetisches Verzeichnis der Berufsbenennungen. Stuttgart: Metzler-Poeschel.

Statistisches Bundesamt (Hrsg.), 1999: Demografische Standards. Methoden - Verfahren - Entwicklungen. Materialien und Berichte. Wiesbaden.

AUSWERTUNG VON ALLBUS-DATEN IM WWW

Einfache Analysen mit ALLBUS-Datensätzen können nun auch online durchgeführt werden. Zusätzlich zur bewährten Online-Codebuchrecherche gibt es somit die Möglichkeit, sich durch Häufigkeitsauszählungen einen Einblick in die Daten zu verschaffen. Der Zugriff auf die Daten erfolgt über einen Webbrowser.

Zur online-Analyse stehen alle ALLBUS-Befragungen seit 1991 zur Verfügung. Bei jeder ALLBUS-Studie kann zwischen dem Datensatz für Westdeutschland (inklusive West-Berlin), für Ostdeutschland (inklusive Ost-Berlin) oder dem Gesamtdatensatz gewählt werden. Für die gesamtdeutschen Datensätze ist eine personenbezogene Ost-West-Gewichtung voreingestellt, welche die Überrepräsentation der ostdeutschen Bevölkerung in der Gesamtstichprobe ausgleicht.

Die Details des Auswertungssystems sind in der Mitteilung „badason - basic data analysis system on-line“ beschrieben (Seite 130-134 in diesem Heft).

Den Zugang zu online-Auswertungen von ALLBUS-Datensätzen mit *badason* finden Sie unter dem entsprechenden Punkt beim WWW-Angebot des ALLBUS:

<http://www.zuma-mannheim.de/data/allbus/>

oder direkt unter:

http://badason.za.uni-koeln.de/Online_Analysis_ALLBUS.htm

Weitere Auskünfte zum ALLBUS erhalten Sie von:

Achim Koch

Zentrum für Umfragen, Methoden und Analysen

E-Mail: allbus@zuma-mannheim.de

Telefon: 0621 / 1246-280

Für Datenbestellungen wenden Sie sich bitte an:

Dr. Michael Tervey

Zentralarchiv für Empirische Sozialforschung

E-Mail: tervey@za.uni-koeln.de

Telefon: 0221 / 47694-22

BADASON - BASIC DATA ANALYSIS SYSTEM ON-LINE

EIN SYSTEM ZUR EINFACHEN DATENAUSWERTUNG ÜBER DEN WEBBROWSER AM BEISPIEL DES ALLBUS 1991-1998²³

Das Ziel der Online-Analyse mit dem Programm BADASON ist die einfache Erstellung von Tabellen bzw. Säulendiagrammen über einen Webbrowser. Damit können einerseits Interessenten einen Datensatz schnell kennenlernen und danach entscheiden, ob dieser Datensatz für komplexere Analysen benötigt wird. Andererseits können Nutzer ohne Erfahrung mit Statistikprogrammen auf diesem Weg einfache Auswertungen vornehmen.

Leistungsumfang

Nach Auswahl der Merkmale aus der Variablenliste können Kreuztabellen (bivariate Tabellen), entsprechende Statistiken und Säulendiagramme erstellt werden. Die Analyseergebnisse werden in üblichen Webdateiformaten angezeigt und können in weiteren Dateiformaten zur Weiterverarbeitung ausgegeben werden. Diese Ergebnisse werden nach den spezifischen Nutzerangaben auf der Serverseite dynamisch erzeugt. Außerdem können Häufigkeitsauszählungen (univariate Tabellen) und entsprechende Säulendiagramme abgerufen werden. Diese Ergebnisse liegen auf der Serverseite als schon erzeugte Dateien vor.

Abbildung 1: Abfrageformular

The screenshot shows a web browser window titled "badason - Basic Data Analysis System on-line - Microsoft Internet Explorer". The address bar shows "http://badason.za.uni-koeln.de/Online_Analysis_ALLBUS.htm". The main content area displays a form for data analysis. At the top, it says "GESIS | ALLBUS Home | ALLBUS 1998" with "Reset" and "Help" buttons. Below this, there is a dropdown menu for selecting a variable, currently set to "ALLBUS 1998 - S3000". There are search fields and buttons for "Text to search", "Search", and "(List all var.)". On the right side, there are options for "Show var. freq.", "Show codebk.", "Select row", and "Select column". The "Show var. freq." section has radio buttons for "East", "West", and "Total", with "Total" selected. The "Show codebk." section has a checked checkbox for "East-West-Weight". The "Select row" section has checkboxes for "% Row", "Col", and "Tot", with "% Row" selected. The "Select column" section has a "COMPUTE TABLE" button. At the bottom left, it says "b a d a s o n - basic data analysis system on-line (0.02b)".

²³ Die Idee zu *badason* ist in einem GESIS-Gemeinschaftsprojekt entstanden. Konzeption und Programmentwicklung wurde von Joachim Wackerow durchgeführt.

Die Funktionen im einzelnen:

Benutzungsmöglichkeiten

- Auswahl der Studie, bei ALLBUS ist dies die Auswahl des Jahrgangs.
- Anzeige der Variablenliste der ausgewählten Studie.
- Suche in der Variablenliste (Zeichenkettensuche).
- Filtermöglichkeit nach einer Subpopulation, bei ALLBUS ist dies Ost- bzw. Westdeutschland.
- Auswahl von Fällen (Filter) mit beliebigen Ausprägungen eines Merkmals.
- Ein- bzw. Ausschalten einer Gewichtungvariable, bei ALLBUS ist dies das personenbezogene Ost-/Westgewicht.

Ausgabemöglichkeiten

- Anzeige der Häufigkeiten einer Variable (univariate Tabelle) mit statistischen Kennzahlen und Säulendiagramm für die Gesamtheit und aufgeteilt nach Subpopulationen, wie z.B. beim ALLBUS in Ost- bzw. Westdeutschland.
- Bei ALLBUS: Anzeige des Codebuchs einer Variable vom WWW-Server des Zentralarchivs (univariate Tabelle) und damit Übergang zur Codebuchrecherche ALLSEA (Terwey 1998, 1999).
- Erstellung einer Kreuztabelle (bivariate Tabelle) und Auswahl der gewünschten Prozentwerte.
- Zu den Kreuztabellen werden übliche Korrelations- bzw. Assoziationsmaße ausgegeben, wie z.B. Chi-Square Tests, Cramer's V, Gamma, Eta und Pearson's R.
- Erstellung eines entsprechenden Säulendiagramms. Die Säulen sind gestapelt nach der Zeilenvariable, mit Prozentwerten in normalisierter Darstellung.
- Ausgabe der Kreuztabelle und der Grafik in üblichen Webdateiformaten, im PDF-Format zum Drucken und auch im Word- und im SPSS-Format zur Weiterverarbeitung.
- Ausgabe der entsprechenden SPSS-Kommandos für weitere Analysen auf dem eigenen Rechner.
- Ausgabe der Daten der Kreuztabelle zur Weiterverarbeitung in anderen Programmen.
- Ausgabe des Säulendiagramms als rotierbare dreidimensionale Grafik für explorative Vorgehensweise (experimentelle Funktion), VRML-Plugin erforderlich.

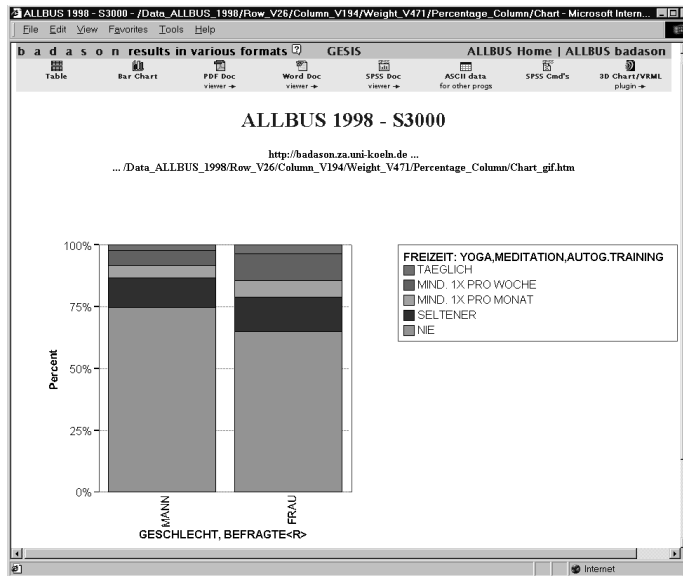
Abbildung 2: Kreuztabelle

ALLBUS 1998 - S3000

V26 FREIZEIT: YOGA,MEDITATION,AUTOG.TRAINING
by V194 GESCHLECHT, BEFRAGTE<R>

		V194 GESCHLECHT, BEFRAGTE<R>		Total	
		1 MANN	2 FRAU		
V26 FREIZEIT: YOGA,MEDITATION,AUTOG.TRAINING	1 TAEGLICH	Count	33	51	84
		% within V194 GESCHLECHT, BEFRAGTE<R>	2.2%	3.0%	2.6%
	2 MIND. 1X PRO WOCHE	Count	91	179	270
		% within V194 GESCHLECHT, BEFRAGTE<R>	6.0%	10.4%	8.4%
	3 MIND. 1X PRO MONAT	Count	75	107	182
		% within V194 GESCHLECHT, BEFRAGTE<R>	5.0%	6.2%	5.6%

Abbildung 3: Säulendiagramm



Anpassungen für den ALLBUS

Die ostdeutsche Bevölkerung ist in der ALLBUS-Stichprobe überrepräsentiert, um ausreichende Fallzahlen für Subgruppenanalysen zu gewährleisten. Das bedeutet, daß die Auswertungen für Gesamtdeutschland gewichtet durchgeführt werden müssen. Sobald ein Gesamtdatensatz für die Online-Auswertung ausgewählt ist, wird die Ost-West-Gewichtung als (abschaltbare) Voreinstellung wirksam. Werden dagegen entweder nur die Ost-Stichprobe oder nur die West-Stichprobe analysiert, wird die Ost-West-Gewichtung deaktiviert.

Ein einfacher Übergang zur Codebuchrecherche ALLSEA (Terwey 1998, 1999) ist realisiert.

Zugang und Bedienung

Die Benutzungsoberfläche von *badason* wurde mit Netscape Navigator Version 4.x und Microsoft Internet Explorer Version 4.x/5.x getestet. Zur Benutzung von *badason* müssen im Webbrowser Javascript (nicht Java) und Stylesheets eingeschaltet sein. Dies muß beim Webbrowser im Einstellungsmenü (Options bzw. Preferences) getan werden, ist aber meistens schon voreingestellt.

Die Ergebnisse werden jeweils in einem neuen Fenster geöffnet. Das bietet den Vorteil, daß mehrere Ergebnisse miteinander verglichen werden können.

Die folgenden drei Abschnitte beschreiben die Seite des Datenanbieters bzw. des Datenverwalters.

Verwaltung des Systems

Weitere SPSS-Datensätze können schnell im System aufgenommen werden. Sie müssen nur durch wenige Parameter in einer zentralen Konfigurationsdatei beschrieben werden.

Es steht ein Verwaltungswerkzeug über den Webbrowser zur Verfügung. Damit können alle wichtigen Konfigurations- und Log-Dateien, sowie die zu *badason* gehörigen Server-Prozesse angezeigt werden. Außerdem kann eine aktuelle Zugriffsstatistik erzeugt werden.

Technische Einzelheiten

Die bivariaten Tabellen und dazugehörigen Grafiken werden dynamisch erzeugt. Die Ergebnisse werden in einem Cache auf der Serverseite gehalten. D.h., wenn ein bestimmter Job einmal gerechnet wurde, können die Ergebnisse ohne erneutes Rechnen

(ohne Wartezeit) wieder abgerufen werden. Der Cache ist natürlich im Platz begrenzt. Diejenigen Dateien werden im Cache gespeichert, die jünger sind bzw. häufiger benutzt werden.

Alle univariaten Tabellen und Säulendiagramme werden statisch in Dateien auf dem Server gespeichert und sind ohne Wartezeit abrufbar.

Bei Einsatz von vielen Datensätzen kann das System schnell um weitere Rechner zu einer Serverfarm erweitert werden. Der Cache-Server wäre dann weiterhin die einzige zentrale Schnittstelle zum WWW für eine Vielzahl von Datensätzen.

Die derzeitige Version von *badason* ist noch in einem Beta-Stadium. Die größte augenblickliche Einschränkung ist die Länge der Rechenzeit für einen Job: ca. 60 Sekunden benötigt der jetzt benutzte Rechner.

Verfügbarkeit des Systems

badason ist frei verfügbar und kann für beliebige SPSS-Datensätze mit numerischen Daten zur Online-Analyse eingesetzt werden. Es werden eine Reihe von frei verfügbaren und von kommerziellen Programmen auf einem Rechner unter Windows NT vorausgesetzt. Die aktuelle Version 0.02b enthält die oben erwähnten ALLBUS-spezifischen Erweiterungen.

Die ALLBUS-Datensätze (1991-1998) können Sie auswerten mit *badason* unter:

http://badason.za.uni-koeln.de/Online_Analysis_ALLBUS.htm

Weitere Informationen zu *badason* können Sie abrufen unter:

<http://www.zuma-mannheim.de/software/en/badason/>

Nähere Auskünfte zu *badason* erhalten Sie von:

Joachim Wackeron
Zentrum für Umfragen, Methoden und Analysen
E-Mail: wackeron@zuma-mannheim.de
Telefon: 0621 / 1246-262

Literatur

Terwey, M., 1998: Eine umfassende Codebuch- und Dokumentationsrecherche für den ALLBUS im WWW. ZA-Information 42: 53-55. Köln.

Terwey, M., 1999: ALLBUS 1998: Erweiterung des Studienangebots. ZA-Information 44: 158-164. Köln.

VERDIENSTKREUZ AM BANDE FÜR PROF. DR. MAX KAASE

Am Dienstag, 12. September 2000, erhielt der Vorsitzende der ZUMA-Mitgliederversammlung, Prof. Dr. Max Kaase, das Verdienstkreuz am Bande der Bundesrepublik Deutschland. Es wurde ihm vom Staatssekretär für Wissenschaft und Forschung des Landes Berlin, Dr. Josef Lange, überreicht.

Nach Abschluß seines Studiums der politischen Wissenschaft war Prof. Dr. Kaase zunächst als Projektleiter Marktforschung in Frankfurt, als wissenschaftlicher Assistent an der Universität zu Köln und dann an der Universität Mannheim tätig. 1993 wurde er als Forschungsprofessor an das Wissenschaftszentrum Berlin für Sozialforschung berufen.

Prof. Dr. Kaase ist einer der führenden deutschen Politik- und Sozialwissenschaftler auf dem Gebiet der politischen Soziologie. Vor allem mit seinen vergleichenden Politik- und Gesellschaftsanalysen trug er wesentlich zur Internationalität der sozialwissenschaftlichen Forschung in Deutschland bei. Verdient gemacht hat sich Prof. Dr. Kaase jedoch nicht nur als Wissenschaftler, sondern auch durch sein Engagement in Vorständen und Beiräten wissenschaftlicher Fachgesellschaften sowie in Gremien, etwa in der Deutschen Forschungsgemeinschaft und im Wissenschaftsrat. Besondere Anerkennung erwarb er sich auf europäischer Ebene. Unter seiner Leitung legte eine international besetzte Wissenschaftlergruppe der Europäischen Kommission 1997 ein weit beachtetes Papier vor, die „Bestandsaufnahme der Stärken und Schwächen der Wirtschafts- bzw. Sozialwissenschaften in Westeuropa“.

Als Zeichen seiner außergewöhnlichen Kompetenzen wurde Prof. Dr. Kaase 1999 zum Vizepräsidenten der europäischen Wirtschaftsstiftung gewählt. Im Mai 2000 beendete Prof. Dr. Kaase seinen Dienst als Forschungsprofessor am Wissenschaftszentrum für Sozialforschung in Berlin. Seit September ist er Vizepräsident und Dean der neugegründeten International University Bremen.