

108

Data Documentation

Anonymisierungskonzept für den Scientific Use File
zur Studie „Erwartungen an Wirtschaft und Gesellschaft“
(LINOS-1)

IMPRESSUM

DIW Berlin, 2024

DIW Berlin

Deutsches Institut für Wirtschaftsforschung

Mohrenstr. 58

10117 Berlin

Tel. +49 (30) 897 89-0

Fax +49 (30) 897 89-200

www.diw.de

ISSN 1861-1532

All rights reserved.

Reproduction and distribution

in any form, also in parts,

requires the express written

permission of DIW Berlin.

Data Documentation 108

Jule Adriaans¹

Tobias Gebel²

Anonymisierungskonzept für den Scientific Use File zur Studie „Erwartungen an Wirtschaft und Gesellschaft“ (LINOS-1)

Berlin, April 2024

¹ Universität Bielefeld, Fakultät für Soziologie, jule.adriaans@uni-bielefeld.de

² DIW Berlin, tgebel@diw.de

Inhalt

1. Ausgangslage	1
2. Kurzbeschreibung der Studie	1
3. Erstellung des Scientific Use Files	2
3.1. Angewandte Regeln zum Zweck der faktischen Anonymisierung	2
3.2. Löschung von Merkmalen	3
3.3. Aggregation von Merkmalen	3
3.4. Änderungen im Einzelnen	4
4. Datenzugang	6

1. Ausgangslage

Bei der Studie „Erwartungen an Wirtschaft und Gesellschaft“³ handelt es sich um eine bundesweite Befragung von 4.731 Personen im Zeitraum Oktober 2012 bis April 2013 zum Thema der Herausbildung von Gerechtigkeitseinstellungen über den Lebensverlauf. Die Erhebung wurde im Rahmen des Sonderforschungsbereichs 882 der Universität Bielefeld im Teilprojekt A6 „Die Legitimation von Ungleichheiten: Strukturelle Bedingungen von Gerechtigkeitseinstellungen über den Lebensverlauf“ durchgeführt. Die Stichprobenziehung basiert auf Meldedaten der Bundesagentur für Arbeit (BA).

Gemäß den Prinzipien guter wissenschaftlicher Praxis⁴ werden die Befragungsdaten durch das Forschungsdatenzentrum Betriebs- und Organisationsdaten (FDZ-BO) am DIW Berlin in anonymisierter Form als Scientific Use File weiteren WissenschaftlerInnen zur Verfügung gestellt.

Das vorliegende Dokument beschreibt das Konzept zur Erstellung des Scientific Use Files.

2. Kurzbeschreibung der Studie

Ziel der Studie ist es zu untersuchen, warum Menschen bestimmte Gerechtigkeitseinstellungen haben. Als entscheidend werden dabei die u.U. wechselnden sozialen Kontexte – insbesondere Haushalte, soziale Netzwerke oder Betriebe – angesehen, in denen Personen über ihren Lebensverlauf eingebunden sind. Denn soziale Kontexte vermitteln Möglichkeiten des sozialen Vergleichs und des sozialen Lernens, die für die Ausbildung von Gerechtigkeitseinstellungen entscheidend sind.

Die Erhebung der Daten erfolgte auf drei Arten:

1. Ohne Interviewer: Papierfragebögen (PAPI, paper and pencil interview)
2. Ohne Interviewer: Web-Interviews (CAWI, computer assisted web interview)
3. Mit Interviewer: Persönliche Interviews (CAPI, computer assisted personal interview)

Der Befragung liegen zwei separate Teilstichproben zugrunde; getrennt für den PAPI/CAWI-Modus und für den CAPI-Modus. Die Ziehung beider Teilstichproben erfolgte aus denjenigen Personen, die am 31.12.2011 sozialversicherungspflichtig beschäftigt sowie im Alter zwischen 19 und 59 Jahren waren. Die Datengrundlage bildete eine 6-Monats-Datei der Meldedaten der Bundesagentur für Arbeit (BA). Für beide Teilstichproben wurde eine disproportionale Schichtung (drei Altersklassen und zwei Betriebszugehörigkeitsklassen) vorgenommen, um in den theoretisch interessanten Befragengruppen im Längsschnitt ausreichende Zellenbesetzungen zu gewährleisten. Beschäftigte mit geringerer Betriebszugehörigkeit erhielten eine höhere Ziehungswahrscheinlichkeit. Auf Basis von 37.986 Adressen wurden 4.731 Interviews realisiert (siehe Tabelle 1).

³ Liebig, Stefan; May, Meike; Sauer, Carsten; Schneider, Simone & Valet, Peter (2014): „Erwartungen an Wirtschaft und Gesellschaft“ SFB 882, doi: 10.4119/unibi/sfb882.2014.9

⁴ Betancort, Noemi; Bongartz, Elke C.; Dörrenbächer, Nora; Goebel, Jan & Kaluza, Harald; Siegers, Pascal (2020): „White Paper on implementing the FAIR principles for data in the Social, Behavioural, and Economic Sciences“ *RatSWD Working Paper Series 274*. URL: https://www.konsortswd.de/wp-content/uploads/RatSWD_WP_274.pdf

Tabelle 1: Bruttostichprobe und realisierte Interviews

	Gelieferte Adressen	Realisierte Interviews
PAPI/CAWI	28.001	3.721
CAPI	9.986	1.010
Gesamt	37.986	4.731

Unterschiede zwischen den beiden Stichproben:

- PAPI/CAWI Stichprobe: bundesweit zufällig gezogene Adressen
- CAPI Stichprobe: Zur Begrenzung der Kosten der persönlichen Befragungen wurde eine regionale Klumpung vorgenommen. In der ersten Stufe wurden die 156 Arbeitsagenturbezirke als Auswahlinheiten festgelegt. Aus diesen wurden 60 Arbeitsagenturbezirke zufällig als Sample-Points gezogen. Die Adressen in der Stichprobe waren über 1.788 Orte und 2.420 Postleitzahlen verteilt.

Eine detaillierte Beschreibung sowohl der Stichprobenziehung als auch der erhobenen Merkmale findet sich in Sauer & Valet (2014).⁵

3. Erstellung des Scientific Use Files

Für die Erstellung des Scientific Use Files wird das Konzept der faktischen Anonymisierung angewandt. Dabei bedeutet faktisch anonym, dass eine Deanonymisierung einzelner im Datensatz enthaltener Individuen nur mit unverhältnismäßig hohem Aufwand an Zeit, Kosten und Arbeit möglich ist und somit die Kosten der Deanonymisierung den Nutzen derselben übersteigt. Anwendung finden dafür die Regeln von Müller et al. (1991)⁶ zur faktischen Anonymisierung von Mikrodaten.

Da der Informationsgehalt für die Genauigkeit der Analysen wichtig ist, sollen die Daten so weit wie möglich erhalten bleiben. Bei der Erstellung des Scientific Use Files kommen deshalb ausschließlich die Regeln der Löschung einzelner Merkmale und Aggregationen von sensiblen Informationen zum Einsatz.

3.1. Angewandte Regeln zum Zweck der faktischen Anonymisierung

Das Ausgangsmaterial der Befragung „Erwartungen an Wirtschaft und Gesellschaft“ weist bereits ein geringes Reidentifikationsrisiko der Befragten auf: Zum einen sind die individuellen Auswahlwahrscheinlichkeiten, die sich aus der breiten Auswahlgrundlage vieler Millionen Beschäftigter und der realisierten Stichprobe von nur 4.731 Befragten ergeben, sehr gering. Zudem wird auf eine regionale Gliederung verzichtet.

Die Anonymisierung erfolgt über die Löschung von Merkmalen und über die Vergrößerung vorhandener Antwortkategorien. Subjektive Angaben und Einschätzungen stellen die Mehrzahl der Merkmale dar. Diese sind im Normalfall als unproblematisch hinsichtlich des Risikos der Reidentifikation zu betrachten und können bleiben erhalten. Darunter fallen auch alle im Rahmen der faktoriellen Surveys⁷ gemachten Angaben. Hingegen können durch die

⁵ Sauer, Carsten & Valet, Peter (2014): „Erwartungen an Wirtschaft und Gesellschaft. Feldbericht und Codebuch zur Erwerbstätigenbefragung. Erste Welle“ SFB 882 Technical Report Series 10. DFG Research Center (SFB) 882 From Heterogeneities to Inequalities, Bielefeld.

⁶ Müller, Walter; Blien, Uwe; Knoche, Peter & Wirth, Heike (1991): *Die faktische Anonymität von Mikrodaten*. Stuttgart: Metzler-Poeschel.

⁷ Faktorielle Surveys bestehen aus fiktiven Beschreibungen von Personen oder Objekten (sog. Vignetten). Diese werden durch verschiedene Attribute (Dimensionen) beschrieben, die in ihren Ausprägungen (Levels) variieren. In dieser Studie wurden zwei verschiedene Vignettenmodule eingesetzt: Einkommensvignetten und Arbeitsangebotsvignetten. Es folgen zwei Beispiele, bei denen die variierten Attribute fett markiert sind.

Kombination von soziodemografischen Merkmalen oder harten Fakten Individuen direkt identifiziert werden, so dass hier eine Anonymisierung durch Löschung oder Aggregation notwendig ist.

Die Löschung oder Aggregation von Merkmalen ist erforderlich, wenn zu erwarten ist, dass in der Grundgesamtheit eine geringe Zellbesetzung vorliegt. Schlüsse über die Zellbesetzung der Grundgesamtheit werden sowohl auf Basis von Auszählungen dieser Befragung als auch auf Basis anderer Datenquellen gemacht.

3.2. Löschung von Merkmalen

Alle Merkmale zum Interviewer (int3 - int21) werden gelöscht, da andernfalls ein Reidentifikationsrisiko für Interviewer besteht und diese Merkmale für inhaltliche Analysen von geringer Bedeutung sind. Lediglich eine systemfreie Interviewer-ID bleibt erhalten. Die Merkmale zur Interviewsituation bei persönlichen Interviews (i1 - i10) bleiben erhalten. Sie stellen kein Reidentifikationsrisiko dar, enthalten aber Informationen zur Datenqualität und erlauben dadurch eine für die Auswertung mitunter wichtige Bewertung unplausibler Antworten.

Alle Merkmale, die Klartextangaben enthalten, werden gelöscht.

3.3. Aggregation von Merkmalen

Soziodemographische Merkmale stellen ein Reidentifikationsrisiko dar, wenn sie Merkmalsausprägungen enthalten, die leicht beobachtbar sind und in der Grundgesamtheit selten vorkommen. Die folgenden Merkmale werden aggregiert:

- Alter: Es liegen das Geburtsjahr und der Geburtsmonat des Befragten vor. Der Geburtsmonat wird gelöscht, Geburtsjahre vor 1954 und nach 1993 werden zusammengefasst.
- Haushaltszusammensetzung, Kinder- und Geschwisterzahl: Bei der Zahl der Personen im Haushalt, der Zahl der Kinder und der leiblichen Kinder sowie der Zahl der Geschwister des Befragten und seiner Position in der Reihenfolge der Geschwister werden hohe Werte zusammengefasst, um die Reidentifikation von Großfamilien zu verhindern.
- Zur aktuellen Erwerbssituation liegen detaillierte Angaben für den Befragten und dessen Partner vor. Die Items entsprechen weitgehend dem Mikrozensus, so dass anhand der Fallzahlen des Mikrozensus auf die Fallzahl in der Grundgesamtheit geschlossen werden kann. Einzelne Kategorien zur Erwerbssituation werden jeweils so zusammengefasst, dass sie nicht feiner sind als im Scientific Use Files des Mikrozensus 2011.
- Sofern der Befragte Vorgesetzter ist, wird der nach der Anzahl der beaufsichtigten Personen befragt. Zum Schutz von Top-Managern werden hohe Werte zusammengefasst.
- Angaben zur aktuellen Tätigkeit des Befragten wurden anhand von offenen Fragen detailliert erfasst und anschließend in Berufsklassifikationen eingeordnet. Im SUF

Einkommensvignette: Eine **30-jährige Frau** ist als **Hilfsarbeiterin** beschäftigt. Sie arbeitet als **Leiharbeiterin** in einem Betrieb und erbringt dort **überdurchschnittliche** Leistung. Die Arbeitslosigkeit in ihrem Beruf ist **hoch**. Ihr monatliches Bruttoeinkommen beträgt **1500 Euro**.

Arbeitsmarktvignette: Eine **30-jährige Frau** mit **abgeschlossener Berufsausbildung** ist nach eigener **Kündigung** arbeitslos. Sie hat sich kürzlich in einem Betrieb als **Produktionsleiterin** beworben. Von diesem Betrieb erhält sie jetzt ein Angebot für einen **unbefristeten** Arbeitsvertrag mit einem Einkommen, das deutlich niedriger ist als sonst üblich. Es gibt eine **Gleitzeitregelung** mit **Arbeitszeitkonto**. Die Belastungen am Arbeitsplatz sind **durchschnittlich hoch**.

Für jedes Fallbeispiel wurde erfragt, ob es als gerecht oder ungerecht beurteilt wird.

bleibt lediglich ein Merkmal erhalten: Beruf nach ISCO-Klassifikation als 4-steller. Dies entspricht dem u.a. dem Vorgehen des Sozio-ökonomischem Panel.

- Erwerbshistorie: Für nicht-erwerbstätige Befragte und deren Partner wurden jeweils der Monat und das Jahr der letzten Erwerbstätigkeit erfragt. Diese Information wird durch das Löschen des Monats vergrößert.
- Da die Daten zur Untersuchung der Herausbildung von Gerechtigkeitseinstellungen verwendet werden sollen, sind die Einkommensangaben des Befragten besonders wichtig und sollen möglichst wenig vergrößert werden. Es werden das persönliche Brutto- und Nettoeinkommen sowie das Nettohaushaltseinkommen erfragt. Die Angaben werden auf einen Maximalwert von 10.000 € zensiert. Neben den realen Einkommen werden auch die jeweils als gerecht empfundenen Einkommen erfragt. Dabei handelt es sich um subjektive Einschätzungen, so dass diese Angaben unverändert erhalten bleiben.

3.4. Änderungen im Einzelnen

Modul- und Merkmalsnamen entsprechen Sauer, Carsten & Valet, Peter (2014): *Erwartungen an Wirtschaft und Gesellschaft. Feldbericht und Codebuch zur Erwerbstätigenbefragung. Erste Welle*. (SFB 882 Technical Report Series 10). DFG Research Center (SFB) 882 From Heterogeneities to Inequalities, Bielefeld.

<https://pub.uni-bielefeld.de/record/2689979>

Modul 1: Lebenssituation und Arbeitsplatz (v1 - v3)

Alle Merkmale sind subjektive Einschätzungen und bleiben erhalten.

Modul 2: Derzeitige Erwerbstätigkeit (v4 - v14)

v4 – Erwerbssituation: vergrößert 3=4 5=6 10=11 (das resultierende Merkmal entspricht den Informationen im Mikrozensus-SUF 2011)

v4_o – Erwerbssituation Sonstige: gelöscht

v5a – Monat frühere Erwerbstätigkeit: gelöscht

v9_o v9a_o v9b_o Freitexte gelöscht, Berufe als ISCO-4-steller

v11a – Anzahl der beaufsichtigten Personen: klassifiziert: 1-9 bleibt, 10-99 in 10er-Schritten, 100-199, 200 und mehr bei größeren Werten

v12_o – Anderer Schulabschluss: gelöscht

v13l_o – Anderer beruflicher Abschluss: gelöscht

Modul 3: Derzeitiger Arbeitsplatz (v15 – v45)

v34 – Tatsächliches Nettoeinkommen: Zensierung auf 10.000 €

v35 – Tatsächliches Bruttoeinkommen: Zensierung auf 10.000 €

Modul 4: Freunde und soziale Beziehungen (v46 – v60)

Es werden Informationen zu drei Personen erfragt, mit denen der Befragte häufig zusammen ist. Diese Informationen sind hinreichend grob, dass sie weder das Reidentifikationsrisiko des Befragten erhöhen noch Aufschluss über die Person geben, von der die Rede ist.

Lediglich Fragen mit offenen Textangaben werden gelöscht.

v49_1_i_o – In welcher Beziehung stehen Sie zu Person 1, Sonstige und zwar: gelöscht

v49_2_i_o – In welcher Beziehung stehen Sie zu Person 2, Sonstige und zwar: gelöscht

v49_3_i_o – In welcher Beziehung stehen Sie zu Person 3, Sonstige und zwar: gelöscht

Modul 5: Persönlicher Hintergrund (v61 - v71)

v61_o – Wohnort seit Geburt, Im Ausland, und zwar in: gelöscht

v63_1_o – Geburtsland, Anderes Land, und zwar: gelöscht

v63_2_o – Geburtsland Vater, Anderes Land, und zwar: gelöscht

v63_3_o – Geburtsland Mutter, Anderes Land, und zwar: gelöscht

v70 Anzahl Geschwister: Zensierung auf 9

v71 Stelle in Altersreihe der Geschwister: Zensierung auf 9

Modul 6: Gesellschaftliche und persönliche Werte (v72 – v79)

Alle Merkmale sind entweder hinreichend grob erhoben oder weisen subjektive Einschätzungen auf, so dass sie erhalten bleiben.

Modul 7: Familiäre Situation (v80 –v104)

v83_o – Partner: Höchster Schulabschluss; Anderer Schulabschluss, und zwar: gelöscht

v84l_o – Partner: beruflicher Ausbildungsabschluss; Anderer Abschluss, und zwar: gelöscht

v85 – Erwerbssituation Partner: vergrößert 3=4 5=6 10=11 (das resultierende Merkmal entspricht den Informationen im Mikrozensus-SUF 2011)

v85_o – Erwerbssituation Partner; Sonstige, und zwar: gelöscht

v87a – Monat frühere Erwerbstätigkeit Partner: gelöscht

v89_o – Hauptsächliche berufliche Tätigkeit Partner: gelöscht

v89a_o – Beschreibung berufliche Tätigkeit Partner: gelöscht

v89b_o – Hauptsächliche berufliche Tätigkeit besonderer Name Partner, Sonstige, und zwar: gelöscht

v93 – Anzahl Personen im Haushalt: auf 7 zensieren

v93a – Anzahl Personen im Haushalt, 15 Jahre und jünger: auf 5 zensieren

v95 – Anzahl leibliche Kinder: auf 5 zensieren

v96 – Anzahl leibliche Kinder im Haushalt: auf 5 zensieren

v98 – Haushaltsnettoeinkommen: Zensierung auf 10.000 €

v100j_o – Bezugsquelle Haushaltseinkommen, Sonstige Quelle, und zwar: gelöscht

v104_o – Wahlabsicht, Andere Partei, und zwar: gelöscht

Modul 8: Angaben zur Person (v105 – v111)

v106a – Geburtsmonat: gelöscht

v106b – Geburtsjahr: zusammenfassen: Zensierung von Extremwerten auf Geburtsjahr vor 1954 und nach 1993

v107_o – Mitglied einer anderen Religionsgemeinschaft: gelöscht

Vignetten (ev- und av-Merkmale)

Vignetten sind subjektive Einschätzungen und bleiben erhalten.

Metavariablen (m1 – m4)

m1 – persönliche Identifikationsnummer: wird durch eine systemfreie ID ersetzt

Merkmale zur Interviewsituation bei CAPI (i1 – i10)

i3a_o – Eingriff ins Interview durch Dritte bei Frage: gelöscht

i10_o – Interviewort; anderer Ort, und zwar: gelöscht

Merkmale des Interviewers/der Interviewerin (int2 – int21)

gelöscht

Generierte Merkmale

Es bleiben drei Gewichtungsfaktoren enthalten: Design-Gewicht (w_design), Anpassungsgewicht (w_zelle) und das kombinierte Gewicht (w_kombi) aus Design- & Anpassungsgewicht.

panel_consent – Einwilligung zur erneuten Kontaktierung (final)

4. Datenzugang

Die Bereitstellung der faktisch anonymisierten Daten und der Materialien zur Studie „Erwartungen an Wirtschaft und Gesellschaft“ als Scientific Use File erfolgt über das Forschungsdatenzentrum Betriebs- und Organisationsdaten (FDZ-BO) am Deutschen Institut für Wirtschaftsforschung (DIW Berlin)

DOI SUF Erwartungen an Wirtschaft und Gesellschaft:

https://doi.org/10.25652/diw_data_S0026.1

Datensatzzitation:

Liebig, Stefan; May, Meike; Sauer, Carsten; Schneider, Simone, Valet, Peter (2014): Erwartungen an Wirtschaft und Gesellschaft. Scientific Use File. FDZ-BO am DIW Berlin. DOI:10.25652/diw_data_S0026.1

Kontakt

Forschungsdatenzentrum Betriebs- und Organisationsdaten (FDZ-BO) am Deutschen Institut für Wirtschaftsforschung (DIW Berlin)

Mohrenstraße 58

10117 Berlin

fdz-bo@diw.de