

SOEP Survey Papers

Series G – General Issues and Teaching Materials

SOEP – The German Socio-Economic Panel Study at DIW Berlin

2019

Wegweiser zur Datenanalyse der IAB-SOEP Migrationsstichproben (M1, M2) und der IAB-BAMF-SOEP Stichproben von Geflüchteten (M3-M5)

Jana Nebelin, Marvin Petrenz und Knut Wenzig

Running since 1984, the German Socio-Economic Panel study (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin.

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing. The SOEP Survey Papers is comprised of the following series:

Series A – Survey Instruments (Erhebungsinstrumente)

Series B – Survey Reports (Methodenberichte)

Series C – Data Documentation (Datendokumentationen)

Series D – Variable Descriptions and Coding

Series E – SOEPmonitors

Series F – SOEP Newsletters

Series G – General Issues and Teaching Materials

The SOEP Survey Papers are available at <http://www.diw.de/soepsurveypapers>

Editors:

Dr. Jan Goebel, DIW Berlin

Prof. Dr. Stefan Liebig, DIW Berlin and Universität Bielefeld

Dr. David Richter, DIW Berlin

Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin

Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Please cite this paper as follows:

Jana Nebelin, Marvin Petrenz und Knut Wenzig. 2019. Wegweiser zur Datenanalyse der IAB-SOEP Migrationsstichproben (M1, M2) und der IAB-BAMF-SOEP Stichproben von Geflüchteten (M3-M5). SOEP Survey Papers 600: Series G. Berlin: DIW/SOEP



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

© 2019 by SOEP

ISSN: 2193-5580 (online)

DIW Berlin

German Socio-Economic Panel (SOEP)

Mohrenstr. 58

10117 Berlin

Germany

soeppapers@diw.de

Wegweiser zur Datenanalyse der IAB-SOEP Migrationsstichproben (M1, M2) und der IAB-BAMF-SOEP Stichproben von Geflüchteten (M3-M5)¹

Jana Nebelin, Marvin Petrenz und Knut Wenzig

¹ Wir sind dankbar für die hilfreichen Anmerkungen und Kommentare aus dem SOEP-Team und von Christina de Paiva Lareiro (Mitarbeiterin des BAMF).

Inhalt

Vorbemerkung	2
Migrations- und Geflüchteten-Samples	2
Einsatz von Fragebögen bei Samples und zugehörige Datensätze	4
Rekonstruktion des Informationsflusses	6
Dokumentation	8
Beispiele	8

Vorbemerkung

Um für Migrationsforschende den Umgang mit Daten sowohl auf Basis der SOEP-Stichproben (A-L3, N) als auch der neu aufgebauten Stichproben IAB-SOEP-Migrationsstichprobe (M1-M2) sowie der IAB-BAMF-SOEP Stichproben Geflüchteter (M3-M5) zu erleichtern und vor allem auch den Vergleich von Fragen der Migrationssamples, die zeit- und sachgleich bei Personen ohne Migrationshintergrund gestellt wurden nutzerfreundlicher zu gestalten, wird auch in diesem Jahr mit der SOEP-Datenweitergabe des Erhebungsjahres 2017 (v34) eine vollständig integrierte Datensatzstruktur angeboten. Um Dopplungen und eine Vielzahl von Datensätzen zu vermeiden, werden ab 2019 erstmals nur die integrierten Datensätze weitergegeben. Dies erleichtert vergleichende Analysen im Quer- und Längsschnitt, ohne den Aufwand für Analysen mit den einzelnen Samples über Gebühr zu erhöhen.

Mit der gestiegenen Anzahl der Fragebögen auf Personen- und Haushaltsebene, sowie weiterer integrierter Fragebögen im Bereich Kinder und Jugend, die bei den Migrations- und Geflüchteten-Samples zu Einsatz kamen, drohte auch die Komplexität der Datenstruktur zuzunehmen. Das hätte gegen den Leitgedanken, den Ansprüchen von Nutzern nachzukommen und die Nutzung der Daten möglichst einfach zu gestalten, verstoßen. So wurden – wie auch schon im letzten Jahr – die Angaben aus den Personenfragebögen, sowie die Personenangaben der Personenbiofragebögen, in **bhp** integriert. Bio-Informationen aus den Personenbiofragebögen finden sich integriert in **biol**. Die Angaben aus den Haushaltsfragebögen wurden zusammengefasst in dem **bhh**-Datensatz.

Die wichtigste Änderung, die erfolgt ist, ist vor allem die geänderte Konvention der Variablennamen in den Datensätzen, die die Querschnittsdaten abbilden.

Migrations- und Geflüchteten-Samples

Folgende Tabelle gibt einen Überblick über die Stichproben. Zu beachten ist, dass es weitere Personen im Haushalt geben kann, die nicht der ursprünglichen Zielpopulation angehören, wenngleich ihre Einbeziehung ins SOEP im Rahmen der Migrations- und Geflüchteten-Samples erfolgte. Da das SOEP jedoch auf dem Haushaltsprinzip beruht, wurden alle Personen des Haushaltes interviewt, ohne zwangsläufig einen Migrationshintergrund zu haben.

Sample	Jahr erste Befragung	Zielpopulation (weitere Informationen)	realisierte Interviews mit Personen älter als 18, im Jahr 2017	realisierte Haushaltsinterviews im Jahr 2017
IAB-SOEP (M1)	2013	Personen, die nach 1995 nach Deutschland eingewandert sind oder Migranten der zweiten Generation (https://www.diw.de/de/diw_01.c.579311.de/soep_iab_soep_mig_2016.html#579306)	2478	1350
IAB-SOEP (M2)	2015	Personen, die zwischen 2010 und 2013 nach Deutschland eingewandert sind (https://www.diw.de/de/diw_01.c.579311.de/soep_iab_soep_mig_2016.html#579306)	930	559
IAB-BAMF-SOEP (M3)	2016	Personen, die vom 01.01.2013 bis 31.01.2016 nach Deutschland eingereist sind, Asylantragsstellung bis 30.04.2016 (https://www.diw.de/de/diw_01.c.538695.de/forschung_beratung/iab_bamf_soep_befragung_gefluechteter_in_deutschland.html)	1562	1102
IAB-BAMF-SOEP (M4)	2016	Personen, die vom 01.01.2013 bis 31.01.2016 nach Deutschland eingereist sind, Asylantragsstellung bis 30.06.2016 (https://www.diw.de/de/diw_01.c.538695.de/forschung_beratung/iab_bamf_soep_befragung_gefluechteter_i)	1907	1263

		n_deutschland.html)		
IAB-BAMF-SOEP (M5)	2017	Personen, die vom 01.01.2013 bis 31.01.2016 nach Deutschland eingereist sind, Asylantragsstellung 01.01.-31.12.2016 (https://www.diw.de/documents/publikationen/73/diw_01.c.../wave_report_2017.pdf)	2252	1519

Einsatz von Fragebögen bei Samples und zugehörige Datensätze

Die folgenden Tabellen zeigen, welche Fragebögen in den einzelnen Samples gestellt wurden und in welchen Datensätzen die jeweils erhobenen Variablen eingeflossen sind. Jedem Erhebungsinstrument wird eine eindeutige ID zugeordnet, mit deren Hilfe die Datennutzer durch die Ausprägungen in der Variable „**instrument**“ auf Fallebene nachvollziehen können welches Instrument genutzt wurde.

Sample	Fragebogen	ID	Datensatz	Long-Datensatz
IAB-SOEP (M1 & M2)	Haushalt	65	bhh	hl
	Personen	54	bhp	pl
	Personenbio	55	bhp/ biol	pl/ biol
	Jugend (16-17 jährige)	72	bhjugend	jugendl
	Frühe Jugend (13-14 jährige)	93	bhschool2	biopupil
	Schüler (11-12 jährige)	90	bhschool	biopupil
	MUKI A (0-1 jährige)	80	bioagel	bioagel

	MUKI B (2-3 jährige)	81		
	MUKI C (5-6 jährige)	82		
	Eltern D (7-8 jährige)	83		
	MUKI E (9-10 jährige)	84		
	Verstorbene		bhvp	vpl
	Lust auf DJ		cogdj	cogdj
	Lücke		bgpluecke	plueckel

Sample	Fragebogen	ID	Datensatz	Long-Datensatz
IAB-BAMF-SOEP (M3 & M4)	Haushalt	66	bhh	hl
	Personenbio II Wiederbefragte	57	bhp/ biol	pl/ biol
	Personenbio I Erstbefragte	56	bhp/ biol	pl/ biol
	Personenbio nicht Geflüchtete	95	bhp/ biol	bl/ biol
	Integrierter Jugendfragebogen (Jugend, Frühe Jugend, Schüler)	73	bhjugend, bhschool, bhschool2	jugendl/ bioagel
	Integrierter Kinderfragebogen (MUKI A, B, C, E & Eltern D)	74	bhmuki-bhmuki5	bioagel
	Kompetenztest für Jugendliche			cog_refu

Sample	Fragebogen	ID	Datensatz	Long-Datensatz
IAB- BAMF- SOEP (M5)	Haushalt	67	bhh	hl
	Personenbio I Erstbefragte	56	bhp/ biol	pl/ biol
	Wohnumfeld (nur durch den Interviewer auszufüllen)		hbrutt17	

Rekonstruktion des Informationsflusses

Weil Datensätze Beobachtungen von Personen aus unterschiedlichen Fragebögen enthalten können, ist es wichtig zu wissen, mit welchem Fragebogen eine Beobachtung erhoben wurde. Die Variable **instrument** hilft dabei, den Datenfluss eindeutig zu rekonstruieren.

Im Datensatz **bhh** zeigt eine Häufigkeitsverteilung der Variable **instrument** folgendes Bild:

Wert	Label	absolute Häufigkeit
-1	keine Angabe	29
61	2017 Fragebogen Haushalt (A-L1 ; PAPI)	3787
63	2017 Fragebogen Haushalt (L2-L3 ; CAWI)	743
64	2017 Fragebogen Haushalt (N; CAPI)	2314
65	2017 Fragebogen Haushalt (M1-M2; CAPI)	1909
66	2017 Fragebogen Haushalt (M3-M4; CAPI)	2365
67	2017 Fragebogen Haushalt (M5; CAPI)	1519

Bei der Arbeit **mit dem auf die jeweilige Erhebungswelle bezogenen Datensatz, also den nicht generierten Querschnittsdaten**, ist zunächst zu beachten, dass aus den Variablennamen im Allgemeinen nicht ersichtlich ist, mit welchen Instrumenten die Informationen erhoben

wurden. Das wäre bei der Anzahl der zugehörigen Fragebögen auch gar nicht möglich. Ableiten lässt sich lediglich anhand der Integrationshierarchie, in welchen Fragebögen die Variable nicht erhoben wurde (s. Beispiel unten).

Sample	Fragebogen auf Personen-ebene	Suffix bei Variablen-namen	Fragebogen auf Haushalts-ebene	Suffix bei Variablen-namen
Haupt-SOEP (A-L1)	50	kein Suffix	61	kein Suffix
Haupt-SOEP (L2-3)	52	_q52	63	_q63
Haupt-SOEP (N)	53	_q53	64	_q64
Migrationssample (M1, M2) Wiederbefragte	54	_q54	65	_q65
Migrationssample (M1, M2) Erstbefragte	55	_q55	65	_q65
Geflüchteten-sample (M3, M4) Wiederbefragte	57	_q57	66	_q66
Geflüchteten-sample (M3, M4, M5) Erstbefragte	56	_q56	67	_q67

Einige Anhaltspunkte kann der Variablenname jedoch liefern.

Einschub: Variablennamen $\$[x]_{\#}_{\#}_{q}\#$

- $\$$ Welle mit 1 bis 2 Buchstaben
- x für Datensatztyp ein- oder zweistellig (**p,b,h,j,m1-m5,v, dj,fj,s,el**)
- $\#$ Frage, ggfs. auch „m01“
- $\#$ Itemnummer, fortlaufend ab 1
- $\#$ Fragebogennummer wie in Instrumentvariable (für „weitere“ Instrumente) (neu)

Im Datensatz **bhh** gehört zu einer Variable ohne Suffix „_q“ eine Frage, die auf jeden Fall im Fragebogen der SOEP-Stichproben (A-L1) gestellt wurde. Eine direkte Zuordnung der Variable zu der Fragennummer der genutzten Fragebögen in den Samples A-L1 ist möglich. Offen bleibt

jedoch, ob die Frage auch in weiteren Fragebögen anderer Samples gestellt wurde. Eine Variable mit dem Suffix „_q66“ wurde nicht in den SOEP-Stichproben (A-L1, L2-3, N) und nicht in den IAB-SOEP- Migrationsstichproben (M1, M2), aber im Geflüchtetenstichprobe (M3, M4) und vielleicht im Geflüchtetenstichprobe (M5) erhoben. Dies ergibt sich aus dem allgemeinen Konzept der Variablennamen in Verbindung mit der gewählten Hierarchisierung der Instrumente bei der Namensvergabe. Die direkte Zuordnung der Variable zu einer Fragennummer trifft für den betreffenden Fragebogen zu, den der Variablenname indiziert, im Beispiel der Fragebogen mit der ID 66 (2017 Fragebogen Haushalt (M3-M4; CAPI)).

Wenn man wissen will, wo eine Variable **var1** erhoben wurde, gibt es folgende Möglichkeiten. Einerseits gibt das metadatenbasierte Codebook (siehe folgenden Abschnitt) eine Antwort auf die Frage, andererseits kann auch folgender Stata-Befehl verwendet werden:

```
fre instrument if var1!=-5
```

Dokumentation

Metadatenbasierte Fragebögen erleichtern erheblich das Auffinden der interessierenden Variablen aus Sicht des Fragebogens. Jedes der erzeugten PDFs spiegelt einen Fragebogen wieder. Mit Hilfe dieser Dokumente erfährt der Nutzer welche Fragen dem jeweiligen Sample gestellt worden sind und in welcher Abfolge. Darüber hinaus wird in den Dokumenten deutlich, wie die zur Frage zugehörige Variable heißt und in welchem Datensatz sie zu finden ist.

Metadatenbasierte Codebooks geben einen Überblick über die enthaltenen Variablen je Datensatz. Diese Dokumente zeigen in welchen Fragebögen die Variable erhoben worden ist und die Formulierung der entsprechenden Frage.

Beispiele

Im Folgenden finden sich zwei Beispiele für die Verwendung von metadatenbasierten Dokumenten. Ein Beispiel konzentriert sich auf eine typisch geflüchtetenstichprobenspezifische Frage. Das andere Beispiel zeigt eine Frage, die in mehreren Samples gestellt worden ist.

(1) Geflüchtetenstichprobenspezifische Frage

Das erste Beispiel zeigt die Frage Q042 aus dem Personfragebogen für Wiederbefragte aus der IAB-BAMF-SOEP Stichprobe für Geflüchtete. Die Angaben zu der Frage sind abgelegt im Datensatz **bhp** unter dem Namen **bhp_42_q57**. Der Name lässt indirekt darauf schließen, dass sich die Frage nicht in den Samples A-M2 befindet, da er das Suffix „_q57“ enthält. Das Suffix

„_q57“ identifiziert den Personenbiographie-Fragebogen für Wiederbefragte der IAB-BAMF-SOEP Stichprobe Geflüchteter (M3, M4). Dies schließt nicht aus, dass die Frage auch in Fragebögen, die in der Hierarchie später kommen, gestellt wurde.

Q042 Wie oft verbringen Sie Zeit mit Personen aus Ihrem Herkunftsland, die nicht mit Ihnen verwandt sind?

Täglich	1	
Mehrmals pro Woche	2	
Jede Woche	3	
Jeden Monat	4	
Seltener	5	
Nie	6	
keine Angabe	99	
Q042:pverz1	bhp	bhp_42_q57
Q042:pverz1	pverz1	

Das Codebook für **bhp** zeigt, dass die Frage „Wie oft verbringen Sie Zeit mit Personen aus Ihrem Herkunftsland, die nicht mit Ihnen verwandt sind?“ hinter der Variable **bhp_42_q57** in zwei Fragebögen abgefragt wird: in Personen(bio) für Wiederbefragte Geflüchtete (soep-core-2017-pb-m34-wieder) und im Personenbiofragebogen für Erstbefragte Geflüchtete (soep-core-2017-pb-m345-erst). Die Fragen sind identisch formuliert. Zusätzlich werden die Antwortkategorien und die Häufigkeitsverteilungen angezeigt.

bhp_42_q57 – Verbrachte Zeit mit Personen des Herkunftslandes

? Wie oft verbringen Sie Zeit mit Personen aus Ihrem Herkunftsland, die nicht mit Ihnen verwandt sind? (from: soep-core/soep-core-2017-pb-m34-wieder/Q042:pverz1[7313])

? Wie oft verbringen Sie Zeit mit Personen aus Ihrem Herkunftsland, die nicht mit Ihnen verwandt sind? (from: soep-core/soep-core-2017-pb-m345-erst/Q269:pverz1[12750])

1	[1] Täglich	1512
2	[2] Mehrmals pro Woche	1018
3	[3] Jede Woche	1035
4	[4] Jeden Monat	549
5	[5] Seltener	791
6	[6] Nie	776
-1		16
-5		26914

(2) Sample übergreifende Frage

Als Beispiel für eine Frage, die in mehreren Samples gestellt wurde, soll die allgemeine Lebenszufriedenheit von Wiederbefragten Geflüchteten betrachtet werden. Gesucht wird über den metadatenbasierten Fragebogen für Wiederbefragte Geflüchtete 2017. Die Frage nach allgemeiner Lebenszufriedenheit wird in Frage Q518 gestellt. Unterhalb der Frage befindet sich die Information in welchem Datensatz die Variable vorkommt und unter welchem Namen sie abgelegt ist.

Q518 Wie zufrieden sind Sie gegenwärtig alles in allem mit Ihrem Leben?

Antworten Sie bitte wieder anhand einer Skala. Der Wert 0 bedeutet: ganz und gar unzufrieden. Der Wert 10 bedeutet: ganz und gar zufrieden. Mit den Werten dazwischen können Sie Ihre Einschätzung abstufen.

0 ganz und gar unzufrieden	1
1	2
2	3
3	4
4	5
5	6
6	7
7	8
8	9
9	10
10 ganz und gar zufrieden	11

Q518:pzule1 **bhp** **bhp_205**
Q518:pzule1 pzule1

Die allgemeine Zufriedenheit mit dem Leben ist zu finden im Datensatz **bhp** unter dem Namen **bhp_205**. Allein der Name lässt darauf schließen, dass die Frage auch in den Samples A-L1 gestellt wird.

Ein Blick in das Codebook bestätigt diese Vermutung. Die Frage wird in vier Personen(bio)bögen gestellt: soep-core-2017-pb-m34-wieder (Personenbiographie-Fragebogen für Wiederbefragte Geflüchtete), soep-core-pb-m345-erst (Personenbiographie-Fragebogen für Erstbefragte Geflüchtete), soep-core-2017-pe (Im Personenfragebogen für die Sample A-L1) und im soep-core-217-pb-m34-erstNG (und im Personenbiographiefragebogen für Erstbefragte Nicht-Geflüchtete Personen aus dem Geflüchtetensample). Die Frage wird weitgehend identisch in den Fragebögen erhoben, weshalb die Information in einer integrierten Variablen (**bhp_205**) abgelegt wird.

bhp_205 – Zufriedenheit gegenwaertig

? Wie zufrieden sind Sie gegenwaertig alles in allem mit Ihrem Leben? (*from: soep-core/soep-core-2017-pb-m34-wieder/Q518:pzule1[12544]*)

? Wie zufrieden sind Sie gegenwaertig alles in allem mit Ihrem Leben? (*from: soep-core/soep-core-2017-pb-m345-erst/Q585:pzule1[12967]*)

? Zum Schluss moechten wir Sie noch nach Ihrer Zufriedenheit mit Ihrem Leben insgesamt fragen. Wie zufrieden sind Sie gegenwaertig, alles in allem, mit Ihrem Leben? (*from: soep-core/soep-core-2017-pe/205:pzule[4306]*)

? Kommen wir nun wieder zu Ihnen persoenlich: Wie zufrieden sind Sie gegenwaertig mit den folgenden Bereichen Ihres Lebens? // alles in allem mit Ihrem Leben? (*from: soep-core/soep-core-2017-pb-m34-erstNG/180:pzule1[8527]*)

0	[0] Ganz und gar unzufrieden	152
1	[1] Zufrieden: Skala 0-Niedrig bis 10-Hoch	125
2	[2] Zufrieden: Skala 0-Niedrig bis 10-Hoch	335
3	[3] Zufrieden: Skala 0-Niedrig bis 10-Hoch	650
4	[4] Zufrieden: Skala 0-Niedrig bis 10-Hoch	800
5	[5] Zufrieden: Skala 0-Niedrig bis 10-Hoch	2916
6	[6] Zufrieden: Skala 0-Niedrig bis 10-Hoch	2914
7	[7] Zufrieden: Skala 0-Niedrig bis 10-Hoch	6590
8	[8] Zufrieden: Skala 0-Niedrig bis 10-Hoch	10512
9	[9] Zufrieden: Skala 0-Niedrig bis 10-Hoch	4955
10	[10] Ganz und gar zufrieden	2595
-1	[-1] keine Angabe	67
-2	[-2] trifft nicht zu	0
-3	[-3] nicht valide	0
-4	[-4] Unzulaessige Mehrfachantwort	0
-5	[-5] In Fragebogenversion nicht enthalten	0
-6	[-6] Fragebogenversion mit geaenderter Filterfuehrung	0