

1183²⁰²²

SOEP Survey Papers
Series D – Variable Descriptions and Coding

SOEP-Core v37 – HPATHL: Household-Related Meta-Dataset

SOEP Group

Running since 1984, the German Socio-Economic Panel (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/soepsurveyspapers>

Editors:

Dr. Jan Goebel, DIW Berlin
Prof. Dr. Stefan Liebig, DIW Berlin and Freie Universität Berlin
Prof. Dr. David Richter, DIW Berlin and Freie Universität Berlin
Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin
Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin
Prof. Dr. Sabine Zinn, DIW Berlin and Humboldt-Universität zu Berlin

Please cite this paper as follows:

SOEP Group, 2022. SOEP-Core v37 – HPATHL: Household-Related Meta-Dataset. SOEP Survey Papers 1183: Series D – Variable Descriptions and Coding. Berlin: DIW Berlin/SOEP



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

© 2022 by SOEP

ISSN: 2193-5580 (online)

DIW Berlin
German Socio-Economic Panel (SOEP)
Mohrenstr. 58
10117 Berlin
Germany

soeppapers@diw.de

SOEP-Core v37 – HPATHL: Household-Related Meta-Dataset

SOEP Group

2022

Contents

1	General Information	3
2	Primary Key, Foreign Keys and Sample Information	3
	hid – Current Wave HH Number	3
	syear – Survey Year	3
	cid – Original Household Number, Case ID	4
	hsample – Sample Member	4
3	Survey History	5
	hnetto – Aktueller Befragungsstatus	5
	hiyear – Interviewjahr (InterviewYear)	5
4	Basic Demographic Information	6
	sampreg – Aktuelle Stichprobenregion	6
	hpop – Aktuelle Populationszugehoerigkeit	6
5	Weighting	7
	hbleib – Inverse Staying Probability	7
	hhrf – Weighting factor	7
	hhrf0 – Weighting factor for new samples (wave 1 of new sample)	7
	hhrf1 – Weighting factor without new samples (wave 1)	7
	hrgroup – Random Groups	7

1 General Information

The path datasets should be the building block of any analysis. Path Files indicate the total population at the household and individual level (over time) and provide all IDs necessary to access further files at different levels (Krause/Glass/Reher 2019a,b). Path-Files are delivered in three data formats – in long-format [H|P-PATHL] (as the most comprehensive version including weighting variables), in wide-format [H|P-PFAD] (the traditional version), and in a short-version [H|P-PATH] as a reduced population file (indicating the total of population of households and individuals) Household Level [HID|SYEAR] {Navigation File: H-PATH-L (long-format)} Individual Level [PID|SYEAR] {Navigation File: P-PATH-L (long-format)} Household Level [HID] {Population File: H-PATH} Individual Level [PID] {Population File: P-PATH} Household Level [HID (HHNRAKT)] {Path File: HPFAD (wide-format)} Individual Level [PID (PERSNR)] {Path File: PPFAD (wide-format)}. The constituting SOEP population considers three levels – cases, households, and individuals. Due to the SOEP sampling and survey process, these levels follow an implicit hierarchy. All samples refer to primary source households – indicated by the household id at the time when the survey starts – the (fixed) Case ID [CID]. New Households may emerge from these original households during the longitudinal survey process by split-offs of family members – all (current) households are therefore indicated by a (variable) Household ID [HID]. IDs for individuals living in the households are derived from the households, where they were living when they were surveyed for the first time – the (fixed) Personal ID [PID].

2 Primary Key, Foreign Keys and Sample Information

hid – Current Wave HH Number

The members and configuration of a household with a given ID (HID) can change between years. Depending on the respective research issue it may be necessary to check alterations using PPATHL, where persons are assigned to a household in each year.

syear – Survey Year

1984	10363
1985	6142
1986	5878
1987	5512
1988	5403
1989	5213
1990	7270
1991	7335
1992	7359
1993	7334
1994	7431
1995	7717
1996	7529
1997	7423
1998	9403
... (7 rows omitted)	101359
2006	16481
2007	13410

2008	12651
2009	16642
2010	21479
2011	25896
2012	22331
2013	31231
2014	19699
2015	28293
2016	43343
2017	28242
2018	29228
2019	45577
2020	57341

Survey year (Reference Year for survey instruments)

cid – Original Household Number, Case ID

Case Id - Household Source Identifier (fixed). The fixed household source id point to the first SOEP ancestor households to which current households can be traced back to. This variable is relevant for sampling information and calculating the weights.

hsample – Sample Member

1	A 1984 Initial Sample (West)	121711
2	B 1984 Migration (until 1983, West)	32033
3	C 1990 Initial Sample (East)	55012
4	D 1994/5 Migration (1984-1994, West)	9830
5	E 1998 Refreshment	13139
6	F 2000 Refreshment	83808
7	G 2002 High Income	20282
8	H 2006 Refreshment	16654
9	I 2009 Innovation Sample	6499
10	J 2011 Refreshment	30393
11	K 2012 Refreshment	14202
12	L1 2010 Birth Cohort (2007-2010)	19825
13	L2 2010 Family Type (Low-Income, Single-Parent, Large Families)	21556
14	L3 2011 Family Type (Single-Parent, Large Families)	7893
15	M1 2013 Migration (1995-2011)	26532
...	(4 rows omitted)	60526
20	N 2017 Refreshment (PIAAC-L)	9950
21	O 2018 Social City	8368
22	P 2019 Top-Shareholder	25242
23	Q 2019 Lesbian-Gay-Bisexual (LGB) persons	1317
24	M6 2020 Refugee	3000
25	M7 2020 Migration	19751
26	M8 2020 Migration Professionals	12992
-1	No Answer	0
-2	Does not apply	0
-3	Answer improbable	0

-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

Unchanging sample membership at first contact with SOEP (at the point of selection of the original household into the sample).

3 Survey History

hnetto – Aktueller Befragungsstatus

1	Successful HH Interview	414759
2	HH In Brutto Only	64098
3	unsuccessful, initial gross sample	141650
-1	No Answer	0
-2	Does not apply	0
-3	Answer improbable	8
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

Wave-specific survey status: for HNETTO=1 there is a HH interview in the respective year, i.e. an entry in the file HL. For HNETTO=2 there is only an entry in the file HBRUTTO, but the HH could not be interviewed (the reason for which is given by the variable HERGS in the file HBRUTTO). HNETTO=3 indicates unrealized Households of new samples in HBRUTT-Files.

hiyear – Interviewjahr (InterviewYear)

1984	5921
1985	6109
1986	5838
1987	5472
1988	5365
1989	5171
1990	7225
1991	7275
1992	7264
1993	7226
1994	7373
1995	7498
1996	7447
1997	7391
1998	8330
... (9 rows omitted)	117473

2008	12600
2009	13170
2010	17355
2011	19232
2012	19223
2013	20861
2014	19572
2015	19267
2016	21065
2017	22148
2018	24109
2019	24588
2020	23501
2021	1503
-2	143943

Interview Year (indicates household interviews realized also outside of standard SYEAR)

4 Basic Demographic Information

sampreg – Aktuelle Stichprobenregion

1	West Germany	437736
2	East Germany	109752
-1	No Answer	0
-2	Does not apply	73019
-3	Answer improbable	8
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

The variable SAMPREG tells whether a HH was located in West or East Germany in the respective year. This variable approximates also former regional differences within Berlin according to zip-codes.

hpop – Aktuelle Populationszugehoerigkeit

1	Private HH, German HH-Head	399742
2	Private HH, Foreign HH-Head	67779
3	Institutionalized HH, German HH-Head	3291
4	Institutionalized HH, Foreign HH-Head	5114
-1	No Answer	0
-2	Does not apply	144589
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0

-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

POP was derived from WUM2 in HBRUTTO as well as PNAT_V* and STELL_V* (nationality and relationship to head of household in PBRUTTO). Missing values were imputed based on the person's history. Thus, the only admissible missing value is -2, meaning not applicable. This variable is therefore particularly important, as it enters into the determination of cross-sectional weights. The variable corresponds with POP in PPATH/L.

5 Weighting

hbleib – Inverse Staying Probability

inverse probability weights

hhrf – Weighting factor

standard household weight

hhrf0 – Weighting factor for new samples (wave 1 of new sample)

household weight for first wave of new samples

hhrf1 – Weighting factor without new samples (wave 1)

household weights without first wave of new samples

hrgroup – Random Groups

0	141943
1	59477
2	59401
3	59762
4	57647
5	60294
6	61477
7	60191
8	60323