SOEP Survey Papers

SOEP-IS 2022 – HPATHL: Household-Related Meta-Dataset (long)

SOEP-IS Group

DIW SOEP

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/soepsurveypapers

SOEP-IS 2022 – HPATHL:
Household-Related Meta-Dataset (long)

SOEP-IS Group

2024

Contents

1 Primary Key, Foreign Keys and Sample Information

hid – Current Wave HH Number

115380		21
115606		10
115614		5
116033		2
116092		16
116165		21
116262		2
116874		10
116882		5
117153		4
117269		20
117668		1
117722		6
118028		20
118036		2
...	(20695 rows omitted)	67508
5566640		1
5566650		1
5566660		1
5566670		1
5566680		1
5566690		1
5566700		1
5566710		1
5566720		1
5566730		1
5566740		1
5566750		1
5566760		1
5566770		1
5566780		1

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

1998	373
1999	388
2000	405
2001	428
2002	432
2003	443
2004	451

2005	463
2006	477
2007	476
2008	471
2009	1960
2010	2223
2011	1840
2012	4610
2013	6972
2014	7720
2015	4578
2016	8680
2017	4877
2018	4467
2019	4054
2020	4687
2021	2558
2022	3635

Survey year (Reference Year for survey instruments)

cid – Original Houshold Number, Case ID

200034	24
200093	15
200140	25
200158	15
200247	19
200328	55
200417	25
200506	25
200581	16
200670	29
200689	23
200808	18
200816	22
200832	31
200840	48
... (19510 rows omitted)	67193
4905836	8
4905844	1
4905852	1
4905860	2
4905879	7
4905887	1
4905895	8
4905909	7
4905917	4
4905925	8

4905933	8
4905941	8
4905950	8
4905968	7
4905976	7

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.

sample1 – Subsample Identifier

1	A German HH Head, West, 1984	0
2	B1 Turkish HH Head, 1984	0
3	B2 Yugoslav HH Head, 1984	0
4	B3 Greek HH Head, 1984	0
5	B4 Italian HH Head, 1984	0
6	B5 Spanish HH Head, 1984	0
7	C German HH Head, East, 1990	0
8	D1 Immigrant W. Germany (1984-1994), 1994/95	0
9	D2 Immigrant W. Germany (1984-1994), 1994/95	0
10	E1 Refreshment (PAPI), 1998	4849
11	E2 Refreshment (CAPI), 1998	4547
12	F1 Refreshment, 2000	0
13	F2 Refreshment, Additional Foreigner Addresses, 2000	0
15	G2 High Income, 2002	0
16	H Refreshment, 2006	0
…	(21 rows omitted)	58272
60	L2/3 FiD Screening, Low Income, 2010/11	0
61	L2/3 FiD Screening, Single Parent, 2010/11	0
62	L2/3 FiD Screening, Families with Several Children, 2010/11	0
63	L1 FiD Birth Cohort 2007, 2010/11	0
64	L1 FiD Birth Cohort 2008, 2010/11	0
65	L1 FiD Birth Cohort 2009, 2010/11	0
66	L1 FiD Birth Cohort 2010, 2010/11	0
-1	No answer	0
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissable 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	0

2 Survey History

hnetto – Aktueller Befragungsstatus

1	Successful HH Interview	46989
2	HH In Brutto Only	20679
3	unsuccessful, initial gross sample	0
-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

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.

3 Basic Demographic Information

sampreg – Aktuelle Stichprobenregion

1	West-Germany	53139
2	East-Germany	11904
-1	No Answer	0
-2	Does not apply	67
-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	2558

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	42518
2	Private HH, Foreign HH-Head	2037
3	Institutionalized HH, German HH-Head	167
4	Institutionalized HH, Foreign HH-Head	14
-1	No Answer	0
-2	Does not apply	22932
-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.

4 Weighting

hbleib – Inverse Staying Probability

0		39938
1		1128
1.00999999046326		832
1.01		251
1.01999998092651		592
1.02		27
1.02999997138977		970
1.03		191
1.03999996185303		1075
1.04		201
1.04499995231628		1005
1.05		184
1.05999994277954		888
1.06		336
1.07		53
...	(945 rows omitted)	19979
5.08		1
5.17		1
5.19		2
5.22		1
5.27		1
5.59		1
5.94		2
6.1		1
6.23		2
6.65		1
6.8		1
6.87		1
7.46		1
7.47		1
7.72		1

inverse probability weights

hhrf – PRELIMINARY: Weighting factor

0	35352
102.554836488493	1
135.522859610515	1
155.990476120212	1
164.040529135787	1
170.969909439076	1
173.070657345641	1
180.394172578254	1
197.159583531232	1
199.598065397166	1
199.786628779685	1
200.40841476272	1
216.06728850827	1
224.505719893504	1
232.541069453447	1
... (30919 rows omitted)	32287
124924.539459815	1
125588.898664034	1
125772.846327675	1
125992.281604959	1
126335.728361698	1
128682.495977036	1
129061.898679802	1
131901.679874408	1
132272.687116397	1
132303.155363102	1
132996.000240848	1
134146.819613007	1
134256.464387417	1
134295.142348304	1
136236.419357491	1

standard household weight