

## SOEP Survey Papers

Series D – Variable Descriptions and Coding

# SOEP-Core v33.1 – Documentation of the Household-related Meta-dataset HPFAD

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/soepsurveyspapers>

Editors:

Dr. Jan Goebel, DIW Berlin

Prof. Dr. Martin Kroh, DIW Berlin and Humboldt Universität 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:

SOEP Group, 2018. SOEP-Core v33.1 – Documentation of the Household-related Meta-dataset HPFAD.

SOEP Survey Papers 484: Series D – Variable Descriptions and Coding. Berlin: DIW Berlin/SOEP



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

© 2018 by SOEP

ISSN: 2193-5580 (online)

DIW Berlin

German Socio-Economic Panel (SOEP)

Mohrenstr. 58

10117 Berlin

Germany

[soepapers@diw.de](mailto:soepapers@diw.de)

# SOEP-Core v33.1 – Documentation of the Household-related Meta-dataset HPFAD

SOEP Group

2018

## Contents

<b>1</b>	<b>General Information</b>	<b>3</b>
<b>2</b>	<b>Variables in File HPFAD</b>	<b>3</b>
	hhnr – Original Household Number . . . . .	3
	hhnrakt – Current Household Number . . . . .	3
	hsample – Sample Member . . . . .	3
	\$HHNR – Current Wave Household Number 1984 [generic] . . . . .	4
	\$HNETTO – Current Wave Survey Status 1984 [generic] . . . . .	4
	\$HPOP – Sample Membership 1984 [generic] . . . . .	5
	\$SAMPREG – Current Wave Sample Region 1991 [generic] . . . . .	5

## 1 General Information

One of the biggest assets of the SOEP data is their longitudinal nature, i.e. repeated observations of the same unit (person or household) over time. There are two datasets which should be the building block of any analysis, as they allow to define longitudinal populations very easily: PPFAD and HPFAD. HPFAD includes all households which have been interviewed successfully at least once. Similarly, PPFAD contains all persons who have ever lived in a household that has participated in the SOEP, i.e. that has been captured in HPFAD, including non-respondents and children. Both datafiles contain one record per household or person, respectively, with wave-specific variables for each year's survey status. In addition to some time-invariant information (like gender, year of birth, migrant status), these files contain all necessary identifiers to combine other files with PPFAD and HPFAD.

## 2 Variables in File HPFAD

**hhnr** - Original Household Number

---

**hhnrakt** - Current Household Number

---

**hsample** - Sample Member

---

1	[1] A 1984 Initial Sample (West)	8309
2	[2] B 1984 Migration (until 1983, West)	2648
3	[3] C 1990 Initial Sample (East)	4101
4	[4] D 1994/5 Migration (1984-1994, West)	891
5	[5] E 1998 Refreshment	1395
6	[6] F 2000 Refreshment	7897
7	[7] G 2002 High Income	1736
8	[8] H 2006 Refreshment	1760
9	[9] I 2009 Innovation Sample	1756
10	[10] J 2011 Refreshment	3493
11	[11] K 2012 Refreshment	1692
12	[12] L1 2010 Birth Cohort (2007-2010)	2273
13	[13] L2 2010 Family Type (Low-Income, Single-Parent, Large Families)	3079
14	[14] L3 2011 Family Type (Single-Parent, Large Families)	1095
15	[15] M1 2013 Migration (1995-2011)	3052
16	[16] M2 2015 Migration (2009-2013)	10021
17	[17] M3 2016 Refugee (2013-2015)	1698
18	[18] M4 2016 Refugee/family (2013-2015)	1638
19	[19] M5 2017 Refugee (2013-2016)	0
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

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

*For more information, contact:* Peter Krause (Tel. 030-89789-690)

**\$HHNR** – Current Wave Household Number 1984 [generic]

---

19		1
27		1
35		1
43		1
51		1
60		1
78		1
86		1
94		1
108		1
116		1
124		1
132		1
140		1
159		1
...	(5892 rows omitted)	5892
59560		1
59579		1
59587		1
59595		1
59609		1
59617		1
59625		1
59633		1
59641		1
59650		1
59668		1
59676		1
59684		1
59692		1
-2		52613

*Waves:* all

current household number = HHNRAKT [This information can be related to a specific variable and is not necessarily generic.]

*For more information, contact:* Peter Krause (Tel. 030-89789-690)

**\$HNETTO** – Current Wave Survey Status 1984 [generic]

---

1	[1] Successful HH Interview	5921
2	[2] HH In Brutto Only	0
-1	[-1] No Answer	0
-2	[-2] Does not apply	52613
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all

Wave-specific survey status: for \$HNETTO=1 there is a HH interview in the respective year, i.e. an entry in the file \$H. 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). [This information can be related to a specific variable and is not necessarily generic.]

For more information, contact: Peter Krause (Tel. 030-89789-690)

### **\$HPOP** – Sample Membership 1984 [generic]

---

1	[1] Private HH, German HH-Head	4513
2	[2] Private HH, Foreign HH-Head	1350
3	[3] Institutionalized HH, German HH-Head	15
4	[4] Institutionalized HH, Foreign HH-Head	43
-1	[-1] No Answer	0
-2	[-2] Does not apply	52613
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all

\$POP was derived from \$WUM2 (in \$HBRUTTO differentiating private from institutional households) as well as \$PNAT and \$STELL (nationality and relationship to HH head in \$PBRUTTO). Missing values are imputed taking into account respondent's history. Thus the only admissible missing value is "-2", which means "does not apply". The variable is therefore of particular importance, as it is used to compute cross-sectional weights. The variable corresponds with \$POP in PPFAD. [This information can be related to a specific variable and is not necessarily generic.]

For more information, contact: Peter Krause (Tel. 030-89789-690)

### **\$SAMPREG** – Current Wave Sample Region 1991 [generic]

---

1	[1] West-Germany	5106
2	[2] East-Germany	2229
-1	[-1] No Answer	0
-2	[-2] Does not apply	51199
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: since 1990

The first East survey (1990=Wave G) of SOEP led to an expansion of the survey area beyond the West sample that was carried out up to 1984. In connection with the SOEP concept of following SOEP respondents continuously from one year to the next, explicitly reflecting spatial mobility, households that leave the "area" of the original sample (former Federal Republic of Germany, within the borders existing up to 3 October 1990 for Samples A and B, or for the former GDR for Sample C) and continue to take part in SOEP are interviewed at their new address. Because of this interregional mobility between West and East Germany, which

has been registered in SOEP since the second East wave (1991=Wave H), using the original sample number (which does not change over time) to distinguish households produces increasingly biased results for analyses differentiating between households in East and West Germany (Variable HSAMPLE; here H for “household” in comparison to PSAMPLE in the file PPFAD for “person”). The wave-specific variable \$SAMPREG thus tells whether a HH was living in West or East Germany in the respective year. [This information can be related to a specific variable and is not necessarily generic.]

*For more information, contact:* Peter Krause (Tel. 030-89789-690)