

SOEP Survey Papers

Series D – Variable Descriptions and Coding

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

2018

Categorizing open answers in the DRM module of SOEP-IS

Tobias Wolf

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:

Tobias Wolf. 2018. Categorizing open answers in the DRM module of SOEP-IS. SOEP Survey Papers 535: Series D. Berlin: DIW/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

soeppapers@diw.de

Categorizing open answers in the DRM module of SOEP-IS

Tobias Wolf*

Abstract

This paper documents the categorization of open answers in the diaries of the Day Reconstruction Method (DRM) module of the SOEP Innovation Sample (SOEP-IS) for employed and unemployed workers. Respondents of the module reported either 1 out of 23/25 listed activities or a self-selected description of an activity during the previous day. By assigning the open answers either to the initially listed categories and by defining new categories for a subset of open answers enriches SOEP-IS DRM by more than 1,700 additional categorized episodes for quantitative analysis.

Keywords: Day Reconstruction Method; DRM; open answers; SOEP-IS

* Contact: Freie Universität Berlin, School of Business and Economics, Boltzmannstraße 20, D-14195 Berlin, Germany. E-mail: tobias.wolf@fu-berlin.de

Comments by Maria Metzing and David Richter are gratefully acknowledged. I also thank Tom Günther for the excellent research assistance.

1. Day reconstruction method in the SOEP innovation sample

The described DRM module is a survey adaptation of the Kahneman et al. (2004) DRM study for the German SOEP Innovation Sample (Richter and Schupp 2015). It is designed to reduce the temporal burden for the respondents while surveying diaries of the previous days accompanying with self-reported affective experiences. The initial DRM of Kahneman et al. (2004) left the construction of the diary open to the respondents. They were asked to split the previous day into episodes that were named by the respondents themselves. For a general population survey is this procedure too time-consuming and the subsequent data processing too resource intensive. In SOEP-IS DRM respondents report all episodes of the previous day by choosing activities out of a list of proposed (pre-set) activities. This procedure makes DRM for population survey studies feasible while it still adds on average 12 minutes individual interview time to SOEP-IS (Anusic, Lucas, and Donnellan 2017).

Choosing from a set of activities instead of free naming each episode of the previous day reduces the temporal burden for interviewer and respondents but imposes restrictions to the set of possible activities to report. SOEP-IS DRM copes with this restriction by giving the opportunity to report for each episode *another activity*. Choosing this option allows respondents to report a freely chosen description of the activity for all distinct episodes. The opportunity to answer open questions was used intensively so that analysis ignoring these open answers lack a relevant part of surveyed information. This documentation describes an ex-post categorization of SOEP-IS DRM open answers to make more episodes feasible for quantitative analysis by categorizing the open answers systematically. In order to make this ex-post categorization accessible to other researchers this paper describes how further episodes can be appended to SOEP-IS DRM data.

In the following the paper describes the pre-set list of activities in SOEP-IS (section 2.1) and describes the method to find suitable categories for activities (section 2.2). The manual correction procedure is described in section 2.3. It follows the overview of the resulting new categorization schema. Section 4 explains how to append the additional categories to the SOEP-IS DRM data.

2. Categorizing open answers in SOEP-IS DRM

2.1 Initial categories and open answers

To mentally reconstruct the previous day, SOEP-IS respondents were initially asking for the time of their getting up and in the following they answer in temporal order what they did in the

following episodes. For each episode the interview comprises the following wording and pre-set activities:

*“[...] we are interested in the most important activity first. Please chose from the following list: On the way to work/ study or way home from there; Way to spare time activity, way back home; Working / studying; Shopping; Preparing food; Eating; Washing / showering / bathing / brushing teeth / dressing oneself / getting ready; Doing housework; Taking care of children, taking them to school, activities; Meet friends; Resting, taking a nap; Relaxing; Intimate relations; Worship, meditation; Watching TV; Reading; Computer/ internet; On the phone; Doing sports; Doctor visits; Gardening; Taking care of pets; **Other activity, namely;** End of the day: beginning of the night rest or 12am.” (SOEP-IS Questionnaire 2012, Question 238)*

Beside the named 23 activity categories respondents could answer *another activity* and state an open answer into the textbox. If a person reports a parallel second activity, the same list is used. In 2014 and 2015 two additional options were added to the list: *drinking tea/coffee* and *listening to the radio*. This reflects that the open answer category was heavily used especially for these activity types in the first two SOEP-IS DRM waves in 2012 and 2013.

2.2 Distinct types of open answers

To categorize the reported open answers, I distinguish between four differ types:

- (1.) Many answers in the open text fields can easily be assigned to one of the pre-set categories from the list. For instance, some respondents stated *working* in the open answer box. Such reports are corrected by assigning the open answer to the correct listed category *working/studying*.
- (2.) Not all open answers are assignable to one of the pre-set categories. Instead, they can constitute a new category of answers. Such answers are identified by the frequency of its appearance in the open text box. The new category comprises activities that are frequently stated but not yet assignable to one existing category. For instance, talking to partners or relatives was reported very often. The newly proposed categories are shown in *Table 1*.
- (3.) Some open answers remain unclear. Neither do they fit into one listed category nor do they qualify for a new to be created category. The reasons for the unclearness are

manifold. Some activities are rare and therefore not assignable, or the meaning of the open answer are unclear.

(4.) A few number of *another activity* reports are falsely made. The open answer box was left blank in these cases or a second activity was reported but no category was chosen.

2.3 Manual correction procedure

We assigned categories to a pooled sample of SOEP-IS DRM 2012-15 of employed and unemployed respondents. Retirees and other employment status individuals were excluded. To qualify for the pooled person-year-activity sample (unique identifier), either the main activity, the parallel activity or both needed to be marked as 23 *other activity*. In total N=1,912 episodes of the main activity are open answers. For the parallel activity were 638 episodes reported as *other activity*. So, overall N=2,503 episodes qualified for the correction procedure. An episode contains a main and a (potential) parallel second activity. As in many cases only main or parallel activity is answered openly, the already categorized episodes are left unchanged. This is the case for the main activity in 591 episodes while for the second activity 1,865 episodes are left unchanged.

The correction was carried out stepwise. Initially, all open answers were read manually and checked for their qualification for the listed 23 answering categories in 2012 and 2013 and the 25 pre-set answering categories in 2014 and 2015. In the case of an unequivocal answer of type (1.), the proposed listed category 1–25 was saved in the new variable ‘activity_new’ for the main activity and ‘activity2_new’ for the second parallel activity. The indicators ‘ind_openact’ (for main activity) and ‘ind_2openact’ (for parallel activity) are set to (1.) to identify this type of categorized episodes.

In a second step, open answers that did not qualify for the listed categories were manually screened for non-neglectable numbers of answers of the same kind. This is the starting point for creating new categories. The aim of this procedure was to keep the number of additionally created categories small while gaining additional data points for analysis. As a rule of thumb, a minimum of 30 open answers should be assignable to a new category. Again, the new categories are saved in ‘activity_new’ for the main activity and ‘activity2_new’ for the second parallel activity. The respective indicator for modification ‘ind_openact’ / ‘ind_2openact’ are set (2.).

The number of open answers that were not clearly assignable to an existing category or to a newly created category was indicated in ‘ind_openact’ / ‘ind_2openact’ as (3.). The meaning

of the reported answers was in many cases unclear or more than one activity was reported as open text. Some episodes were set (4.) as they did not contain an open answer.

In total, 956 episodes are assigned to the previously listed categories (main + second activity) and 960 activities are assigned to newly created activity groups (main + second activity).

3. New categorization scheme

The initial questionnaire in 2012 lists 23 activities to choose, one open answer category and an end of the day option. In the third wave of SOEP-IS DRM in 2014 this list was extended with to additional categories. The author proposes further 11 activity categories. Table 1 shows initial pre-set categories and new created categories. In the dataset the newly created variables ‘activity_new’ and ‘activity2_new’ contain these information for all checked episodes.

Among the initially listed main activities, *working/studying* (127 times), and *doing housework* (70 times) and *Taking care of children* (73 times) were reported most frequently. *Working/studying* open answers are often more in detail than the pre-set categories like *office work* or specific tasks like *preparing the bill for a customer*. Very specific tasks were cross checked with the two-digit NACE industry. For instance, farmers reporting tasks like *feeding the chicken* are distinguished from persons having pets. Open answers assigned to *doing housework* were in the most cases tasks that are clearly identifiable as housework (e.g. recycling, tidying up, ...). The most frequent main activities that are assigned to new categories are *meeting / talking to partner or relatives* (213 times), *DIY, handicrafts, renovate* (129 times), and *walking / stroll* (96 times).

The most frequently reported second activity was *listening to radio/music*. It was correct predominantly in 2012 and 2013 where this category was not available. And, among newly created categories *meeting / talking to partner or relatives* (133 times) could be assigned most frequently.

Table 1: Categories in activity_new & activity2_new

#	Name of activity (English translation)	Source	Comment on content
1.	On the way to work/ study or way home from there	SOEP-IS 2012	
2.	Way to spare time activity, way back home	SOEP-IS 2012	
3.	Working / studying	SOEP-IS 2012	
4.	Shopping	SOEP-IS 2012	
5.	Preparing food	SOEP-IS 2012	
6.	Eating	SOEP-IS 2012	
7.	Washing / showering / bathing / brushing teeth / dressing oneself / getting ready	SOEP-IS 2012	<i>Getting ready</i> comprises any preparation like <i>packing a bag for a journey</i>
8.	Doing housework	SOEP-IS 2012	
9.	Taking care of children, taking them to school, activities	SOEP-IS 2012	
10.	Meet friends	SOEP-IS 2012	
11.	Resting, taking a nap	SOEP-IS 2012	
12.	Relaxing	SOEP-IS 2012	Including <i>relaxing</i> in breaks of work.
13.	Intimate relations	SOEP-IS 2012	
14.	Worship, meditation	SOEP-IS 2012	
15.	Watching TV	SOEP-IS 2012	
16.	Reading	SOEP-IS 2012	
17.	Computer/ internet	SOEP-IS 2012	
18.	On the phone	SOEP-IS 2012	
19.	Doing sports	SOEP-IS 2012	
20.	Doctor visits	SOEP-IS 2012	
21.	Gardening	SOEP-IS 2012	
22.	Taking care of pets	SOEP-IS 2012	
23.	Other activity, namely: [<i>open answer</i>]	SOEP-IS 2012	
24.	End of the day: beginning of the night rest or 12am.	SOEP-IS 2012	
25.	Drinking coffee / tea	New: SOEP-IS '14	Newly created category in wave 2014.
26.	Listening to radio / music	New: SOEP-IS '14	Newly created category in wave 2014.
27.	Care giving to relatives	Created by author	
28.	Volunteering	Created by author	Directly stated <i>Ehrenamt</i> or <i>volunteering</i> in political parties, religious groups, and associations.
29.	Walking / stroll	Created by author	<i>Spazieren gehen</i> is walking about in a leisurely manner. It's not specifically directed to a pre-/post activity.
30.	Job search / job center	Created by author	<i>Job search</i> activities like writing an application letter or CV. Further, being at the job center.
31.	Meeting / talking to partner or relatives	Created by author	Meeting, talking, visiting, etc. – not “bring sb. to...”.
32.	Artistic activity: painting, drawing, singing, playing an instrument	Created by author	Carrying out activity by oneself, not visiting cultural activity as audience.
33.	Service of hairdresser, manicure, pedicure, cosmetician, masseur/masseuse	Created by author	Person-to-person service, no individual wellness (like sauna or spa).
34.	At party / events (“Fest”) / going out	Created by author	No soccer matches.
35.	Doing DIY, handicrafts, renovate	Created by author	
36.	Playing (board) games, solving quizzes	Created by author	Not computer/video games.
37.	Drinking alcoholic drinks, smoking	Created by author	

Source: SOEP-IS DRM data; own assignment.

4. Merging the categories to SOEP-IS DRM

SOEP-IS DRM data can easily be merged with the additional categorized activities described above. Every episode is uniquely identified by the variables for the person ID *pid*, the survey year *syear* and the running sequence number *sequence* within one specific DRM diary. This constitutes the variable key for merge to the distributed SOEP-IS DRM. After merging, the researcher can decide whether to use information from ‘activity_new’ and ‘activity2_new’ completely instead of the open answers in the variables ‘activity’ = 23 and ‘activity2’ = 23 or just a subset which can be restricted by ‘ind_openact’ and ‘ind_2openact’.

Literature

- Anusic, Ivana, Richard E. Lucas, and M. Brent Donnellan (2017): "The Validity of the Day Reconstruction Method in the German Socio-economic Panel Study", *Social Indicators Research* 130(1), pp. 213–232.
- Kahneman, Daniel, Alan B. Krueger, David A. Schkade, Norbert Schwarz, and Arthur A. Stone (2004): "A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method", *Science* 306(December), pp. 1776–1781.
- Richter, David and Jürgen Schupp (2015): "The SOEP Innovation Sample (SOEP IS)", *Schmollers Jahrbuch: Journal of Applied Social Science Studies* 135(3), pp. 389–399.