

## EU-Labour Force Survey November 2021 release

### Setup for importing the anonymised special data sets for 2002-2020

#### Content

- I. Overview
- II. Explanatory notes on the recoding of alphanumeric variables into numeric variables
- III. Explanatory notes on the recoding of missing values
- IV. Explanatory notes on specific recoding, renaming and label definitions
- V. Notes on irregularities not mentioned in the User Guide
- VI. Classifications

#### I. Overview

The following information will help you to import the EU-LFS 2002-2020 special data. The data you received from Eurostat are formatted as comma separated values (csv). With the routines provided by GESIS's German Microdata Lab, these data can be converted into SPSS and Stata.

The routines to convert the data into SPSS or Stata consist of a setup and a universal label syntax:

Setup\_EULFS\_2002-2020\_s.sps  
Labels\_EULFS\_1983-2020.sps

Setup\_EULFS\_2002-2020\_s.do  
Labels\_EULFS\_1983-2020.do

The routines achieve the following objectives:

- Transformation of EU-LFS special data from csv into sav (SPSS) and dta (Stata)
- Recoding of alphanumeric variables into numeric variables
- Recoding of missing values
- Assigning a release variable (RELEASE)
- Definition of variable and value labels (by executing the label syntax)

For a first check of the converted data, you can compare the number of cases per country with the number of cases listed in Table 2. Consistent numbers indicate that the execution of the setup was successful.

Note 1: The variable HATLEVEL has been renamed into HAT97LEV since the December 2014 release of the data (compared to older releases).

Note 2: The variables HAT11LEV, HATVOC, and EDUCVOC have been added with the November 2015 release.

Note 3: The variables ESEG1D and ESEG2D have been added with the December 2018 release.

## II. Explanatory notes on the recoding of alphanumeric variables into numeric variables

### – HHNUM, QHHNUM, QUARTER

These variables have a large number of values or – in case of QUARTER – only one value. Thus, they remain alphanumeric.

### – YEARESID

The alphanumerically coded 5-year groups (11–14, 15–19 etc.) are recoded into numeric codes in line with the standard aggregation of AGE.

### – COUNTRYW, COUNTR1Y

Numeric 2-digit codes are assigned for European countries. These codes display the alphabetical order of the country as given by the ISO codes (e.g. 06 for BG representing Bulgaria). The alphanumeric codes in the EU-LFS differ from the ISO codes with regard to the United Kingdom (UK instead of GB). In this case, the numeric code is assigned in accordance with the code used in the EU-LFS (i.e. 50 for UK).

Numeric 4-digit codes are assigned for countries outside Europe. The first two digits indicate the region, according to the country codification from 2020 (e.g. 05 for North Africa).<sup>1</sup> The last two digits display the alphabetical order of the country in the corresponding region as given by the ISO codes (e.g. 02 for EG representing Egypt, so that the whole code for Egypt is 0502).<sup>2</sup>

### – REGION, REGIONW, REGION1Y

The values of these variables have country-specific meanings (according to the NUTS classification).<sup>3</sup> That is why they are attached to the consequent country variables (COUNTRY, COUNTRYW, COUNTR1Y), excluding region variables with the code "00", which stands for "no information delivered" (see User Guide 2021). The new codes consist of 4 digits. Digits 1 and 2 are similar to the codification of COUNTRY/W/1Y. Digits 3 and 4 are similar to the original codification of REGION/W/1Y.

---

<sup>1</sup> [https://ec.europa.eu/eurostat/documents/1978984/6037342/Country\\_codification\\_from\\_2020\\_onwards.pdf](https://ec.europa.eu/eurostat/documents/1978984/6037342/Country_codification_from_2020_onwards.pdf)

<sup>2</sup> <https://www.iso.org/iso-3166-country-codes.html>

<sup>3</sup> [http://ec.europa.eu/eurostat/statistics-explained/index.php/EU\\_labour\\_force\\_survey\\_-\\_methodology#Regions](http://ec.europa.eu/eurostat/statistics-explained/index.php/EU_labour_force_survey_-_methodology#Regions)

### III. Explanatory notes on the recoding of missing values

Coding standard for missing values:

in SPSS and Stata	in Stata (optional)	Labels
-1	.a	"No answer"
-2	.b	"Not applicable"
-3	.c	"Not available"
-4	.d	"Not specified"
-5	.e	"No information delivered"
-9	.j	"Otherwise" in exceptional cases (only LEAVCLAS)

The Code -3/.c for "Not available" is assigned, if a variable has no valid cases in a country. Some of these variables have only system missings (i.e. blanks) in the raw data of the respective country, others have only cases with "No answer" and/or "Not applicable".

#### Please note:

Optionally, missing values in Stata can be encoded as the extended missing values .a, .b, .c, etc. by reinserting the block "Declare User-Missing Values" at the end of the respective setup.

Specifics and exceptions:

#### – NATIONAL, COUNTRYB

The code -1/.a includes not only cases without a valid answer, but also suppressed cases, cases with the original code "Other and Stateless", and in some countries certain other cases (see User Guide 2021, p. 75-77).

### IV. Explanatory notes on specific recoding, renaming and label definitions

#### – HHCOMP

The variable contains the value 50, which is not mentioned in the User Guide. All persons with this value are children with no adult household members present in the dataset. Thus, we labelled the value "No adult household members present in the dataset".

#### – COURFIELD, HATFIELD

The codes of the variables COURFIELD and HATFIELD changed in the year 2016 (see User Guide 2021). If (and only if) the same code is assigned in the code version until 2015 as well as in the version from 2016, the first part of the value label refers to the current version (codes from 2016), whereas the latter part refers to codes until 2015. For instance, the value label of code 100 is "Services / until 2015: Teacher training and education science".

#### – IS883D, IS88PR3D

The variables IS883D and IS88PR3D include the unknown value 11 in the years 2002-2010 which are not mentioned in the User Guide and which are not recoded.

## V. Notes on irregularities not mentioned in the User Guide

Note that the weighting factor COEFF can be zero or missing in some cases. This is correct and guarantees appropriate weighting.

Moreover, across years and countries, some variables contain system missings and irregular values (deviating from the User Guide) of unclear origin. These cases are not changed with the setup files. Tables 1 shows the number of unexplainable system missing values.

## VI. Classifications

For information on the classifications used in the EU-LFS, see

[http://ec.europa.eu/eurostat/statistics-explained/index.php/EU\\_labour\\_force\\_survey\\_-\\_methodology#Classifications\\_in\\_the\\_EU-LFS](http://ec.europa.eu/eurostat/statistics-explained/index.php/EU_labour_force_survey_-_methodology#Classifications_in_the_EU-LFS)

---

GESIS - Leibniz-Institute for the Social Sciences  
German Microdata Lab  
B6, 4-5  
68159 Mannheim  
Germany  
Phone: +49 (0)621-1246-265  
Fax: +49 (0)621-1246-100  
<http://www.gesis.org/en/gml/>  
E-Mail: [gml@gesis.org](mailto:gml@gesis.org)

---

Table 1: Number of unexplainable system missing values per variables and years in the EU-Labour Force Survey 2002-2020 special data

Variables	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011
coeff	.	.	.	.	.	.	750	.	.	.
degurba	.	.	.	.	.	.	.	.	.	.
hhlink	.	.	.	.	.	.	.	.	.	.
hhseqnum	.	.	.	.	.	.	.	.	.	.
ilostat	.	.	.	.	.	.	.	.	.	.
intwave	.	.	.	.	.	.	.	.	.	.
intweek	6,078	6,403	6,246	6,180	7,751	6,753	7,326	8,078	8,063	7,699
methoda	.	.	.	.	.	.	.	.	.	.
methodb	.	.	.	.	.	.	.	.	.	.
methodc	.	.	.	.	.	.	.	.	.	.
methodd	.	.	.	.	.	.	.	.	.	.
methode	.	.	.	.	.	.	.	.	.	.
methodf	.	.	.	.	.	.	.	.	.	.
methodg	.	.	.	.	.	.	.	.	.	.
methodh	.	.	.	.	.	.	.	.	.	.
methodi	.	.	.	.	.	.	.	.	.	.
methodj	.	.	.	.	.	.	.	.	.	.
methodk	.	.	.	.	.	.	.	.	.	.
methodl	.	.	.	.	.	.	.	.	.	.
methodm	.	.	.	.	.	.	.	.	.	.
na111y1d	.	.	.	.	.	.	.	.	.	.
nowkreas	.	.	.	.	.	.	.	.	.	.
refweek	.	.	.	.	.	.	.	.	.	.

Variables in	2010	2009	2008	2007	2006	2005	2004	2003	2002
coeff	.	16	27	120	.	.	.	.	.
degurba	.	.	.	.	.	112	105	.	.
hhlink	.	.	.	.	.	.	.	.	.
hhseqnum	.	16	27	.	.	.	.	.	.
ilostat	.	.	.	.	.	.	.	.	.
intwave	.	.	.	.	.	.	.	.	.
intweek	45,524	.	.	.	.	.	.	.	.
methoda	.	.	.	.	.	.	.	.	.
methodb	.	.	.	.	.	.	.	.	.
methodc	.	.	.	.	.	.	.	.	.
methodd	.	.	.	.	.	.	.	.	.
methode	.	.	.	.	.	.	.	.	.
methodf	.	.	.	.	.	.	.	.	.
methodg	.	.	.	.	.	.	.	.	.
methodh	.	.	.	.	.	.	.	.	.
methodi	.	.	.	.	.	.	.	.	.
methodj	.	.	.	.	.	3,615	3,711	3,918	.
methodk	.	.	.	.	.	3,615	3,711	3,918	.
methodl	.	.	.	.	.	.	.	.	.
methodm	.	.	.	.	.	3,615	3,711	3,918	.
na111y1d	.	.	.	.	.	.	.	.	.
nowkreas	.	.	.	.	.	.	.	.	.
refweek	.	.	.	.	.	.	.	.	.

Table 2: Number of cases per country in the EU-Labour Force Survey 2002-2020 special data

Year	N (without selection)			
	DK Denmark	FI Finland	LU Luxembourg	SE Sweden
2002	37,504	.	.	.
2003	36,407	63,667	.	.
2004	35,129	62,591	.	.
2005	34,337	62,422	.	.
2006	34,368	61,068	.	.
2007	52,338	59,942	.	.
2008	46,108	60,310	.	.
2009	46,327	58,686	.	48,623
2010	87,583	57,114	.	44,167
2011	36,958	54,892	.	41,207
2012	39,979	54,177	.	39,596
2013	40,303	52,503	.	36,113
2014	38,097	53,158	.	33,340
2015	36,787	51,391	11,475	29,812
2016	39,523	50,012	20,340	19,647
2017	31,401	48,920	18,806	19,224
2018	32,032	46,724	17,959	12,115
2019	32,577	45,490	19,568	10,795
2020	31,697	44,292	18,717	13,787
$\Sigma$	769,455	987,359	106,865	348,426