

EU-Labour Force Survey December 2009 release

Setup for importing the Anonymised Yearly Data Sets for 1998–2007

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 and the definition of variable and value labels
- V. Notes on irregularities not mentioned in the User Guide
- VI. Classifications

I. Overview

The following information is a guideline on how to import the EU-LFS 1998–2007 yearly data. The data supplied by Eurostat are formatted as comma separated values (csv). With the routines provided by GESIS' German Microdata Lab, these data can be converted into SPSS.

The routine to convert the data into SPSS consists of two files:

Setup_EULFS_1998–2007_y.sps

Labels_EULFS_1998–2009.sps

The setup provided by the German Microdata Lab accomplishes the following tasks:

- Transformation of EU-LFS 1998–2007 yearly data from csv into sav (SPSS).
- Recoding of alphanumeric variables into numeric variables.
- Recoding of missing values.
- Definition of variable and value labels (by executing Labels_EULFS_1998–2009.sps).

For a first check of the converted data, the number of cases per country can be compared with the number of cases listed in table 1. Consistent numbers indicate that the execution of the setup was successful.

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

– COUNTRYW, COUNTR1Y

The first two numbers indicate the region, according to the country codification from 2007.¹ Countries not explicitly assigned to one region by this classification have been classified by GESIS.

The last two digits indicate the ranking of the country within the alphabetical order according to the ISO 3166-1-alpha-2 code elements (e.g. United Arab Emirates=AE=1101).² The last two digits for surveyed countries remain the same as in the COUNTRY Variable (e.g. Bulgaria=0103). The alphanumerical code from the EU-LFS deviates from the ISO classification in regard to two countries: XS instead of RS for Serbia and UK instead of GB for the United Kingdom.

In Slovenia, the coding of COUNTRYW in the raw data differs from the other countries (see User Guide 2009: 59). Thus the recoding within the setup is different as well. The cases of "EU15" and "Non EU15" in the years from 1998 until 2003 have been assigned the values "1501" and "1502".

– HHNUM, QHHNUM, QUARTER, REGION, REGIONW, REGION1Y

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

III. Explanatory notes on the recoding of missing values

Coding standard for missing values:

- -1 for "No answer"
- -2 for "Not applicable"
- -3 for "Not available"
- in exceptional cases -9 for "Otherwise" (only LEAVCLAS), -101 for "No answer", -102 for "Not applicable" and -103 for "Not available (only LEAVTIME and STARTIME, see below)

In SPSS these missing values are declared as user-missing values.

The Code -3 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, while others only have cases with "No answer" and/or "Not applicable". For detailed

¹ http://circa.europa.eu/irc/dsis/employment/info/data/eu_lfs/LFS_MAIN/Coding_lists_explanatory_notes_and_classifications/Country_codification_from_2007_onwards.pdf

² http://www.iso.org/iso/country_codes/iso_3166_code_lists/country_names_and_code_elements.htm

information about which variables have no valid cases in certain years in any country and have thus been recoded to -3, see table 2.

Specifications and exceptions:

– COUNTRYB, NATIONAL

The code -1 includes not only cases without a valid answer, but also suppressed cases, cases with the original code "Other or Stateless", and in some countries certain other cases (see User Guide 2009: 63).

– LEAVTIME, STARTIME

These variables have valid negative values. Thus the missing value category is set to -101 (for "No answer"), -102 (for "Not applicable") and -103 (for "Not available").

IV. Explanatory notes on specific recoding and the definition of variable and value labels

– ISCO3D, ISCOPR1D, ISCOPR3D

These variables contained a number of coding errors. This became obvious while cross-tabulating the single digit with the three digit ISCO code. Thus the erroneous cases have been recoded.

– LEAVTIME, STARTIME

There were no cases containing the value 999 for "Not applicable" in the raw data, as mentioned in the User Guide 2009. Thus the "Not applicable" category has been derived with help of the filtering variables EXISTPR (for LEAVTIME) and WSTATOR (for STARTIME) (see User Guide 2009: 32f).

– RELEASE

A new variable is added by the setup to indicate the release date of the Dataset (i.e. DEC 2009).

– REGION, REGIONW, REGION1Y

The values of these variables have country-specific meanings (according to the NUTS classification)³. Thus they remain unlabelled.

V. Notes on irregularities not mentioned in the User Guide

– gr2005_y

There are seven cases with defective raw data. Due to a supernumerary comma for the variable COURPURP in the raw data, the values of all following variables are mismatched. To exclude these cases from the analysis, they can be identified as follows:

(QHHNUM=Q121686 and HHSEQNUM=1), (QHHNUM=Q123430 and HHSEQNUM=4),
(QHHNUM=Q127129 and HHSEQNUM=1), (QHHNUM=Q146354 and HHSEQNUM=3),
(QHHNUM=Q149929 and HHSEQNUM=3), (QHHNUM=Q149929 and HHSEQNUM=4) and
(QHHNUM=Q152057 and HHSEQNUM=4).

³ http://circa.europa.eu/irc/dsis/employment/info/data/eu_lfs/LFS_MAIN/Related_documents/Info_users_Regions_EN.xls

- COEFF
There are a number of cases with a weighting factor of zero in the data.
- COEFF, HHSEQNUM, HHLINK, ILOSTAT, INTWEEK, NA111Y1D, SEEKWORK, METHODA – METHODM
These Variables contain system-missings in some countries.
- HHSEQNUM
In countries where the final sampling unit are persons⁴ and not households, this variable does not contain any useful values.
- MARSTAT
In certain years in some countries, children under the age of 15 were not asked about their marital status, but falsely coded -1 for "No answer" instead of -2 for "Not applicable".

VI. Classifications

- ISCED classification (HATLEVEL):
http://circa.europa.eu/irc/dsis/employment/info/data/eu_lfs/LFS_MAIN/Related_documents/ISCED_EN.htm
- other classifications used in the EU LFS:
http://circa.europa.eu/irc/dsis/employment/info/data/eu_lfs/LFS_MAIN/Coding_lists_explanatory_notes_and_classifications/Index_EU_LFS_coding_%20lists_explanatory_notes_and_classifications.htm

GESIS - Leibniz-Institute for the Social Sciences
 German Microdata Lab
 Square B2, 1
 68159 Mannheim
 Tel: 0621-1246-265
 Fax: 0621-1246-100
<http://www.gesis.org/gml>
 E-Mail: gml@gesis.org

⁴ See section: Characteristics of the national surveys at http://circa.europa.eu/irc/dsis/employment/info/data/eu_lfs/lfs_main/lfs/lfs_publications.htm

Table 1: Number of cases per country in the EU-Labour Force Survey 1998-2007, yearly data

Country	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
01 AT Austria	61.902	61.618	59.146	59.296	58.717	56.268	45.185	205.058	201.083	201.771
02 BE Belgium	80.066	27.479	27.446	26.391	27.834	27.565	27.739	109.126	114.278	111.978
03 BG Bulgaria	-	-	53.404	59.991	57.549	38.524	36.540	140.035	134.298	130.964
04 CY Cyprus	-	10.135	10.301	10.596	10.667	10.779	10.615	39.514	38.227	38.462
05 CZ Czech Republic	70.721	68.824	65.464	63.964	62.091	60.973	62.807	249.737	253.254	250.602
06 DE Germany	-	-	-	-	328.074	330.297	327.089	477.276	49.760	49.216
07 DK Denmark	18.233	17.895	17.993	15.986	16.081	15.471	15.445	59.792	58.899	117.997
08 EE Estonia	16.861	16.361	4.676	5.179	4.948	4.735	4.533	17.904	20.189	23.342
09 ES Spain	190.911	196.532	180.853	173.643	172.552	175.209	175.159	610.851	106.279	107.224
10 FI Finland	17.063	35.071	42.146	42.712	42.743	40.782	39.680	39.500	38.353	60.517
11 FR France	183.072	182.155	182.066	178.143	175.939	87.413	88.185	343.958	339.775	350.757
12 GR Greece	84.007	82.921	81.264	80.282	77.451	73.017	80.992	314.576	303.621	296.383
13 HU Hungary	84.980	87.541	85.080	84.671	82.904	88.212	82.310	312.190	311.397	305.048
14 IE Ireland	111.342	109.768	106.306	105.405	105.569	101.500	86.545	91.174	78.599	85.133
15 IS Iceland	3.755	3.656	3.697	3.685	3.646	3.186	3.110	12.424	12.550	12.435
16 IT Italy	201.835	200.625	199.367	196.236	194.041	192.359	172.264	704.372	684.303	677.746
17 LT Lithuania	7.543	7.558	7.583	7.788	12.993	12.219	12.149	47.947	44.958	63.022
18 LU Luxembourg	17.326	16.095	15.257	14.814	13.429	16.394	21.189	90.024	85.080	21.178
19 LV Latvia	18.756	18.701	19.193	18.834	5.943	6.022	6.020	21.950	18.640	37.822
20 NL Netherlands	55.959	51.929	66.992	90.265	97.594	96.740	112.913	469.514	108.545	106.472
21 NO Norway	20.664	20.717	21.095	20.213	20.838	20.931	21.306	85.331	20.663	20.785
22 PL Poland	54.545	53.300	46.295	57.575	58.623	58.303	57.433	225.325	216.840	206.875
23 PT Portugal	50.067	47.315	45.626	45.681	45.617	46.385	50.714	190.128	179.534	171.960
24 RO Romania	47.718	46.261	44.805	43.742	41.757	41.556	68.596	269.408	257.982	249.521
25 SE Sweden	15.810	17.909	17.321	50.512	55.699	56.316	52.185	148.124	207.115	205.302
26 SI Slovenia	17.976	19.627	18.751	19.607	19.766	19.881	18.871	71.406	69.813	67.877
27 SK Slovak Republic	31.301	30.442	30.846	30.554	29.420	28.889	28.772	116.250	113.417	110.377
28 UK United Kingdom	144.979	143.058	140.067	135.887	136.156	130.417	125.610	123.141	120.651	120.379
Total	1.607.392	1.573.493	1.593.040	1.641.652	1.958.641	1.840.343	1.833.956	5.586.035	4.188.103	4.201.145

^a see also "datafileinfo_year_1998-2007.csv" on the CD you received from Eurostat

Table 2: Variables without valid cases (-3) in certain years^a, EU LFS yearly data

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
COURATT	X	X	X	X	X					
COURFILD	X	X	X	X	X					
COURLEN	X	X	X	X	X					
COURPURP	X	X	X	X	X					
COURWORH	X	X	X	X	X					
EDUCFILD	X	X	X	X	X					
EDUCSTAT	X	X	X	X	X					
HATFIELD	X	X	X	X	X					
HWOVERP	X	X	X	X	X	X	X	X		
HWOVERPU	X	X	X	X	X	X	X	X		
HWUSTRAI						X	X	X	X	X
LENT4W							X	X	X	X
NACE1D	X	X	X	X	X	X	X	X	X	X
NACE1Y1D	X	X	X	X	X	X	X	X	X	X
NACE2J1D	X	X	X	X	X	X	X	X	X	X
NACEPR1D	X	X	X	X	X	X	X	X	X	X
NEEDCARE	X	X	X	X	X	X	X	X		
PURP4W									X	X
SIGNISAL	X	X	X	X	X	X	X	X		
SUPVISOR	X	X	X	X	X	X	X	X		
TEMPAGCY	X	X	X	X	X	X	X	X		
WAYJFOUN	X	X	X	X	X	X	X	X		

^a Variables with either only system-missings (i.e. blanks) in the raw data or only cases with "No answer" and/or "Not applicable".