UK Data Service

Teaching Dataset

Quarterly Labour Force Survey, January - March, 2015

User Guide

Author: Margherita Ceraolo

Version: 2

Date: 26th August 2016

Contents

About this User Guide

This is a user guide for the Quarterly Labour Force Survey (QLFS), January-March, 2015: Unrestricted Access Teaching Dataset. This teaching dataset has been created for teaching related purposes only such as for teaching basic statistics or for students to write report for coursework. Those wishing to use the Labour Force Survey (LFS) for research purposes are advised to use the original SN7725 Quarterly Labour Force Survey, January-March, 2015. The QLFS January-March 2015: Unrestricted Access Teaching Dataset contains a reduced number of variables from the original QLFS January-March 2015. It gives students the opportunity to gain experience of using nationally representative UK data.

Introduction to the Labour Force Survey

The Labour Force Survey (LFS) is the largest regular social survey in the United Kingdom. It is a unique source of information using international definitions of employment and unemployment and economic inactivity, together with a wide range of related topics such as occupation, training, hours of work and personal characteristics of household members aged 16 years and over living at private addresses in the United Kingdom.

Its purpose is to provide information on the UK labour market that can then be used to develop, manage, evaluate and report on labour market policies. The survey is managed by the Social Surveys division of the Office for National Statistics (ONS) in Great Britain and by the Central Survey Unit of the Department of Finance and Personnel in Northern Ireland. Data from each Quarterly Labour Force survey are anonymised and deposited with the UK Data Archive at the University of Essex and are distributed through the UK Data Service.

Main Topics

The LFS questionnaire comprises a 'core' of questions which are included in every quarter of the survey, together with 'non-core' questions which vary from quarter to quarter. The core questions cover respondent's household, family structure, basic housing information and

demographic details of household members. The 'non-core' questions cover economic activity, education and health. They might also include a few questions asked on behalf of other government departments such as on childcare or regional mobility and might be included in the survey only every 2 or 3 years.

The Datasets

There are a number of different types of LFS datasets available from the UK Data Service such as quarterly dataset, household data, local area data, Eurostat data, two-quarterly and five-quarterly longitudinal datasets. The UK Data Service website contains fuller explanations of each of the datasets. The teaching dataset is based on the January-March 2015 Quarterly LFS, which contains data from that quarter only.

The Sample Design

Since 1992, the LFS has had a single stage, stratified random sample design drawn from the Postcode Address File (PAF). A household, once initially selected for interview, is retained in the sample for a total of five consecutive quarters. The first interview is face-to-face and subsequent ones are by telephone. Interviewers can accept proxy information for household members who are unavailable when the interview takes place.

Other Useful Sources of Information

Further information about the methodology of the LFS can be found in the LFS user guide – Volume 1 - Background and Methodology on the UK Data Service website. This contains detailed explanations of the survey design, sampling errors and weighting, among other things.

There is also a series of UK Data Service guides, including introductory guides to using the large scale government surveys and to the analysis packages SPSS and Stata.

Additional information about LFS, including publications, summaries of labour market statistics, guidance and methodology is available on the ONS webpage.

Introduction to the QLFS January-March 2015: Unrestricted Access Teaching Dataset

The QLFS January-March 2015: Unrestricted Access Teaching Dataset contains 13 variables. The variables have been chosen to enable users to explore the data and apply some basic statistical techniques such as regression analysis, ANOVA, comparing means and cross-tabs.

The total sample size of the Unrestricted Access Teaching Dataset is 22,428, which is a sub-sample of the original QLFS dataset. It is based on a random selection of 30 per cent of the main QLFS sample, excluding children aged 0-15. The syntax for this is in Appendix II.

Details of the variables selected for the Unrestricted Access Teaching Dataset are given below including a data dictionary which lists information about variable names, values, labels, missing values and frequency. The name of the variables and their labels remain the same or very close to those in the original LFS dataset. However, due to concerns about statistical disclosure, some variables have been recoded and their level of detail reduced for this teaching dataset. Variables that differ from the original LFS have been suffixed with "r".[5] The syntax for how these variables were recoded is in Appendix II. In addition, a new ID variable was created in order to avoid linking the Unrestricted Access Teaching Dataset with the original dataset.

Weighting the Dataset

Respondents to the survey did not have an equal chance of selection. There were also differences in response rates between groups. To adjust for the unequal selection probabilities and non-response, users will usually need to use the appropriate weight when analysing the LFS.
This dataset includes two weight variables:

- pwt14– is the original person weight used to reduce the effect of non-response bias and produce population totals and means. This weight is designed to gross the frequencies to the population of England and Wales. A 'grossing weight' is useful for calculating the prevalence of phenomena at the nation level. However, users should aware that grossing weights can cause errors when calculating the precision of estimates

- pwt14r – has been calculated for this teaching dataset. It is the original annual weight scaled to have a mean of 1. This weight variable will be particularly useful for new users exploring statistics in statistical packages such as SPSS.

---

[5] With the exception of "ilodefr" (economic activity –reported) which is the original variable name.

List of Variables

Observations: 22,428

Variables: 13

Variable name	Variable label	Variable type

CASENEW[6]	New random ID number	Scalar
PWT14	Person weight	Scalar
SEX	Sex of respondent	Nominal
AGEEULR[7]	Age bands in 5 year intervals – 12 categories (recoded)	Ordinal
MARSTA3R[8]	Marital status – 3 categories (recoded)	Nominal
HIQUL15D	Highest qualification – detailed grouping	Nominal
ETHUK7R[9]	Ethnicity –7 categories (recoded)	Nominal
ILODEFR	Economic activity (reported)	Nominal
STAT3R[10]	Employment status – 3 categories (recoded)	Nominal
FTPTWK	Full-time or part-time in main job	Nominal
TOTHRS	Total hours worked in reference week	Scalar
NSECMJ3R[11]	NS-SEC 3 (main job) (SOC2010) (recoded)	Ordinal
GOVTOF2	Government Office Region 2 and 3 combined	Nominal

---

[6] New computed variable
[7] Derived from "AGEEUL" (28 categories)
[8] Derived from "MARSTA" and "mardy6"
[9] Derived from "ETHUKEUL" (9 categories)
[10] Derived from "STAT" (4 categories)
[11] Derived from "NSECMJ10" (8 categories)

Missing Values

Missing values occur when a respondent does not answer a particular question. This may happen for a number of reasons, including refusal or inability to answer a question, or where it does not apply to them. It is common practice to exclude missing values from analyses, although they may occasionally be of interest. Both SPSS and Stata have different conventions for missing values:

	SPSS dataset	Stata dataset
No answer	-8	.a
Does not apply	-9	.b
System missing	.	.

Downloading the Data

Unlike other government surveys the QLFS January-March 2015: Unrestricted Access Teaching Dataset does not require any registration and can be downloaded via the UK Data Service website.

The dataset is available in two formats: SPSS and STATA.

Appendix I: Data Dictionary

CASENEW

N	Valid	22428
	Missing	0
Mean		15019.8455
Std. Deviation		8672.81083
Range		29999.00
	10	3049.0000
	25	7551.2500
Percentiles	50	14920.5000
	75	22559.0000
	90	27100.0000

PWT14

N	Valid	22428
	Missing	0
Range		7442

SEX

		Frequency	Percent	Valid Percent	Cumulative Percent
	Male	10656	47.5	47.5	47.5
Valid	Female	11772	52.5	52.5	100.0
	Total	22428	100.0	100.0	

AGEEULR

		Frequency	Percent	Valid Percent	Cumulative Percent
	15-19	1334	5.9	5.9	5.9
	20-24	1455	6.5	6.5	12.4
	25-29	1652	7.4	7.4	19.8
	30-34	1850	8.2	8.2	28.0
	35-39	1717	7.7	7.7	35.7
Valid	40-44	1946	8.7	8.7	44.4
	45-49	2084	9.3	9.3	53.7
	50-54	2121	9.5	9.5	63.1
	55-59	1978	8.8	8.8	72.0
	60-64	1798	8.0	8.0	80.0

		Frequency	Percent	Valid Percent	Cumulative Percent
	65-69	1961	8.7	8.7	88.7
	70 and over	2532	11.3	11.3	100.0
	Total	22428	100.0	100.0	

MARSTA3R

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Single, never married	7179	32.0	32.0	32.0
	Married/ Cohabiting/ Civil Partner	12303	54.9	54.9	86.9
	Divorced/ Widowed/ Previously in Civil Partnership	2946	13.1	13.1	100.0
	Total	22428	100.0	100.0	

HIQUL15D

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Degree or equivalent	5224	23.3	26.1	26.1
	Higher education	1937	8.6	9.7	35.8
	GCE, A-level or equivalent	4539	20.2	22.7	58.5
	GCSE grades A*-C or equivalent	4195	18.7	21.0	79.5
	Other qualifications	1820	8.1	9.1	88.6
	No qualification	2096	9.3	10.5	99.0
	Don't know	194	.9	1.0	100.0
	Total	20005	89.2	100.0	
Missing	Does not apply	2353	10.5		
	No answer	70	.3		
	Total	2423	10.8		
Total		22428	100.0		

ETHUK7R

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	White	20151	89.8	89.9	89.9
	Mixed/Multiple ethnic groups	201	.9	.9	90.8
	Indian	495	2.2	2.2	93.0
	Pakistani or Bangladeshi	436	1.9	1.9	95.0
	Chinese or any other Asian background	294	1.3	1.3	96.3

	Black/African/Caribbean/Black British	515	2.3	2.3	98.6
	Other ethnic group (includes White - Gypsy or Irish Traveller and White - Polish)	319	1.4	1.4	100.0
	Total	22411	99.9	100.0	
Missing	No answer	17	.1		
Total		22428	100.0		

ILODEFR

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	In employment	13751	61.3	61.3	61.3
	ILO unemployed	754	3.4	3.4	64.7
	Inactive	7923	35.3	35.3	100.0
	Total	22428	100.0	100.0	

STAT3R

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Employee	14818	66.1	85.7	89.0
	Self-employed	2385	10.6	13.8	99.6
	Government scheme or unpaid family worker	81	.4	.5	100.0
	Total	17284	77.1	100.0	
Missing	Does not apply	5144	22.9		
Total		22428	100.0		

FTPTWK

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Full-time	11999	53.5	69.5	69.5
	Part-time	5264	23.5	30.5	100.0
	Total	17263	77.0	100.0	
Missing	Does not apply	5150	23.0		
	No answer	15	.1		
	Total	5165	23.0		
Total		22428	100.0		

TOTHRS

N	Valid	13471
	Missing	8957
Mean		31.68
Std. Deviation		17.602
Range		97
Minimum		0
Maximum		97

NSECMJ3R

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Higher managerial, administrative and professional occupations	6922	30.9	30.9	30.9
	Intermediate occupations and small employers	4029	18.0	18.0	48.8
	Routine and manual occupations	5448	24.3	24.3	73.1
	Never worked, unemployed, and nec	6029	26.9	26.9	100.0
	Total	22428	100.0	100.0	

GOVTOF2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	North East	976	4.4	4.4	4.4
	North West (inc Merseyside)	2478	11.0	11.0	15.4
	Yorkshire and Humberside	1982	8.8	8.8	24.2
	East Midlands	1624	7.2	7.2	31.5
	West Midlands	1870	8.3	8.3	39.8
	Eastern	2080	9.3	9.3	49.1
	London	2387	10.6	10.6	59.7
	South East	3113	13.9	13.9	73.6
	South West	1970	8.8	8.8	82.4
	Wales	1125	5.0	5.0	87.4
	Scotland	1914	8.5	8.5	95.9
	Northern Ireland	909	4.1	4.1	100.0
	Total	22428	100.0	100.0	

Appendix II: SPSS Syntax for variables derived for the QLFS January-March 2015:

Unrestricted Access Teaching Dataset

Compute new ethnicity variable: recode ETHUKEUL into ETHUK7R

```
RECODE ETHUKEUL (-8=-8) (-9=-9) (1=1) (2=2) (3=3) (4
thru 5=4) (6 thru 7=5) (8=6) (9=7) INTO ETHUK7R.

VARIABLE LABELS   ETHUK7R 'Ethnicity UK 7 cat. (recoded)'.

VALUE LABELS ETHUK7R   -8 'No answer' -9'Does not apply' 1'White'
2'Mixed/Multiple ethnic groups' 3'Indian' 4'Pakistani or Bangladeshi'
5'Chinese or any other Asian background'
6'Black/African/Caribbean/Black British' 7'Other ethnic group (includes
Northern Irish Gypsy, Traveller or Irish Traveller)'.

EXECUTE.
```

Compute new marital status variable: compute "MARSTA3R" from "MARSTA" (Marital status) and "MARDY6" (Married/co-habiting/civil partners)

```
COMPUTE MARSTA3R=0.

IF  (MARSTA=1)  MARSTA3R=1.

IF  (MARSTA=2)  MARSTA3R=2.

IF  (MARSTA=3)  MARSTA3R=2.

IF  (MARSTA=4)  MARSTA3R=3.

IF (MARSTA=5) MARSTA3R=3.

IF (MARSTA=6 & MARDY6=2) MARSTA3R=3.

IF (MARSTA=6 & MARDY6=1) MARSTA3R=2.

VARIABLE LABELS MARSTA3R 'Marital status (recoded)'.

VALUE LABELS MARSTA3R 1'Single, never married' 2 'Married/ Cohabiting/
Civil Partner' 3 'Divorced/ Widowed/ Previously in Civil Partnership'.

EXECUTE.
```

Compute new NS-SEC 3 variable: recode NSECMJ10 into NSECMJ3R

```
RECODE NSECMJ10 (-9=-9) (-8=-8) (8=4) (1 thru 2=1) (3 thru 4=2) (5 thru
7=3) INTO nsecmj3r.

VARIABLE LABELS NSECMJ3R'NS-SEC 3 class (main job, SOC2010, recoded)'.
```

```
VALUE LABELS NSECMJ3R -8 'No answer' -9'Does not apply' 1'Higher
managerial, administrative and professional occupations' 2'Intermediate
occupations and small employers' 3'Routine and technical occupations'
4'Never worked, unemployed, and nec'.
EXECUTE.
```

Compute new employment status variable: recode STAT into STAT3R

```
RECODE STAT (-9=-9) (-8=-8) (1=1) (2=2) (3 thru 4=3) INTO STAT3R.

VARIABLE LABELS STAT3R 'Employment status recoded'.

VALUE LABELS STAT3R -8 'No answer' -9'Does not apply' 1'Employee'
2'Self-employed' 3'Government scheme or unpaid family worker'.

EXECUTE.
```

Compute new age variable: recode AGEEUL (Age band 28 cat.) into AGEEULR (Age band 12 cat.)

```
RECODE AGEEUL (14=3) (15=4) (16=5) (17=6) (18=7) (19=8) (20=9) (21=10)
(22=11) (5 thru 8=1) (9 thru 13=2) (23 thru 28=12) INTO AGEEULR.

VARIABLE LABELS AGEEULR ' Age bands in 5 year intervals (recoded)'.

VALUE LABELS AGEEULR 1 '16-19' 2 '20-24' 3 '25-29' 4 '30-34' 5 '35-39'
6 '40-44' 7 '45-49' 8 '50-54' 9 '55-59' 10 '60-64' 11 '65-69' 12 '70
and over'.

EXECUTE.
```

Select only cases age 16+ and delete unselected cases i.e. ages 0-15.

```
FILTER OFF.
USE ALL.
SELECT IF (ageeul>=5).
EXECUTE.
```

Reduce the sample size: select only 30% of the cases and delete unselected cases

```
FILTER OFF.
USE ALL.
SAMPLE   .30.
```

```
EXECUTE.
```

Compute a new ID variable with random numbers (from 0-15000) and then sort dataset by the new ID variable ascending.

```
COMPUTE casenew=TRUNC (UNIFORM(30000)) +1.

VARIABLE LABELS   casenew 'New random ID number'.

EXECUTE.

SORT CASES BY casenew (A).

EXECUTE.
```