The standardisation routines for highest level of completed education

When harmonising household surveys, LIS has to deal with all the typical national educational systems in the different member countries.

The slots in which information on educational attainment can be stored are:
 - PEDUC for general education, and/or
 - PTOCC for vocational training.

When filling these variables, the rule is that all original detail is preserved.

As such, no information is lost, but the comparability between countries suffers.

An internationally comparable variable is needed, when for instance, running regressions. Therefore, LIS supplies a routine that derives such a variable.

Because this new variable comes in addition to the two original ones, users now have a choice to either work with the full detail, or to use standardised categories.

Methods

- The concept for the standardised variable is based on the International Standard Classification of Education from UNESCO, ISCED97 (see references below), and uses highest attained level of education.
- Only completed training is taken into account. If the original survey contains information about incomplete educational attainments (e.g. "started" "incomplete, "without leaving certificate") then the standardised variable is coded to the next lower level.

LIS standardised levels

The levels are expressed as low, medium and high, and correspond to ISCED97 in the following manner:

1. LOW, including	ISCED levels 1 and 2
no education	0
pre-primary	0
primary	1
lower secondary education	2A
compulsory education	2A
initial vocational education	2B, 2C
2. MEDIUM, including	ISCED levels 3 and 4
upper secondary general education	3A
basic vocational education	3B, 3C
secondary vocational education	4B, 4C
post-secondary education *(including either shorter vocational courses or programs preparing for courses on tertiary level)*	4A
3. HIGH, including	ISCED levels 5 and 6
specialized vocational education	5B
university/college education	5A
(post)-doctorate and equivalent degrees	6
9. All else, including	
missing	
unknown	
level undefined	
other school degree	
not asked	
still in school	
These vague labels are kept in a separate category to keep the true levels free from pollution.	

Recoding

The recoding is done using the variables PEDUC or PTOCC or both, depending on availability. The ISCED97 manual serves as the main guide for all OECD countries. For other countries, the LIS contact person is consulted. In addition, we make use of several web-based documents to ensure that the recoding is as comparable as possible.

LIS attempts to derive the standardised variable for as many datasets as possible. However, it will be clear that standardisation may not be possible for several reasons. For instance, when the survey did not collect levels of education, but instead contains years in education, or age when leaving full-time education, standardisation is not possible.

A problem that we regularly face is when original categories contain a mix of educational levels. If a value of PEDUC corresponds to both ISCED levels 2 and 3, for example, the enrollment figures from ISCED are used to see if the vast majority can be assigned to either of the two.

As a rule, recoding is only attempted in clear situations. In other words, we do not make assumptions when the information is not clear.

The details on the recodes are available for each dataset we provide either the concordance table, or explain why the recode is not carried out.

For those datasets where no recoding is made available, LIS users may of course feel free to use their own judgement and pursue a recode at their own responsibility.

References

Classifying Educational Programs, Manual for ISCED-97, Implementation in OECD Countries 1999 Edition. Paris: UNESCO.

Education at a Glance: OECD Indicators 2005 Edition, Paris: OECD.

UNESCO World Education Forum Education for all , Paris, UNESCO

Link to UNESCO ISCED97 classification :

http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm

Instructions

The standardisation routine is available both for the person-level and household files.
A) <u>person file:</u>
The routine EDUCRECODEPP standardises the person-level education and uses the variables COUNTRY, PEDUC and PTOCC.
Users should be careful to include these variables in the keep-statement.
The outcome of the routine is a new variable EDUC.

B) <u>household file:</u>
The routine EDUCRECODEHH standardises the level of education of head and spouse, and uses the variables COUNTRY, D10, D11, D12 and D13.
Users should be careful to include these variables in the keep-statement.
The outcome of the routine are two new variables EDUCHD and EDUCSP.

Variables EDUC, EDUCHD and EDUCSP all have the following codes and labels:
1. low
2. medium
3. high
9. all else

Syntax that executes the routine

The routine is available in each of the three statistical packages that the LISSY remote access system supports.

SAS
- person file   : %include "&myincl.educrecodepp.sas" ;
- household file : %include "&myincl.educrecodehh.sas" ;

SPSS
- person file   : include file edurecop .
- household file : include file edurecoh .

Stata
- person file   : run $myincl\educrecodepp.do
- household file : run $myincl\educrecodehh.do

Note :
The routines are not full programs, but merely contain the necessary steps to recode the LIS variables.