Janez Štebe:

Introductory Exercise: Establish extent of precarious employment in EU countries and explore potential for comparative analysis

Training Course on EU-LFS, September 17th-20th 2014, Ljubljana

Content:

I. Explore the data set
II. Prepare the working data set
III. Precarious employment in different countries – separate analysis by countries

IV. Including the macro level variable into explanation – joint analysis

Before you start

Select only working age population (15-74) and respondents living in private households for analysis.

I. Explore the data set

Start the analysis by checking the structure of the data file. Does it contain the expected variables? Do they contain the definitions of the missing values? What is the order of the variables in a data set? What are the units of the analysis?

(*Display, frequencies, codebook routines in SPSS*)

II. Prepare the working data set

a) **Select the relevant population you need to work with. In order to analyse the forms of precarious employment we will limit to currently employed population.**

Solution (Tip: Use Format Painter to make the solution visible):

freq WSTATOR .

* You can either type the command into syntax or use the menu Data-->Select... and then paste and execute it.

DATASET COPY working_pop.

DATASET ACTIVATE working_pop.

FILTER OFF.

USE ALL.

select if (WSTATOR = 1 or WSTATOR=2) .

EXECUTE.

*Check the result.

freq WSTATOR .

freq country.

End of Solution

b) **While comparing countries you may wish to obtain equal sample size of the selected population by countries.**

Question: Do the data set contains weights of some kind? Which shall we use?

Solution:

See explanation in EUROSTAT (2013): Quality report of the European Union Labour Force Survey 2012 - 2014 edition
http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-TC-14-001

Weights usually express the inverse probability of selection. You can multiply different (independent) type of weights if exists. Since no weight variable is present in a training data set, we will create a constant one to begin with.

End of Solution

Task

Prepare the weighting variable and activate it to obtain in each country equal sample size.

Solution:

compute COEFF=1.

* make the weight active.

WEIGHT by coeff.

*Check the result. With the COEFF=1 nothing should happen.

freq country.

* obtain the values for the new weight coeff that will adjust sample size.

* for safety reasons, some commands require file to be sorted on key variables, therefore we will sort by country.

SORT CASES BY COUNTRY.

AGGREGATE

 /OUTFILE=* MODE=ADDVARIABLES

 /PRESORTED

 /BREAK=COUNTRY

 /N_BREAK=N.

AGGREGATE

 /OUTFILE=* MODE=ADDVARIABLES

 /PRESORTED

 /N_tot=N.

freq N_tot.

cross N_BREAK by country.

*produce weight coeff in order to have equal sample size: N_tot/ number of countries .

compute coeff= coeff*((N_tot/21)/N_BREAK).

 *Check the result. all sample size have to be equal. .

freq country.

End of Solution

III. Precarious employment in different countries

a) Identify the variables that could be used for analysis of precarious employment

Candidates are STARTIME temp ftpt .

b) prepare some variables for further descriptive analysis

Task

Start with some basic descriptive analysis. Compare countries to establish differences. Include also some bivariate analysis comparing the subpopulations in countries in order to see, if correlations among variables shows any differences depending of institutional context.

In our example we choose the STARTIME as the dependent variable and temp, ftpt, sex and age as independent. At the end of the exercise we will pursue the linear regression analysis. For that purpose we will prepare in advance and create dummy variables where apply.

Solution:

freq STARTIME temp ftpt .

* create dummy variables.

RECODE TEMP (2=1) (else=0) INTO temp_lm.

recode ftpt (1=1) (2=0) (MISSING=SYSMIS) INTO FT.

VARIABLE LABELS temp_lm 'Permanent_dummy'.

VARIABLE LABELS FT 'Full_time_dummy'.

EXECUTE.

*check.

cross temp by temp_lm.

cross ftpt by ft.

* prepare age for descriptive analysis.

RECODE age (17 22=20) (27 =27) (32 37=35) (42 THRU 52= 47) (57 thru 72=65) INTO age5.

var lab age5 'Lifecycle - 5 groups seniority levels (recode age)'.

val lab age5 20 'up to 22 years old' 27 'up to 29 years old' 35 'up to 40 years old' 47 'up to 54 years old'   65 'up to 72 years old' .

format age5 (f2.0).

freq age5.

* create dummy variables.

recode sex (1=1) (else=0) into sex_male .

cross sex by sex_male.

freq sex_male temp ftpt ft .

End of Solution

Ideas for thinking: Why did we handle missing differently whyle recoding 'TEMP'?

c) Use a limited set of countries for country level oriented exploratory analysis.

Task

Select countries that have representatives among workshop participants. It is more practical to do the exploratory analysis on a limited set of countries. Present some descriptive statistics on a country level and on the descriptive independent variables by country level.

Note that we will save the current data set with all the countries for further analysis at the end of session.

Solution:

DATASET COPY country_sel.

DATASET ACTIVATE country_sel.

FILTER OFF.

USE ALL.

SELECT IF (COUNTRY= 7 | COUNTRY=15 | COUNTRY=16 | COUNTRY=18 | COUNTRY=23 | COUNTRY=25 | COUNTRY=27 | COUNTRY=29 | COUNTRY=31).

EXECUTE.

* chec select.

freq country.

*Select countries that have representatives among workshop participants. It is more practical to do the exploratory analysis on a limited set of countries.

means   STARTIME temp_lm ft sex_male by country

 /CELLS MEAN MEDIAN COUNT STDDEV

 /STATISTICS ANOVA .

*Select countries that have representatives among workshop participants. It is more practical to do the exploratory analysis on a limited set of countries.

MEANS TABLES=STARTIME by COUNTRY BY age5 sex temp ftpt

 /CELLS MEAN MEDIAN COUNT STDDEV.

End of Solution

d) Display separate analysis by country

Task

Split file into portions by country and perform some further exploratory analysis. Conclude with the linear regression analysis of STARTIME including the set of individual level independent variables.

Solution:

SPLIT FILE LAYERED BY COUNTRY.

MEANS TABLES=STARTIME by age5 sex temp ftpt
 /CELLS MEAN MEDIAN COUNT STDDEV
 /STATISTICS ANOVA .

corr STARTIME with temp_lm ft sex_male .

regression VARIABLES = STARTIME sex_male age temp_lm ft
 /DEPENDENT STARTIME
 /METHOD enter sex_male age temp_lm ft .
End of Solution

Ideas for thinking: How would you test if one country is statistically different from another?

IV. Including the macro level variable into explanation

a) Aggregate the information from individual level data into country level table

Task

Pull country rate of unemployment out of data and store it in a table. Add contextual variable into individual level file.

Solution:

DATASET ACTIVATE DataSet1.

freq ILOSTAT .

freq country.

* Create dummy unemploy.

Recode ILOSTAT (2= 1) (1=0) (else=sysmiss) into unemploy.

* check.

cross ilostat by unemploy / missing = include.

* display.

means unemploy by country

 /CELLS MEAN   COUNT

 /STATISTICS ANOVA.

* put unemploy rates to table.

DATASET DECLARE unemploy_mean.

* if not presorted it' requred to sort.

SORT CASES BY COUNTRY.

AGGREGATE

 /OUTFILE='unemploy_mean'

 /PRESORTED

 /BREAK=COUNTRY

 /unemploy_mean=MEAN(unemploy).

DATASET ACTIVATE unemploy_mean .

*see result.

list .

End of Solution

b) Perform regression analysis that includes the macro level variable

Task

Repeat the regression from before and add the aggregate unemployment information.

Solution:

* open the working_pop data set.

DATASET ACTIVATE working_pop .

*note that all countries are included.

freq country.

* add the values from the table on the country level.

MATCH FILES /FILE=*

 /TABLE='unemploy_mean'

 /BY COUNTRY.

EXECUTE.

*check.

means unemploy_mean by country

 /CELLS MEAN   COUNT STDDEV

 /STATISTICS ANOVA.

* include macro_level variable into regression.

regression VARIABLES = STARTIME sex_male age temp_lm ft unemploy_mean

 /DEPENDENT STARTIME

 /METHOD enter sex_male age temp_lm ft unemploy_mean .

End of Solution

Ideas for thinking: Adding additional macro level variables; which sources to use? How to include them into data set?

A typology of countries might be sought. How to build one?

Literature:

- EUROSTAT (2013): Quality report of the European Union Labour Force Survey 2012 - 2014 edition
  http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-TC-14-001

- EUROSTAT (2013): EU LABOUR FORCE SURVEY EXPLANATORY NOTES
  http://epp.eurostat.ec.europa.eu/portal/page/portal/employment_unemployment_lfs/documents/EU_LFS_explanatory_notes_from_2014_onwards.pdf

- MIMAS, The University of Manchester: Countries and Citizens. Linking International Macro and Micro Data. Unit 4: Study Gide. An Introduction to combining macro and micro data. **https://www.esds.ac.uk/international/elearning/limmd/materials/studyguides/unit4-studyguide.pdf**

- Jelle Visser Database on Institutional Characteristics of Trade Unions, Wage Setting, State Intervention and Social Pacts, ICTWSS. Amsterdam Institute for Advanced Labour Studies (AIAS) University of Amsterdam, Version 4 – April 2013 http://www.uva-aias.net/207