gesis

Leibniz Institute
for the Social Sciences

PIAAC Data Analysis in Stata:
A practical guide

Video 4: repest macro

GESIS – Leibniz Institute for the Social Sciences

Leibniz
Association

In a Nutshell

Goal:

- Practical guide for three existing tools to perform PIAAC data analysis in Stata

Target Group:

- Researchers with some experience in Stata, but little or no experience on analysis of PIAAC data in Stata

Structure:

- General overview of how the three tools work, as well as several exemplary analyses with the PIAAC data

Video 4:
repest macro

Repest macro: Overview

Chapter 7
Analysing PIAAC Data with Stata

François Keslair

Abstract This chapter explains the basics of analysing data from the Programme for the International Assessment of Adult Competencies (PIAAC) with Stata. It describes how to import the PIAAC datasets into Stata, gives an overview of the different categories of variables available in these datasets, and mentions a number of features of some types of variables about which users should be aware. The different types of missing values are explained. Routines frequently used with PIAAC datasets are presented using examples. Furthermore, the chapter is devoted to the use of plausible values variables and to the computation of imputation errors and sampling errors. In particular, it presents *repest*, a Stata ado file written to facilitate the analysis of international skills assessments, such as PIAAC.

Stata is an integrated statistical analysis package designed for research professionals. It is particularly well suited for analysing the Organisation for Economic Co-operation and Development's (OECD) Programme for the International Assessment of Adult Competencies (PIAAC) survey (OECD 2013, 2016b, c). Among existing statistical software packages, Stata stands out as it is designed to operate on one dataset at a time, using a dataset that has been previously loaded in memory. With a one-dataset survey such as PIAAC, it brings a simplicity of use and computation speeds difficult to find elsewhere. Moreover, Stata users can benefit from *repest*, a Stata *ado* file developed at the OECD and designed to facilitate the analysis of international skills assessments.

Stata works as a command-line-driven software. It also includes a graphic user interface. Commands can be run—one command at a time—from a prompt located below the results window. This makes the preliminary exploration of a dataset both simple and interactive, in particular, because another window is dedicated to displaying the list of all variables. Commands can be regrouped and saved in

F. Keslair (✉)
Organisation for Economic Co-operation and Development, OECD, Paris, France
e-mail: francois.keslair@oecd.org

D. B. Maehler, B. Rammstedt (eds.), *Large-Scale Cognitive Assessment*,
Methodology of Educational Measurement and Assessment,
https://doi.org/10.1007/978-3-030-47515-4_7

Authors:

Francesco Avvisati and François Keslair

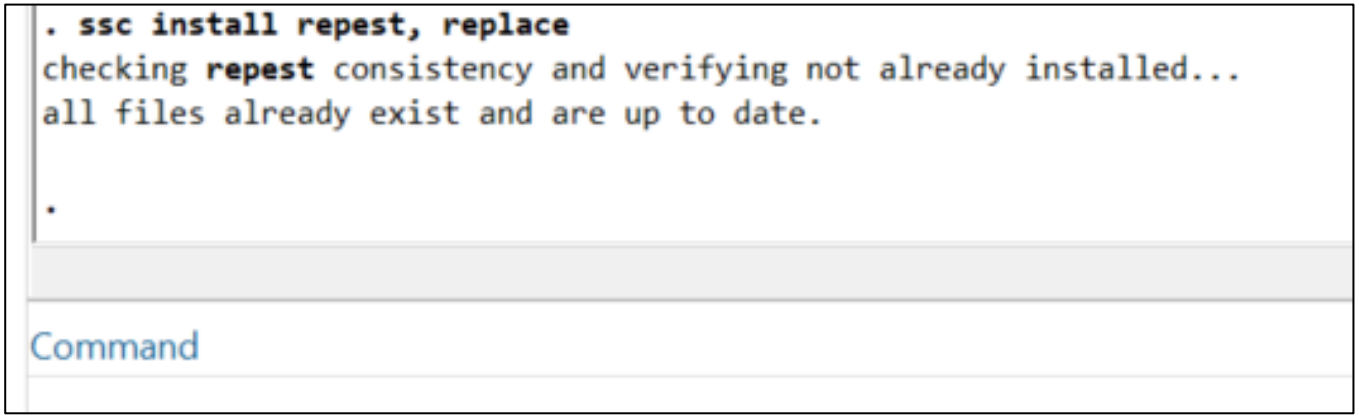
References:

- Avvisati, F. & Keslair, F. 2014. *REPEST: Stata module to run estimations with weighted replicate samples and plausible values*, Boston College Department of Economics, revised 06 Jan 2020.
- Keslair, F. (2020). Analysing PIAAC Data with Stata. In B. Rammstedt, D. Maehler (Eds.) *Large-Scale Cognitive Assessment*. Springer, Cham. p. 149-164.

repest macro: Installation

- Install package:

```
ssc install repest, replace
```



- Data preparation: all variables names should be in lower cases:

```
rename * , lower
```

repest macro: General syntax I

```
repest PIAAC [if] [in], estimate(command
[,command_options]) [repest_options]
```

- `PIAAC`: keyword to "activate" parameters associated with the PIAAC survey design
  → Weights, variance estimation, and plausible values are taken into account
- `if/in`: enable sample restrictions
- `command`: (built-in) Stata command
  → `means, freq, summarize, corr`
- `command_options`: specifies command specific options
- `repest_options`: specifies all further options (see slides 8+9)

repest macro: General syntax II

```
repest PIAAC [if] [in], estimate(stata: command
[,command_options]) [repest_options]
```

- `PIAAC`: keyword to "activate" parameters associated with the PIAAC survey design
  → Weights, variance estimation, and plausible values are taken into account
- `if/in`: enable sample restrictions
- `command`: e-class command (typically regression commands)
  → `reg, logit, qreg`
- `command_options`: specifies command specific options
- `repest_options`: specifies all further options (see slides 8+9)

repest options I

```
repest PIAAC [if] [in], estimate(stata: command
[,command_options]) [repest_options]
```

1. `by(variable)` : separate estimates over the categories of a specified variables
   → possible to estimate the average over certain categories
2. `over(variable)` : jointly estimates over the categories of a specified variables
   → possible to test for differences between categories
3. `results`: keep, add, and combine estimation results
   → keep or add statistics calculated during estimation; create new results with the combine function

repest options II

```
repest PIAAC [if] [in], estimate(stata: command
[,command_options]) [repest_options]
```

4. `outfile`: creates a Stata dataset containing estimation results
   - → outfile contains one observation per country
   - → results are not displayed in the Stata output window
5. `display`: displays results in Stata output window
   - → only necessary when “outfile” is specified

Examples descriptive statistics

I. Average years of education and literacy skills; overall and separately for men and women

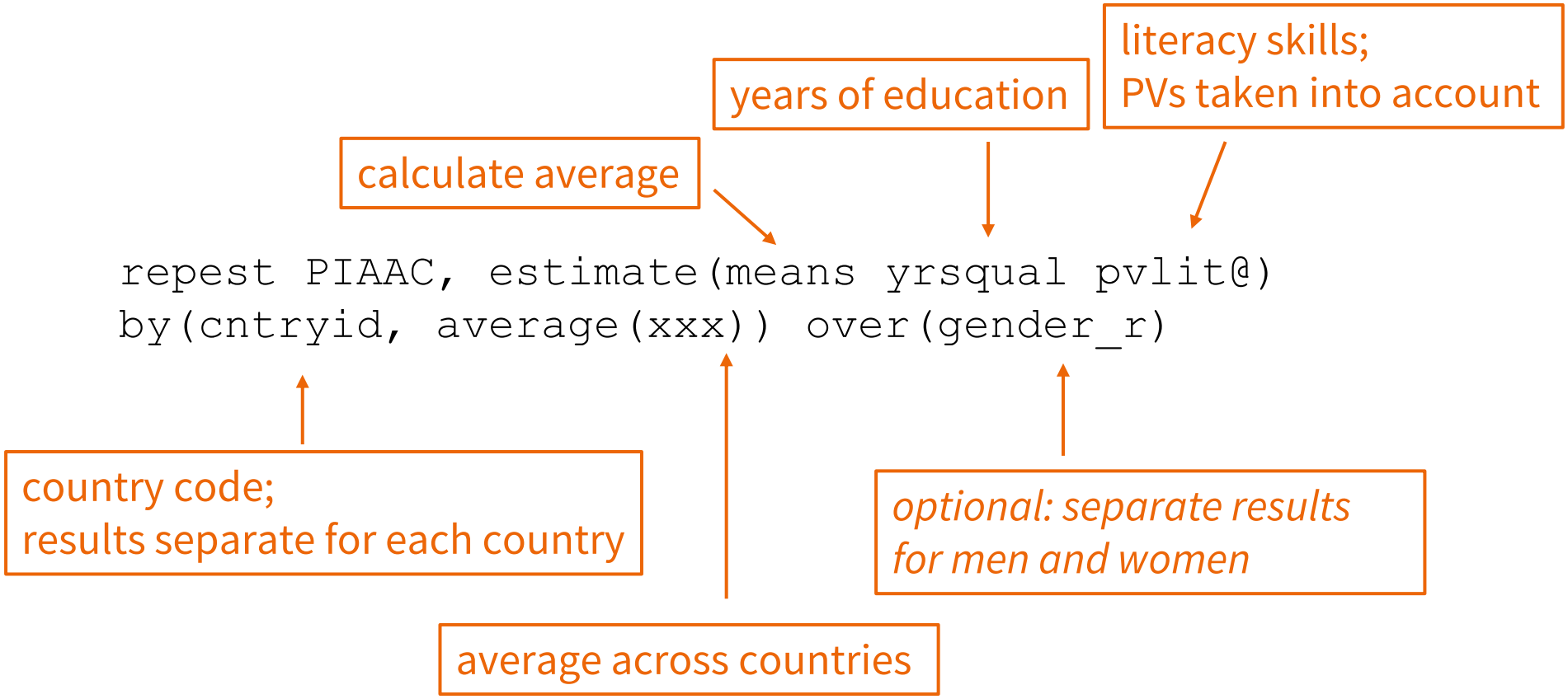
II. Dispersion of literacy skills (5^{th}, 25^{th}, 75^{th}, 95^{th} quantile); overall and for men between 16 and 34 years

III. Percentages of respondent at each numeracy level; overall and for the employed population

IV. Crosstable of numeracy skills (levels) and native language

V. Correlations between literacy, numeracy, and problem-solving skills

Example I
Average years of education and literacy



Example II
Dispersion of literacy skills

Optional: results for men between 16 and 34 years

```
repest PIAAC if gender_r == 1 & ageg10lfs == 1,
estimate(summarize pvlit@, stats(p5 p25 p75 p90))
by(cntryid, levels(xxx))
```

country code; results separate for each country

literacy skills; PVs taken into account

5^{th}, 25^{th}, 75^{th}, 95^{th} percentiles

results only reported for certain countries

gesis Leibniz Institute for the Social Sciences

Example III
Numeracy levels

optional: results for employed respondents

calculates frequencies

```
repest PIAAC if c_d05 == 1, estimate(freq numlevel@)
by(cntryid) outfile("repest_Ex3")
```

numeracy levels; PVs taken into account

country code;
results separate for each country

output file; repest_Ex3.dta

Example IV a)
Numeracy and native language

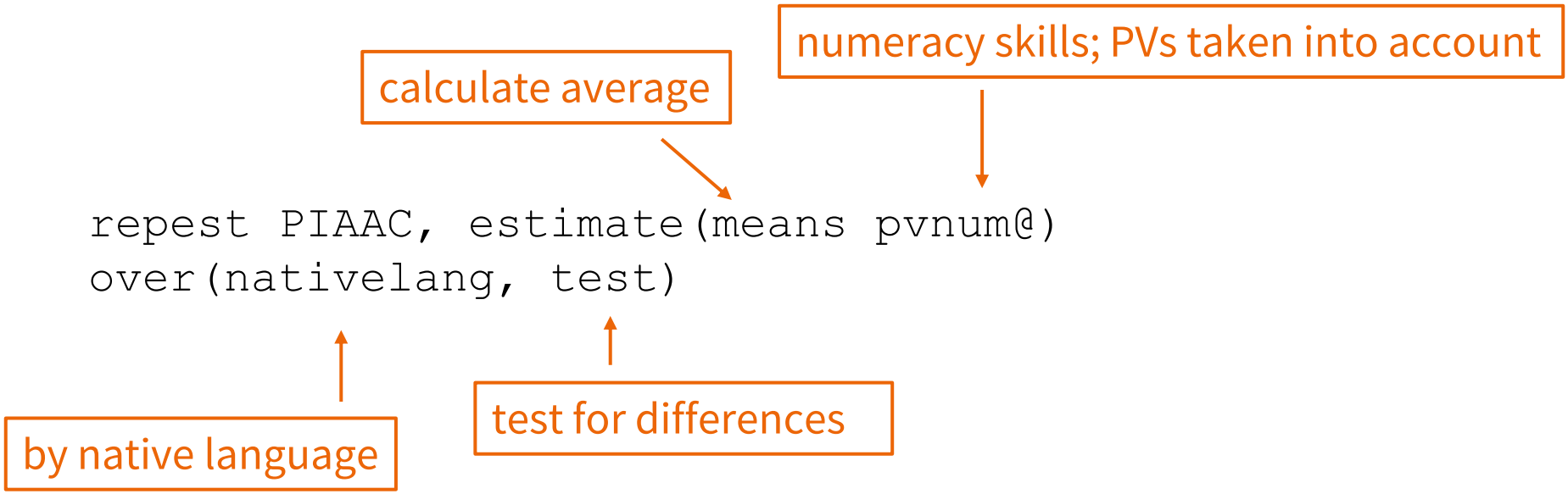
numeracy levels; PVs taken into account

calculate frequencies

```
repest PIAAC, estimate(freq numlevel@)
over(nativelang)
```

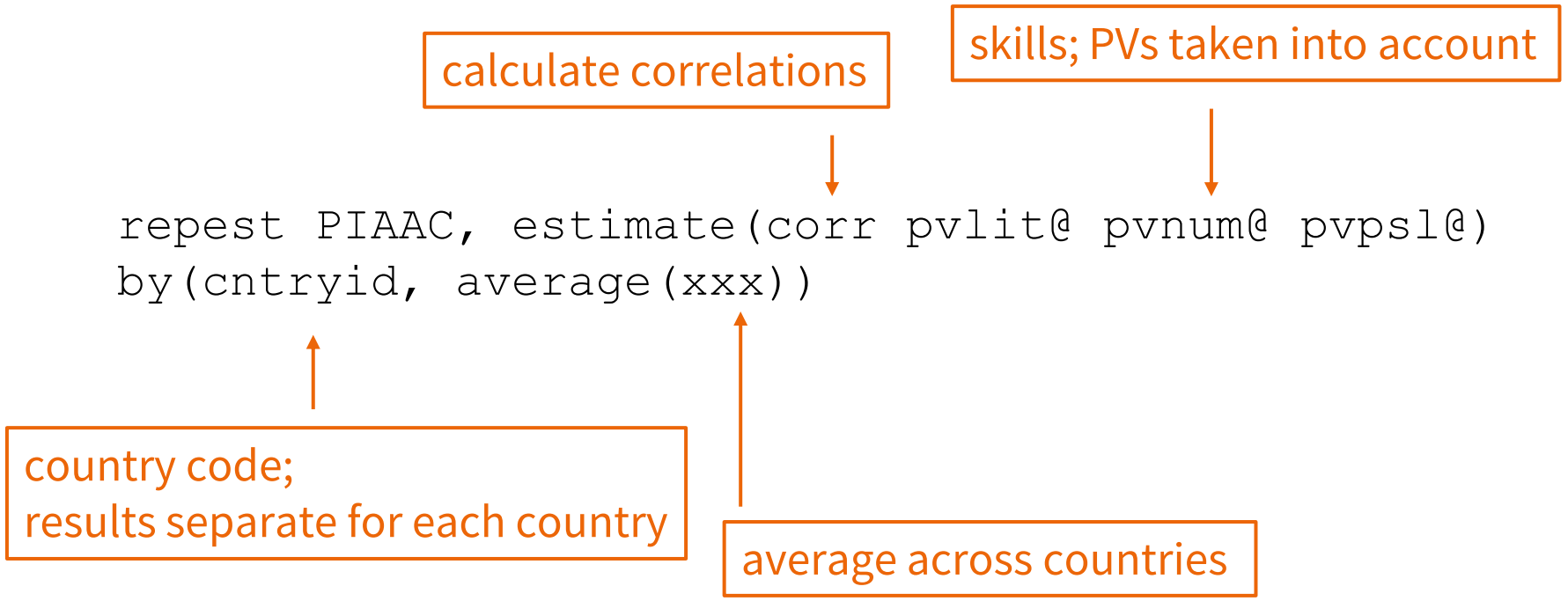
by native language

Example IV b)
Numeracy and native language

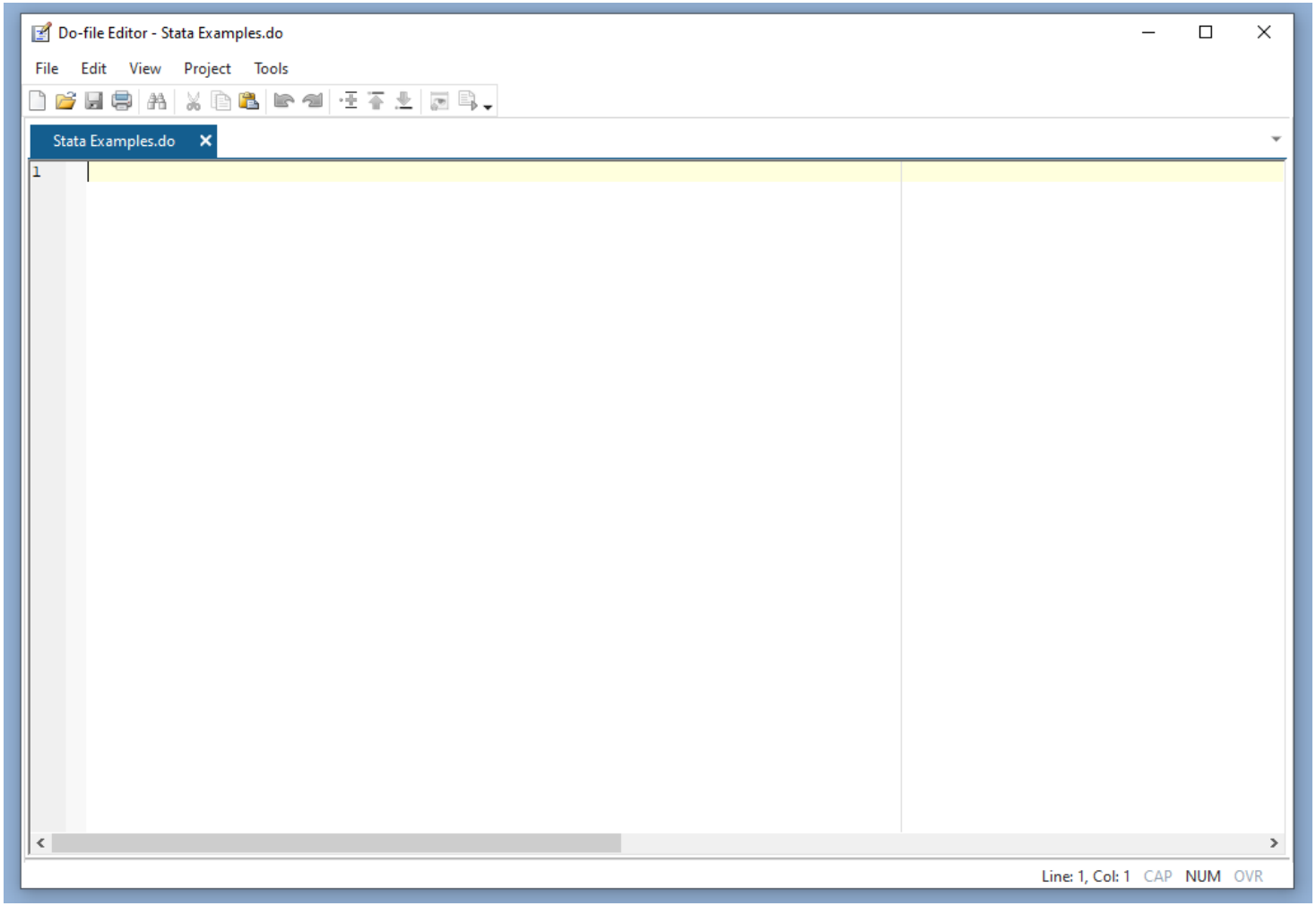


gesis Leibniz Institute for the Social Sciences

Example V
Correlations between skills



Let‘s go to Stata

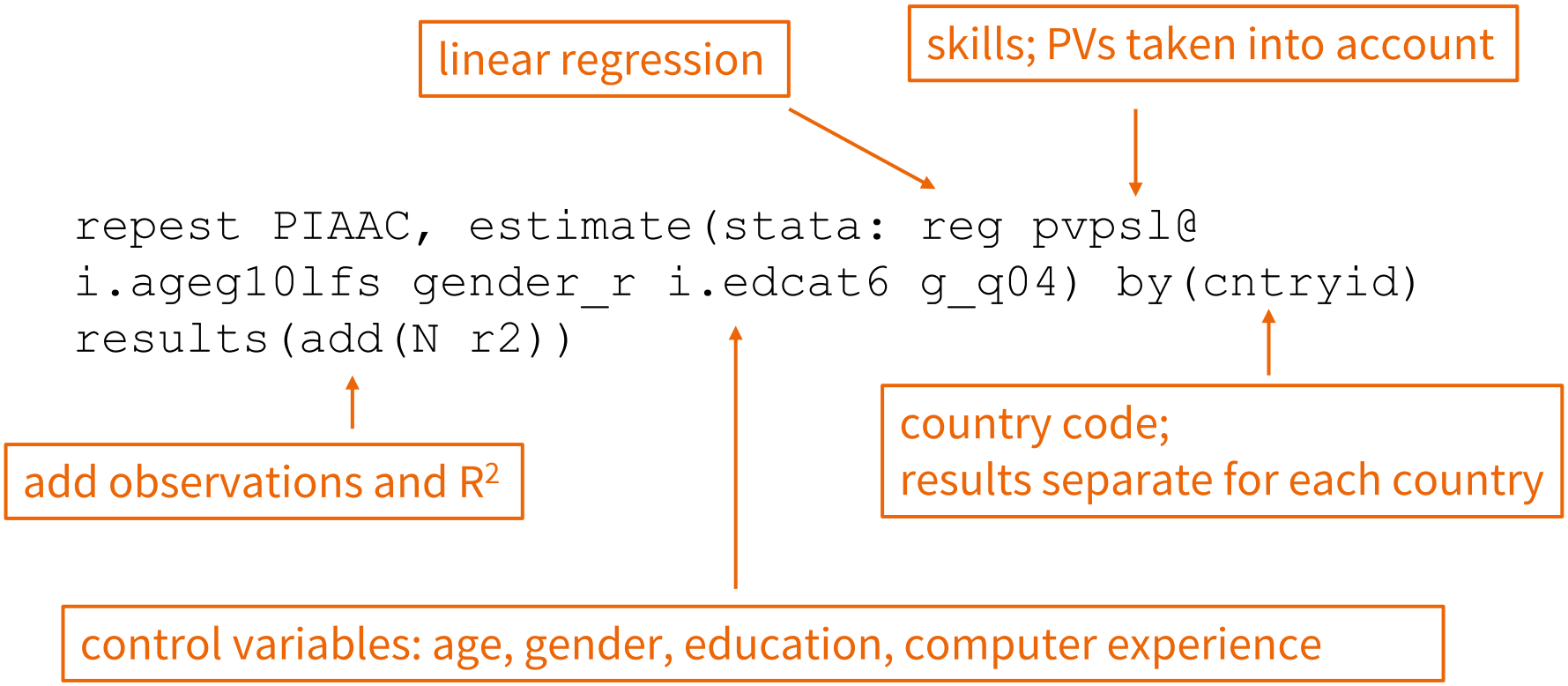


Examples regression analyses

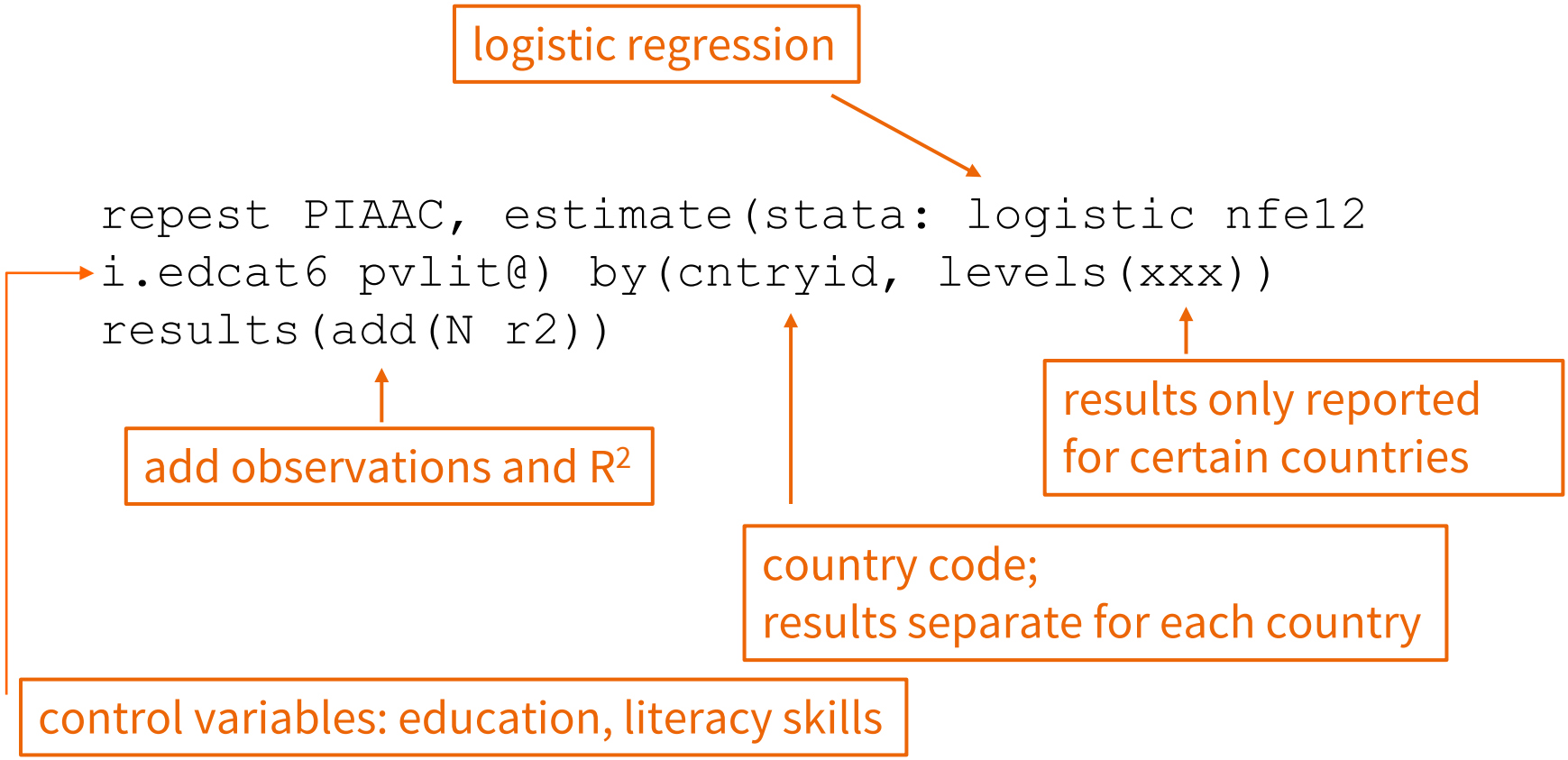
I. Linear regression: Are age, gender, formal education, and computer experience in the workplace related to PS-TRE skills?

II. Logistic regression: Do literacy skills and formal education determine participation in adult education for women between 35 and 54 years?

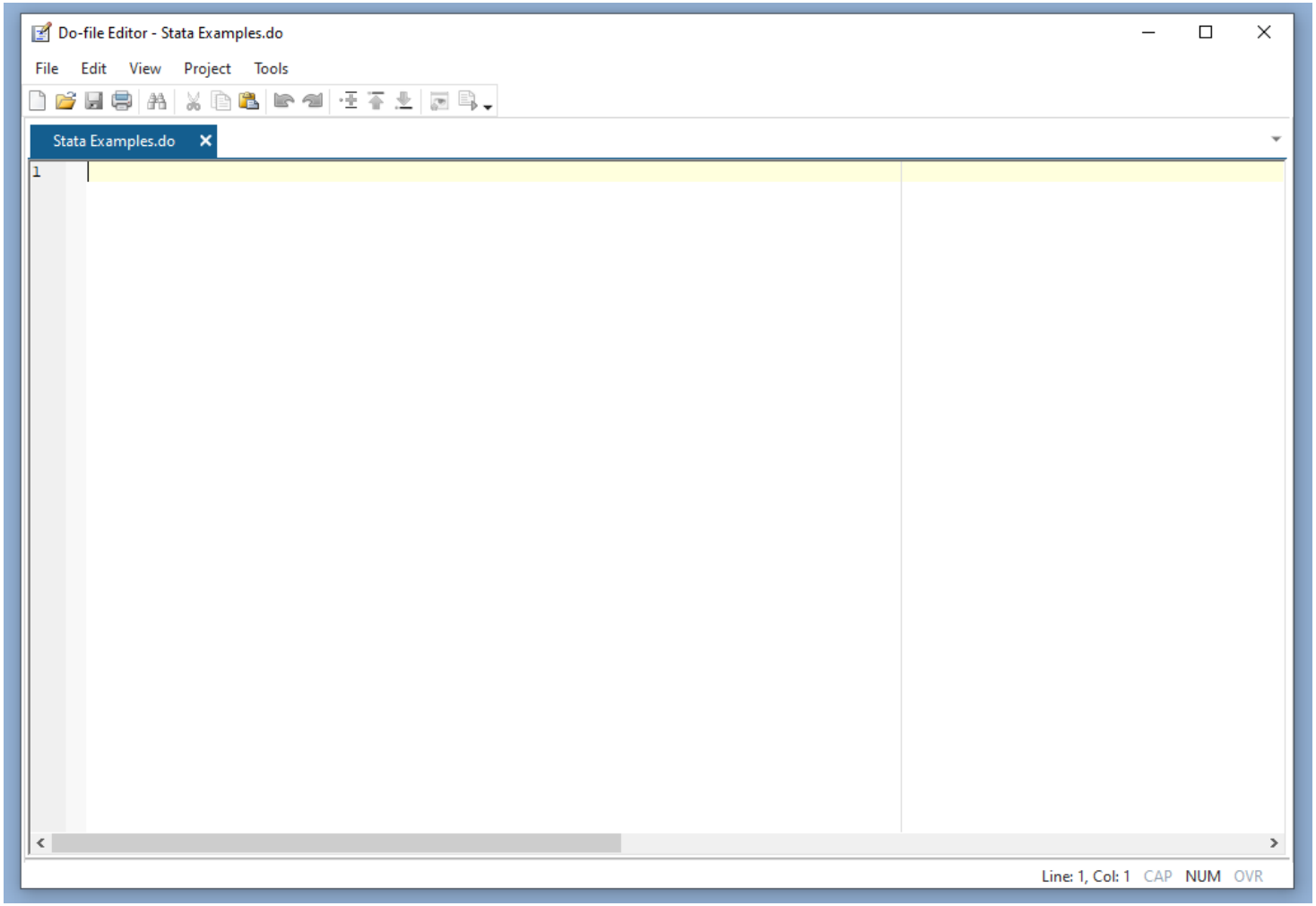
Example I
Linear Regression of PS-TRE skills



Example II
Logistic regression of training



Let‘s go to Stata



gesis

Leibniz Institute
for the Social Sciences

Leibniz
Association