

# gesis

Leibniz Institute  
for the Social Sciences



## PIAAC Data Analysis in Stata: A practical guide

*Video 3: piaactools*

*GESIS – Leibniz Institute for the Social Sciences*

*Video 3:  
piaactools*

# piaactools: Overview

The Stata Journal (yyyy) vt, Number ii, pp. 1–16

## PIAACTOOLS: Stata programs for data analysis with PIAAC data

Maciej Jakubowski  
Faculty of Economic Sciences,  
University of Warsaw, Poland  
Evidence Institute, Poland  
mjakubowski@uw.edu.pl

Artur Pokropek  
Institute of Philosophy and Sociology,  
Polish Academy of Science,  
Institute of Educational Research (IBE),  
Warsaw, Poland  
artur.pokropek@gmail.com

**Abstract.** The OECD Programme for the International Assessment of Adult Competencies (PIAAC) is currently the only international survey of adult skills. It provides rich data on skills, work and life situation, earnings and attitudes. To ensure representativeness and high reliability, the study is based on a complex survey design and advanced statistical methods. To obtain correct results from publicly available microdata one needs to use special methods, which are often too advanced for less experienced researchers. This paper presents PIAACTOOLS – a package of three Stata commands that facilitate analysis with PIAAC data. The command `-piaacdes-` allows for the calculation of basic statistics, while `-piaactab-` computes frequencies of adults at each proficiency level. The command `-piaacreg-` allows for the use of several regression models with PIAAC data. Output is saved as HTML files that can be opened in most spreadsheets and as Stata matrices that can be further processed in Stata. The paper also explains how to use these commands and provides examples that can be easily modified for usage with different models and variables.

**Keywords:** st0001, piaactools, Stata, PIAAC, OECD, regression, plausible values, replicate weights, adult survey, skills, proficiency, education

### 1 Introduction

The OECD *Programme for the International Assessment of Adult Competencies* (PIAAC) is a household survey that aims at measuring literacy, numeracy and problem solving skills. The assessment was launched in 2008, with first round finalized in 2013 covering 24 countries and economies. Round 2 was finalized in 2016, bringing in 9 new countries for the dataset, and round 3 is scheduled to be finished in 2019. The study uses a complex survey design and advanced statistical methods to improve the representativeness and reliability of the results. As the only up-to-date source of internationally comparable information about adult skills, PIAAC is highly popular with researchers across the world.

However, in order to analyse microdata from the PIAAC study one needs to use relatively complex methods. In Stata, it is not possible to obtain correct results from the PIAAC data without some programming, which is a barrier difficult to overcome for some researchers. In many cases, researchers use wrong approaches and have later

Authors:

Maciej Jakubowski & Artur Pokropek

Reference:

Jakubowski, M., & Pokropek, A. (2019). PIAACTOOLS: A program for data analysis with PIAAC data. *The Stata Journal*, 19(1), 112-128.

# piaactools: Overview

- Three main commands:
  - ▶ `piaacdes`  
descriptive statistics (mean, variance, percentiles)
  - ▶ `piaactab`  
tables and crosstables
  - ▶ `piaacreg`  
linear and logistic regressions

## piaactools: Installation

- Install package:

```
ssc install piaactools, replace
```

```
. ssc install piaactools, replace  
checking piaactools consistency and verifying not already installed...  
installing into c:\ado\plus\...  
installation complete.
```

```
Command
```

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

```
rename * , lower
```

## piaacdes: General Syntax

```
piaacdes <variable1>, pv(<skills>)  
countryid(<countrycode>) stats(<statistic>)  
centile(<percentile>) round(<decimalplace>)  
save(<outputfile>)
```

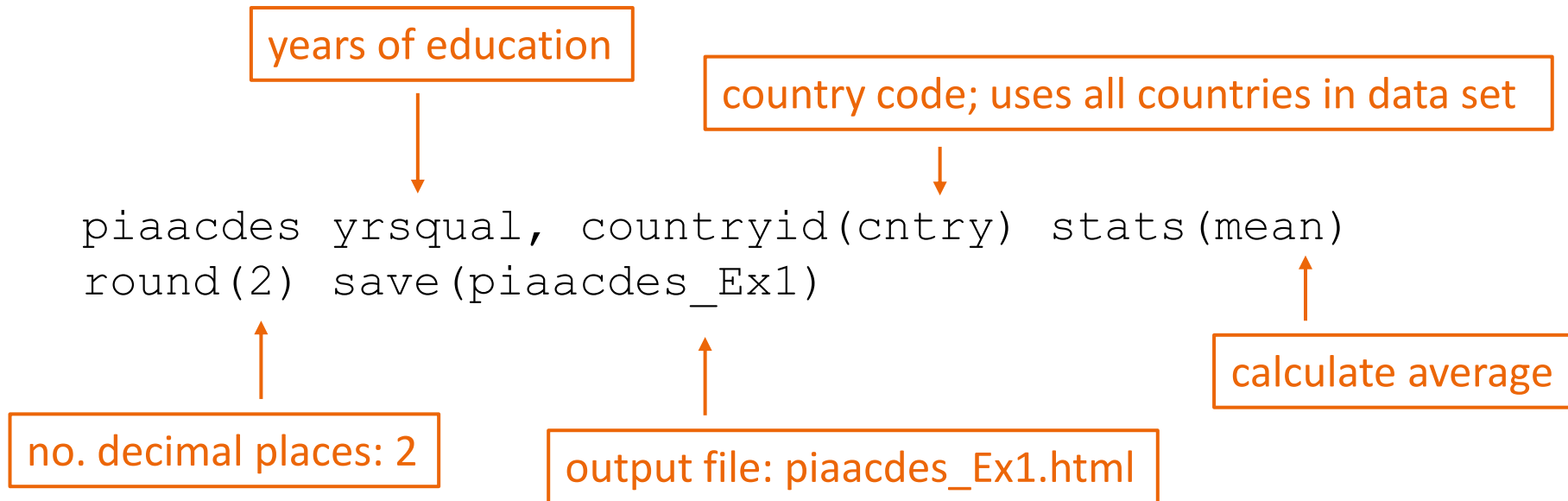
- variable1: any variable
- skills: literacy, numeracy, or PS-TRE; plausible values will automatically be taken into account
- countrycode: specifies countries for which results are obtained; default: “cntry” variable
- statistic: mean or standard deviation (sd)
- percentile: percentiles between 1 and 100
- decimalplace: specifies number of decimal places in results tables
- outputfile: specifies name of output files used

## piaacdes: Examples

- I. Average years of education
- II. Average literacy skills;  
overall and separately for men and women
- III. Dispersion of literacy skills (5<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup>, 95<sup>th</sup> quantile);  
overall and for men between 16 and 34 years

# piaacdes: Example I

## Average years of education





# piaacdes: Example II

## Average literacy skills

literacy skills; PVs taken into account

country code; uses all countries in data set

```
piaacdes, pv(pvlit) countryid(cntry) stats(mean)  
over(gender_r) round(2) save(piaacdes_Ex2)
```

calculate average

no. decimal places: 2

output file: piaacdes\_Ex2.html

*Optional: separate results for men and women*

# piaacdes: Example III

## Dispersion of literacy skills

literacy skills; PVs taken into account

*Optional: results for men between 16 and 34 years*

```
piaacdes if gender_r == 1 & ageg10lfs < 3, pv(pvlit)  
countryid(cntry) centile(5 25 75 95) round(2)  
save(piaacdes_Ex3)
```

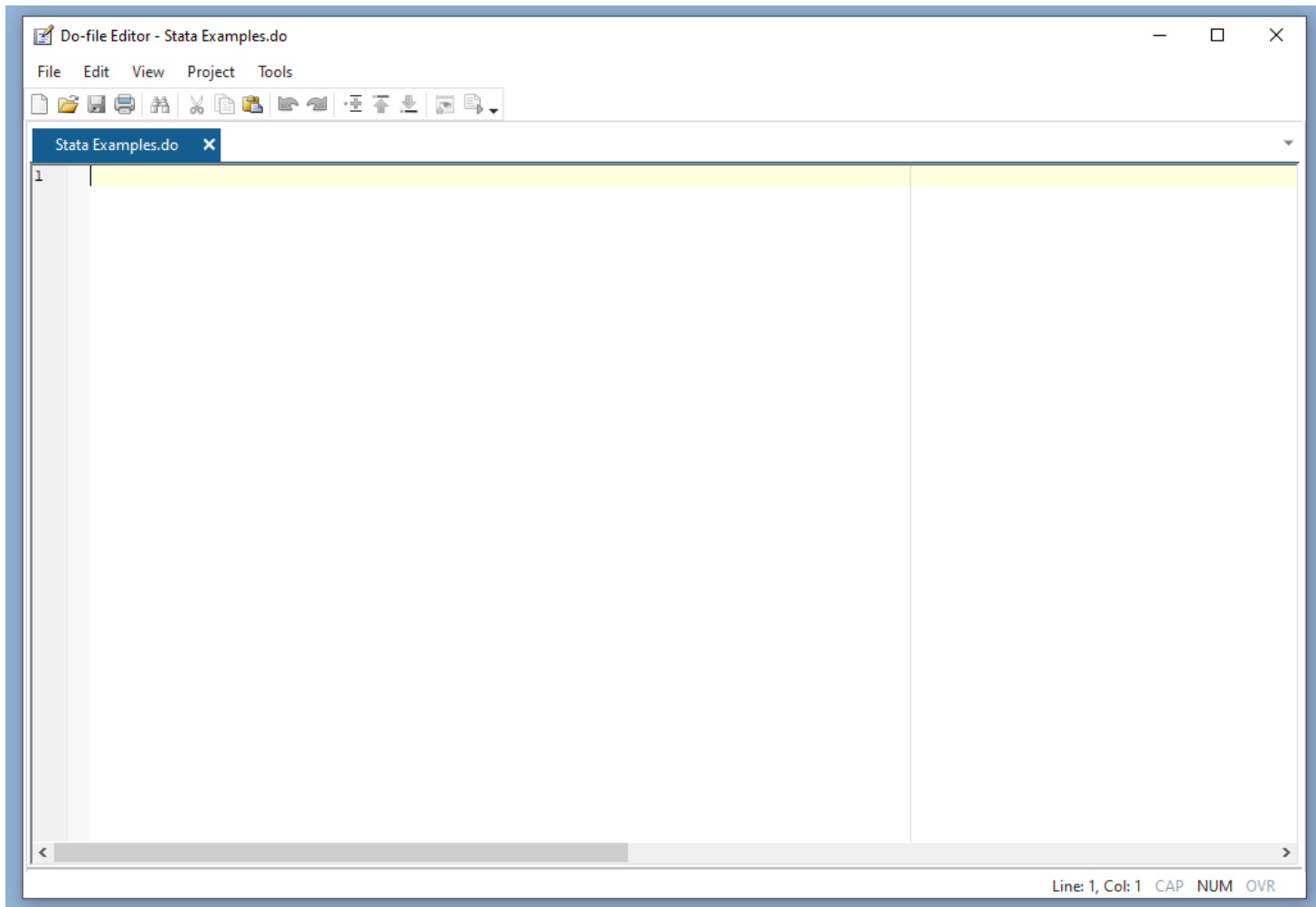
5<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup>, 95<sup>th</sup> percentiles

country code; uses all countries in data set

no. decimal places: 2

output file: piaacdes\_Ex3.html

# Let's go to Stata



## piaactab: General Syntax

```
piaactab <variable1>, over(<variable2>)  
countryid(<countrycode>)  
round(<decimalplace>) save(<outputfile>)
```

- variable1: any categorical variable; if skills are included, they will be considered as proficiency levels; plausible values will automatically be taken into account
- variable2: any categorical variable; allows tabulation over subgroups, skills included as stated above
- countrycode: specifies countries for which results are obtained; default: “cntry” variable
- decimalplace: specifies number of decimal places in results tables
- outputfile: specifies name of output files used

## piaactab: Examples

- I. Percentages of educational qualifications of respondents' mother
- II. Percentages of respondent at each numeracy level; overall and for the employed population
- III. Crosstable of numeracy skills (levels) and native language

# piaactab: Example I

## Percentages mothers' education

Mothers' education

country code; uses all countries in data set

```
piaactab j_q06b, countryid(cntry) round(1)  
save(piaactab_Ex1)
```

no. decimal places: 1

output file; piaactab\_Ex1.html

## piaactab: Example II Numeracy levels

numeracy levels

*optional: results for employed respondents*

```
piaactab pvnum if c_d05 == 1, countryid(cntry)  
round(1) save(piaactab_Ex2)
```

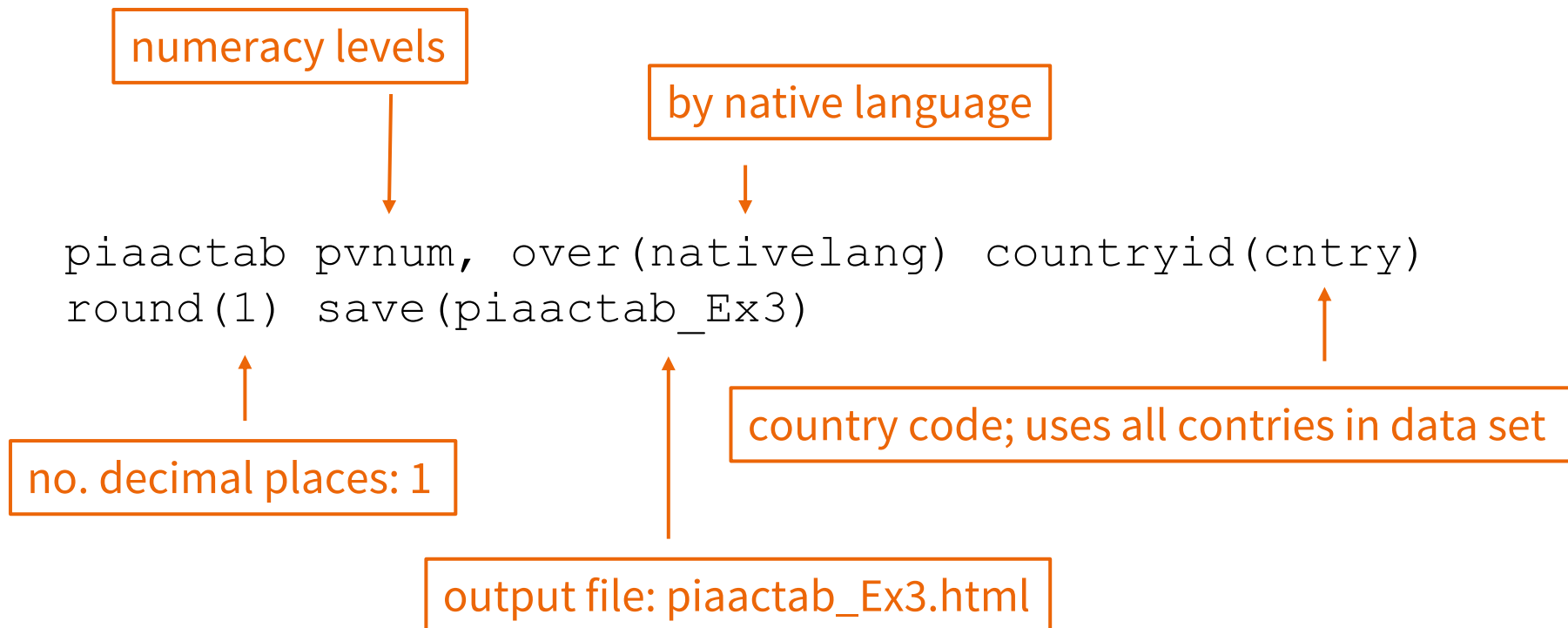
output file; piaactab\_Ex2.html

no. decimal places: 1

country code; uses all countries in data set

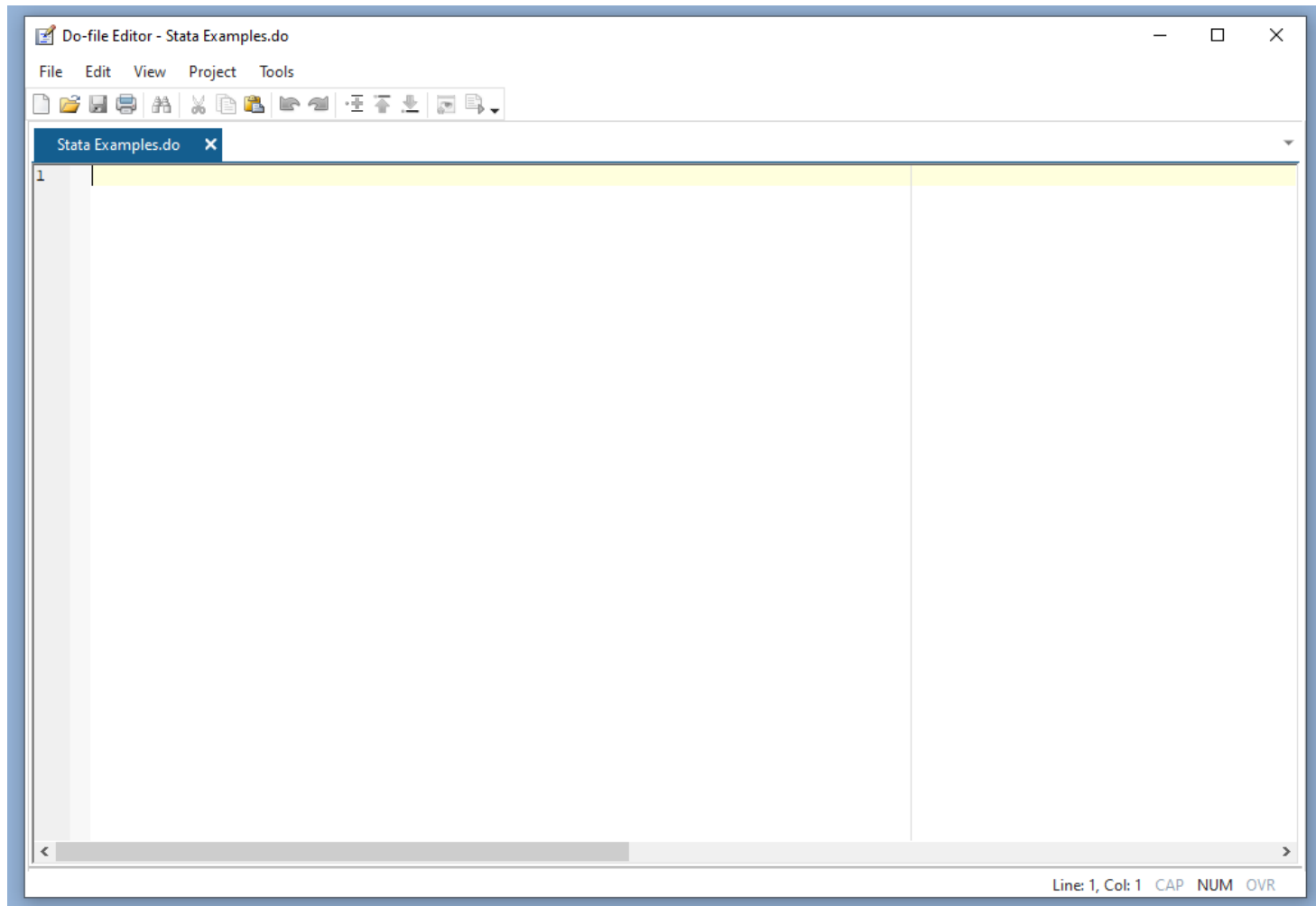
# piaactab: Example III

## Numeracy and native language





# Let's go to Stata



## piaacreg: General Syntax

```
piaacreg <depvar> <indepvar>, pvdep(<skills>)  
pvindep1(<skills>) cons countryid(<countrycode>)  
round(<decimalplace>) save(<outputfile>)
```

- depvar: dependent variable
- indepvar: (list of) independent / control variable(s)
- skills: can be included as dependent or independent variable(s)
- cons: if specified, constant is reported in regression tables
- countrycode: specifies countries for which results are obtained;  
default: “cntry” variable
- decimalplace: specifies number of decimal places in results tables
- outputfile: specifies name of output files used

## piaacreg: Examples

- 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?

# piaactab: Example I

## Linear Regression of PS-TRE skills

control variables: age, gender, education, computer experience

```
piaacreg ageg10lfs gender_r edcat6 g_q04,  
pvdep(pvpsl) countryid(cntry) round(2)  
save(piaacreg_Ex1)
```

output file: piaacreg\_Ex1.html

no. decimal places: 2

skills as dependent variable

country code; uses all countries in data set

# piaactab: Example II

## Logistic regression of training

dependent variable: training participation

control variables: education, literacy

```
piaacreg nfe12 edcat6, pvindep1(pvlit)  
cmd("logit") countryid(cntry) round(2)  
save(piaacreg_Ex2)
```

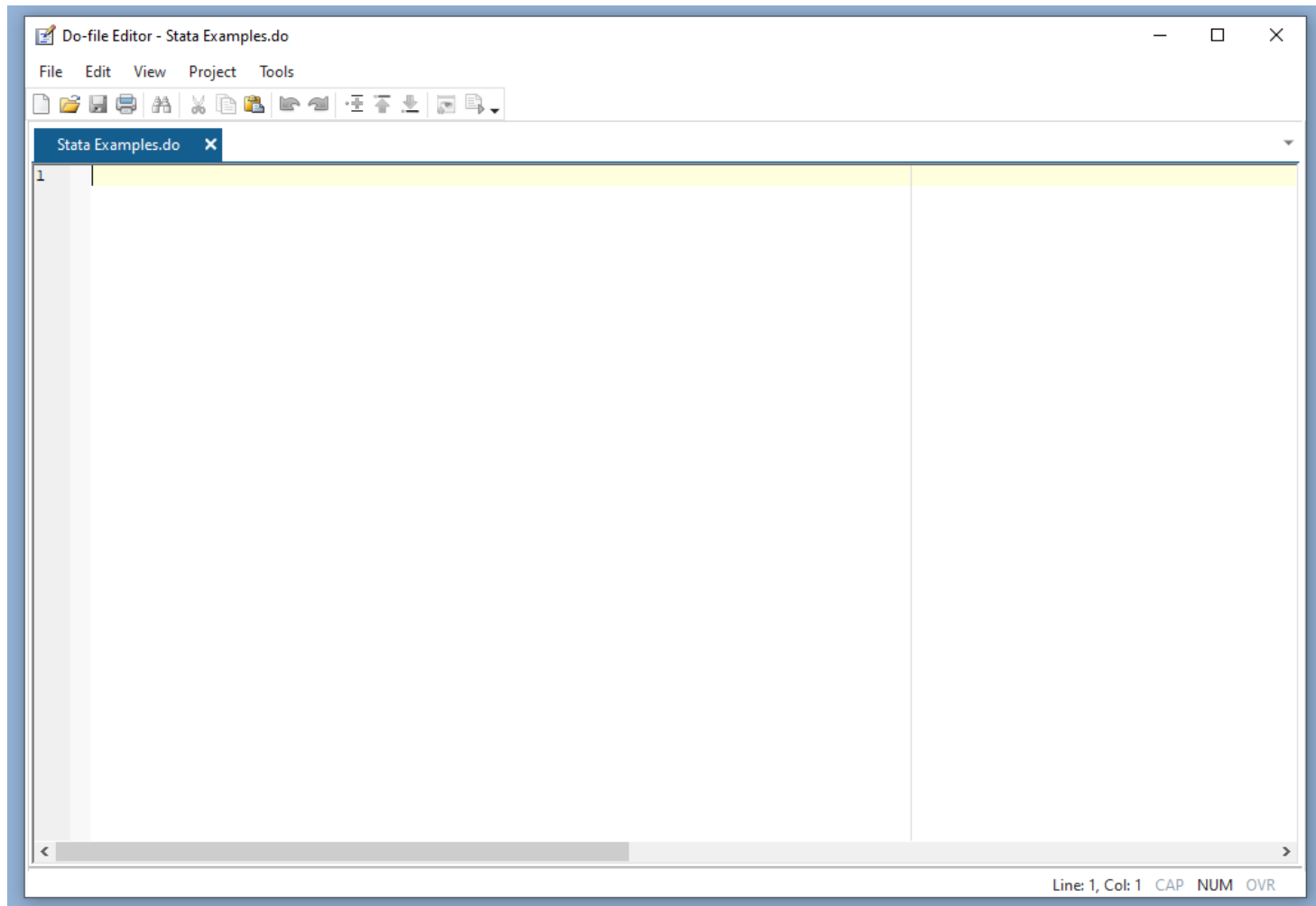
logistic regression

no. decimal places: 2

country code; uses all countries in data set

output file: piaacreg\_Ex2.html

# Let's go to Stata



gesis

Leibniz Institute  
for the Social Sciences

Leibniz  
Leibniz  
Association