PIAAC Data Analysis in Stata: A practical guide

Video 3: piaactools
GESIS – Leibniz Institute for the Social Sciences
Video 3: piaactools
piaactools: Overview

Authors:
Maciej Jakubowski & Artur Pokropek

Reference:
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
  ```

- 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 (5th, 25th, 75th, 95th quantile); overall and for men between 16 and 34 years
piaacdes: Example I
Average years of education

```
piaacdes yrsqual, countryid(cntry) stats(mean) round(2) save(piaacdes_Ex1)
```

country code; uses all countries in data set

Years of education

No. decimal places: 2

Output file: piaacdes_Ex1.html

Calculate average
piaacdes: Example II
Average literacy skills

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

- **literacy skills; PVs taken into account**
- **country code; uses all countries in data set**
- **no. decimal places: 2**
- **Optional: separate results for men and women**
- **Calculate average**
- **Output file: piaacdes_Ex2.html**
piaacdes: Example III
Dispersion of literacy skills

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)

5th, 25th, 75th, 95th percentiles

country code; uses all countries in data set

literacy skills; PVs taken into account

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)

output file; piaactab_Ex1.html

no. decimal places: 1
piaactab: Example II
Numeracy levels

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

output file: piaactab_Ex2.html

- numeracy levels
- country code: uses all countries in data set
- no. decimal places: 1
- optional: results for employed respondents
piaactab: Example III
Numeracy and native language

numeracy levels

by native language

piaactab pvnum, over(nativelang) countryid(cntry) round(1) save(piaactab_Ex3)

no. decimal places: 1

country code; uses all countries in data set

output file: piaactab_Ex3.html
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

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

- **Output file:** piaacreg_Ex1.html
- **Country code:** uses all countries in data set
- **Decimal places:** 2
**piaactab: Example II**

**Logistic regression of training participation**

- **Dependent variable:** training participation
- **Control variables:** education, literacy

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

- **Logistic regression**
- **Output file:** piaacreg_Ex2.html
- **Country code:** uses all countries in data set
- **No. decimal places:** 2
Let’s go to Stata