R package `intsvy`

Demonstration

Dr. Daniel Caro

The use of test scores in secondary analysis
OECD Paris, 14 June 2019

What is `intsvy`?

- R package that provides tools for the analysis of ILSA data
  - Merge and import data
  - Analyze data (eg. means, percentages, regression)
  - Visualize results
- ILSA studies
  - PISA, TIMSS, PIRLS, PIAAC, ICILS, and extendable

Online resources for `intsvy`

- Learning resources are available at http://danielcaro.net/r-intsvy/
  - Video tutorials
  - pdf tutorial (article published in *Journal of Statistical Software*)
  - Data analysis examples
  - R-bloggers articles

Video tutorials

Video tutorials

Video 1: Installing R ‘intsvy’

Video 2: Directory paths and operating systems

Video 3: Printing data labels

Video 4: Importing data

Video 5: Calculating average student performance

Video 6: Frequency tables

Video 7: Proficiency levels

Video 8: Regression analysis

JSS article

Journal of Statistical Software

November 2017, Volume 81, Issue 7. *doi: 10.18637/jss.v081.i07*

intsvy: An R Package for Analyzing International Large-Scale Assessment Data

Daniel H. Caro
University of Oxford

Przemysław Biecek
Warsaw University of Technology

Abstract

This paper introduces **intsvy**, an R package for working with international assessment data (e.g., PISA, TIMSS, PIRLS). The package includes functions for importing data, performing data analysis, and visualizing results. The paper describes the underlying methodology and provides real data examples. Tools for importing data allow useRs to select variables from student, home, school, and teacher survey instruments as well as for specific countries. Data analysis functions take into account the complex sample design (with replicate weights) and rotated test forms (with plausible values of achievement scores) in the calculation of point estimates and standard errors of means, standard deviations, regression coefficients, correlation coefficients, and frequency tables. Visualization tools present data aggregates in standardized graphical form.

Keywords: international assessments, complex survey analysis, replicate weights, plausible values.

PISA 2015 data analysis examples

Examples with PISA 2015

The data from PISA 2015 can be analysed with the development version of intsvy. The development version of the intsvy package can be installed and loaded with:

```
library("devtools")
install_github("eldafani/intsvy")
```

```
library('intsvy')
```

Importing the data

You can create an object with the data directory and then apply the `pisa.select.merge` function.

```
dir <- "/home/eldani/MEGA/Work/international LSA/PISA/2015/Data" # your data directory (eg C:/PISA 2015/Data)

pisa15 <- pisa.select.merge(folder = dir,
                            student.file="CY6_MS_CMB_STU_QQQ.sav",
                            school.file="CY6_MS_CMB_SCH_QQQ.sav",
                            student= c("ESCS", "ST004D01T", "TEACHSUP", "JOYSCIE",
                                       "IBTEACH", "TDTEACH", "DISCLISCI"),
                            school = c("EDUSHORT", "SCHLTYPE"))
```

The object pisa15 is a data frame with selected variables for all education systems participating in PISA 2015. Now you can start to analyse the data. Below are some examples.

Science average performance (PISA 2015 report, Table I.2.3)

```
pisa2015.mean.pv(pvlabel = "SCIE", by = "CNT", data = pisa15)
```

```
##                                  CNT  Freq   Mean s.e.     SD  s.e
## 1                            Albania  5215 427.22 3.28  78.48 1.45
## 2                United Arab Emirates 14167 436.73 2.42  99.14 1.06
## 3                          Australia 14530 509.99 1.54 102.30 0.92
## 4                            Austria  7007 495.04 2.44  97.34 1.31
## 5                            Belgium  9651 502.00 2.29 100.19 1.24
## 6                           Bulgaria  5928 445.77 4.35 101.52 2.10
## 7                             Brazil 23141 400.68 2.30  89.15 1.27
## 8                             Canada 20058 527.70 2.08  92.37 0.88
## 9                        Switzerland  5860 505.51 2.90  99.52 1.55
```

PISA 2015 in R-bloggers

R news and tutorials contributed by (750) R bloggers

PISA 2015 – how to read/process/plot the data with R

How to install `intsvy`?

- Stable version from CRAN

```
install.packages("intsvy")
library('intsvy')
```

- Development version from Github

```
install.packages('devtools')
library('devtools')
install_github("eldafani/intsvy")
library('intsvy')
```

Examples with PISA 2015

Read data

```
dir <-  "C:/PISA15/DATA"


pisa <- pisa.select.merge(folder = dir,
        student.file = "CY6_MS_CMB_STU_QQQ.sav",
        school.file= "CY6_MS_CMB_SCH_QQQ.sav",
        student =c('ST004D01T','REPEAT', 'ESCS', 'BELONG', 'ANXTEST', 'ST001D01T'),
        school = c('PROSTCE', 'PROATCE', 'EDUSHORT', 'STAFFSHORT', 'STUBEHA', 'TEACHBEHA'),
        countries = c("HUN", "POL", "ROU", "RUS", "SVK"))


## re-encoding from CP1252
## re-encoding from CP1252
```

Average with plausible values

Science performance by country

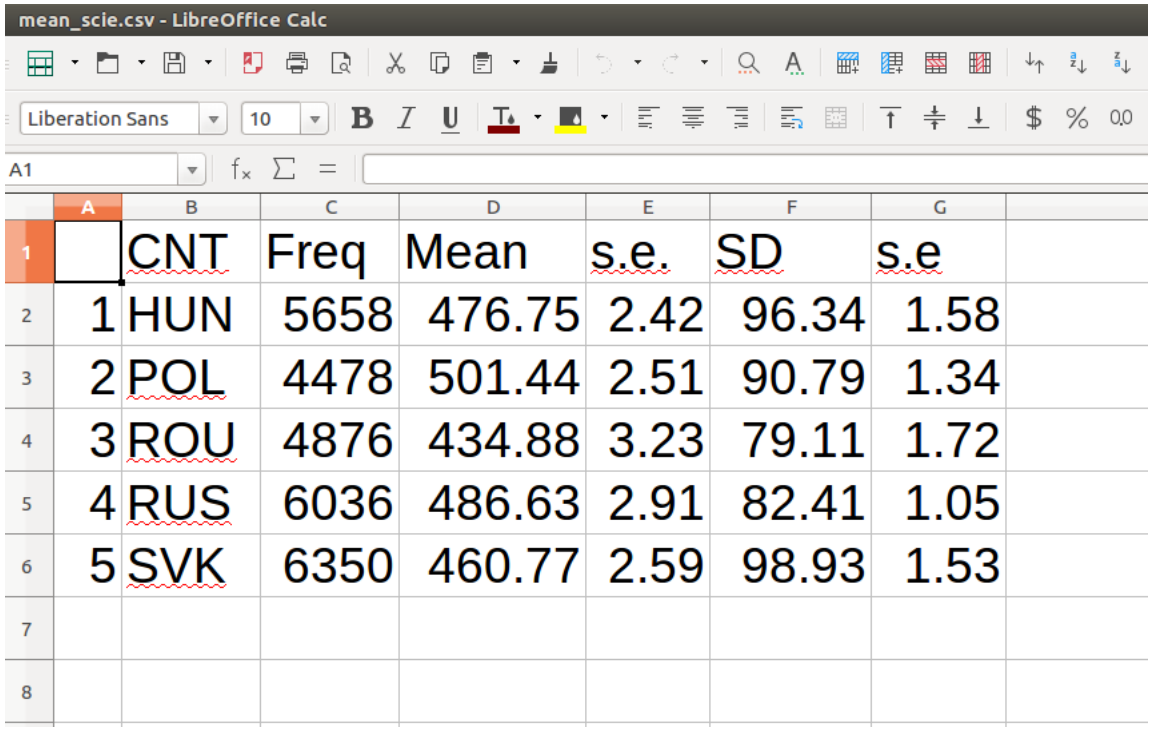
```
pisa2015.mean.pv(pvlabel = "SCIE", by = "CNT", data = pisa)
```

```
##   CNT Freq   Mean s.e.    SD  s.e
## 1 HUN 5658 476.75 2.42 96.34 1.58
## 2 POL 4478 501.44 2.51 90.79 1.34
## 3 ROU 4876 434.88 3.23 79.11 1.72
## 4 RUS 6036 486.63 2.91 82.41 1.05
## 5 SVK 6350 460.77 2.59 98.93 1.53
```

Exporting results to spreadsheet

```
pisa2015.mean.pv(pvlabel = "SCIE", by = "CNT", data = pisa,
                 export=TRUE, folder=dir, name = 'mean_scie')
```

Exporting results to spreadsheet



	A	B	C	D	E	F	G
1		CNT	Freq	Mean	s.e.	SD	s.e
2	1	HUN	5658	476.75	2.42	96.34	1.58
3	2	POL	4478	501.44	2.51	90.79	1.34
4	3	ROU	4876	434.88	3.23	79.11	1.72
5	4	RUS	6036	486.63	2.91	82.41	1.05
6	5	SVK	6350	460.77	2.59	98.93	1.53
7							
8							

Science performance by country and sex

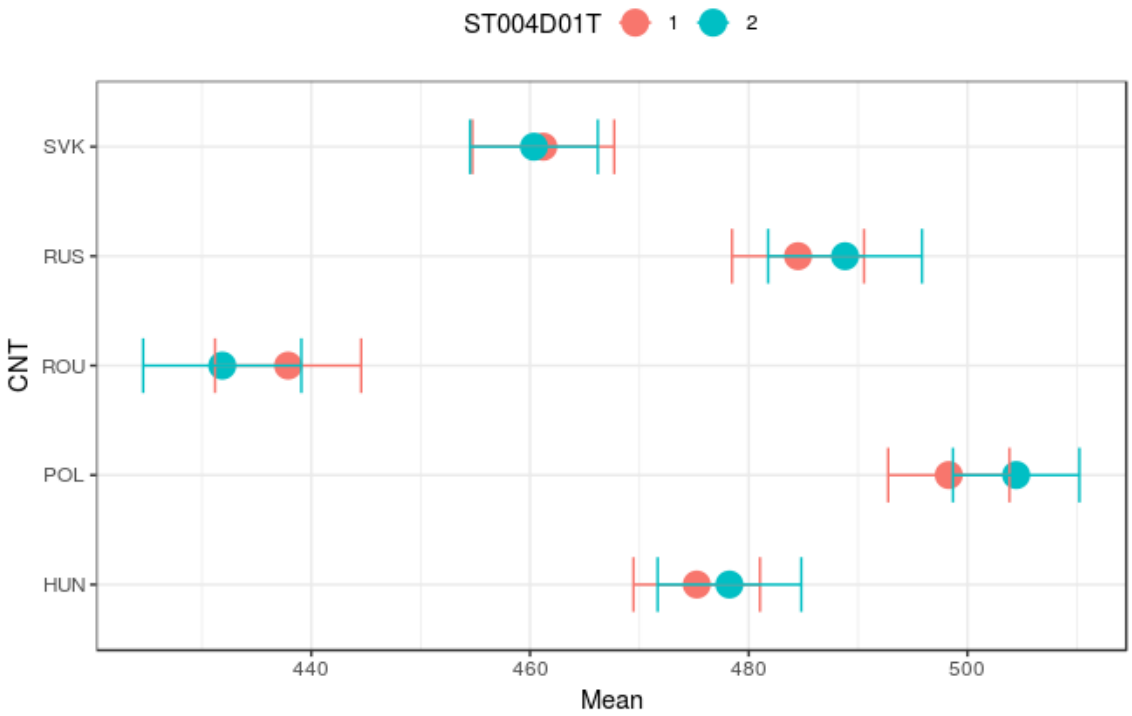
```
pisa2015.mean.pv(pvlabel = "SCIE", by = c("CNT", "ST004D01T"), data = pisa)
```

```
##    CNT ST004D01T Freq   Mean s.e.     SD  s.e
## 1  HUN         1 2850 475.25 2.95  94.46 1.85
## 2  HUN         2 2808 478.24 3.35  98.14 1.83
## 3  POL         1 2209 498.30 2.83  86.87 1.75
## 4  POL         2 2269 504.46 2.95  94.32 1.81
## 5  ROU         1 2466 437.88 3.41  78.24 1.83
## 6  ROU         2 2410 431.86 3.69  79.86 2.07
## 7  RUS         1 3107 484.51 3.08  79.76 1.04
## 8  RUS         2 2929 488.81 3.59  85.00 1.50
## 9  SVK         1 3035 461.22 3.31  96.15 2.04
## 10 SVK         2 3315 460.36 2.98 101.47 1.71
```

Visualizing results

```
# Store results in object
scie_sex <- pisa2015.mean.pv(pvlabel = "SCIE", by = c("CNT", "ST004D01T"), data = pisa)
# Plot object
plot(scie_sex)
```

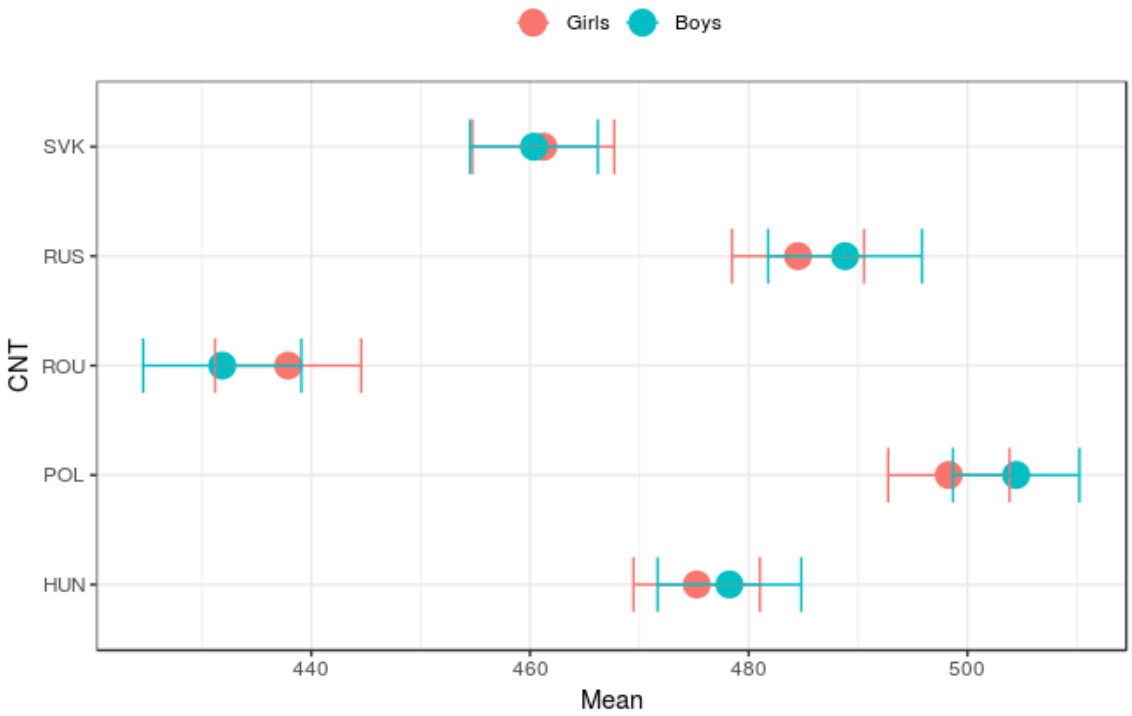
Visualizing results



Visualizing results

```
# Use ggplot2 to customize graph
library('ggplot2')
# Change labels in legends
plot(scie_sex) +
  scale_color_discrete(name = "", labels = c("Girls", "Boys"))
```

Visualizing results



Average without plausible values

Economic, social and cultural status (ESCS)

```
pisa2015.mean(variable = "ESCS", by = "CNT", data = pisa)
```

```
##   CNT Freq  Mean s.e.   SD  s.e
## 1 HUN 5570 -0.23 0.02 0.95 0.01
## 2 POL 4446 -0.39 0.02 0.82 0.01
## 3 ROU 4873 -0.58 0.04 0.87 0.02
## 4 RUS 5789  0.05 0.02 0.73 0.01
## 5 SVK 6257 -0.11 0.02 0.94 0.02
```

Frequency table

Percentage of students by school grade (ST001D01T)

```
pisa2015.table(variable="ST001D01T", by= "CNT", data = pisa)
```

```
##    CNT ST001D01T Freq Percentage Std.err.
## 1  HUN         7   40       1.71     0.33
## 2  HUN         8  209       8.49     0.48
## 3  HUN         9 4542      75.80     0.67
## 4  HUN        10  867      13.99     0.49
## 5  POL         7   22       0.64     0.15
## 6  POL         8  170       4.91     0.32
## 7  POL         9 4274      93.82     0.41
## 8  POL        10   12       0.63     0.19
## 9  ROU         7   45       1.43     0.28
## 10 ROU         8  266       8.90     0.54
## 11 ROU         9 3810      74.79     0.89
## 12 ROU        10  755      14.88     0.66
## 13 RUS         7   10       0.18     0.08
## 14 RUS         8  409       6.62     0.32
## 15 RUS         9 4849      79.73     1.47
```

Proficiency levels

Science performance proficiency levels

```
pisa2015.ben.pv(pvlabel="SCIE",
                cutoff = c(260.54, 334.94, 409.54, 484.14, 558.73, 633.33, 707.93),
                by="CNT", data=pisa)
```

```
##    CNT       Benchmarks Percentage Std. err.
## 1  HUN        <= 260.54       0.81      0.19
## 2  HUN (260.54, 334.94]       6.82      0.63
## 3  HUN (334.94, 409.54]      18.38      0.86
## 4  HUN (409.54, 484.14]      25.45      0.85
## 5  HUN (484.14, 558.73]      27.32      0.88
## 6  HUN (558.73, 633.33]      16.62      0.77
## 7  HUN (633.33, 707.93]       4.26      0.41
## 8  HUN         > 707.93       0.34      0.12
## 9  POL        <= 260.54       0.32      0.11
## 10 POL (260.54, 334.94]       2.60      0.36
## 11 POL (334.94, 409.54]      13.33      0.70
## 12 POL (409.54, 484.14]      26.60      0.93
## 13 POL (484.14, 558.73]      29.92      0.93
## 14 POL (558.73, 633.33]      19.89      0.78
## 15 POL (633.33, 707.93]       6.30      0.55
```

Regression analysis with plausible values

Science performance on ESCS and school grade

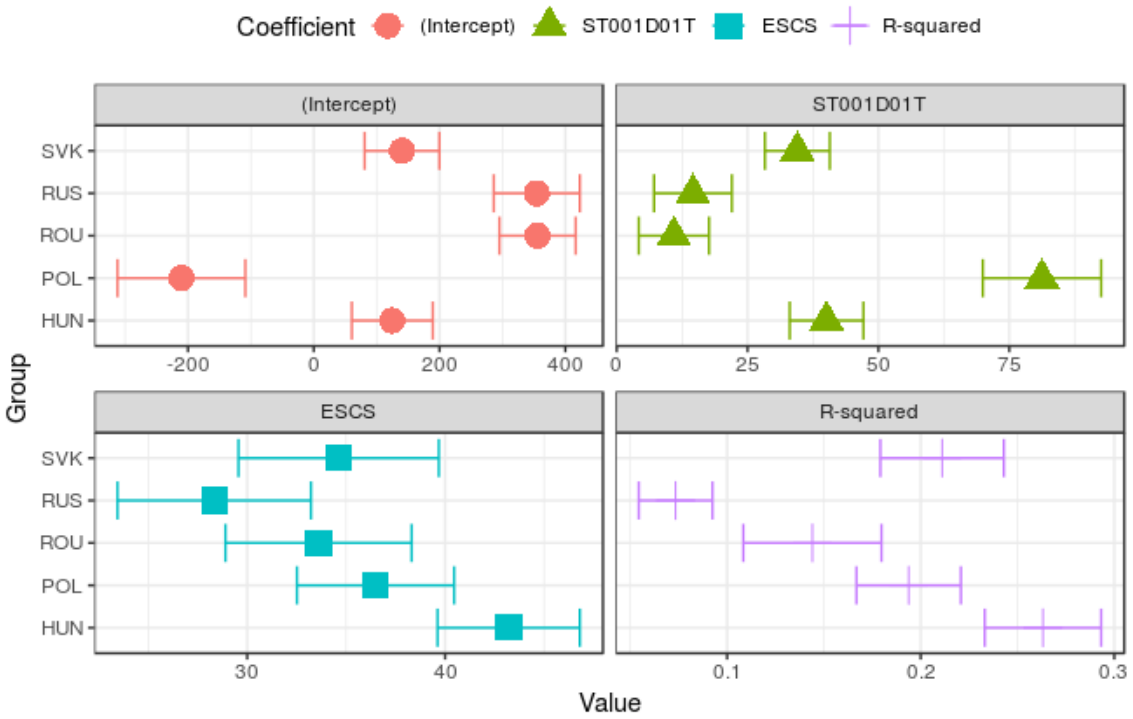
```
pisa2015.reg.pv(pvlabel="SCIE", x=c("ST001D01T", "ESCS"), by= "CNT", data = pisa)
```

```
## $HUN
##             Estimate Std. Error t value
## (Intercept)   124.89      32.86    3.80
## ST001D01T      40.09       3.59   11.16
## ESCS           43.21       1.83   23.57
## R-squared       0.26       0.02   17.16
##
## $POL
##             Estimate Std. Error t value
## (Intercept)  -210.38      51.85   -4.06
## ST001D01T      81.25       5.77   14.09
## ESCS           36.49       2.03   18.01
## R-squared       0.19       0.01   14.08
##
## $ROU
##             Estimate Std. Error t value
## (Intercept)   355.81      30.95   11.50
## ST001D01T      10.93       3.42    3.19
```

Visualizing results

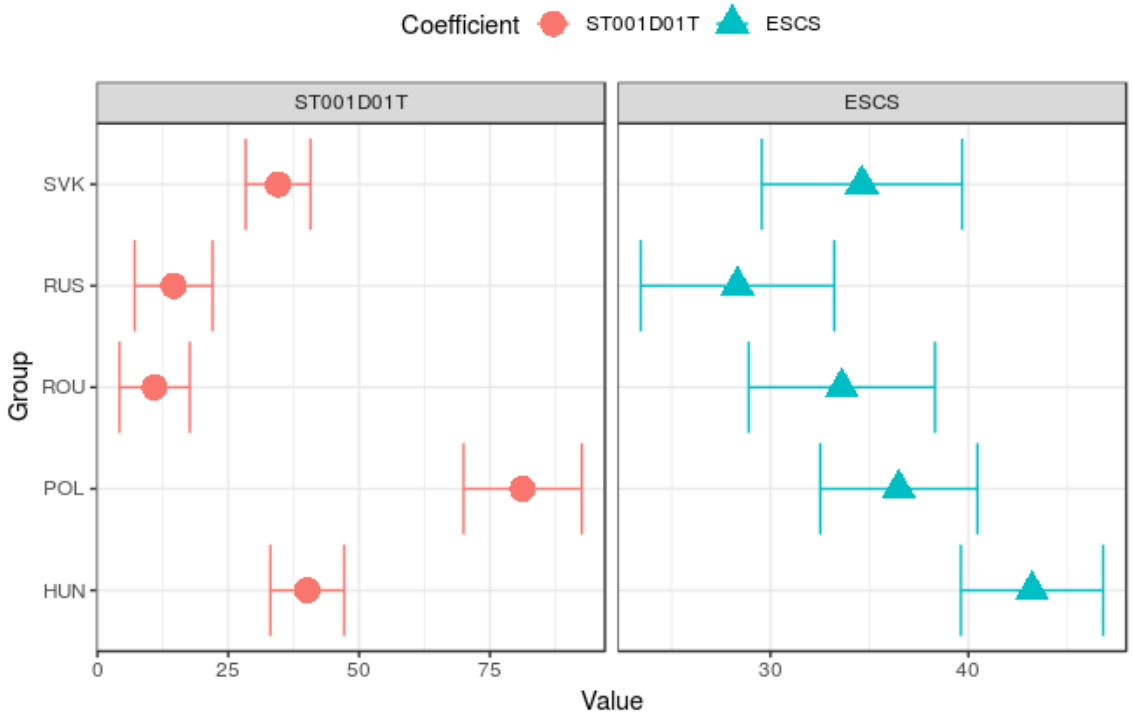
```
# Store results in object
scie_reg <- pisa2015.reg.pv(pvlabel="SCIE", x=c("ST001D01T", "ESCS"), by= "CNT", data = pisa)
# Plot object
plot(scie_reg)
```

Visualizing results



Selected coefficients

```
plot(scie_reg, vars = c("ST001D01T", "ESCS"))
```



Regression analysis without plausible values

Sense of belonging to school on ESCS

```
pisa2015.reg(y="BELONG", x="ESCS", by= "CNT", data = pisa)
```

```
## $HUN
##             Estimate Std. Error t value
## (Intercept)     0.09       0.02    4.81
## ESCS            0.12       0.02    7.48
## R-squared       0.01       0.00    3.85
##
## $POL
##             Estimate Std. Error t value
## (Intercept)    -0.24       0.02  -14.59
## ESCS            0.04       0.02    2.38
## R-squared       0.00       0.00    1.19
##
## $ROU
##             Estimate Std. Error t value
## (Intercept)     0.04       0.02    1.43
## ESCS            0.06       0.02    3.86
## R-squared       0.00       0.00    1.87
##
```

Thank you!