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.

1. Introduction

International large-scale assessments (LSA) studies measure student performance through standardized achievement tests and administer questionnaires to collect data on students, their families, and schools that shed light on the mechanisms responsible for student performance in a number of countries. The results have received a great deal of attention from researchers and policymakers around the world and have had significant impact on educational policy and on the educational debate. The Programme for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS), and the Progress in International Reading Literacy Study (PIRLS) stand out for their impact, comparative trend data, and number of participating countries. More recently, attention is directed as well towards the International Computer and Information Literacy Study (ICILS)
and the Programme for the International Assessment of Adult Competencies (PIAAC). The data from PISA, TIMSS, PIRLS, ICILS, and PIAAC are publicly available, but its use is somewhat limited by available analytical tools for handling the complex design of LSA studies.

The design of international LSA studies involves complex sampling and testing procedures that have consequences on the analysis stage. Sampling is conducted in two stages: Schools are selected in the first stage and students in the second stage. Testing uses a rotated design consisting of different test versions comparable through a common core of items. Datasets contain sampling variables (e.g., replicate weights) and plausible values of achievement scores in order to account for the complex sampling and test design, respectively. Traditional statistical procedures cannot handle these design complexities. Further, the organization of public datasets from TIMSS and PIRLS in a large number of files by country and survey instrument is not straightforward for users and requires commercial software alternatives (e.g., IDB Analyzer, IEA Hamburg 2017, in combination with SPSS, IBM Corporation 2015) in order to merge and select data. The R (R Core Team 2017) package intsvy (Caro and Biecek 2017) facilitates access to international assessment data by providing tools for importing data and conducting analysis while soundly considering the sample and test design in the calculation of statistics and associated standard errors. intsvy is an acronym for international surveys and the package is available from the Comprehensive R Archive Network (CRAN) at https://CRAN.R-project.org/package=intsvy.

2. Complex design of international LSA

Obtaining point estimates of any statistic of interest $\theta$ (e.g., mean, correlation, percentage, regression coefficient) is not particularly complicated with international assessment data. Standard procedures weighted by the total sampling weight can be used to calculate $\theta$ for the observed data. For student performance, the average of plausible values estimates yields the estimate of group-level student performance,

$$\theta = \frac{1}{M} \sum_{i=1}^{M} \theta_i,$$

where $M$ is the number of imputations, typically 5 in international assessments, $\theta_i$ is the average score for plausible value $M$, and $\theta$ is the average estimate of student performance.

What is particularly challenging is the calculation of the standard error of $\theta$, that is, the uncertainty associated with its estimation. This is because the complex test and sampling design introduce two sources of error in the estimation of $\theta$: imputation error and sampling error, respectively. And these errors cannot be calculated with standard routines of statistical software. The calculation of correct standard errors is important for making valid comparisons of performance between countries or boys and girls, for example. It is for this reason that specialized tools like the intsvy package are required.

2.1. Rotated test design

The total item pool of international assessments consists of hundreds of items that demand hours of testing time in order to produce valid and reliable measures of student achievement
constructs. Clearly, it is not feasible to administer a test including the entire item pool for logistic, fatigue, and testing time issues in general. International assessments employ a rotated design form in order to achieve a balance between validity and reasonable testing time. Test items are arranged into clusters that in turn are distributed between booklets administered to students. Clusters are distributed such that it is possible to link test booklets through clusters in common. Cluster linkage between booklets ensures the comparability of results between students and reporting on the same scale. Rotated test forms introduce technical complexities in the estimation of student performance, since students respond only to a subset of items, the ones in the booklet, but inferences on student performance are made as if the students had responded to the entire assessment through plausible value techniques (von Davier, Gonzalez, and Mislevy 2009).

The plausible values approach combines item response theory and latent regression techniques to produce unbiased estimates of student performance at the population level. Plausible values are random draws from the estimated posterior distribution of student performance given student responses to the subset of test items and background information collected in questionnaires. Importantly, plausible values are not used to infer performance at the individual level, since students responded only to a subset of the items and measurement errors at the individual level tend to be large. The average of plausible values estimates was calculated in Equation 1. The variance reflects uncertainty in the estimation associated with making multiple imputations of plausible values based on the posterior distribution of student performance. The formula of the imputation variance, \( \text{VAR}_{\text{imp}}[\theta] \), is as follows (Little and Rubin 2002):

\[
\text{VAR}_{\text{imp}}[\theta] = \frac{1}{M-1} \sum_{i=1}^{M} (\theta_i - \bar{\theta})^2.
\]

### 2.2. Complex sample design

Student samples in international LSA are selected in two stages: Schools are sampled in the first stage and students within the school in the second stage. For example, 15-year-olds are sampled randomly within schools in PISA and intact classes within schools are sampled randomly in TIMSS and PIRLS. The sampling error takes into account the uncertainty related with the sample selection, as different samples of schools and students from the population not necessarily yield the same estimates. The sampling error formula under two-stage sampling cannot assume that observations are independent as in random sampling because students within schools tend to share similar characteristics, for example, family socio-economic status (SES) and the instructional setting. Compared to random sampling, the dependency of observations within schools in two-stage sampling tends to reduce the amount of information and increase the uncertainty of estimates, that is, the standard error. For example, a two-stage sample of 100 students per school in 10 schools will likely yield less information than a random sample of 1000 students. In one extreme scenario, if all students within schools are identical, the two-stage sample will represent 10 students and not 1000. In the other extreme, if all students within schools are uncorrelated, the two-stage sample size will be 1000. In real data the dependency of observations lies between these two scenarios (i.e., a sample size of 10 and 1000 students).

Replicate weights are used in international LSA to calculate sampling errors. A replicate
weight represents a sample of schools. The PISA dataset, for example, contains 80 replicate weights that represent 80 different school samples. An estimate (e.g., mean, percentage, regression coefficient) can be obtained for each sample. The variability of estimates across samples or replicate weights indicates the uncertainty due to school sample selection or the sampling error. Like multilevel models, replicate weights estimation introduces randomness in the selection of schools. Multilevel models do it by introducing random effects and replicate weights estimation by creating different samples in the data while maintaining the traditional ordinary least squares (OLS) model. From this perspective, replicate weights can be regarded as a case of adapting the data to the model and multilevel models as one of adapting the model to the data. Further, school sample variation with replicate weights of international LSA is not entirely random but takes into account stratification (e.g., one school is selected at random within each stratum for each replicate weight). As a result, multilevel models and replicate weights estimation do not yield exactly the same results. To the extent that multilevel models do not take into account stratification information in random effects, they tend to produce standard errors that are larger than for regression analysis using replicate weights. There are different replication techniques for two-stage sampling. TIMSS and PIRLS employ jackknife repeated replication (JRR) and PISA employs balanced repeated replication (BRR) with Fay’s modification. The principles underpinning these techniques and worked examples are presented in technical reports of international assessments (e.g., OECD 2014b). Here we will just present the formulas.

The sampling variance for PIRLS and TIMSS is:

$$\text{VAR}_{\text{sml}}[\theta] = \sum_{j=1}^{R} (\theta_j - \theta)^2.$$ 

The sampling variance in PISA is:

$$\text{VAR}_{\text{sml}}[\theta] = \frac{1}{G(1-k)} \sum_{j=1}^{R} (\theta_j - \theta)^2.$$ 

$R$ is the number of replicate weights, 75 jackknife replicate weights in PIRLS and TIMSS and 80 BRR replicate weights in PISA. For PIAAC estimation is slightly more complicated because different replication methods and numbers of replications were used in different countries. Thus the general formula for the sampling variance in PIAAC is:

$$\text{VAR}_{\text{sml}}[\theta] = c \sum_{j=1}^{R} (\theta_j - \theta)^2,$$

where $c = \frac{G-1}{G}$ (so called random groups (delete-one) approach) for Australia, Austria, Canada, Denmark and Germany while $c = 1$ (so called paired jackknife) for other countries. See the `intsvy::piaacReplicationScheme` table or the PIAAC Technical Report (OECD 2013b) for more details.

For student performance data, the sampling variance is the average across the 5 plausible values:

$$\text{VAR}_{\text{sml}}[\theta] = \frac{1}{5} (\text{VAR}_1[\theta] + \text{VAR}_2[\theta] + \text{VAR}_3[\theta] + \text{VAR}_4[\theta] + \text{VAR}_5[\theta]).$$
TIMSS and PIRLS, however, use an unbiased shortcut for calculating the sampling variance. Instead of the average, the sampling variance is equal to the sampling variance for the first plausible value, $\text{VAR}_1[\theta]$.

### 2.3. Standard error formula

The total standard error for single observed variables in international assessment data is equal to the sampling error. For the plausible values of student performance the standard error additionally takes into account imputation error. The total variance formula combines the sampling error and the imputation error as follows:

$$\text{VAR}_{\text{tot}}[\theta] = \text{VAR}_{\text{svl}}[\theta] + \left(1 + \frac{1}{M}\right) \times \text{VAR}_{\text{imp}}[\theta].$$

The standard error is the square root:

$$\text{SE}[\theta] = \sqrt{\text{VAR}_{\text{tot}}[\theta]}.$$  \hspace{1cm} (2)

### 3. Overview of the package

There are different statistical tools for conducting analysis with international assessment data while handling replicate weights and plausible values. The IEA has produced the International Database **IDB Analyzer**, an **SPSS** add-on application for importing and analyzing data from IEA studies (e.g., PIRLS, TIMSS) and PISA. The National Center for Education Statistics (NCES) has developed the **International Data Explorer** (National Center for Education Statistics 2017), a web-based tool for creating tables and charts with data from PISA, PIRLS, TIMSS, and PIAAC. The OECD has published **SPSS** and **SAS** (SAS Institute Inc. 2013) macros for conducting analysis with PISA (OECD 2009). **Mplus** (Muthén and Muthén 1998–2017) is able to perform structural equation modeling while incorporating replicate weights. In **Stata** (StataCorp. 2015), **REPEST** (Avvisati and Keslair 2014) and **PV** (Macdonald 2008) modules handle plausible values and replicate weights with IEA and OECD data. Non-commercial alternatives in **R** to analyze survey data include packages **survey** (Lumley 2004), **BIFIEsurvey** (BIFIE 2017), **lavaan.survey** (Oberski 2014), and the [http://www.asdfree.com/](http://www.asdfree.com/) code repository (Damico 2015). Moreover packages **DAKS** (Ünlü and Sargin 2010) and **multilevelPSA** (Bryer and Pruzek 2011) include additional functionalities for psychometric analyses.

Package **intsvy** provides a non-commercial and extendible alternative to the **IDB Analyzer**. Unlike available packages in **R** for survey analysis, **intsvy** is tailored towards the analysis of international assessment data specifically. For example, as with the **IDB Analyzer**, an important purpose of the package is to provide functions to import data from studies conducted by the International Association for the Evaluation of Educational Achievement (IEA), such as TIMSS and PIRLS. Also, analysis functions calculate estimates by education system, percentages of students by international benchmarks (e.g., TIMSS and PIRLS) and proficiency levels (e.g., PISA), estimate percentiles for achievement scores with plausible values, and implicitly assume the replication method used, for example BRR for PISA and JRR with one plausible value used for estimation of sampling error in TIMSS and PIRLS. That is, the user is not
required to enter study-specific parameters (e.g., the replication method, names of weight variables and plausible values) in the analysis or to know in-depth study-specific estimation procedures. With that, `intsvy` facilitates access and analysis of international assessments. At the same time, study-specific parameters can be modified and the package can be extended to handle data from other studies.

Package `intsvy` includes functions for importing data and for data analysis. Data importation functions include `intsvy.var.label` for printing variable names and variable labels by instrument as well as names of participating countries, and `intsvy.select.merge` for selecting and merging data into a single data frame. Analysis functions include `intsvy.mean.pv` for calculating means with plausible values, `intsvy.mean` for calculating means, `intsvy.table` for producing frequency tables, `intsvy.log.pv` for estimating logistic regression with plausible values, `intsvy.log` for estimating logistic regression, `intsvy.per.pv` for calculating percentiles with plausible values, `intsvy.ben.pv` for calculating percentages of students at each benchmarks or proficiency levels, `intsvy.reg` for running regression, and `intsvy.reg.pv` for running regression with plausible values.

Alternatively, study-specific functions (e.g., `pisa.reg.pv`, `timss.table`) that call generic functions (e.g., `intsvy.reg.pv`, `intsvy.table`) can be used. For example, the following functions produce the same output of average mathematics scores by country using PISA data, one using the study-specific function `pisa.mean.pv` and the other with the generic function `intsvy.mean.pv`.

```r
R> pisa.mean.pv(pvlabel = "MATH", by = "CNT", data = pisa)
R> intsvy.mean.pv(pvnames = paste0("PV", 1:5, "MATH"), by = "CNT",
+ data = pisa, config = pisa_conf)
```

The argument `config = pisa_conf` supplies study-specific parameters (e.g., replication method, name of weight variables) for the analysis. Study-specific parameters (e.g., `pisa_conf`, `pirls_conf`) are contained in a script that is part of the package. The script and therefore package `intsvy` can be extended to handle data from other international assessment studies with the `intsvy.config()` function.

The architecture of the package is presented in Table 1. For example, the output of functions `piaac.table`, `timms.table`, `pirls.table`, `pisa.table`, or the generic `intsvy.table` is an object of the class ‘`intsvy.table`’, and a plot can be produced with the associated `plot` method.

Below data analysis examples are presented for the different functions. More examples alongside video tutorials for `intsvy` can be found at [http://users.ox.ac.uk/~educ0279/](http://users.ox.ac.uk/~educ0279/).

## 4. Applied examples

Package `intsvy` uses the formulas above to calculate point estimates (e.g., Equation 1) and correct standard errors (see Equation 2) for different statistics, including means, standard deviations, percentages, correlations, and regression coefficients with data from observed variables or plausible values of student performance. As usual, the package can be installed and loaded into R by running:

```r
R> install.packages("intsvy")
R> library("intsvy")
```
Table 1: Analytical functions implemented in the intsvy package are presented in the first column. The second column presents the classes of the returned objects. For each class, a plot() method has been implemented.

4.1. Select and merge data

Package intsvy provides tools for selecting and importing data into R. Data can be imported in two steps. First, the generic function intsvy.var.label facilitates data selection by reporting variable names, variable labels, and names of participating countries in available datasets. Secondly, the generic function intsvy.select.merge produces a single data frame for selected variables and countries. Sampling variables (i.e., replicate weights and total weights) and plausible variables are selected automatically. Alternatively, study-specific functions (e.g., pisa.var.label, pirls.select.merge) can be used.

TIMSS, PIRLS, and ICILS

Variable names, variable labels, and name abbreviations of countries in the PIRLS 2011 datasets are printed with

```R
R> pirls.var.label(folder = file.path(getwd(), "PIRLS 2011"))
```

The folder argument indicates where the multiple data files are located. The output is automatically stored in a text file located in the working directory (i.e., getwd()). The location and name of the output file can be modified with the output and name arguments. Alternatively, the same output with data characteristics can be produced with the generic intsvy.var.label function,

```R
R> intsvy.var.label(folder = file.path(getwd(), "PIRLS 2011"),
  + config = pirls_conf)
```

where the argument config = pirls_conf provides specific parameters for the PIRLS study. Similarly, the data from TIMSS and ICILS can be described with

```R
R> intsvy.var.label(folder = file.path(getwd(), "TIMSS 2011"),
  + config = timss8_conf)
R> intsvy.var.label(folder = file.path(getwd(), "ICILS 2013"),
  + config = icils_conf)
```
where again `config = timss8_conf` and `icils_conf` contain specific parameters for the data of TIMSS Grade 8 and ICILS.

Subsequently, selected data of specific variables and countries can be imported into a single data frame using `intsvy.select.merge` or study-specific functions (such as, for example, `timssg8.select.merge`, `timssg4.select.merge`, and `pirls.select.merge`). Data importing tools are particularly useful for TIMSS, PIRLS, and ICILS because original datasets available from the IEA Data Repository ([http://rms.iea-dpc.org/](http://rms.iea-dpc.org/)) are organized in a large number of data files by country, school grade, and survey instrument (e.g., student questionnaire, home questionnaire, teacher questionnaire) and users are usually not familiar with the data administrative structure.

For example, selected variables from the student and school questionnaire in TIMSS 2011 Grade 8 for Australia, Bahrain, Armenia, and Chile are imported by

```r
R> timss8g <- intsvy.select.merge(
+   folder = file.path(getwd(), "TIMSS 2011"),
+   countries = c("AUS", "BHR", "ARM", "CHL"),
+   student = c("BSDGEDUP", "ITSEX", "BSDAGE", "BSBGSLM", "BSDGSLM"),
+   school = c("BCBGDAS", "BCDG03"),
+   config = timss8_conf)
```

It is assumed that TIMSS data files were downloaded from the IEA Data Repository and stored in the location of `folder`. The same dataset can be imported using the function `timssg8.select.merge` with

```r
R> timss8g <- timssg8.select.merge(
+   folder = file.path(getwd(), "TIMSS 2011"),
+   countries = c("AUS", "BHR", "ARM", "CHL"),
+   student = c("BSDGEDUP", "ITSEX", "BSDAGE", "BSBGSLM", "BSDGSLM"),
+   school = c("BCBGDAS", "BCDG03"))
```

The resulting data frame `timss8g` contains the selected data. A country identifier variable, `IDCNTRYL`, is created automatically in IEA datasets. Number of boys (`ITSEX = 2`) and girls (`ITSEX = 1`) by education system can be calculated with

```r
R> with(timss8g, table(IDCNTRYL, ITSEX))
```

| ITSEX | 1 | 2 |
|-------|---|---|
| Armenia | 2894 | 2952 |
| Australia | 3747 | 3809 |
| Bahrain | 2288 | 2352 |
| Chile | 3133 | 2702 |

Data from the mathematics teacher questionnaire or the science teacher questionnaire can be selected using the arguments `math.teacher` or `science.teacher`. For example, the data frame `timss_nt` contains variables "BTBG02", "BTBG04", "BTBGTC" from the mathematics teacher questionnaire in addition to selected data from the student and school questionnaire.
R> timss_mt <- timssg8.select.merge(
+ folder = file.path(getwd(), "TIMSS 2011"),
+ countries = c("AUS", "BHR", "ARM", "CHL"),
+ student = c("BSDGEDUP", "ITSEX", "BSDAGE", "BSBGSML", "BSDGSLM"),
+ math.teacher = c("BTBG02", "BTBG04", "BTBGTC5"),
+ school = c("BCBGDAS", "BCDG03"))

The data frame timss_st contains the same teacher variables but for the science teacher.

R> timss_st <- timssg8.select.merge(
+ folder = file.path(getwd(), "TIMSS 2011"),
+ countries = c("AUS", "BHR", "ARM", "CHL"),
+ student = c("BSDGEDUP", "ITSEX", "BSDAGE", "BSBGSML", "BSDGSLM"),
+ science.teacher = c("BTBG02", "BTBG04", "BTBGTC5"),
+ school = c("BCBGDAS", "BCDG03"))

As before, it is assumed that teacher data was downloaded in SPSS format and stored in the directory specified in folder or subfolders of this directory. Variable selection is facilitated by intsvy.var.label.

Selected PIRLS 2011 data from the student, home, and school questionnaires can be imported into a single data frame with the pirls.select.merge function

R> pirls <- pirls.select.merge(folder = file.path(getwd(), "PIRLS 2011"),
+ countries = c("AUS", "AUT", "AZE", "BFR"),
+ student = c("ITSEX", "ASDAGE", "ASBGSMR"),
+ home = c("ASDHEDUP", "ASDHCCP", "ASDHEL", "ASBHEL"),
+ school = c("ACDGDAS", "ACDGCM", "ACDGO3"))

or alternatively with the generic intsvy.select.merge function

R> pirls <- intsvy.select.merge(folder = file.path(getwd(), "PIRLS 2011"),
+ countries = c("AUS", "AUT", "AZE", "BFR"),
+ student = c("ITSEX", "ASDAGE", "ASBGSMR"),
+ home = c("ASDHEDUP", "ASDHCCP", "ASDHEL", "ASBHEL"),
+ school = c("ACDGDAS", "ACDGCM", "ACDGO3"), config = pirls_conf)

A cross-table of parental education levels (ASDHEDUP; 1 = university or higher, 2 = post-secondary, 3 = upper secondary, 4 = lower secondary, 5 = some primary or no school, 6 = NA) by education system can be produced with the selected pirls data

R> with(pirls, table(ASDHEDUP, IDCNTRYL))

| IDCNTRYL | ASDHEDUP | Australia | Austria | Azerbaijan | Belgium (French) |
|----------|----------|-----------|---------|------------|-----------------|
| 1        | 1336     | 1005      | 1296    | 1631       |
| 2        | 1243     | 881       | 1175    | 401        |
| 3        | 449      | 2281      | 1393    | 607        |
| 4        | 125      | 156       | 479     | 338        |
| 5        | 9        | 42        | 171     | 160        |
| 6        | 16       | 35        | 17      | 41         |
It is also possible to import data from the teacher questionnaire in PIRLS using the argument `teacher`, for example:

```r
R> pirls_teach <- pirls.select.merge(
+  folder = file.path(getwd(), "PIRLS 2011"),
+  countries = c("AUS", "AUT", "AZE", "BFR"),
+  student = c("ITSEX", "ASDAGE", "ASBGSMR"),
+  home = c("ASDHEDUP", "ASDHOCMP", "ASDHELA", "ASBHELA"),
+  teacher = c("ATBG01", "ATBG02", "ATBG03"),
+  school = c("ACDGDAS", "ACDGCMP", "ACDG03"))
```

Also ICILS data for selected countries and variables can be imported as follows:

```r
R> icils <- intsvy.select.merge(folder = file.path(getwd(), "ICILS 2013"),
+  countries = c("AUS", "POL", "SVK"),
+  student = c("S_SEX", "S_TLANG", "S_MISEI"),
+  school = c("IP1G02J", "IP1G03A"), config = icils_conf)
```

**PISA and PIAAC**

The data from PISA has a different structure. Original datasets available from the OECD [here](http://www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm) are organized in large files for the student, school, and parent questionnaire containing data for all participating countries. Accordingly, study-specific functions to describe (i.e., `pisa.var.label`) and import (i.e., `pisa.select.merge`) the data have a different structure with arguments for entering names of original data files directly.

For PISA, names of variables and participating countries can be printed with:

```r
R> pisa.var.label(folder = file.path(getwd(), "PISA 2012"),
+  school.file = "INT_SCQ12_DEC03.sav",
+  student.file = "INT_STU12_DEC03.sav")
```

where arguments `school.file`, `student.file`, and `parent.file` indicate the names of the original files located in the folder.

The function `pisa.select.merge` can be used to create a data frame with selected data. For example, selected data from the student and school questionnaire can be imported for Hong Kong, the United States, Sweden, Poland, and Peru, as follows:

```r
R> pisa <- pisa.select.merge(folder = file.path(getwd(), "PISA 2012"),
+  school.file = "INT_SCQ12_DEC03.sav",
+  student.file = "INT_STU12_DEC03.sav",
+  student = c("ST01Q01", "ST04Q01", "ST08Q01", "ST09Q01", "ST115Q01",
+              "ESCS", "PARED"),
+  school = c("CLSIZE", "TCSHORT"),
+  countries = c("HKG", "USA", "SWE", "POL", "PER"))
```

An alternative way to access data from PIAAC or PISA studies is by using R packages with converted data. Since these datasets have a significant size, up to few hundreds MB, they
are not available on CRAN. But they can be downloaded from the pbiecek account on github (https://github.com/pbiecek).

Packages with consecutive releases of PISA data are named PISA2000lite, PISA2003lite, PISA2006lite, PISA2009lite, PISA2012lite, while the package with PIAAC data is named PIAAC. For example, the following code installs the package with the PISA 2012 data

```r
R> library("devtools")
R> install_github("pbiecek/PISA2012lite")
```

Dictionaries with variable names are available in student2012dict, school2012dict and parent2012dict vectors. With aid of the grep function it is possible to find a desired variable. Here is an example for finding the variable with the number of books at home.

```r
R> data("student2012dict", package = "PISA2012lite")
R> grep(student2012dict, pattern = "books", value = TRUE)
```

```
ST26Q10
"Possessions - textbooks"
ST26Q11
"Possessions - <technical reference books>"
ST28Q01
"How many books at home"
```

Variable names, such as ST28Q01 can be used to extract information of specific variables from data frames student2012, school2012 and parent2012. For example:

```r
R> data("student2012", package = "PISA2012lite")
R> table(student2012["ST28Q01"])
```

```
0-10 books 11-25 books 26-100 books
94016 96371 133686
101-200 books 201-500 books More than 500 books
67538 48633 28293
```

For PIAAC, the corresponding package can be installed using:

```r
R> library("devtools")
R> install_github("pbiecek/PIAAC")
```

A single data frame with PIAAC data is available in the piaac data frame while a dictionary for variable names is stored in the piaacdict vector.

```r
R> data("piaacdict", package = "PIAAC")
R> grep(piaacdict, pattern = "Number of books", value = TRUE)
```

```
J_Q08
"Background - Number of books at home"
```
A frequency table with number of books at home is produced by

\begin{verbatim}
R> data("piaac", package = "PIAAC")
R> table(piaac["J_Q08"])
\end{verbatim}

| Books at Home          | Frequency |
|------------------------|-----------|
| 10 books or less      | 21590     |
| 11 to 25 books        | 23069     |
| 26 to 100 books       | 47999     |
| 101 to 200 books      | 25938     |
| 201 to 500 books      | 20125     |
| More than 500 books   | 10760     |

4.2. Average achievement scores with plausible values

Functions `pisa.mean.pv`, `piaac.mean.pv`, `timss.mean.pv`, and `pirls.mean.pv` calculate average estimates and associated standard errors for achievement variables with plausible values. Three main arguments are supplied by the `useR`: `pvlabel`, `by`, and `data`. Argument `pvlabel` indicates the part of the label in common for the plausible values variables (e.g., "READ", "MATH"). Argument `by` defines the level of grouping for the analysis (e.g., "CNT") and may contain more than one level (e.g., `c("CNT", "SEX")`). And argument `data` defines the dataset to be used in the analysis. Alternatively, the generic function `intsy.mean.pv` can be used.

**PISA and PIAAC**

For example, in PISA 2012, the average math performance by education system and associated standard errors can be calculated as follows (see OECD 2014a, p. 305):

\begin{verbatim}
R> pisa.mean.pv(pvlabel = "MATH", by = "CNT", data = pisa)
\end{verbatim}

| Education System | Freq | Mean  | s.e. | SD   | s.e  |
|------------------|------|-------|------|------|------|
| HKG              | 4670 | 561.24| 3.22 | 96.31| 1.92 |
| PER              | 6035 | 368.10| 3.69 | 84.36| 2.20 |
| POL              | 4607 | 517.50| 3.62 | 90.37| 1.89 |
| SWE              | 4736 | 478.26| 2.26 | 91.75| 1.28 |
| USA              | 4978 | 481.37| 3.60 | 89.86| 1.30 |

The argument `pvlabel = "MATH"` refers to the name suffix in common of the variables containing the plausible values variables: `PV1MATH`, `PV2MATH`, `PV3MATH`, `PV4MATH`, and `PV5MATH`. For science and reading, this argument should be changed to `pvlabel = "READ"` and `pvlabel = "SCIE"`, for example.

The same output can be produced with:

\begin{verbatim}
R> intsvy.mean.pv(pvnames = paste0("PV", 1:5, "MATH"), by = "CNT",
+                 data = pisa, config = pisa_conf)
\end{verbatim}

where the structure is similar to `pisa.mean.pv` but names of plausible values are entered directly in `pvnames` and specific parameters for the PISA dataset are entered in the `config` argument.
More levels of grouping can be included in the analysis. For example the following code produces results by education system (CNT) and the student’s sex (ST04Q01; 1 = female, 2 = male), while exporting results (export = TRUE) into a comma-separated value (CSV) file (see OECD 2014a, p. 305).

```R
R> pisa.mean.pv(pvlabel = "MATH", by = c("CNT", "ST04Q01"),
+ data = pisa, export = TRUE, name = "PISA mean by sex",
+ folder = "C:/PISA/PISA 2012/Results")
```

| CNT | ST04Q01 | Freq | Mean   | s.e. | SD   | s.e  |
|-----|---------|------|--------|------|------|------|
| HKG | 1       | 2161 | 552.96 | 3.94 | 90.51| 2.23 |
| HKG | 2       | 2509 | 568.38 | 4.55 | 100.49| 2.18 |
| PER | 1       | 3118 | 358.92 | 4.75 | 83.44| 2.61 |
| PER | 2       | 2917 | 377.82 | 3.65 | 84.24| 2.51 |
| POL | 1       | 2388 | 515.53 | 3.76 | 86.38| 1.59 |
| POL | 2       | 2219 | 519.56 | 4.25 | 94.32| 2.65 |
| SWE | 1       | 2378 | 479.63 | 2.41 | 87.60| 1.60 |
| SWE | 2       | 2358 | 476.92 | 2.97 | 95.63| 1.88 |
| USA | 1       | 2453 | 479.00 | 3.91 | 87.08| 1.71 |
| USA | 2       | 2525 | 483.65 | 3.81 | 92.40| 1.61 |

The name of the resulting CSV file is `PISA mean by sex.csv` and it is located in the folder `C:/PISA/PISA 2012/Results`. It can be imported directly into a spreadsheet for further analysis or for formatting for publication.

For PIAAC, numeracy average performance can be calculated with `piaac.mean.pv` function with:

```R
R> head(piaac.mean.pv(pvlabel = "NUM", by = "CNTRYID", data = piaac))
```

| CNTRYID | Freq | Mean   | s.e. | SD   | s.e  |
|---------|------|--------|------|------|------|
| Austria | 5130 | 275.04 | 0.88 | 48.84| 0.64 |
| Belgium | 5463 | 280.39 | 0.83 | 49.27| 0.67 |
| Canada  | 26683| 265.24 | 0.70 | 55.60| 0.54 |
| Czech Republic | 6102 | 275.73 | 0.93 | 43.59| 0.78 |
| Germany | 5465 | 271.73 | 1.00 | 52.68| 0.74 |
| Denmark | 7328 | 278.28 | 0.73 | 51.13| 0.59 |

or with the generic `intsvy.mean.pv` function:

```R
R> head(intsvy.mean.pv(pvnames = paste0("PVNUM", 1:10), by = "CNTRYID",
+ data = piaac, config = piaac_conf))
```

Results by country and age group can be produced with:

```R
R> head(piaac.mean.pv(pvlabel = "NUM", by = c("CNTRYID", "AGEG10LFS"),
+ data = piaac))
```
\textit{ints
vy}: Analyzing International Large-Scale Assessment Data in \texttt{R}

\begin{verbatim}
   CNTRYID  AGEG10LFS  Freq  Mean  s.e.  SD  s.e
1  Austria  24 or less  898  279.27  1.63  46.15  1.82
2  Austria      25-34  958  282.06  1.73  49.98  1.63
3  Austria      35-44 1117  281.35  2.01  50.26  1.40
4  Austria      45-54 1188  274.48  1.67  46.49  1.24
5  Austria      55 plus  969  257.48  1.74  46.83  1.47
6  Belgium  24 or less  994  282.82  1.74  45.07  1.63
\end{verbatim}

\textbf{TIMSS, PIRLS, and ICILS}

A similar analysis can be conducted with TIMSS and PIRLS data. Mathematics average performance by education system in TIMSS 2011, Grade 8 can be calculated with (see Foy, Arora, and Stanco 2013, p. 15):

\begin{verbatim}
R> timss.mean.pv(pvlabel = "BSMMAT", by = "IDCNTRYL", data = timss8g)

   IDCNTRYL  Freq  Mean  s.e.  SD  s.e
1   Armenia 23384  466.59  2.73  90.68  1.73
2   Australia 30224  504.80  5.09  85.42  3.36
3   Bahrain 18560  409.22  1.96  99.57  1.72
4    Chile  23340  416.27  2.59  79.65  1.85
\end{verbatim}

or using \texttt{ints
vy}.\texttt{mean.pv}:

\begin{verbatim}
R> intsvy.mean.pv(pvnames = paste0("BSMMAT0", 1:5), by = "IDCNTRYL",
                   data = timss8g, config = timss8_conf)

Unlike PISA, the argument \texttt{pvlabel} in study-specific functions for TIMSS and PIRLS refers to the prefix of the variable names containing the plausible values. For example, variable names of math plausible values in TIMSS are \texttt{BSMMAT01}, \texttt{BSMMAT02}, \texttt{BSMMAT03}, \texttt{BSMMAT04}, and \texttt{BSMMAT01} and variable names of reading plausible values in PIRLS are \texttt{ASRREA01}, \texttt{ASRREA02}, \texttt{ASRREA03}, \texttt{ASRREA04}, and \texttt{ASRREA05}. When using the generic \texttt{ints
vy}.\texttt{mean.pv}, names of plausible values are entered directly in the argument \texttt{pvnames}, for example for mathematics in TIMSS \texttt{pvnames = paste0("BSMMAT0", 1:5)}, where

\begin{verbatim}
R> paste0("BSMMAT0", 1:5)

[1] "BSMMAT01" "BSMMAT02" "BSMMAT03" "BSMMAT04" "BSMMAT05"
\end{verbatim}

As with other functions, results can be exported into a CSV file using the \texttt{export = TRUE} argument.

TIMSS results by education system and student’s sex (1 = female, 2 = male) can be calculated with (see Foy et al. 2013, p. 18):

\begin{verbatim}
R> timss.mean.pv(pvlabel = "BSMMAT", by = c("IDCNTRYL", "ITSEX"),
                data = timss8g)
\end{verbatim}
In PIRLS 2011, reading performance results by country can be calculated equally with the following two commands (see Foy and Drucker 2013, p. 15):

```r
R> pirls.mean.pv(pvlabel = "ASRREA", by = "IDCNTRYL", data = pirls)
R> intsvy.mean.pv(pvnames = paste0("ASRREA0", 1:5), by = "IDCNTRYL", + data = pirls, config = pirls_conf)
```

Reading performance by country and student’s sex (1 = female, 2 = male) can be calculated by (see Foy and Drucker 2013, p. 18):

```r
R> pirls.mean.pv(pvlabel = "ASRREA", by = c("IDCNTRYL", "ITSEX"), + data = pirls)
```

ICILS average performance results by education system can be calculated with:

```r
R> intsvy.mean.pv(pvnames = paste0("PV", 1:5, "CIL"), by = "IDCNTRY", + data = icils, config = icils_conf)
```
4.3. Average estimates without plausible values

Means and standard errors for variables without plausible values, that is, for all of the other variables in the datasets, can be calculated with functions `pisa.mean`, `piaac.mean`, `timss.mean`, `pirls.mean` or with the generic function `intsy.mean`.

**PISA and PIAAC**

For example, the following code calculates the average highest level of education of parents in years of schooling (PARED) by education system in PISA 2012 (see OECD 2013a, p. 183):

```R
R> pisa.mean(variable = "PARED", by = "CNT", data = pisa)

  CNT  Freq  Mean  s.e.  SD  s.e
1  HKG 4477 11.41 0.14 3.02 0.05
2  PER 5960 11.46 0.14 4.06 0.04
3  POL 4481 12.68 0.06 2.09 0.03
4  SWE 4496 14.09 0.04 2.27 0.04
5  USA 4869 13.65 0.09 2.63 0.07
```

The same output can be produced with the generic function:

```R
R> intsvy.mean(variable = "PARED", by = "CNT", data = pisa, + config = pisa_conf)
```

The following example with PIAAC data calculates the average score in the index of use of reading skills at home (READHOME) by country:

```R
R> head(piaac.mean(variable = "READHOME", by = "CNTRYID", data = piaac))

      CNTRYID  Freq  Mean  s.e.
1     Austria   4962 2.15 0.01
2   Belgium    4945 1.94 0.01
3   Canada   26508 2.27 0.01
4 Czech Republic   6051 1.86 0.02
5   Germany    5357 2.28 0.02
6   Denmark   7226 2.18 0.01
```

The same output can be produced with:

```R
R> head(intsvy.mean(variable = "READHOME", by = "CNTRYID", data = piaac, + config = piaac_conf))
```

**TIMSS and PIRLS**

For TIMSS 2011, the following code calculates the average of the index Students Like Learning Mathematics (BSBGLM) by education system (see Foy et al. 2013, p. 27):

```R
R> head(timss.mean(variable = "BSBGLM", by = "EDUC_SYS", data = timss))

      EDUC_SYS  Freq  Mean  s.e.
1   Byeon 12270 2.67 0.01
2  Canada  63098 2.12 0.01
3 Deutschland 6233 2.70 0.01
4  Denmark 52426 2.29 0.01
5  USA 93819 2.31 0.01
6  China  92799 2.37 0.01
```

The same output can be produced with:

```R
R> head(intsvy.mean(variable = "BSBGLM", by = "EDUC_SYS", data = timss, + config = timss_conf))
```
R> timss.mean(variable = "BSBGSLM", by = "IDCNTRYL", data = timss8g)

IDCNTRYL  Freq   Mean   s.e.
1 Armenia 5626 10.87 0.05
2 Australia7389  9.32 0.06
3 Bahrain 4581  9.77 0.03
4 Chile  5772  9.76 0.04

For PIRLS 2011, the following calculates the average of the index Early Literacy Activities before Beginning Primary School by education system (see Foy and Drucker 2013, p. 28):

R> pirls.mean(variable = "ASBHELA", by = "IDCNTRYL", data = pirls)

IDCNTRYL  n   Mean   Std.err.
1 Australia3232 10.84 0.06
2 Austria 4393  9.98 0.03
3 Azerbaijan4509  9.47 0.07
4 Belgium (French)3383  9.69 0.04

As before, the generic function intsvy.mean can be used to reproduce the same output.

4.4. Regression analysis

Regression analysis is performed by functions pisa.reg.pv, timss.reg.pv, pirls.reg.pv, and the generic function intsvy.reg.pv.

PISA and PIAAC

Differences in mean performance calculated previously for boys and girls can be tested for statistical significance using a regression approach. For example, significance tests can be conducted in PISA 2012 as follows (see OECD 2014a, p. 305):

R> pisa$SEX[pisa$ST04Q01 == 1] <- "female"
R> pisa$SEX[pisa$ST04Q01 == 2] <- "male"
R> pisa.reg.pv(pvlabel = "MATH", x = "SEX", by = "CNT", data = pisa)

$HKG

|                     | Estimate | Std. Error | t value |
|---------------------|----------|------------|---------|
| (Intercept)         | 552.96   | 3.94       | 140.18  |
| SEXmale             | 15.42    | 5.69       | 2.71    |
| R-squared           | 0.01     | 0.00       | 1.31    |

$PER

|                     | Estimate | Std. Error | t value |
|---------------------|----------|------------|---------|
| (Intercept)         | 358.92   | 4.75       | 75.53   |
| SEXmale             | 18.90    | 3.92       | 4.82    |
| R-squared           | 0.01     | 0.01       | 2.33    |
**$POL**

| Estimate | Std. Error | t value |
|----------|------------|---------|
| (Intercept) | 515.53 | 3.76 | 137.28 |
| SEXmale | 4.03 | 3.42 | 1.18 |
| R-squared | 0.00 | 0.00 | 0.59 |

**$SWE**

| Estimate | Std. Error | t value |
|----------|------------|---------|
| (Intercept) | 479.63 | 2.41 | 199.08 |
| SEXmale | -2.71 | 2.98 | -0.91 |
| R-squared | 0.00 | 0.00 | 0.41 |

**$USA**

| Estimate | Std. Error | t value |
|----------|------------|---------|
| (Intercept) | 479.00 | 3.91 | 122.52 |
| SEXmale | 4.65 | 2.80 | 1.66 |
| R-squared | 0.00 | 0.00 | 0.81 |

The same output can be produced with the generic function:

```r
R> intsvy.reg.pv(pvlabel = "MATH", x = "SEX", by = "CNT", + data = pisa, config = pisa_conf)
```

Argument `x` defines the independent variable(s), in this case `SEX`, but more variables can be included separated by commas (e.g., `x = c("SEX", "ESCS")`). The output is a list with regression results by education system. Coefficient `SEXmale` captures differences between boys and girls and its `t` value indicates whether they are statistically significant.

Regression results including replicate estimates and residuals can be stored in an object and retrieved for further analysis. For example, `pisa_ses` contains results of a regression of mathematics performance on the student’s sex and the index of economic, social, and cultural status (`ESCS`):

```r
R> (pisa_ses <- pisa.reg.pv(pvlabel = "MATH", x = c("SEX", "ESCS"), + by = "CNT", data = pisa))
```

**$HKG**

| Estimate | Std. Error | t value |
|----------|------------|---------|
| (Intercept) | 576.70 | 3.78 | 152.71 |
| SEXmale | 13.97 | 4.85 | 2.88 |
| ESCS | 26.63 | 2.64 | 10.09 |
| R-squared | 0.08 | 0.01 | 5.47 |

**$PER**

| Estimate | Std. Error | t value |
|----------|------------|---------|
| (Intercept) | 400.25 | 4.64 | 86.18 |
| SEXmale | 17.94 | 2.70 | 6.65 |
ESCS 33.06  2.03  16.25  
R-squared 0.25  0.02  10.37

$POL

|              | Estimate | Std. Error | t value |
|--------------|----------|------------|---------|
| (Intercept)  | 524.71   | 3.40       | 154.16  |
| SEXmale      | 3.08     | 2.90       | 1.06    |
| ESCS         | 40.94    | 2.43       | 16.85   |
| R-squared    | 0.17     | 0.02       | 9.99    |

$SWE

|              | Estimate | Std. Error | t value |
|--------------|----------|------------|---------|
| (Intercept)  | 472.28   | 2.15       | 219.20  |
| SEXmale      | -1.63    | 2.82       | -0.58   |
| ESCS         | 35.88    | 1.93       | 18.60   |
| R-squared    | 0.11     | 0.01       | 9.86    |

$USA

|              | Estimate | Std. Error | t value |
|--------------|----------|------------|---------|
| (Intercept)  | 473.44   | 3.06       | 154.53  |
| SEXmale      | 5.35     | 2.76       | 1.94    |
| ESCS         | 35.40    | 1.67       | 21.25   |
| R-squared    | 0.15     | 0.01       | 11.15   |

The internal structure of the object is displayed with:

R> str(pisa_ses)

The object contains a list with five elements, one for each education system. In turn, each element is a list containing other five elements, for example:

R> names(pisa_ses[["POL"]])

[1] "replicates" "residuals" "var.w" "var.b" "reg"

where var.w and var.b contain the variance within (i.e., sampling error) and between (i.e., imputation error) of regression coefficients, reg is a data frame with final regression results, replicates and residuals are lists again with five elements, one for each plausible value, containing replicate estimates and residuals. pisa_ses[["POL"]][["replicates"]][[1]], for example, is a matrix with 80 rows (replicate estimates) and 4 columns (two independent variables plus the intercept and the $R^2$ estimate). We could extract replicate estimates of the ESCS coefficient for the first plausible value in Poland as follows:

R> (ses_poland <- pisa_ses[["POL"]][["replicates"]][[1]][, "ESCS"])

[1]  42.07649  40.98270  39.14176  38.98344  41.59449  42.05496  40.19260  40.06118
[9]  39.42540  42.82519  42.53080  41.71617  40.35559  39.4029  39.4687  39.60190
The distribution of replicate estimates can be visualized with `hist(ses_poland)` or with `ggplot(as.data.frame(ses_poland), aes(x = ses_poland)) + geom_density()` if package `ggplot2` (Wickham 2009) is available. It indicates sampling error in the estimation of the ESCS coefficient.

Logistic regression can be performed with and without plausible values with the functions `intsvy.log.pv` and `intsvy.log`.

With plausible values, the following code estimates the probability of being above proficiency level 5 in mathematics as a function of ESCS. The argument `cutoff` in `intsvy.log.pv` defines the level at which the plausible values are dichotomized, in this case 606.99, the lowest score at proficiency level 5. The binary dependent variable takes the value of one for scores above the `cutoff` and the value of zero for scores below or equal to the `cutoff`.

```R
R> intsvy.log.pv(pvlabel = "MATH", cutoff = 606.99, x = "ESCS", by = "CNT", + data = pisa, config = pisa_conf)
```

|  | Coef. Std. Error t value OR CI95low CI95up |
|---|----------------------------------------|
| HKG (Intercept) | -0.28 0.07 -4.22 0.67 0.86 |
| ESCS | 0.52 0.06 9.30 1.51 1.87 |

|  | Coef. Std. Error t value OR CI95low CI95up |
|---|----------------------------------------|
| PER (Intercept) | -5.17 0.37 -13.92 0.00 0.01 |
| ESCS | 1.97 0.41 4.86 3.24 15.85 |

|  | Coef. Std. Error t value OR CI95low CI95up |
|---|----------------------------------------|
| POL (Intercept) | -1.61 0.09 -18.70 0.17 0.24 |
| ESCS | 0.86 0.06 14.78 2.11 2.66 |

|  | Coef. Std. Error t value OR CI95low CI95up |
|---|----------------------------------------|
| SWE (Intercept) | -2.91 0.10 -29.00 0.04 0.07 |
| ESCS | 0.95 0.09 11.07 2.19 3.07 |

|  | Coef. Std. Error t value OR CI95low CI95up |
|---|----------------------------------------|
| USA (Intercept) | -2.91 0.10 -29.00 0.04 0.07 |
| ESCS | 0.95 0.09 11.07 2.19 3.07 |
The output reports odds ratios and associated confidence intervals in addition to coefficients, standard errors, and \(t\) values. The same output can be produced with:

\begin{verbatim}
R> pisa.log.pv(pvlabel = "MATH", cutoff = 606.99, x = "ESCS", +       by = "CNT", data = pisa)
\end{verbatim}

It is also possible to run a logistic regression without plausible values. We could for example estimate a regression of skipping class or school on having arrived late for school. The dependent binary variable is \(\text{SKIP}\):

\begin{verbatim}
R> pisa$SKIP <- ifelse(!(pisa$ST09Q01 == 1 & pisa$ST115Q01 == 1), 1, 0)
\end{verbatim}

The independent variable is \(\text{LATE}\):

\begin{verbatim}
R> pisa$LATE <- ifelse(!pisa$ST08Q01 == 1, 1, 0)
\end{verbatim}

The logistic regression model can be estimated with the generic \texttt{intsvy.log} or with:

\begin{verbatim}
R> pisa.log(y = "SKIP", x = "LATE", by = "CNT", data = pisa)
\end{verbatim}

\begin{tabular}{lccccc}
\hline
$\text{HKG}$ & Coef. & Std. Error & \textit{t} value & OR CI95low & CI95up \\
\hline
(Intercept) & -3.08 & 0.08 & -37.98 & 0.04 & 0.05 \\
LATE & 1.40 & 0.14 & 10.29 & 4.07 & 5.31 \\
\hline
$\text{PER}$ & Coef. & Std. Error & \textit{t} value & OR CI95low & CI95up \\
\hline
(Intercept) & -1.93 & 0.08 & -24.49 & 0.13 & 0.17 \\
LATE & 0.91 & 0.07 & 12.47 & 2.48 & 2.87 \\
\hline
$\text{POL}$ & Coef. & Std. Error & \textit{t} value & OR CI95low & CI95up \\
\hline
(Intercept) & -1.79 & 0.07 & -26.72 & 0.15 & 0.19 \\
LATE & 1.59 & 0.09 & 18.03 & 4.89 & 5.81 \\
\hline
$\text{SWE}$ & Coef. & Std. Error & \textit{t} value & OR CI95low & CI95up \\
\hline
(Intercept) & -2.14 & 0.08 & -26.26 & 0.12 & 0.14 \\
LATE & 1.41 & 0.09 & 15.33 & 4.08 & 4.89 \\
\hline
$\text{USA}$ & Coef. & Std. Error & \textit{t} value & OR CI95low & CI95up \\
\hline
(Intercept) & -1.24 & 0.05 & -25.55 & 0.26 & 0.32 \\
LATE & 0.86 & 0.06 & 13.29 & 2.36 & 2.68 \\
\hline
\end{tabular}
The following provides an example of regression with literacy scores as dependent variable and the participant’s sex as independent variable for PIAAC data.

R> rmodelLG <- piaac.reg.pv(pvlabel = "LIT", x = "GENDER_R", +   by = "CNTRYID", data = piaac)
R> head(summary(rmodelLG))

$Austria
(Intercept) 271.53 1.04 259.90
GENDER_RFemale -4.14 1.32 -3.13
R-squared 0.00 0.00 1.58

$Belgium
(Intercept) 278.09 0.97 287.08
GENDER_RFemale -5.27 1.21 -4.36
R-squared 0.00 0.00 2.17

$Canada
(Intercept) 274.49 0.86 317.75
GENDER_RFemale -2.30 1.20 -1.92
R-squared 0.00 0.00 1.04

$Czech Republic
(Intercept) 275.68 1.26 219.47
GENDER_RFemale -3.36 1.63 -2.06
R-squared 0.00 0.00 1.04

$Germany
(Intercept) 272.35 1.17 233.35
GENDER_RFemale -5.13 1.49 -3.46
R-squared 0.00 0.00 1.73

$Denmark
(Intercept) 270.58 1.03 262.31
GENDER_RFemale 0.43 1.36 0.31
R-squared 0.00 0.00 0.21

**TIMSS and PIRLS**

Tests of mean differences between boys and girls in TIMSS 2011, Grade 8 can be performed using a regression approach (see Foy et al. 2013, p. 21):
R> timss8g$SEX[timss8g$ITSEX == 1] <- "female"
R> timss8g$SEX[timss8g$ITSEX == 2] <- "male"
R> timss.reg.pv(pvlabel = "BSMMAT", by = "IDCNTRYL", x = "SEX",
+ data = timss8g)

$Armenia

| Estimate | Std. Error | t value |
|----------|------------|---------|
| (Intercept) | 471.52 | 3.07 | 153.75 |
| SEXmale | -9.66 | 3.10 | -3.12 |
| R-squared | 0.00 | 0.00 | 1.61 |

$Australia

| Estimate | Std. Error | t value |
|----------|------------|---------|
| (Intercept) | 500.41 | 4.72 | 105.93 |
| SEXmale | 8.75 | 6.90 | 1.27 |
| R-squared | 0.00 | 0.00 | 0.83 |

$Bahrain

| Estimate | Std. Error | t value |
|----------|------------|---------|
| (Intercept) | 430.78 | 2.51 | 171.50 |
| SEXmale | -42.89 | 3.99 | -10.74 |
| R-squared | 0.05 | 0.01 | 5.44 |

$Chile

| Estimate | Std. Error | t value |
|----------|------------|---------|
| (Intercept) | 409.46 | 3.23 | 126.86 |
| SEXmale | 14.48 | 3.63 | 3.99 |
| R-squared | 0.01 | 0.00 | 1.89 |

The same mean differences test can be performed for PIRLS 2011 with a regression (see Foy and Drucker 2013, p. 21):

R> pirls$SEX[pirls$ITSEX == 1] <- "female"
R> pirls$SEX[pirls$ITSEX == 2] <- "male"
R> pirls.reg.pv(pvlabel = "ASRREA", by = "IDCNTRYL", x = "SEX",
+ data = pirls)

$Australia

| Estimate | Std. Error | t value |
|----------|------------|---------|
| (Intercept) | 535.79 | 2.67 | 200.57 |
| SEXmale | -16.58 | 3.11 | -5.33 |
| R-squared | 0.01 | 0.00 | 2.69 |

$Austria

| Estimate | Std. Error | t value |
|----------|------------|---------|
| (Intercept) | 532.76 | 2.18 | 244.47 |
| SEXmale | -7.58 | 2.31 | -3.28 |
Or, alternatively the generic function `intsy.reg.pv` can be used. Estimates of the student’s sex coefficient and its significance indicate whether differences in performance are significant or not.

As before, regression results can be stored in an object for further analysis. We will run the previous regressions again adding one independent variable, `BSBGSMLM` in TIMSS, which is an index of how much students like learning mathematics, and `ASBHELA` in PIRLS which is the index of early literacy activities at home.

```R
R> timss_like <- timss.reg.pv(pvlabel = "BSMMAT", by = "IDCNTRYL", + x = c("SEX", "BSBGSMLM"), data = timss8g)
R> pirls_ela <- pirls.reg.pv(pvlabel = "ASRREA", by = "IDCNTRYL", + x = c("SEX", "ASBHELA"), data = pirls)
```

Regression output is stored in `timss_like` and `pirls_ela`. Each object contains a list with 4 elements, one for each education system, and each element contains subsequently a list with 5 elements, "replicates", "residuals", "var.w", "var.b", and "reg", which were defined before. For example, the following code retrieves replicate estimates of the `BSBGSMLM` coefficient in Armenia:

```R
R> timss_like[['Armenia']][['replicates']][["BSBGSMLM", ]
```

And replicate estimates in of `ASBHELA` in the PIRLS are
The distribution indicates variability due to sampling error and can be used in further analysis. Note that unlike the example above with PISA, it is not necessary to indicate the plausible value because TIMSS and PIRLS always use the first plausible value to calculate the sampling error.

Function `summary` can be used to print regression results without rounding output, for example:

```r
R> summary(timss_like)
```

### Armenia

|                | Estimate | Std. Error | t value  |
|----------------|----------|------------|----------|
| (Intercept)    | 311.1680384 | 10.28824804 | 30.244998 |
| SEXmale        | -5.5578132  | 3.01928392  | -1.840772  |
| BSBGSLM       | 14.8104129  | 0.88127636  | 16.805640  |

### Australia

|                | Estimate | Std. Error | t value  |
|----------------|----------|------------|----------|
| (Intercept)    | 360.6344877 | 10.51957182 | 34.2822402 |
| SEXmale        | 4.4935709  | 6.37453920  | 0.7049248  |
| BSBGSLM       | 15.2874963  | 1.08093043  | 14.1429049  |

### Bahrain

|                | Estimate | Std. Error | t value  |
|----------------|----------|------------|----------|
| (Intercept)    | 302.5794155 | 9.80668067 | 30.854417 |
| SEXmale        | -41.7903743 | 4.05984207 | -10.293596 |
| BSBGSLM       | 13.1924987  | 0.97558460  | 13.522660  |

### Chile

|                | Estimate | Std. Error | t value  |
|----------------|----------|------------|----------|
| (Intercept)    | 319.68963174 | 6.64649043 | 48.098987 |

```
A logistic regression with TIMSS data for performance above the international benchmark (i.e., cutoff = 550) is produced by:

```R
R> timss.log.pv(pvlabel = "BSMMAT", cutoff = 550, by = "IDCNTRYL", + x = c("SEX", "BSBGSMLM"), data = timss8g)
```

$Armenia

Coef. Std. Error t value OR CI95low CI95up
(Intercept) -5.64 0.41 -13.66 0.00 0.00 0.01
SEXmale 0.04 0.10 0.38 1.04 0.85 1.27
BSBGSMLM 0.36 0.04 10.35 1.44 1.34 1.54

$Australia

Coef. Std. Error t value OR CI95low CI95up
(Intercept) -4.56 0.31 -14.95 0.01 0.01 0.02
SEXmale 0.10 0.18 0.57 1.11 0.78 1.56
BSBGSMLM 0.38 0.03 11.66 1.46 1.37 1.55

$Bahrain

Coef. Std. Error t value OR CI95low CI95up
(Intercept) -5.33 0.43 -12.44 0.00 0.00 0.01
SEXmale -0.23 0.19 -1.20 0.79 0.54 1.16
BSBGSMLM 0.29 0.04 6.51 1.34 1.23 1.46

$Chile

Coef. Std. Error t value OR CI95low CI95up
(Intercept) -6.04 0.33 -18.24 0.00 0.00 0.00
SEXmale 0.15 0.22 0.70 1.17 0.76 1.79
BSBGSMLM 0.30 0.03 9.97 1.35 1.27 1.43

Using PIRLS data, the following code estimates a logistic regression of reading performance above the high international benchmark on the student’s sex and the index of early literacy activities.

```R
R> pirls.log.pv(pvlabel = "ASRREA", cutoff = 550, by = "IDCNTRYL", + x = c("SEX", "ASBHELA"), data = pirls)
```

$Australia

Coef. Std. Error t value OR CI95low CI95up
(Intercept) -1.88 0.32 -5.82 0.05 0.08 0.29
SEXmale -0.10 0.13 -0.75 0.47 0.71 1.17
ASBHELA 0.17 0.03 6.59 1.19 1.13 1.25
$\text{Austria}$

| Coef.  | Std. Error | t value | OR CI95low | CI95up  |
|--------|------------|---------|------------|---------|
| (Intercept) | -2.19 | 0.30 | -7.39 | 0.11 | 0.06 | 0.20 |
| SEXmale | -0.10 | 0.07 | -1.37 | 0.90 | 0.78 | 1.05 |
| ASBHELA | 0.18 | 0.03 | 6.75 | 1.20 | 1.14 | 1.27 |

$\text{Azerbaijan}$

| Coef.  | Std. Error | t value | OR CI95low | CI95up  |
|--------|------------|---------|------------|---------|
| (Intercept) | -2.78 | 0.56 | -4.97 | 0.06 | 0.02 | 0.19 |
| SEXmale | -0.37 | 0.17 | -2.24 | 0.69 | 0.50 | 0.96 |
| ASBHELA | 0.07 | 0.06 | 1.26 | 1.07 | 0.96 | 1.20 |

$\text{Belgium (French)}$

| Coef.  | Std. Error | t value | OR CI95low | CI95up  |
|--------|------------|---------|------------|---------|
| (Intercept) | -2.96 | 0.42 | -7.00 | 0.03 | 0.02 | 0.12 |
| SEXmale | 0.00 | 0.10 | -0.03 | 1.00 | 0.81 | 1.22 |
| ASBHELA | 0.20 | 0.04 | 4.77 | 1.22 | 1.12 | 1.32 |

Also, functions intsvy.log, pisa.reg, timss.reg, pirls.reg, and the generic intsvy.reg perform regression analysis for observed variables without plausible values.

4.5. Frequency tables

Functions pisa.table, piaac.table, timss.table, and pirls.table produce frequency tables including percentages and associated standard errors.

For example, the following code produces the frequency and percentage of students in each school grade level (i.e., variable = "ST01Q01") by education system in PISA 2012 (see OECD 2014a, p. 274).

```R
R> pisa.table(variable = "ST01Q01", by = "CNT", data = pisa)

    CNT ST01Q01 Freq Percentage Std.err.
  1  HKG     7  51  1.06 0.14
  2  HKG     8 300  6.47 0.41
  3  HKG    12 1205 25.94 0.72
  4  HKG   10 3088 65.01 0.91
  5  HKG    11  26  1.51 1.36
  6  PER     7 150  2.69 0.44
  7  PER     8 466  7.79 0.54
  8  PER    9 1056 18.10 0.67
  9  PER    10 2907 47.68 0.95
 10  PER   11 4566 23.74 0.82
 11  POL     7  20  0.53 0.13
 12  POL     8 158  4.08 0.37
 13  POL    9 4416 94.89 0.42
 14  POL    10  13  0.50 0.22
 15  SWE     7  1  0.03 0.03
```
intsvy: Analyzing International Large-Scale Assessment Data in R

With PIAAC data, the percentages of age groups by country can be calculated as follows:

```R
R> head(piaac.table(variable = "AGEG10LFS", by = "CNTRYID", data = piaac))
   CNTRYID AGEG10LFS Freq Percentage Std.err.
   1 Austria 24 or less 898 16.00 0.04
   2 Austria 25-34 958 19.11 0.06
   3 Austria 35-44 1117 22.18 0.07
   4 Austria 45-54 1188 23.83 0.07
   5 Austria 55 plus 969 18.89 0.04
   6 Belgium 24 or less 994 15.33 0.03
```

With TIMSS data, it is possible to calculate the percentage of students according to how much they like learning mathematics (1 = like learning mathematics; 2 = somewhat like learning mathematics; 3 = do not like learning mathematics) reported by own students (see Foy et al. 2013, p. 29):

```R
R> timss.table(variable = "BSDGSLM", by = "IDCNTRYL", data = timss8g)
   IDCNTRYL BSDGSLM Freq Percentage Std.err.
   1 Armenia 1 2421 42.92 0.97
   2 Armenia 2 2181 39.48 0.76
   3 Armenia 3 1024 17.60 0.97
   4 Australia 1 1068 15.67 0.94
   5 Australia 2 2985 39.81 1.41
   6 Australia 3 3336 44.53 1.41
   7 Bahrain 1 1072 23.75 0.64
   8 Bahrain 2 1756 38.37 0.86
   9 Bahrain 3 1753 37.88 0.84
  10 Chile 1 1289 22.06 0.86
  11 Chile 2 2291 40.21 0.89
  12 Chile 3 2192 37.73 0.97
```

And using school level data, we can calculate the percentage of students in schools classified by the socio-economic composition (1 = more affluent, 2 = neither more affluent nor more disadvantaged; 3 = more disadvantaged) reported by principals (see Foy et al. 2013, p. 36):

```R
R> timss.table(variable = "BCDG03", by = "IDCNTRYL", data = timss8g)
```
As before, the same tables can be produced with the generic `intsvy.table` function.

### 4.6. Performance benchmarks

Functions `pisa.ben.pv`, `timss.ben.pv`, and `pirls.ben.pv` calculate percentages of students in each proficiency level and associated standard errors. Proficiency levels are defined by PISA, TIMSS, and PIRLS studies and can be modified by the user.

For example, in PISA 2012 the percentage of students in each math proficiency level can be calculated as follows (see OECD 2014a, p. 298):

```r
R> pisa.ben.pv(pvlabel = "MATH", cutoff = c(357.77, 420.07, 482.38,
+ 544.68, 606.99, 669.30), by = "CNT", data = pisa)
```

| CNT     | Benchmarks | Percentage | Std. err. |
|---------|------------|------------|-----------|
| HKG <= 357.77 | 2.57 | 0.36 |
| HKG (357.77, 420.07] | 5.94 | 0.61 |
| HKG (420.07, 482.38] | 12.02 | 0.77 |
| HKG (482.38, 544.68] | 19.69 | 0.97 |
| HKG (544.68, 606.99] | 26.07 | 1.09 |
| HKG (606.99, 669.3] | 21.45 | 0.96 |
| HKG > 669.3 | 12.26 | 0.95 |
| PER <= 357.77 | 46.97 | 1.79 |
| PER (357.77, 420.07] | 27.61 | 0.88 |
| PER (420.07, 482.38] | 16.13 | 1.00 |
| PER (482.38, 544.68] | 6.66 | 0.68 |
| PER (544.68, 606.99] | 2.06 | 0.38 |
| PER (606.99, 669.3] | 0.55 | 0.20 |
| PER > 669.3 | 0.03 | 0.03 |
| POL <= 357.77 | 3.28 | 0.38 |
| POL (357.77, 420.07] | 11.10 | 0.77 |
| POL (420.07, 482.38] | 22.08 | 0.93 |
| POL (482.38, 544.68] | 25.46 | 0.94 |
| POL (544.68, 606.99] | 21.34 | 1.12 |
### **intsvy**: Analyzing International Large-Scale Assessment Data in R

|   | POL         | Benchmark | Percentage | Std. err. |
|---|-------------|-----------|------------|-----------|
| 20| (606.99, 669.3] |           | 11.74      | 0.78      |
| 21| > 669.3     |           | 5.00       | 0.80      |
| 22| <= 357.77   |           | 9.55       | 0.68      |
| 23| (357.77, 420.07] |       | 17.53      | 0.76      |
| 24| (420.07, 482.38] |       | 24.69      | 0.92      |
| 25| (482.38, 544.68] |       | 23.93      | 0.78      |
| 26| (544.68, 606.99] |       | 16.30      | 0.69      |
| 27| (606.99, 669.3] |           | 6.46       | 0.49      |
| 28| > 669.3     |           | 1.55       | 0.25      |
| 29| USA         |           |            |           |
| 30| (357.77, 420.07] |       | 7.96       | 0.73      |
| 31| (420.07, 482.38] |       | 17.89      | 0.98      |
| 32| (482.38, 544.68] |       | 26.25      | 0.84      |
| 33| (544.68, 606.99] |       | 23.34      | 0.93      |
| 34| (606.99, 669.3] |           | 15.79      | 0.91      |
| 35| > 669.3     |           | 6.58       | 0.61      |

The argument `cutoff` specifies proficiency levels for math performance in PISA 2012. These values are the default, can be omitted for 2012 data, and should be modified for data with different proficiency levels. The same output can be produced with

```r
R> intsvy.ben.pv(pvlabel = "MATH", by = "CNT", data = pisa,
+               config = pisa_conf)
```

Likewise, `intsrvy.ben.pv` calculates the percentage of students according to performance levels established by TIMSS and PIRLS. For example, for TIMSS 2011, Grade 8 (see Foy *et al.* 2013, p. 24):

```r
R> timss.ben.pv(pvlabel = "BSMMAT", by = "IDCNTRYL",
+               cutoff = c(400, 475, 550, 625), data = timss8g)
```

| IDCNTRYL                  | Benchmark | Percentage | Std. err. |
|---------------------------|-----------|------------|-----------|
| 1 Armenia At or above 400 |           | 76.38      | 1.16      |
| 2 Armenia At or above 475 |           | 49.02      | 1.37      |
| 3 Armenia At or above 550 |           | 17.65      | 0.88      |
| 4 Armenia At or above 625 |           | 3.23       | 0.40      |
| 5 Australia At or above 400 |       | 89.17      | 1.08      |
| 6 Australia At or above 475 |       | 62.94      | 2.40      |
| 7 Australia At or above 550 |       | 28.65      | 2.63      |
| 8 Australia At or above 625 |       | 8.68       | 1.68      |
| 9 Bahrain At or above 400  |           | 53.49      | 0.79      |
| 10 Bahrain At or above 475 |           | 26.19      | 0.65      |
| 11 Bahrain At or above 550 |           | 7.97       | 0.68      |
| 12 Bahrain At or above 625 |           | 1.26       | 0.25      |
| 13 Chile At or above 400   |           | 56.86      | 1.57      |
| 14 Chile At or above 475   |           | 22.95      | 1.11      |
| 15 Chile At or above 550   |           | 5.35       | 0.62      |
| 16 Chile At or above 625   |           | 0.56       | 0.16      |
And for PIRLS 2011 (see Foy and Drucker 2013, p. 24):

```r
R> pirls.ben.pv(pvlabel = "ASRREA", by = "IDCNTRYL", data = pirls)
```

| IDCNTRYL | Benchmark Percentage Std. err. |
|----------|--------------------------------|
| 1        | Australia At or above 400 92.93 0.67 |
| 2        | Australia At or above 475 75.62 1.03 |
| 3        | Australia At or above 550 41.91 1.14 |
| 4        | Australia At or above 625 9.93 0.65 |
| 5        | Austria At or above 400 97.10 0.35 |
| 6        | Austria At or above 475 80.38 0.94 |
| 7        | Austria At or above 550 39.05 1.50 |
| 8        | Austria At or above 625 5.22 0.54 |
| 9        | Azerbaijan At or above 400 81.86 1.60 |
| 10       | Azerbaijan At or above 475 45.16 2.10 |
| 11       | Azerbaijan At or above 550 8.94 0.93 |
| 12       | Azerbaijan At or above 625 0.44 0.28 |
| 13       | Belgium (French) At or above 400 93.79 1.08 |
| 14       | Belgium (French) At or above 475 70.39 1.67 |
| 15       | Belgium (French) At or above 550 25.50 1.39 |
| 16       | Belgium (French) At or above 625 2.25 0.49 |

As before, the argument `cutoff` can be omitted since these are the benchmark levels established by PIRLS 2011. For different benchmarks, the cut-off values can be modified. Also, more grouping levels for the analysis can be added with `by` and the same results can be reproduced with the generic `intsvy.ben.pv` function.

### 4.7. Calculating percentiles

Percentiles and associated standard errors can be calculated with study-specific functions `pisa.per.pv`, `pirls.per.pv`, `timss.per.pv` or with the generic function `intsvy.per.pv`. For example, the 10th, 25th, 75th, and 90th percentile in mathematics achievement can be calculated with (see OECD 2014a, p. 309):

```r
R> pisa.per.pv(pvlabel = "MATH", per = c(10, 25, 75, 90), by = "CNT", + data = pisa)
```

| CNT | Percentiles Score Std. err. |
|-----|-----------------------------|
| 1   | HKG 10 430.48 6.16          |
| 2   | HKG 25 498.84 4.69          |
| 3   | HKG 75 628.59 3.47          |
| 4   | HKG 90 679.44 4.20          |
| 5   | PER 10 264.04 3.38          |
| 6   | PER 25 310.55 3.61          |
| 7   | PER 75 421.14 4.90          |
| 8   | PER 90 477.75 6.74          |
| 9   | POL 10 401.80 2.77          |
Or, alternatively, the same table can be produced with:

```r
R> intsvy.per.pv(pvlabel = "MATH", per = c(10, 25, 75, 90),
+     by = "CNT", data = pisa, config = pisa_conf)
```

The following code calculates specific percentiles for reading achievement in PIRLS:

```r
R> pirls.per.pv(pvlabel = "ASRREA", per = c(5, 10, 25, 50, 75, 90, 95),
+     by = "IDCNTRYL", data = pirls)
```

And the following code calculates specific percentiles for mathematics achievement in TIMSS:

```R
R> timss.per.pv(pvlabel = "BSMMAT", per = c(5, 10, 25, 50, 75, 90, 95),
+     by = "IDCNTRYL", data = timss8g)
```

| IDCNTRYL | Percentiles | Score | Std. err. |
|----------|------------|-------|-----------|
| Armenia  | 5          | 310.38| 5.89      |
| Armenia  | 10         | 344.07| 4.55      |
| Armenia  | 25         | 404.74| 4.65      |
| Armenia  | 50         | 472.69| 3.32      |
| Armenia  | 75         | 530.60| 2.64      |
| Armenia  | 90         | 577.63| 3.91      |
| Armenia  | 95         | 607.55| 3.61      |
| Australia| 5          | 368.61| 4.82      |
| Australia| 10         | 396.72| 3.35      |
| Australia| 25         | 444.93| 5.09      |
| Australia| 50         | 502.64| 6.03      |
| Australia| 75         | 559.54| 6.98      |
| Australia| 90         | 617.77| 7.80      |
| Australia| 95         | 652.46| 12.02     |
| Bahrain  | 5          | 246.26| 6.02      |
| Bahrain  | 10         | 279.18| 5.70      |
| Bahrain  | 25         | 339.02| 3.35      |
| Bahrain  | 50         | 409.02| 2.09      |
| Bahrain  | 75         | 478.77| 1.99      |
| Bahrain  | 90         | 538.51| 3.55      |
| Bahrain  | 95         | 570.44| 4.00      |
| Chile    | 5          | 290.42| 8.12      |
| Chile    | 10         | 314.53| 3.60      |
| Chile    | 25         | 360.52| 3.09      |
| Chile    | 50         | 413.67| 4.02      |
| Chile    | 75         | 468.99| 4.09      |
| Chile    | 90         | 521.81| 4.16      |
| Chile    | 95         | 552.64| 4.34      |

As before, the same results can be reproduced with the `intsyv.per.pv` function.

### 4.8. Data visualization

The functions presented above allow to precisely estimate averages, frequencies or regression coefficients together with their standard errors. Since large tables filled with numbers could be difficult to understand at first sight, `intsyv` provides functions for data visualization that facilitate interpretation of results.
**intsyv**: Analyzing International Large-Scale Assessment Data in R

Figure 1: Graphical summary of a frequency table. This example presents the structure of age groups in the PIAAC dataset.

**Frequency tables**

The `plot` method for ‘`intsyv.table`’ objects produces a ggplot2 based barplot that summarizes frequency tables. Optional arguments for this `plot` method are `stacked` (should bars be stacked or not) and `se` (should standard error be plotted or not).

The following example calculates and plots two tables using the PIAAC dataset. The first is a plot of the age structure (see Figure 1) and the second a plot of the age structure by country and gender (see Figure 2).

```r
R> data("piaac", package = "PIAAC")
R> ptable <- piaac.table(variable = "AGEG10LFS", data = piaac)
R> plot(ptable)
R> ptableCA <- piaac.table(variable = "AGEG10LFS", + by = c("CNTRYID", "GENDER_R"), data = piaac)
R> plot(na.omit(ptableCA), stacked = TRUE)
```

It is common that items in surveys have ordered values in a Likert scale or a similar scale. In such cases a useful graphical summaries are floating barplots, i.e. barplots centered around specified value, usually the middle of scale. Such plots are available for `intsyv` frequency tables through argument `centered = TRUE`. For example, Figure 3 is generated with following code.

```r
R> ptableC <- piaac.table(variable = "AGEG10LFS", by = "CNTRYID", + data = piaac)
R> plot(na.omit(ptableC), centered = TRUE)
```

The following commands produce plots of parental education levels in PIRLS (see Figure 4) and the percentage of students who like learning mathematics in TIMSS (see Figure 5) using the `plot` function in combination with `intsyv.table`.

```r
R> pirls$PARED <- factor(pirls$ASDHEDUP, levels = 1:5, labels = + c("university or higher", "post-secondary", "upper secondary", + "lower secondary", "some primary or no school"))
```
Figure 2: Graphical summary of a frequency table with three grouping variables, age groups by country and gender.

Figure 3: Graphical summary of a frequency table for two variables, country and age groups. Setting the argument `center = TRUE` results in floating barplots centered around the middle value.
**Figure 4:** Graphical summary of a frequency table. This example presents the parental education levels in the PIRLS dataset.

**Figure 5:** Graphical summary of a frequency table with grouping variable. This example presents the percentage of students who like learning mathematics by country in the TIMSS Grade 8 dataset.

```r
R> plot(pirls.table(variable = "PARED", data = pirls))
R> timss8g$LIKE <- factor(timss8g$BSDGSLM, labels =
+ c("like learning mathematics", "somewhat like learning mathematics",
+ "do not like learning mathematics"))
R> plot(intsvy.table(variable = "LIKE", by = "IDCNTRYL", data = timss8g,
+ config = timss8_conf))
```
Average achievement scores

Functions `intsuy.mean.pv` and `intsuy.mean`, as well as associated study-specific functions (e.g., `pisa.mean.pv`, `timss.mean`), produce objects of the class ‘`intsuy.mean`’. The associated `plot` method produces a `ggplot2` based dotplot that resents calculated averages and their standard errors.

Optional arguments for the `plot` method are `sort` (should groups be sorted along the average or not) and `se` (should standard error be plotted or not).

The following example calculates and plots average numeracy performance by country (see Figure 6) and by country and age group (see Figure 7) based on the PIAAC dataset.

```r
R> pmeansNC <- piaac.mean.pv(pvlabel = "NUM", by = "CNTRYID", + data = piaac, export = FALSE)
R> plot(pmeansNC, sort = TRUE)
R> pmeansNCA <- piaac.mean.pv(pvlabel = "NUM", + by = c("CNTRYID", "AGEG10LFS"), data = piaac, export = FALSE)
R> plot(pmeansNCA, sort = TRUE)
```

The following code produces two plots. Figure 8 shows average mathematics scores in PISA by education system and gender and Figure 9 displays average mathematics scores in TIMSS by education system and the extent to which students like mathematics.

```r
R> plot(pisa.mean.pv(pvlabel = "MATH", by = c("CNT", "SEX"), + data = pisa))
```
**intsvis**: Analyzing International Large-Scale Assessment Data in R

Figure 7: Graphical summary of averages in groups and their standard errors. This example presents average numeracy scores and their standard errors for different countries and age groups based on the PIAAC dataset.

Figure 8: Graphical summary of averages and their standard errors. This example presents average mathematics scores and their standard errors by education system and gender based on the PISA dataset.
Figure 9: Graphical summary of averages in groups and their standard errors. This example presents average mathematics scores and their standard errors by education system and the extent to which students like mathematics based on the TIMSS Grade dataset.

R> plot(na.omit(timss.mean.pv(pvlabel = "LIKE", + by = c("IDCNTRYL", "BSDGSLM"), data = timss8g)))

Regression analysis

Functions intsvy.reg.pv and intsvy.reg produce objects of the class ‘intsvy.reg’. The associated plot method produces a ggplot2 based dotplot that summarizes regression based model coefficients and their standard errors.

Optional arguments for the plot method are sort (should groups be sorted along the average or not) and se (should standard error be plotted or not).

The following example calculates and plots regression coefficients based on the PIAAC dataset (see Figure 10).

R> rmodelLG <- piaac.reg.pv(pvlabel = "LIT", x = "GENDER_R", + by = "CNTRYID", data = piaac, export = FALSE)
R> plot(rmodelLG, vars = c("GENDER_RFemale"), se = TRUE, sort = TRUE)

The following code produces a plot with results of a regression of mathematics scores on gender (SEX) and the economic, social, and cultural status index (ESCS) (see Figure 11).

R> pisa_ses <- pisa.reg.pv(pvlabel = "MATH", x = c("SEX", "ESCS"), + by = "CNT", data = pisa)
R> plot(pisa_ses, vars = c("SEXmale", "ESCS"))

The following code plots regression results with TIMSS Grade 8 data (see Figure 12). A single variable is selected in the plot command, the index of students liking mathematics, BSBGSLM.

R> timss_like <- timss.reg.pv(pvlabel = "BSMMAT", by = "IDCNTRYL", + x = c("SEX", "BSBGSLM"), data = timss8g)
R> plot(timss_like, vars = "BSBGSLM")
Figure 10: Graphical summary of regression models. This example presents outcomes for regression models with literacy scores as dependent variable and gender as independent variable.

Figure 11: Graphical summary of regression models. This example presents outcomes for regression models with mathematics scores as dependent variable and gender (SEX) and the economic, social, and cultural status index (ESCS) as independent variables based on the PISA dataset.
Figure 12: Graphical summary of regression models. This example shows the coefficient of the index of students liking mathematics (BSBGSLM) in a regression of mathematics scores on gender (ITSEX) and BSBGSLM based on the TIMSS Grade 8 dataset.

Figure 13: Graphical summary of regression models. This example presents outcomes for regression models with reading scores as dependent variable and gender (SEX) and the index of early literacy activities at home (ASBHELA) as independent variables based on the PIRLS dataset.

Finally, a example is presented with PIRLS data (see Figure 13):

```R
R> pirls_ela <- pirls.reg.pv(pvlabel = "ASRREA", by = "IDCNTRYL", + x = c("SEX", "ASBHELA"), data = pirls)
R> plot(pirls_ela, vars = c("SEXmale", "ASBHELA"))
```
5. Summary

This article introduced \texttt{intsy} and demonstrated its use with data from PISA, PIRLS, TIMSS, ICILS, and PIAAC. Package \texttt{intsy} provides another alternative within \texttt{R} to soundly handle data from international LSA. In addition to analysis and visualization tools, the package includes functions for merging and importing data, which are particularly handy for TIMSS and PIRLS. The package can be extended to handle datasets from different international assessment studies. There are several limitations and plans for incorporating new features in future releases of this package. Currently \texttt{intsy} handles missing data using listwise deletion, cannot analyze trend data from international LSA, cannot perform tests of statistical significance beyond those provided by regressions, to mention some limitations.

References

Avvisati F, Keslair F (2014). \textit{“REPEST: Stata Module to Run Estimations with Weighted Replicate Samples and Plausible Values.”} Statistical Software Components, Boston College Department of Economics. URL \url{https://ideas.repec.org/c/boc/bocode/s457918.html}.

BIFIE (2017). \textit{BIFIESurvey: Tools for Survey Statistics in Educational Assessment}. \texttt{R} package version 2.1-6, URL \url{https://CRAN.R-project.org/package=BIFIESurvey}.

Bryer JM, Pruzek RM (2011). “Abstract: An International Comparison of Private and Public Schools Using Multilevel Propensity Score Methods and Graphics.” \textit{Multivariate Behavioral Research}, \textbf{46}(6), 1010–1011. \doi{10.1080/00273171.2011.636693}.

Caro DH, Biecek P (2017). \textit{intsy: International Assessment Data Manager}. \texttt{R} package version 2.1, URL \url{https://CRAN.R-project.org/package=intsy}.

Damico A (2015). “Analyze Survey Data for Free.” URL \url{http://www.asdfree.com}.

Foy P, Arora A, Stanco GM (2013). \textit{TIMSS 2011 User Guide for the International Database}. TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA), Chestnut Hill.

Foy P, Drucker KT (2013). \textit{PIRLS 2011 User Guide for the International Database}. TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA), Chestnut Hill.

IBM Corporation (2015). \textit{IBM SPSS Statistics 23}. IBM Corporation, Armonk. URL \url{http://www.ibm.com/software/analytics/spss/}.

IEA Hamburg (2017). \textit{IDB Analyzer: IEA International Database (IDB) Analyzer}. URL \url{http://www.iea.nl/eula.html}.

Little RJA, Rubin D (2002). \textit{Statistical Analysis with Missing Data}. 2nd edition. John Wiley & Sons, New York. \doi{10.1002/9781119013563}. 
Lumley T (2004). “Analysis of Complex Survey Samples.” *Journal of Statistical Software*, 9(8), 1–19. doi:10.18637/jss.v009.i08.

Macdonald K (2008). “PV: Stata Module to Perform Estimation with Plausible Values.” Statistical Software Components, Boston College Department of Economics. URL https://ideas.repec.org/c/boc/bocode/s456951.html.

Muthén LK, Muthén BO (1998–2017). *Mplus* User’s Guide. 8th edition. Muthén & Muthén, Los Angeles.

National Center for Education Statistics (2017). *International Data Explorer*. URL https://nces.ed.gov/surveys/international/ide/.

Oberski D (2014). “lavaan.survey: An R Package for Complex Survey Analysis of Structural Equation Models.” *Journal of Statistical Software*, 57(1), 1–27. doi:10.18637/jss.v057.i01.

OECD (2009). *PISA Data Analysis Manual. SPSS*. 2nd edition. OECD Publishing, Paris. doi:10.1787/9789264056275-en.

OECD (2013a). *PISA 2012 Results: Excellence Through Equity: Giving Every Student the Chance to Succeed*, volume II. OECD Publishing, Paris.

OECD (2013b). *Technical Report of the Survey of Adult Skills (PIAAC)*. OECD Publishing, Paris.

OECD (2014a). *PISA 2012 Results: What Students Know and Can Do: Student Performance in Mathematics, Reading and Science*, volume I. Revised edition. OECD Publishing, Paris.

OECD (2014b). *PISA 2012 Technical Report*. OECD Publishing, Paris.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

SAS Institute Inc (2013). *The SAS System, Version 9.4*. SAS Institute Inc., Cary. URL http://www.sas.com/.

StataCorp (2015). *Stata Data Analysis Statistical Software: Release 14*. StataCorp LP, College Station. URL http://www.stata.com/.

Ünlü A, Sargin A (2010). “DAKS: An R Package for Data Analysis Methods in Knowledge Space Theory.” *Journal of Statistical Software*, 37(2), 1–31. doi:10.18637/jss.v037.i02.

von Davier M, Gonzalez E, Mislevy RJ (2009). “What Are Plausible Values and Why Are They Useful?” In *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, volume 2, pp. 9–36. IEA-ETS Research Institute.

Wickham H (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York. doi:10.1007/978-0-387-98141-3.
Affiliation:
Daniel H. Caro
Oxford University Centre for Educational Assessment
University of Oxford
15 Norham Gardens, OX2 6PY Oxford, United Kingdom
E-mail: daniel.caro@education.ox.ac.uk

Przemyslaw Biecek
Faculty of Mathematics and Information Science
Warsaw University of Technology
Koszykowa 75, 00-662 Warsaw, Poland
E-mail: Przemyslaw.Biecek@gmail.com