Description for the PISA 2012 data included in the package pairwise

Joerg-Henrik Heine

26.03.2014

1 Loading data

First install and load the package 'pairwise'.

```
> library(pairwise)
```

Once you have installed the package 'pairwise' the selected PISA 2012 Data for the German sub sample can be loaded by simply typing the following command into the R console.

```
> data(DEU_PISA2012)
```

2 Structure of the data

The data is organized as an multiple nested list.

The respective name of every list entry at any level should give an idea of its content. For a first overview look at the names of the first levels of the list object.

```
> names(DEU_PISA2012)
```

```
[1] "id"        "covariate" "cog"       "ncog"      "weights"
```

On the first list level of the object `DEU_PISA2012` there are 5 list entries. For a first closer inspection of the content check the object class for every list entry

```
> lapply(DEU_PISA2012,class)
```

```
$id
[1] "data.frame"

$covariate
[1] "data.frame"
```

```
$cog
[1] "list"

$ncog
[1] "list"

$weights
[1] "data.frame"
```

Three of the five list entries at first level are data frames, which hold the following kind of data. ID-Variables in the first list entry, additional variables like gender in the second list entry and case weights and replicate weights in the last list entry.

2.1 ID variables

```
> names(DEU_PISA2012$id)

[1] "INT_MACH_ID" "SCHOOLID"    "StIDStd"     "BOOKID"      "QuestID"
```

- the variable INT_MACH_ID is an unique identifier for every case in regard to the international data set – which is not included in this package.
- the variable SCHOOLID groups every case into the school level when drawing the sample using school lists for the German sub sample.
- the variable StIDStd is an unique identifier for every case in regard to the international German sub sample.
- the variable BOOKID indicates which booklet each participant from the German sub sample was assigned to. The following command returns the frequencies for the booklet variable:

  ```
  > table(DEU_PISA2012$id$BOOKID)

    1   2   3   4   5   6   7   8   9  10  11  12  13  20
  379 385 353 351 374 361 386 390 387 367 377 370 382 139
  ```

  There are 14 booklets. Each booklet from "1" to "13" is part of the rotated design for the assessment of the three competencies (math, reading, science). Booklet "20" is the so called UH-booklet, containing items from the regular item set but is a shortened version of the regular booklets.

- the variable QuestID indicates which of the rotated questionnaire-booklets, assessing the non cognitive constructs, was assigned to each participant from the German sub sample.

  ```
  > table(DEU_PISA2012$id$QuestID)
  ```

```
   1    2    3    5
1627 1654 1581  139
```

There are 4 different questionnaire-booklets.
Questionaire-booklets "1" to "3" are regular versions and "5" was a short version (UH-version).

2.2 Additional variables (covariate)

```
> names(DEU_PISA2012$covariate)
```

```
[1] "EASY"   "GRADE"  "PROGN"  "REPEAT" "AGE"    "SEX"
```

- EASY indicates whether an easier booklet, containing only easy items from the regular item set, was given to the participant. As there was no easier booklet assigned to any participant in the international German sub sample, this variable is rather a constant.
- GRADE indicates the relative grade of each participant, relating to the target population.
- PROGN indicates national study program, which is 'Schulart' in German language.
- REPEAT indicates whether an participant repeated a class (at any grade).
- AGE the age of the participant.
- SEX the gender.

2.3 Cognitive variables (PISA competencies)

```
> names(DEU_PISA2012$cog)
```

```
[1] "pv"  "dat"
```

This list level contains the plausible values `pv`, drawn from the international scalig procedure as well as the scored responses `dat` of the participants answering to the questions coresponding to the three PISA competencies (**math**, **reading** and **science**).

2.3.1 Plausible values (pv)

```
> names(DEU_PISA2012$cog$pv)
```

```
[1] "MATH" "READ" "SCIE"
```

for any of the three PISA competencies 5 plausible values were drawn, which are stored as a list with length 5.

```
> names(DEU_PISA2012$cog$pv$MATH)

[1] "PV1MATH" "PV2MATH" "PV3MATH" "PV4MATH" "PV5MATH"

> names(DEU_PISA2012$cog$pv$READ)

[1] "PV1READ" "PV2READ" "PV3READ" "PV4READ" "PV5READ"

> names(DEU_PISA2012$cog$pv$SCIE)

[1] "PV1SCIE" "PV2SCIE" "PV3SCIE" "PV4SCIE" "PV5SCIE"
```

Any of the respective list entries covers all participants in the German sub sample and therefor has a lenght of 5001 (cases) – e.g.:

```
> length(DEU_PISA2012$cog$pv$MATH$PV1MATH)

[1] 5001
```

2.3.2 Scored data and missing incidenz matrices (dat)

```
> names(DEU_PISA2012$cog$dat)

[1] "MATH" "READ" "SCIE"
```

The list level `DEU_PISA2012$cog$dat` contains the scored data and missing incidence matrices for any of the three PISA competencies.

The complete (sub-) structure of the list level `DEU_PISA2012$cog$dat` is returned by the following R-command.

```
> rapply(DEU_PISA2012$cog$dat,names,classes = "list",how="list")

$MATH
$MATH$resp
NULL

$MATH$inc7
NULL

$MATH$inc8
NULL


$READ
$READ$resp
NULL

$READ$inc7
NULL
```

```
$READ$inc8
NULL


$SCIE
$SCIE$resp
NULL

$SCIE$inc7
NULL

$SCIE$inc8
NULL
```

As an example we demonstrate the structure using the `MATH` domain:

```
> names(DEU_PISA2012$cog$dat$MATH)

[1] "resp" "inc7" "inc8"
```

For any of the three competence domains there are three list entries at that list level, which contain the scored responses (`resp`), an icidence matrix for missing by design (`inc7`) and an icidence matrix for not reached items (`inc8`). All of the three entries are stored as matrices, which have the same dimensionality.

```
> dim(DEU_PISA2012$cog$dat$MATH$resp)

[1] 5001   84

> dim(DEU_PISA2012$cog$dat$MATH$inc7)

[1] 5001   84

> dim(DEU_PISA2012$cog$dat$MATH$inc8)

[1] 5001   84
```

The values in `DEU_PISA2012$cog$dat$MATH$resp` range from 0 to 2 – with `NA` values removed. These vlues in `DEU_PISA2012$cog$dat$MATH$resp` represent the scored responses of the participants following an partial-credit-scoring approach – with the folowing meaning:

- 0 = wrong
- 1 = correct / or partial correct (*if there is a third category for the respective item*)
- 2 = correct (*if there is a third category for the respective item*)

The `NA` values in the `resp` matrix have two meanings: either missing by (rotated) design or not reached. Either meaning of an respective `NA` value is coded in the two incidence matices named `inc7` and `inc8` .

2.4 "Non cognitive" variables

```
> names(DEU_PISA2012$ncog)
```

```
[1] "wle" "dat"
```

this list level contains the scales and item responses of the "non-cognitive" constructs, assessed in the student-questionnaire.

2.4.1 Weighted likelihood estimates for constructs (WLE)

```
> names(DEU_PISA2012$ncog$wle)
```

```
 [1] "TEACHSUP" "COGACT"   "DISCLIMA" "MTSUP"    "CLSMAN"   "STUDREL"
 [7] "FAMCON"   "INTMAT"   "MATHEFF"  "SCMAT"    "ANXMAT"   "SUBNORM"
[13] "FAILMAT"  "MATINTFC" "MATWKETH" "MATBEH"
```

The level `DEU_PISA2012$ncog$wle` contains the weighted likelihood estimates (WLE) for a selection of 16 constructs out of 52 constructs beeing assesed in PISA 2012. The selection of the 16 constructs is more or less arbitrary, following the personal intresst of the author of the package 'pairwise'. The WLE-estimates are based on the responses on the respective items in the student-questionnaire, as a result of the international scaling procedure.

2.4.2 Item responses for "non-cognitive" constructs (dat)

```
> names(DEU_PISA2012$ncog$dat)
```

```
 [1] "TEACHSUP" "COGACT"   "DISCLIMA" "MTSUP"    "CLSMAN"   "STUDREL"
 [7] "FAMCON"   "INTMAT"   "MATHEFF"  "SCMAT"    "ANXMAT"   "SUBNORM"
[13] "FAILMAT"  "MATINTFC" "MATWKETH" "MATBEH"
```

For any of the 16 "non-cognitive" constructs there are four list entries at that list level.

As an example we demontstrate the further structure using the "non-cognitive" construct `CLSMAN`:

```
> names(DEU_PISA2012$ncog$dat$CLSMAN)
```

```
[1] "resp" "inc7" "inc8" "inc9"
```

These 4 matrices contain the scored responses `resp` , an icidence matrix for missing (by design) `inc7`, an icidence matrix for invalid responses `inc8` and an icidence matrix for missing (by testee) responses `inc9`. All of the three entries are stored as matrices, which have the same dimensionality.

```
> lapply(DEU_PISA2012$ncog$dat$CLSMAN,dim)
```

```
$resp
[1] 5001    4

$inc7
[1] 5001    4

$inc8
[1] 5001    4

$inc9
[1] 5001    4
```

For any of the 16 "non-cognitive" constructs the list level entry `resp` is a matrix with named columns giving the international PISA item names.

```
> colnames(DEU_PISA2012$ncog$dat$CLSMAN$resp)

[1] "ST85Q01" "ST85Q02" "ST85Q03" "ST85Q04"
```

2.5 Case weights and replicate weights

The last list level `DEU_PISA2012$weights` finaly contains the FINAL STUDENT WEIGHT (`W_FSTUWT`) and the FINAL STUDENT REPLICATE BRR-FAY WEIGHTS (`W_FSTR1` to `W_FSTR80`) and few other weighting variables used in the international analysis procedure.

```
> names(DEU_PISA2012$weights)

 [1] "W_FSTUWT"   "W_FSTR1"    "W_FSTR2"    "W_FSTR3"    "W_FSTR4"
 [6] "W_FSTR5"    "W_FSTR6"    "W_FSTR7"    "W_FSTR8"    "W_FSTR9"
[11] "W_FSTR10"   "W_FSTR11"   "W_FSTR12"   "W_FSTR13"   "W_FSTR14"
[16] "W_FSTR15"   "W_FSTR16"   "W_FSTR17"   "W_FSTR18"   "W_FSTR19"
[21] "W_FSTR20"   "W_FSTR21"   "W_FSTR22"   "W_FSTR23"   "W_FSTR24"
[26] "W_FSTR25"   "W_FSTR26"   "W_FSTR27"   "W_FSTR28"   "W_FSTR29"
[31] "W_FSTR30"   "W_FSTR31"   "W_FSTR32"   "W_FSTR33"   "W_FSTR34"
[36] "W_FSTR35"   "W_FSTR36"   "W_FSTR37"   "W_FSTR38"   "W_FSTR39"
[41] "W_FSTR40"   "W_FSTR41"   "W_FSTR42"   "W_FSTR43"   "W_FSTR44"
[46] "W_FSTR45"   "W_FSTR46"   "W_FSTR47"   "W_FSTR48"   "W_FSTR49"
[51] "W_FSTR50"   "W_FSTR51"   "W_FSTR52"   "W_FSTR53"   "W_FSTR54"
[56] "W_FSTR55"   "W_FSTR56"   "W_FSTR57"   "W_FSTR58"   "W_FSTR59"
[61] "W_FSTR60"   "W_FSTR61"   "W_FSTR62"   "W_FSTR63"   "W_FSTR64"
[66] "W_FSTR65"   "W_FSTR66"   "W_FSTR67"   "W_FSTR68"   "W_FSTR69"
[71] "W_FSTR70"   "W_FSTR71"   "W_FSTR72"   "W_FSTR73"   "W_FSTR74"
[76] "W_FSTR75"   "W_FSTR76"   "W_FSTR77"   "W_FSTR78"   "W_FSTR79"
[81] "W_FSTR80"   "WVARSTRR"   "VAR_UNIT"   "senwgt_STU"
```

3 Analysis with the German PISA 2012 data set

Compute the percentge of missing by design:
for mathmatic competenc assessment

```
> (sum(DEU_PISA2012$cog$dat$MATH$inc7))/(prod(dim(DEU_PISA2012$cog$dat$MATH$inc7)))*100
[1] 70.03314
```

for reading competenc assessment

```
> (sum(DEU_PISA2012$cog$dat$READ$inc7))/(prod(dim(DEU_PISA2012$cog$dat$READ$inc7)))*100
[1] 70.43182
```

for science competenc assessment

```
> (sum(DEU_PISA2012$cog$dat$SCIE$inc7))/(prod(dim(DEU_PISA2012$cog$dat$SCIE$inc7)))*100
[1] 70.23652
```