Package 'timss8reg'

November 18, 2026

Type Package

Title Helper Functions for Regression Analysis of TIMSS Grade 8 Student Data

Version 0.1.0

Description Tools for reading Trends in International Mathematics and Science Study (TIMSS) international database files stored in R format <https://timss2023.org/data/>, merging the student questionnaire and student achievement files, and preparing questionnaire variables (missing-value codes, reversed agreement scales and yes/no items). The package also fits a multiple linear regression of the average of the five plausible values on selected predictors, reporting standardized coefficients, correlations and variance inflation factors, and provides heteroscedasticity diagnostics (Breusch-Pagan test and HC3 robust standard errors). Sampling weights and replicate weights are not applied; for design-based estimation see the packages 'intsvy' and 'RALSA'.

License GPL-3

Encoding UTF-8

Language en

LazyData true

Depends R (>= 4.0.0)

Imports lmtest, sandwich, stats

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

RoxygenNote 7.3.1

NeedsCompilation no

Author Elfin Harse Saputra Harefa [aut, cre], Fadhlul Mubarak [aut]

Maintainer Elfin Harse Saputra Harefa <elfinharsesaputraharefaa@gmail.com>

Repository CRAN

Date/Publication 2026-11-18 10:00:02 UTC

Contents

timss_baca . . . 2
timss_balik_skala . . . 3
timss_contoh . . . 4
timss_deskriptif . . . 4
timss_diagnostik . . . 5
timss_gabung . . . 6
timss_kode_kosong . . . 6
timss_nilai_rata . . . 7
timss_pilih_variabel . . . 8
timss_regresi . . . 8
timss_ya_tidak . . . 9

Index **10**

`timss_baca` *Read a TIMSS data file in R format*

Description

Reads one TIMSS International Database file stored in R format (`.RData`), converts all variable names to upper case and removes the SPSS value labels (class `haven_labelled`) that otherwise stop functions such as `merge` from working.

Usage

```
timss_baca(folder, prefix, negara, siklus = "m8")
```

Arguments

`folder`	Folder that contains the TIMSS `.RData` files.
`prefix`	File prefix, for example `"bsg"` (student questionnaire) or `"bsa"` (student achievement).
`negara`	Three-letter country code used in the file name, for example `"bhr"`.
`siklus`	Cycle suffix of the file name. Default `"m8"` (TIMSS 2023).

Value

A data frame with upper-case variable names.

Examples

```
dir_tmp <- tempfile("timss")
dir.create(dir_tmp)
bsgxxxm8 <- timss_contoh
save(bsgxxxm8, file = file.path(dir_tmp, "bsgxxxm8.RData"))
d <- timss_baca(dir_tmp, "bsg", "xxx")
dim(d)
unlink(dir_tmp, recursive = TRUE)
```

---

timss_balik_skala *Reverse the direction of a response scale*

Description

TIMSS codes agreement items as 1 = agree a lot to 4 = disagree a lot and frequency items in a similar decreasing order. This function reverses the scale (new value = `maks + 1 - value`) so that a larger number means more agreement or higher frequency.

Usage

```
timss_balik_skala(data, variabel, maks)
```

Arguments

`data`	Data frame.
`variabel`	Character vector of numeric variables to reverse.
`maks`	Highest code of the scale, for example 4 for a four-point scale.

Value

`data` with the selected variables reversed.

Examples

```
d <- timss_balik_skala(timss_contoh, c("BSBS28A", "BSBS28C"), maks = 4)
head(cbind(timss_contoh$BSBS28A, d$BSBS28A))
```

timss_contoh *Simulated example data in TIMSS style*

Description

A small simulated data set (300 students) that mimics the structure of the TIMSS grade 8 student files. It is generated by random numbers, contains no real TIMSS data and is used only in the examples and tests of this package. Codes `9` in `BSBG04` and `BSBG07` represent omitted answers.

Usage

```
timss_contoh
```

Format

A data frame with 300 rows and 14 variables:

IDCNTRY Country code (999 for simulated).

IDSTUD Student identifier.

BSBG01 Student sex code (1 or 2).

BSBG04 Number of books at home (1 to 5, 9 = omitted).

BSBG05A Own computer (1 = yes, 2 = no).

BSBG07 Expected education level (1 to 6, 9 = omitted).

BSBS28A Agreement: I usually do well in science (1 = agree a lot to 4 = disagree a lot).

BSBS28C Agreement: Science is not one of my strengths (1 to 4).

BSBS29I Agreement: It is important to do well in science (1 to 4).

BSSSCI01, BSSSCI02, BSSSCI03, BSSSCI04, BSSSCI05 Five simulated plausible values for science.

Source

Simulated; see `data-raw/timss_contoh.R` in the package source.

timss_deskriptif *Descriptive statistics table*

Description

Computes the number of valid values, mean, standard deviation, minimum, maximum and percentage of missing values of numeric variables.

Usage

```
timss_deskriptif(data, variabel = names(data))
```

Arguments

`data`	Data frame.
`variabel`	Character vector of variable names. Non-numeric variables are skipped.

Value

A data frame with one row per numeric variable.

Examples

```
timss_deskriptif(timss_contoh, c("BSBG04", "BSBS28A"))
```

`timss_diagnostik` *Heteroscedasticity diagnostics for a regression model*

Description

Runs the studentized Breusch-Pagan test and computes coefficient tests with HC3 heteroscedasticity-consistent standard errors.

Usage

```
timss_diagnostik(model)
```

Arguments

`model`	An object of class `lm`, for example the `model` element returned by `timss_regresi`.

Value

A list with `bp` (a named vector with the test statistic, degrees of freedom and p-value) and `robust` (a coefficient table with HC3 standard errors).

Examples

```
d <- timss_nilai_rata(timss_contoh)
d <- timss_kode_kosong(d, c("BSBG04", "BSBG07"))
hasil <- timss_regresi(d, "SAINS", c("BSBG04", "BSBG07", "BSBS28A"))
timss_diagnostik(hasil$model)$bp
```

timss_gabung *Merge the student questionnaire and achievement files*

Description

Merges the student questionnaire data (bsg file) and the student achievement data (bsa file) by country code and student identifier. Only columns of the achievement file that are not in the questionnaire file are added.

Usage

```
timss_gabung(siswa, nilai, kunci = c("IDCNTRY", "IDSTUD"))
```

Arguments

siswa	Data frame of the student questionnaire.
nilai	Data frame of the student achievement file.
kunci	Character vector of key variables present in both data frames.

Value

A data frame with one row per student present in both files.

Examples

```
siswa <- timss_contoh[, c("IDCNTRY", "IDSTUD", "BSBG04")]
nilai <- timss_contoh[, c("IDCNTRY", "IDSTUD", "BSSSCI01")]
dim(timss_gabung(siswa, nilai))
```

timss_kode_kosong *Convert missing-value codes to NA*

Description

Converts the selected variables to numeric and replaces the codes that TIMSS uses for omitted or invalid answers (for example 9, 99, 999) by NA.

Usage

```
timss_kode_kosong(data, variabel, kode = c(9, 99, 999, 9999))
```

Arguments

data	Data frame.
variabel	Character vector of variable names to convert.
kode	Numeric vector of codes to be treated as missing.

Value

`data` with the selected variables converted.

Examples

```
d <- timss_kode_kosong(timss_contoh, c("BSBG04", "BSBG07"))
sum(is.na(d$BSBG04))
```

`timss_nilai_rata` *Average of the plausible values*

Description

Adds a column with the row mean of the plausible values of one subject. Averaging is a simplification; a full analysis uses every plausible value separately and combines the results.

Usage

```
timss_nilai_rata(data, awalan = "BSSSCI", jumlah = 5L, nama = "SAINS")
```

Arguments

`data`	Data frame that contains the plausible value columns.
`awalan`	Prefix of the plausible value columns, for example `"BSSSCI"` (science) or `"BSMMAT"` (mathematics).
`jumlah`	Number of plausible values. Default 5.
`nama`	Name of the new column.

Value

`data` with one extra column.

Examples

```
d <- timss_nilai_rata(timss_contoh)
summary(d$SAINS)
```

timss_pilih_variabel *Select usable predictor variables*

Description

Keeps the variables whose share of missing values is below a threshold and that have more than one distinct value.

Usage

```
timss_pilih_variabel(data, variabel, batas_na = 0.25)
```

Arguments

data	Data frame.
variabel	Character vector of candidate variable names.
batas_na	Largest allowed proportion of missing values (exclusive).

Value

A list with elements `terpilih` (kept variables), `dibuang` (dropped variables) and `proporsi_na` (proportion of missing values of every candidate found in `data`).

Examples

```
d <- timss_kode_kosong(timss_contoh, c("BSBG04", "BSBG07"))
timss_pilih_variabel(d, c("BSBG04", "BSBG07", "BSBS28A"))$terpilih
```

timss_regresi *Multiple linear regression with standardized coefficients and VIF*

Description

Fits an ordinary least squares regression after removing rows with missing values, and returns the model together with standardized coefficients, correlations of every predictor with the response and variance inflation factors. Sampling weights are not used.

Usage

```
timss_regresi(data, y = "SAINS", x)
```

Arguments

data	Data frame.
y	Name of the response variable.
x	Character vector with at least two predictor names.

Value

A list with elements `model` (an `lm` object), `data` (the complete cases used), `n` (number of rows used), `beta_terstandar`, `korelasi` and `vif`.

Examples

```
d <- timss_nilai_rata(timss_contoh)
d <- timss_kode_kosong(d, c("BSBG04", "BSBG07"))
hasil <- timss_regresi(d, "SAINS", c("BSBG04", "BSBG07", "BSBS28A"))
summary(hasil$model)
hasil$vif
```

`timss_ya_tidak` *Recode yes/no items to 1/0*

Description

TIMSS codes yes/no items as 1 = yes and 2 = no. This function recodes them to 1 = yes and 0 = no, which is convenient for regression.

Usage

```
timss_ya_tidak(data, variabel)
```

Arguments

`data`	Data frame.
`variabel`	Character vector of numeric variables coded 1 and 2.

Value

`data` with the selected variables recoded.

Examples

```
d <- timss_ya_tidak(timss_contoh, "BSBG05A")
table(d$BSBG05A)
```

Index

∗ **datasets**
    timss_contoh, 4

merge, *2*

timss_baca, 2
timss_balik_skala, 3
timss_contoh, 4
timss_deskriptif, 4
timss_diagnostik, 5
timss_gabung, 6
timss_kode_kosong, 6
timss_nilai_rata, 7
timss_pilih_variabel, 8
timss_regresi, *5*, 8
timss_ya_tidak, 9