

Package ‘simputation’

July 27, 2016

Maintainer Mark van der Loo <mark.vanderloo@gmail.com>

License GPL-3

Title Simple Imputation

LazyData no

Type Package

LazyLoad yes

Description

Easy to use interfaces to a number of imputation methods that fit in the not-a-pipe operator of the 'magrittr' package.

Version 0.0.0.1

Imports stats, MASS, rpart

URL <https://github.com/markvanderloo/simputation>

BugReports <https://github.com/markvanderloo/simputation/issues>

Date 2016-07-21

Suggests testthat

RoxygenNote 5.0.1

R topics documented:

impute_rhd	2
Index	4

impute_rhd

*Impute missing data***Description**

Use to fit and impute missing data.

Usage

```
impute_rhd(dat, model, donor_set = c("per-variable", "per-record"), prob, ...)

impute_shd(dat, model, donor_set = c("per-variable", "per-record"),
  order = c("locf", "nocb"), ...)

impute_lm(dat, model, add_residual = c("none", "observed", "normal"), ...)

impute_rlm(dat, model, add_residual = c("none", "observed", "normal"), ...)

impute_const(dat, model, add_residual = c("none", "observed", "normal"), ...)

impute_median(dat, model, add_residual = c("none", "observed", "normal"), ...)

impute_proxy(dat, model, add_residual = c("none", "observed", "normal"), ...)

impute_glm(dat, model, ...)

impute_cart(data, model, add_residual = c("none", "observed", "normal"), ...)
```

Arguments

<code>dat</code>	A data.frame, with variables to be imputed
<code>model</code>	[formula] imputation model description (see Details below).
<code>donor_set</code>	Determine donors set for each variable or once for all imputed variables.
<code>prob</code>	[numeric] Sampling probability weights (passed through to sample). Must be of length <code>nrow(dat)</code> .
<code>...</code>	further arguments passed to <ul style="list-style-type: none"> • lm for <code>impute_lm</code> • rlm for <code>impute_rlm</code> • order for <code>impute_shd</code>
<code>order</code>	Last Observation Carried Forward or Next Observation Carried Backward
<code>add_residual</code>	[character] Type of residual to add. "normal" means that the imputed value is drawn from $N(\mu, sd)$ where μ and sd are estimated from the model's residuals (μ should equal zero in most cases). If <code>add_residual = "observed"</code> , residuals are drawn (with replacement) from the model's residuals. Ignored for non-numeric predicted variables.

Value

data, imputed according to model.

Details

Model specification works as usual, except that it is possible to impute multiple variables based on the same model. To specify the same model for multiple variables, simply add variables to the left-hand side of the formula using `+`. Also see the examples.

If a value cannot be imputed because one of its predictors is missing, the value will remain missing after imputation.

If a model cannot be fitted, e.g. because the imputed model is missing, a warning is emitted and for that variable no imputation will take place.

Model descriptions

Model	description
<code>impute_lm</code>	Use <code>stats::lm</code> to train the imputation model.
<code>impute_rlm</code>	Use <code>MASS::rlm</code> to train the imputation model.
<code>impute_median</code>	Median imputation. Predictors are treated as grouping variables for computing medians.
<code>impute_const</code>	Impute a constant value
<code>impute_proxy</code>	Copy a value from the predictor variable.
<code>impute_cart</code>	Use <code>rpart::rpart</code> to train a CART model.

See Also

[lm](#) [rlm](#) [rpart](#) [ctree](#)

Examples

```
data(iris)
irisNA <- iris
irisNA[1:4, "Sepal.Length"] <- NA
irisNA[3:7, "Sepal.Width"] <- NA

# impute a single variable (Sepal.Length)
i1 <- impute_lm(irisNA, Sepal.Length ~ Sepal.Width + Species)

# impute both Sepal.Length and Sepal.Width, using robust linear regression
i2 <- impute_rlm(irisNA, Sepal.Length + Sepal.Width ~ Species + Petal.Length)
```

Index

ctree, [3](#)

impute_ (impute_rhd), [2](#)
impute_cart (impute_rhd), [2](#)
impute_const (impute_rhd), [2](#)
impute_glm (impute_rhd), [2](#)
impute_lm (impute_rhd), [2](#)
impute_median (impute_rhd), [2](#)
impute_proxy (impute_rhd), [2](#)
impute_rhd, [2](#)
impute_rlm (impute_rhd), [2](#)
impute_shd (impute_rhd), [2](#)

lm, [2](#), [3](#)

order, [2](#)

rlm, [2](#), [3](#)

rpart, [3](#)

sample, [2](#)