Package 'simputation'

July 28, 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 *Impute missing data*

Description

Use to fit and impute missing data.

Usage

```
impute_rhd(dat, model, pool = c("univariate", "multivariate"), prob, ...)

impute_shd(dat, model, order = c("locf", "nocb"), pool = c("univariate",
  "multivariate"), ...)

impute_pmm(dat, model, predictor = impute_lm, ...)

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

Arguments

dat	[data.frame], with variables to be imputed and their predictors.
model	[formula] imputation model description (see Details below).
pool	Create a donor pool for each variable ("single") or create a donor pool for each missingess pattern ("multiple").
prob	[numeric] Sampling probability weights (passed through to sample). Must be of length nrow(dat).
...	further arguments passed to • lm for impute_lm • rlm for impute_rlm • order for impute_shd • The predictor for impute_pmm
order	Last Observation Carried Forward or Next Observarion Carried Backward

`predictor` [function] Imputation to use for predictive part in predictive mean matching. Any of the `impute_` functions of this package.

`add_residual` [character] Type of residual to add. `"normal"` means that the imputed value is drawn from `N(mu,sd)` where `mu` and `sd` are estimated from the model's residuals (`mu` should equal zero in most cases). If `add_residual = "observed"`, 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 fot htat variable no imputation will take place.

Model descriptions

Model	description
`impute_lm`	Use `stats::lm` to train the imputation model.
`impute_rlm`	Use `MASS::rlm` to train the imputation model.
`impute_median`	Median imputation. Predictors are treated as grouping variables for computing medians.
`impute_const`	Impute a constant value
`impute_proxy`	Copy a value from the predictor variable.
`impute_rhd`	Random hot deck. Predictors are used to group the donors.
`impute_shd`	Sequential hot deck. Predictors sort the data.
`impute_pmm`	Predictive mean matching.
`impute_cart`	Use `rpart::rpart` to train a CART model.

See Also

`lm` `rlm` `rpart`

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)
```

---

simputation *simputation*

Description

simputation

Index