Description

Takes a data.frame (ds) and using the variables specified in x_vars, selects matches from the control group (group_var == 0) for members of the treatment group (group_var == 1) where possible. It returns a data.frame containing only rows which are part of a match.

The caliper width for propensity scores filters candidates prior to calculating distances, these can be widened to allow more but poorer matches. The distance measure can be one of "mahal" (default), "euclid", "norm_euclid" or "sad".

max_candidates allows the user to limit the number of matches within the calipers, effectively narrowing the calipers temporarily for treatment cases that have a large number of candidate matches.

The default seed argument ensures that given the exact same dataset, the function will return the same matches, this is because the algorithm is greedy and matches are assigned in random order.

n_per_match can be used to assign more than one control case to each treatment case and may be useful when the treatment group is small but the control group is large.

If loud is TRUE, progress updates and some summary information are printed to the console, otherwise the function prints nothing.

Usage

```r
mmatcher(ds, group_var, x_vars = "_all_", id_var = NA, distance = "mahal", caliper = 0.10, seed = 12345, max_candidates = 1000, n_per_match = 1, loud = TRUE)
```

Arguments

ds data.frame containing at least a group (0/1) variable and others to calculate distance

group_var variable with 0=control and 1=treatment in ds

x_vars list of variables to use in distance calculation

id_var name of ID variable in ds (if present)

distance one of "mahal", "euclid", "norm_euclid" or "sad"

caliper proportionate width for propensity score calipers

seed initial random seed value

max_candidates maximum number of candidates within calipers per match

n_per_match number of control cases to match to each treatment case

loud print update bars and stats

Examples

```r
treat_n <- 100
control_n <- 300
n <- treat_n + control_n
set.seed(123)

df <- data.frame(age = round(c(rnorm(control_n, 40, 15), rnorm(treat_n, 60, 15)), 2),
male = c(rbinom(control_n, 1, 0.4), rbinom(treat_n, 1, 0.6)),
grp = c(rep(0, control_n), rep(1, treat_n)))
```
ruler

Rcpp distance calculator

Description

Returns a vector of distances from all rows in vR to the single row uR using ciR as the inverted covariance matrix.

Usage

ruler(vR, uR, ciR)

Arguments

uR     a vector of length k containing a list of values for all features (k) for the target. Numeric and dense.

vR     an n x k matrix containing a matrix of values for all features (k) for all candidates (n). Numeric and dense.

ciR    a square k x k matrix containing the inverted covariance matrix for all features (k). Numeric and dense.

Examples

set.seed(123)
df <- data.frame(x = rpois(10, 20), y = rnorm(10, 50, 10))
cov_inv <- MASS::ginv(cov(df))
mmsample::ruler(as.matrix(df[2:10, ]), as.numeric(df[1, ]), cov_inv)
Index

mmatcher, 1
ruler, 3