

Package ‘multid’

October 22, 2021

Title Multivariate Difference Between Two Groups

Version 0.2.0

Description Estimation of multivariate differences between two groups (e.g., multivariate sex differences) with regularized regression methods and predictive approach. See Lönnqvist & Ilmarinen (2021) <[doi:10.1007/s11109-021-09681-2](https://doi.org/10.1007/s11109-021-09681-2)> and Ilmarinen et al. (2021) <[doi:10.31234/osf.io/j59bs](https://doi.org/10.31234/osf.io/j59bs)>.

License GPL-3

Encoding UTF-8

BugReports <https://github.com/vjilmari/multid/issues>

RoxygenNote 7.1.2

Imports dplyr (>= 1.0.7), glmnet (>= 4.1.2), stats (>= 4.0.2), pROC (>= 1.18.0)

NeedsCompilation no

Author Ville-Juhani Ilmarinen [aut, cre]
(<<https://orcid.org/0000-0001-9493-379X>>)

Maintainer Ville-Juhani Ilmarinen <vj.ilmarinen@gmail.com>

Repository CRAN

Date/Publication 2021-10-22 08:50:02 UTC

R topics documented:

colwise_pool	2
d_pooled_sd	2
D_regularized	3
D_regularized_fold	6
D_regularized_fold_out	7
D_regularized_out	9
D_regularized_vanilla	10
pcc	12

Index	13
--------------	-----------

colwise_pool	<i>Column-wise pooling of standard deviations</i>
--------------	---

Description

Column-wise pooling of standard deviations

Usage

```
colwise_pool(data, n1, n2, m1, m2, sd1, sd2)
```

Arguments

data	Data frame of d_pooled_sd output for multiple samples.
n1	Sample sizes of group1.
n2	Sample sizes of group2.
m1	Means of group1.
m2	Means of group2.
sd1	Standard deviations of group1.
sd2	Standard deviations of group2.

Value

pooled SDs for groups and across groups

d_pooled_sd	<i>Standardized mean difference with pooled standard deviation</i>
-------------	--

Description

Standardized mean difference with pooled standard deviation

Usage

```
d_pooled_sd(  
  data,  
  var,  
  group.var,  
  group.values,  
  rename.output = TRUE,  
  infer = FALSE  
)
```

Arguments

data	A data frame.
var	A continuous variable for which difference is estimated.
group.var	The name of the group variable.
group.values	Vector of length 2, group values (e.g. c("male", "female) or c(0,1)).
rename.output	Logical. Should the output values be renamed according to the group.values? Default TRUE.
infer	Logical. Statistical inference with Welch test? (default FALSE)

Value

Descriptive statistics and mean differences

Examples

```
d_pooled_sd(iris[iris$Species == "setosa" | iris$Species == "versicolor", ],
  var = "Petal.Length", group.var = "Species",
  group.values = c("setosa", "versicolor"), infer = TRUE
)
```

D_regularized	<i>Multivariate group difference estimation with regularized binomial regression</i>
---------------	--

Description

Multivariate group difference estimation with regularized binomial regression

Usage

```
D_regularized(
  data,
  mv.vars,
  group.var,
  group.values,
  alpha = 0.5,
  nfolds = 10,
  s = "lambda.min",
  type.measure = "deviance",
  rename.output = TRUE,
  out = FALSE,
  size = NULL,
  fold = FALSE,
  fold.var = NULL,
  pcc = FALSE,
  auc = FALSE,
```

```

  pred.prob = FALSE,
  prob.cutoffs = seq(0, 1, 0.2)
)

```

Arguments

<code>data</code>	A data frame.
<code>mv.vars</code>	Character vector. Variable names in the multivariate variable set.
<code>group.var</code>	The name of the group variable.
<code>group.values</code>	Vector of length 2, group values (e.g. <code>c("male", "female")</code> or <code>c(0,1)</code>).
<code>alpha</code>	Alpha-value for penalizing function ranging from 0 to 1: 0 = ridge regression, 1 = lasso, 0.5 = elastic net (default).
<code>nfolds</code>	Number of folds used for obtaining lambda (range from 3 to n-1, default 10).
<code>s</code>	Which lambda value is used for predicted values? Either "lambda.min" (default) or "lambda.1se".
<code>type.measure</code>	Which measure is used during cross-validation. Default "deviance".
<code>rename.output</code>	Logical. Should the output values be renamed according to the <code>group.values</code> ? Default TRUE.
<code>out</code>	Logical. Should results and predictions be calculated on out-of-bad data set? (Default FALSE)
<code>size</code>	Integer. Number of cases in regularization data per each group. Default 1/4 of cases.
<code>fold</code>	Logical. Is regularization applied across sample folds with separate predictions for each fold? (Default FALSE)
<code>fold.var</code>	Character string. Name of the fold variable. (default NULL)
<code>pcc</code>	Logical. Include probabilities of correct classification? Default FALSE.
<code>auc</code>	Logical. Include area under the receiver operating characteristics? Default FALSE.
<code>pred.prob</code>	Logical. Include table of predicted probabilities? Default FALSE.
<code>prob.cutoffs</code>	Vector. Cutoffs for table of predicted probabilities. Default <code>seq(0,1,0.20)</code> .

Value

<code>D</code>	Multivariate descriptive statistics and differences.
<code>pred.dat</code>	A data.frame with predicted values.
<code>cv.mod</code>	Regularized regression model from <code>cv.glmnet</code> .
<code>P.table</code>	Table of predicted probabilities by cutoffs.

See Also

[cv.glmnet](#)

Examples

```

D_regularized(
  data = iris[iris$Species == "setosa" | iris$Species == "versicolor", ],
  mv.vars = c("Sepal.Length", "Sepal.Width", "Petal.Length", "Petal.Width"),
  group.var = "Species", group.values = c("setosa", "versicolor")
)$D

# out-of-bag predictions
D_regularized(
  data = iris[iris$Species == "setosa" | iris$Species == "versicolor", ],
  mv.vars = c("Sepal.Length", "Sepal.Width", "Petal.Length", "Petal.Width"),
  group.var = "Species", group.values = c("setosa", "versicolor"),
  out = TRUE, size = 15, pcc = TRUE, auc = TRUE
)$D

# separate sample folds
# generate data for 10 groups
set.seed(34246)
n1 <- 100
n2 <- 10
d <-
  data.frame(
    sex = sample(c("male", "female"), n1 * n2, replace = TRUE),
    fold = sample(x = LETTERS[1:n2], size = n1 * n2, replace = TRUE),
    x1 = rnorm(n1 * n2),
    x2 = rnorm(n1 * n2),
    x3 = rnorm(n1 * n2)
  )

# Fit and predict with same data
D_regularized(
  data = d,
  mv.vars = c("x1", "x2", "x3"),
  group.var = "sex",
  group.values = c("female", "male"),
  fold.var = "fold",
  fold = TRUE,
  rename.output = TRUE
)$D

# Out-of-bag data for each fold
D_regularized(
  data = d,
  mv.vars = c("x1", "x2", "x3"),
  group.var = "sex",
  group.values = c("female", "male"),
  fold.var = "fold",
  size = 17,
  out = TRUE,
  fold = TRUE,
  rename.output = TRUE
)$D

```

D_regularized_fold	<i>Use manually defined data folds for regularization and obtain estimates for each separately.</i>
--------------------	---

Description

Use manually defined data folds for regularization and obtain estimates for each separately.

Usage

```
D_regularized_fold(
  data,
  mv.vars,
  group.var,
  group.values,
  alpha = 0.5,
  s = "lambda.min",
  type.measure = "deviance",
  rename.output = TRUE,
  fold.var
)
```

Arguments

data	A data frame.
mv.vars	Character vector. Variable names in the multivariate variable set.
group.var	The name of the group variable.
group.values	Vector of length 2, group values (e.g. c("male", "female") or c(0,1)).
alpha	Alpha-value for penalizing function ranging from 0 to 1: 0 = ridge regression, 1 = lasso, 0.5 = elastic net (default).
s	Which lambda value is used for predicted values? Either "lambda.min" (default) or "lambda.1se".
type.measure	Which measure is used during cross-validation. Default "deviance".
rename.output	Logical. Should the output values be renamed according to the group.values? Default TRUE.
fold.var	Character string. Name of the fold variable.

Value

D	Multivariate descriptive statistics and differences.
pred.dat	A data.frame with predicted values.
cv.mod	Regularized regression model from cv.glmnet.

See Also[cv.glmnet](#)**Examples**

```
set.seed(34246)
n1 <- 100
n2 <- 10
d <-
  data.frame(
    sex = sample(c("male", "female"), n1 * n2, replace = TRUE),
    fold = sample(x = LETTERS[1:n2], size = n1 * n2, replace = TRUE),
    x1 = rnorm(n1 * n2),
    x2 = rnorm(n1 * n2),
    x3 = rnorm(n1 * n2)
  )

D_regularized_fold(
  data = d,
  mv.vars = c("x1", "x2", "x3"),
  group.var = "sex",
  group.values = c("female", "male"),
  fold.var = "fold"
)$D
```

D_regularized_fold_out

Use separate data partitions for regularization and estimation across defined data folds.

Description

Use separate data partitions for regularization and estimation across defined data folds.

Usage

```
D_regularized_fold_out(
  data,
  mv.vars,
  group.var,
  group.values,
  alpha = 0.5,
  s = "lambda.min",
  type.measure = "deviance",
  rename.output = TRUE,
  size = NULL,
  fold.var,
  pcc = FALSE,
```

```

auc = FALSE,
pred.prob = FALSE,
prob.cutoffs = seq(from = 0, to = 1, by = 0.2)
)

```

Arguments

<code>data</code>	A data frame.
<code>mv.vars</code>	Character vector. Variable names in the multivariate variable set.
<code>group.var</code>	The name of the group variable.
<code>group.values</code>	Vector of length 2, group values (e.g. <code>c("male", "female")</code> or <code>c(0,1)</code>).
<code>alpha</code>	Alpha-value for penalizing function ranging from 0 to 1: 0 = ridge regression, 1 = lasso, 0.5 = elastic net (default).
<code>s</code>	Which lambda value is used for predicted values? Either "lambda.min" (default) or "lambda.1se".
<code>type.measure</code>	Which measure is used during cross-validation. Default "deviance".
<code>rename.output</code>	Logical. Should the output values be renamed according to the <code>group.values</code> ? Default TRUE.
<code>size</code>	Integer. Size of regularization data per each group. Default 1/4 of cases.
<code>fold.var</code>	Name of the fold variable.
<code>pcc</code>	Logical. Include probabilities of correct classification? Default FALSE.
<code>auc</code>	Logical. Include area under the receiver operating characteristics? Default FALSE.
<code>pred.prob</code>	Logical. Include table of predicted probabilities? Default FALSE.
<code>prob.cutoffs</code>	Vector. Cutoffs for table of predicted probabilities. Default <code>seq(0,1,0.20)</code> .

Value

<code>D</code>	Multivariate descriptive statistics and differences.
<code>pred.dat</code>	A data.frame with predicted values.
<code>cv.mod</code>	Regularized regression model from <code>cv.glmnet</code> .
<code>P.table</code>	Table of predicted probabilities by cutoffs.

Examples

```

set.seed(34246)
n1 <- 100
n2 <- 10
d <-
  data.frame(
    sex = sample(c("male", "female"), n1 * n2, replace = TRUE),
    fold = sample(x = LETTERS[1:n2], size = n1 * n2, replace = TRUE),
    x1 = rnorm(n1 * n2),
    x2 = rnorm(n1 * n2),
    x3 = rnorm(n1 * n2)
  )

```



```

)
D_regularized_fold_out(
  data = d,
  mv.vars = c("x1", "x2", "x3"),
  group.var = "sex",
  group.values = c("female", "male"),
  fold.var = "fold",
  size = 17,
  pcc = TRUE
)$D

```

D_regularized_out *Use separate data partition for regularization and estimation.*

Description

Use separate data partition for regularization and estimation.

Usage

```

D_regularized_out(
  data,
  mv.vars,
  group.var,
  group.values,
  alpha = 0.5,
  nfolds = 10,
  s = "lambda.min",
  type.measure = "deviance",
  rename.output = TRUE,
  size = NULL,
  pcc = FALSE,
  auc = FALSE,
  pred.prob = FALSE,
  prob.cutoffs = seq(from = 0, to = 1, by = 0.2)
)

```

Arguments

data	A data frame.
mv.vars	Character vector. Variable names in the multivariate variable set.
group.var	The name of the group variable.
group.values	Vector of length 2, group values (e.g. c("male", "female") or c(0,1)).
alpha	Alpha-value for penalizing function ranging from 0 to 1: 0 = ridge regression, 1 = lasso, 0.5 = elastic net (default).
nfolds	Number of folds used for obtaining lambda (range from 3 to n-1, default 10).

s	Which lambda value is used for predicted values? Either "lambda.min" (default) or "lambda.1se".
type.measure	Which measure is used during cross-validation. Default "deviance".
rename.output	Logical. Should the output values be renamed according to the group.values? Default TRUE.
size	Integer. Size of regularization data per each group. Default 1/4 of cases.
pcc	Logical. Include probabilities of correct classification? Default FALSE.
auc	Logical. Include area under the receiver operating characteristics? Default FALSE.
pred.prob	Logical. Include table of predicted probabilities? Default FALSE.
prob.cutoffs	Vector. Cutoffs for table of predicted probabilities. Default seq(0,1,0.20).

Value

D	Multivariate descriptive statistics and differences.
pred.dat	A data.frame with predicted values.
cv.mod	Regularized regression model from cv.glmnet.
P.table	Table of predicted probabilities by cutoffs.

Examples

```
D_regularized_out(
  data = iris[iris$Species == "setosa" |
    iris$Species == "versicolor", ],
  mv.vars = c(
    "Sepal.Length", "Sepal.Width",
    "Petal.Length", "Petal.Width"
  ),
  group.var = "Species",
  group.values = c("setosa", "versicolor"),
  size = 40,
  pcc = TRUE
)$D
```

D_regularized_vanilla *Use same data partition for regularization and estimation.*

Description

Use same data partition for regularization and estimation.

Usage

```
D_regularized_vanilla(
  data,
  mv.vars,
  group.var,
  group.values,
  alpha = 0.5,
  nfolds = 10,
  s = "lambda.min",
  type.measure = "deviance",
  rename.output = TRUE
)
```

Arguments

data	A data frame.
mv.vars	Character vector. Variable names in the multivariate variable set.
group.var	The name of the group variable.
group.values	Vector of length 2, group values (e.g. c("male", "female") or c(0,1)).
alpha	Alpha-value for penalizing function ranging from 0 to 1: 0 = ridge regression, 1 = lasso, 0.5 = elastic net (default).
nfolds	Number of folds used for obtaining lambda (range from 3 to n-1, default 10).
s	Which lambda value is used for predicted values? Either "lambda.min" (default) or "lambda.1se".
type.measure	Which measure is used during cross-validation. Default "deviance".
rename.output	Logical. Should the output values be renamed according to the group.values? Default TRUE.

Value

D	Multivariate descriptive statistics and differences.
pred.dat	A data.frame with predicted values.
cv.mod	Regularized regression model from cv.glmnet.

See Also

[cv.glmnet](#)

Examples

```
D_regularized(
  data = iris[iris$Species == "setosa" | iris$Species == "versicolor", ],
  mv.vars = c("Sepal.Length", "Sepal.Width", "Petal.Length", "Petal.Width"),
  group.var = "Species", group.values = c("setosa", "versicolor")
)$D
```

pcc	<i>Returns probabilities of correct classification for both groups in independent data partition.</i>
-----	---

Description

Returns probabilities of correct classification for both groups in independent data partition.

Usage

```
pcc(data, pred.var, group.var, group.values)
```

Arguments

data	Data frame including predicted values (e.g., pred.dat from D_regularized_out).
pred.var	Character string. Variable name for predicted values.
group.var	The name of the group variable.
group.values	Vector of length 2, group values (e.g. c("male", "female) or c(0,1)).

Value

Vector of length 2. Probabilities of correct classification.

Examples

```
D_out <- D_regularized_out(  
  data = iris[iris$Species == "versicolor" | iris$Species == "virginica", ],  
  mv.vars = c("Sepal.Length", "Sepal.Width", "Petal.Length", "Petal.Width"),  
  group.var = "Species", group.values = c("versicolor", "virginica"),  
  size = 15  
)  
  
pcc(  
  data = D_out$pred.dat,  
  pred.var = "pred",  
  group.var = "group",  
  group.values = c("versicolor", "virginica")  
)
```

Index

`colwise_pool`, [2](#)

`cv.glmnet`, [4](#), [7](#), [11](#)

`d_pooled_sd`, [2](#)

`D_regularized`, [3](#)

`D_regularized_fold`, [6](#)

`D_regularized_fold_out`, [7](#)

`D_regularized_out`, [9](#)

`D_regularized_vanilla`, [10](#)

`pcc`, [12](#)