Package ‘predtools’

June 6, 2023

Title Prediction Model Tools

Version 0.0.3

Description Provides additional functions for evaluating predictive models, including plotting calibration curves and model-based Receiver Operating Characteristic (mROC) based on Sadatsafavi et al (2021) <arXiv:2003.00316>.

License GPL

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

URL https://github.com/resplab/predtools

BugReports https://github.com/resplab/predtools/issues

Depends R (>= 3.6)

Imports Rcpp, pROC, stats, graphics, RConics, ggplot2, dplyr, magrittr, mvtnorm

LinkingTo Rcpp

Suggests rmarkdown, knitr, spelling

VignetteBuilder knitr

Language en-US

NeedsCompilation yes

Author Mohsen Sadatsafavi [aut, cph] (<https://orcid.org/0000-0002-0419-7862>),
Amin Adibi [cre] (<https://orcid.org/0000-0003-2748-4781>),
Abdollah Safari [aut],
Tae Yoon Lee [aut]

Maintainer Amin Adibi <adibi@alumni.ubc.ca>

Repository CRAN

Date/Publication 2023-06-05 23:10:03 UTC

R topics documented:

`calc_mROC_stats` *Calculates the absolute surface between the empirical and expected ROCs*

Description

Calculates the absolute surface between the empirical and expected ROCs

Usage

```
calc_mROC_stats(y, p, ordered = FALSE, fast = TRUE)
```

Arguments

`y`	y vector of binary responses
`p`	p vector of predicted probabilities (same length as y)
`ordered`	defaults to false
`fast`	defaults to true

Value

Returns a list with the A (mean calibration statistic) and B (mROC/ROC equality statistic) as well as the direction of potential miscalibration (sign of the difference between the actual and predicted mean risk) calc_NB_moments *Calculates the first two moments of the bivariate distribution of NB_model and NB_all*

Description

Calculates the first two moments of the bivariate distribution of NB_model and NB_all

Usage

```
calc_NB_moments(Y, pi, z, weights = NULL)
```

Arguments

Y	Vector of the binary response variable
pi	Vector of predicted risks
z	Decision threshold at which the NBs are calculated
weights	Optinal - observation weights

Value

Two means, two SDs, and one correlation coefficient. First element is for the model and second is for treating all

calibration_plot *Title Create calibration plot based on observed and predicted outcomes.*

Description

Title Create calibration plot based on observed and predicted outcomes.

Usage

```
calibration_plot(
  data,
  obs,
  follow_up = NULL,
  pred,
  group = NULL,
  nTiles = 10,
  legendPosition = "right",
  title = NULL,
  x_lim = NULL,
  y_lim = NULL,
```

```
  xlab = "Prediction",
  ylab = "Observation",
  points_col_list = NULL,
  data_summary = FALSE
)
```

Arguments

`data`	Data include observed and predicted outcomes.
`obs`	Name of observed outcome in the input data.
`follow_up`	Name of follow-up time (if applicable) in the input data.
`pred`	Name of first predicted outcome in the input data.
`group`	Name of grouping column (if applicable) in the input data.
`nTiles`	Number of tiles (e.g., 10 for deciles) in the calibration plot.
`legendPosition`	Legend position on the calibration plot.
`title`	Title on the calibration plot.
`x_lim`	Limits of x-axis on the calibration plot.
`y_lim`	Limits of y-axis on the calibration plot.
`xlab`	Label of x-axis on the calibration plot.
`ylab`	Label of y-axis on the calibration plot.
`points_col_list`	Points' color on the calibration plot.
`data_summary`	Logical indicates whether a summary of the predicted and observed outcomes. needs to be included in the output.

Value

Returns calibration plot (a ggplot object) and a dataset including summary statistics of the predicted and observed outcomes (if data_summary set to be TRUE).

Examples

```
library(predtools)
library(dplyr)
x <- rnorm(100, 10, 2)
y <- x + rnorm(100,0, 1)
data <- data.frame(x, y)
calibration_plot(data, obs = "x", pred = "y")
```

dev_data — *model development data*

Description

A dataset containing sample model development data

Format

A data frame with 500 rows and 5 variables:

- ageage
- severitywhether or not the disease was severe
- sexbinary sex variable, 1 for female and 0 for male
- comorbiditywhether or not comorbidities are present
- yresponse variable

Source

Simulated

evpi_val — *EVPI (Expected Value of Perfect Information) for validation Takes a vector of mean and a 2X2 covariance matrix*

Description

EVPI (Expected Value of Perfect Information) for validation Takes a vector of mean and a 2X2 covariance matrix

Usage

```
evpi_val(
  Y,
  pi,
  method = c("bootstrap", "bayesian_bootstrap", "asymptotic"),
  n_sim = 1000,
  zs = (0:99)/100,
  weights = NULL
)
```

Arguments

Y	Binary response variable
pi	Mean of the second distribution
method	EVPI calculation method
n_sim	Number of Monte Carlo simulations (for bootstrap-based methods)
zs	vector of risk thresholds at which EVPI is to be calculated
weights	(optional) observation weights

Value

Returns a data frame containing thresholds, EVPIs, and some auxilary output.

---

gusto *Anonymized data from the gusto trial*

Description

A dataset containing anonymized data from the gusto trial

Format

A data frame with 40830 rows and 29 variables:

- day30whether death happened by day 30 after intervention
- showhether cardiac shock was present
- higwhether the patient hat high blood pressure
- diawhether the patient had diabetes
- hrtwhether the patient was on hormone replacement therapies

Source

Internet

`mAUC` *Takes in a mROC object and calculates the area under the curve*

Description

Takes in a mROC object and calculates the area under the curve

Usage

```
mAUC(mROC_obj)
```

Arguments

mROC_obj	An object of class mROC

Value

Returns the area under the mROC curve

`mROC` *Calculates mROC from the vector of predicted risks Takes in a vector of probabilities and returns mROC values (True positives, False Positives in an object of class mROC)*

Description

Calculates mROC from the vector of predicted risks Takes in a vector of probabilities and returns mROC values (True positives, False Positives in an object of class mROC)

Usage

```
mROC(p, ordered = FALSE)
```

Arguments

p	A numeric vector of probabilities.
ordered	Optional, if the vector p is ordered from small to large (if not the function will do it; TRUE is to facilitate fast computations).

Value

This function returns an object of class mROC. It has three vectors: thresholds on predicted risks (which is the ordered vector of input probabilities), false positive rates (FPs), and true positive rates (TPs). You can directly call the plot function on this object to draw the mROC

mROC_analysis *Main eROC analysis that plots ROC and eROC*

Description

Main eROC analysis that plots ROC and eROC

Usage

```
mROC_analysis(y, p, inference = 0, n_sim, fast = TRUE)
```

Arguments

`y`	y vector of observed responses.
`p`	p vector of predicted probabilities (the same length as observed responses)
`inference`	0 for no inference, 1 for p-value only, and 2 for p-value and 95 percent CI.
`n_sim`	number of simulations
`fast`	defaults to true

Value

returns a list containing the results of mROC analysis.

mROC_inference *Statistical inference for comparing empirical and expected ROCs. If CI=TRUE then also returns pointwise CIs*

Description

Statistical inference for comparing empirical and expected ROCs. If CI=TRUE then also returns pointwise CIs

Usage

```
mROC_inference(y, p, n_sim = 1e+05, CI = FALSE, aux = FALSE, fast = TRUE)
```

Arguments

`y`	vector of binary response values
`p`	vector of probabilities
`n_sim`	number of Monte Carlo simulations to calculate p-value
`CI`	optional. Whether confidence interval should be calculated for each point of mROC. Default is FALSE.

aux	aux optional. whether additional results (component-wise p-values etc) should be written in the package's aux variable. Default is FALSE.
fast	fast optional. Whether the fast code (C++) or slow code (R) should be called. Default is TRUE (R code will be slow unless the dataset is small)

Value

Returns an object of type mROC_inference containing the results of statistical inference for the mROC curve

---

mu_max_trunc_bvn — *Calculates the expected value of the maximum of two random variables with zero-truncated bivariate normal distribution Takes a vector of mean and a 2X2 covariance matrix*

Description

Calculates the expected value of the maximum of two random variables with zero-truncated bivariate normal distribution Takes a vector of mean and a 2X2 covariance matrix

Usage

```
mu_max_trunc_bvn(
  mu1,
  mu2,
  sigma1,
  sigma2,
  rho,
  precision = .Machine$double.eps
)
```

Arguments

mu1	Mean of the first distribution
mu2	Mean of the second distribution
sigma1	SD of the first distribution
sigma2	SD of the second distribution
rho	Correlation coefficient of the two random variables
precision	Numerical precision value

Value

A scalar value for the expected value

odds_adjust

Title Update a prediction model for a binary outcome by multiplying a fixed odd-ratio to the predicted odds.

Description

Title Update a prediction model for a binary outcome by multiplying a fixed odd-ratio to the predicted odds.

Usage

```
odds_adjust(p0, p1, v)
```

Arguments

p0	Mean of observed risk or predicted risk in development sample.
p1	Mean of observed risk in target population.
v	Variance of predicted risk in development sample.

Value

Returns a correction factor that can be applied to the predicted odds in order to update the predictions for a new target population.

pred_summary_stat

Title Estimate mean and variance of prediction based on model calibration output.

Description

Title Estimate mean and variance of prediction based on model calibration output.

Usage

```
pred_summary_stat(calibVector)
```

Arguments

calibVector	Vector of predicted probability of risk per decile or percentile (e.g., from a calibration plot).

Value

Returns mean and variance of predictions based on the predicted probabilities.

val_data *model validation data*

Description

A dataset containing sample model validation data

Format

A data frame with 400 rows and 5 variables:

- ageage of the patient
- severitywhether or not the disease was severe
- sexbinary sex variable, 1 for female and 0 for male
- comorbiditywhether or not comorbidities are present
- yresponse variable

Source

Simulated

Index