

# Package ‘RBPcurve’

January 27, 2015

**Type** Package

**Title** The Residual-based Predictiveness Curve

**Version** 1.0-20

**Date** 2015-01-09

**Author** Giuseppe Casalicchio, Bernd Bischl

**URL** <https://github.com/giuseppec/RBPcurve>

**BugReports** <https://github.com/giuseppec/RBPcurve/issues>

**Maintainer** Giuseppe Casalicchio <giuseppe.casalicchio@stat.uni-muenchen.de>

**Description** The package provides a visual tool, the RBP curve, to assess the performance of prediction models.

**License** GPL-3

**Depends** mlr

**Imports** BBmisc, checkmate, shape, TeachingDemos

**Suggests** testthat, mlbench, mboost

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2015-01-09 21:57:51

## R topics documented:

addGoodCalib . . . . .	2
addPEV . . . . .	2
addPrevalence . . . . .	3
addRates . . . . .	4
addWellCalib . . . . .	4
makeRBPObj . . . . .	5
plotRBPcurve . . . . .	6

<b>Index</b>	<b>8</b>
--------------	----------

---

addGoodCalib	<i>Visualizes a measure for good calibration on the RBP curve.</i>
--------------	--

---

### Description

The integral of the RBP curve is a measure for good calibration. If the sum of the two integrals (below and above the RBP curve) is close to 0, good calibration is satisfied and the prevalence is close to the average predicted probabilities.

### Usage

```
addGoodCalib(obj, plot.values = TRUE, show.info = TRUE, col = rgb(0, 0, 0,
  0.25), border = NA, ...)
```

### Arguments

obj	[RBPObj] Data container for RBP curve.
plot.values	[logical(1)] Whether the values of the corresponding measure should be added to the plot? Default is FALSE.
show.info	[logical(1)] Print more information for the respective measure on console? Default is TRUE.
col	[vector(1)] Color for filling the polygon, as in <a href="#">polygon</a> . Default is “grey”.
border	[vector(1)] Color to draw the borders, as in <a href="#">polygon</a> . Default is NA to omit borders.
...	[any] Passed to <a href="#">polygon</a> .

### Value

invisible(NULL) .

---

addPEV	<i>Visualize the PEV on the RBP curve.</i>
--------	--

---

### Description

The PEV measure is the difference between the conditional expectation of the predicted probabilities (conditional on the two groups that are determined by the target variable). The PEV measure can be visually obtained by the RBP curve, namely by the difference of the two areas that are Highlighted with addPEV.

**Usage**

```
addPEV(obj, plot.values = TRUE, show.info = TRUE, text.col = "black",
        col = rgb(0, 0, 0, 0.25))
```

**Arguments**

obj	[RBPObj] Data container for RBP curve.
plot.values	[logical(1)] Whether the values of the corresponding measure should be added to the plot? Default is FALSE.
show.info	[logical(1)] Print more information for the respective measure on console? Default is TRUE.
text.col	[character(1)   numeric(1)] Text color, used when plot.values = TRUE, otherwise ignored. Default is "black".
col	[character(1)   numeric(1)] A specification for the plotting color.

**Value**

invisible(NULL) .

---

addPrevalence	<i>Visualizes the prevalence on the RBP curve.</i>
---------------	--

---

**Description**

The prevalence is the proportion of a population having a specific condition. In binary classification, the condition refers to whether the target variable has the value 1, that is, whether the target variable corresponds to the positive class.

**Usage**

```
addPrevalence(obj, plot.values = TRUE, col = "grey")
```

**Arguments**

obj	[RBPObj] Data container for RBP curve.
plot.values	[logical(1)] Whether the values of the corresponding measure should be added to the plot? Default is FALSE.
col	[character(1)   numeric(1)] A specification for the plotting color.

**Value**

invisible(NULL) .

---

addRates	<i>Visualizes the TPR and FPR on the RBP curve.</i>
----------	---

---

**Description**

For a given threshold `tresh`, the true positive rate (TPR) and the false positive rate (FPR) can be visually assessed by the RBP curve by the intersection of the RBP curve with the horizontal lines at `-thresh` and `1 - thresh`, respectively.

**Usage**

```
addRates(obj, plot.values = TRUE, thresh = obj$prev, col = "black")
```

**Arguments**

<code>obj</code>	[RBPObj] Data container for RBP curve.
<code>plot.values</code>	[logical(1)] Whether the values of the corresponding measure should be added to the plot? Default is FALSE.
<code>thresh</code>	[numeric(1)] Threshold that is used to compute the true positive and false positive rate. Default is prevalence.
<code>col</code>	[character(1)   numeric(1)] A specification for the plotting color.

**Value**

invisible(NULL) .

---

addWellCalib	<i>Visualizes a measure for well calibration on the RBP curve.</i>
--------------	--

---

**Description**

A measure for a well calibrated model can be obtained by grouping the predicted probabilities via deciles yielding 10 groups. The equally colored areas belong to a specific group. When each of the two equally colored areas are similar, the model is well calibrated.

**Usage**

```
addWellCalib(obj, plot.values = TRUE, col = shape::greycol(10L),
  pos = "topleft")
```

**Arguments**

obj	[RBPObj] Data container for RBP curve.
plot.values	[logical(1)] Whether the values of the corresponding measure should be added to the plot? Default is FALSE.
col	[character   numeric] A specification for the the plotting color for the areas.
pos	[character(1)] Determines the position of the subplot that is plotted when plot.values = TRUE. Can be either pos = "topleft" or pos = "bottomright". Default is pos = "topleft".

**Value**

A matrix that contains the average of the “probabilities within deciles” conditional on Y.

---

makeRBPObj	<i>Create data container for RBP curve.</i>
------------	---

---

**Description**

Must be created for all subsequent plot function calls.

**Usage**

```
makeRBPObj(pred, y, positive = NULL)
```

**Arguments**

pred	[numeric] Predicted probabilities for each observation.
y	[numeric   factor] Class labels of the target variable. Either a numeric vector with values 0 or 1, or a factor with two levels.
positive	[character(1)] Set positive class label for target variable which is transformed as 1 to compute. Only needed when y is a "factor".

**Value**

Object members:

`n` [numeric(1) ] Number of observations.

`pred` [numeric(n) ] Predicted probabilities.

`y` [numeric(n) ] Target variable having the values 0 and 1.

`positive` [character(1) ] Positive class label of target variable. Only present when `y` is a factor.

`e0` [numeric(1) ] Average of the predicted probabilities conditional on  $y=0$ .

`e1` [numeric(1) ] Average of the predicted probabilities conditional on  $y=1$ .

`pev` [numeric(1) ] Proportion of explained variation measure. Computed as  $e1-e0$ .

`tpr` [numeric(1) ] True positive rate.

`fpr` [numeric(1) ] False positive rate.

`prev` [numeric(1) ] Prevalence.

`one.min.prev` [numeric(1) ] One minus the value of the prevalence.

`axis.x` [numeric(n) ] Values for the X-Axis of the RBP curve.

`axis.y` [numeric(n) ] Values for the Y-Axis of the RBP curve.

---

plotRBPCurve

*Plot residual-based predictiveness (RBP) curve.*

---

**Description**

Plot residual-based predictiveness (RBP) curve.

**Usage**

```
plotRBPCurve(obj, main = "RBP Curve", xlab = "Cumulative Percentage",
  ylab = "Estimated Residuals", type = "l", ylim = c(-1, 1.2),
  cond.axis = FALSE, title.line = ifelse(cond.axis, 3, 2), add = FALSE,
  ...)
```

**Arguments**

<code>obj</code>	[RBPObj] Data container for RBP curve.
<code>main</code>	[character(1)] An overall title for the plot.
<code>xlab</code>	[character(1)] Label for X-axis. Default is "Cumulative Percentage".
<code>ylab</code>	[character(1)] Label for Y-axis. Default is "Estimated Residuals".

type	[character(1)] The plot type that should be drawn, see <a href="#">plot</a> for all possible types. Default is type = "l" for lines.
ylim	[numeric(2)] Limits for Y-axis. Default is c(-1, 1.1).
cond.axis	[logical(1)] Should an additional axis be plotted reflecting residuals conditional on y? Default is FALSE.
title.line	[integer(1)] Where to plot the title, see <a href="#">title</a> .
add	[logical(1)] Should RBP plot be added to current plot? Default is FALSE.
...	[any] Passed to <a href="#">plot</a> or <a href="#">lines</a> , depending on add.

### Examples

```
# Download data
mydata = read.csv("http://www.ats.ucla.edu/stat/data/binary.csv")
head(mydata)

# Build logit model and plot RBP curve
mylogit <- glm(admit ~ ., data = mydata, family = "binomial")
y = mydata$admit
pred1 = predict(mylogit, type="response")
obj1 = makeRBPobj(pred1, y)
plotRBPCurve(obj1, cond.axis = TRUE, type = "b")

# Build logit model using mlr and plot RBP curve
task = makeClassifTask(data = mydata, target = "admit", positive = 1)
lrn = makeLearner("classif.logreg", predict.type = "prob")
tr = train(lrn, task)
pred2 = getProbabilities(predict(tr, task))
obj2 = makeRBPobj(pred2, y)
plotRBPCurve(obj2, cond.axis = TRUE, type = "b", col = 2)
```

# Index

`addGoodCalib`, 2  
`addPEV`, 2  
`addPrevalence`, 3  
`addRates`, 4  
`addWellCalib`, 4

`lines`, 7

`makeRBPObj`, 5

`plot`, 7  
`plotRBPCurve`, 6  
`polygon`, 2

`RBPObj`, 2–6  
`RBPObj (makeRBPObj)`, 5

`title`, 7