Package: separationplot (via r-universe)

September 9, 2024

Type Package

Title Separation Plots

Version 1.4

Date 2023-04-08

Author Brian D. Greenhill, Michael D. Ward and Audrey Sacks

Maintainer Brian Greenhill <bgreenhill@albany.edu>

Depends RColorBrewer, Hmisc, MASS, foreign

Description Visual representations of model fit or predictive success
in the form of ``separation plots.'' See Greenhill, Brian,
Michael D. Ward, and Audrey Sacks. ``The separation plot: A new
visual method for evaluating the fit of binary models.''
American Journal of Political Science 55.4 (2011): 991-1002.

License Artistic-2.0

NeedsCompilation no

Repository CRAN

Date/Publication 2023-04-09 18:20:02 UTC

Contents

---

separationplot-package

Separation Plots

Description

Visual representations of model fit or predictive success in the form of "separation plots." See Greenhill, Brian, Michael D. Ward, and Audrey Sacks. "The separation plot: A new visual method for evaluating the fit of binary models." American Journal of Political Science 55.4 (2011): 991-1002.

Details

The DESCRIPTION file:

Package:	separationplot
Type:	Package
Title:	Separation Plots
Version:	1.4
Date:	2023-04-08
Author:	Brian D. Greenhill, Michael D. Ward and Audrey Sacks
Maintainer:	Brian Greenhill <bgreenhill@albany.edu>
Depends:	RColorBrewer, Hmisc, MASS, foreign
Description:	Visual representations of model fit or predictive success in the form of "separation plots." See Greenhill, Brian,
License:	Artistic-2.0

Index of help topics:

```
separationplot          Generate a Separation Plot
separationplot-package

                        Separation Plots
sp.categorical          Separation plots for variables with more than
                        two outcome levels
```

A package to create "separation plots" as described in Greenhill, Ward, and Sacks (2011).

Author(s)

Brian D. Greenhill, Michael D. Ward and Audrey Sacks

Maintainer: Brian Greenhill <bgreenhill@albany.edu>

References

Greenhill, Brian, Michael D. Ward, and Audrey Sacks. "The separation plot: A new visual method for evaluating the fit of binary models." American Journal of Political Science 55.4 (2011): 991-1002.

Examples

```
# Very simple example from the introduction to the paper:
separationplot(pred=c(0.774, 0.364, 0.997, 0.728, 0.961, 0.422),
actual=c(0,0,1,0,1,1), shuffle=FALSE,
line=FALSE, type="rect", rectborder=1)
```

---

separationplot   *Generate a Separation Plot*

Description

This is the core function for the generation of a separation plot.

Usage

```
separationplot(pred, actual, type = "line", line = T, lwd1 = 0.5, lwd2 = 0.5,
heading = "",  xlab = "", shuffle = T, width = 9, height = 1.2, col0 = "#FEF0D9",
col1 = "#E34A33", flag = NULL, flagcol = 1, file = NULL, newplot = T, locate = NULL,
rectborder = NA, show.expected = F, zerosfirst = T, BW=F)
```

Arguments

pred	Vector of predicted probabilities (on a continuous 0 to 1 scale).
actual	Vector of actual outcomes (each element must be either 0 or 1).
type	Should the individual lines on the separation plot be plotted as line segments (type="line") or rectangles (type="rect"), or should the probabilities in different regions of the plot be grouped into distinct bands (type="bands")?
line	Should a trace line be added to the plot?
lwd1	The width of the individual line segments (only when type="line").
lwd2	The width of the trace line (only when line=T).
heading	An optional title for the plot.
xlab	An option x-axis label.
shuffle	If shuffle=T, the order of rows in the results data is randomized to break up any pre-existing patterns that may distort the appearance of the results in the special case where many of the observations share the same fitted values. This happens, for example, when the original dataframe is organized in such a way that all the cases with the event of interest come before the cases without the event. Note that when shuffle=T, the random number seed is reset to 1 each time this function is called. This ensures that replicable results can be obtained even when the order of observations is randomized.
width	Width of the plot space (in inches).
height	Height of the plot space (in inches).

col0	Color of the predicted probabilities corresponding to 0s in the `actual` vector. The default color has been chosen from one of the palettes on https://colorbrewer2.org/.
col1	Color of the predicted probabilities corresponding to 1s in the `actual` vector. The default color has been chosen from one of the palettes on https://colorbrewer2.org/.
flag	A vector of row number(s) in the `actual` vector corresponding to the observations to flag.
flagcol	A vector of colors for the flags.
file	The name and file path of where the pdf output should be written, if desired. If `file=NULL` the output will be written to the screen.
newplot	Should a new plotting space be opened up for the separation plot? Select `newplot=F` if you want the separation plot to be added to currently open output device.
locate	Number of lines (if any) on the separation plot that you want to identify with the mouse using the `locator` function.
rectborder	When `type="rect"`, the value of this argument is passed to the `border` argument of the `rect` function used to draw the line segments. The default setting (`rectborder=NA`) suppresses the drawing of borders around the individual segments of the plot.
show.expected	If `show.expected=T`, a marker is added to the plot showing the expected total number of events. The expected number of events is calculated by simply summing (and rounding) the predicted probabilities over all observations.
zerosfirst	When `type="line"`, should the 0s be plotted in the background, and the 1s in the foreground, or vice-versa? This will affect the output when the number of observations is very large relative to the size of the plot.
BW	Should the Black and White color scheme be implemented?

Details

Please see the paper by Greenhill, Ward and Sacks (2011) for more information on the features of the separation plot.

Value

resultsmatrix	The dataframe containing the data used to generate the separation plot. The first column is the vector of predicted probabilities, the second is the vector of actual outcomes, the third indicates which observations have been flagged using the `flag` argument above, the fourth gives the position of each observation on the horizontal axis of the separation plot, and the fifth gives the color used to plot each observation.

Author(s)

Brian Greenhill <<bgreenhill@albany.edu>>

References

Greenhill, Brian, Michael D. Ward, and Audrey Sacks. "The separation plot: A new visual method for evaluating the fit of binary models." American Journal of Political Science 55.4 (2011): 991-1002.

See Also

See sp.categorical for plotting separation plots for models with polytomous dependent variables.

Examples

```
# Create a separation plot for a simple logit model:

library(MASS)
set.seed(1)
Sigma <- matrix(c(1,0.78,0.78,1),2,2)
a<-(mvrnorm(n=500, rep(0, 2), Sigma))
a[,2][a[,2]>0.75]<-1
a[,2][a[,2]<=0.75]<-0
a[,1]<-a[,1]-min(a[,1])
a[,1]<-a[,1]/max(a[,1])

cor(a) # should be 0.55

model1<-glm(a[,2]~a[,1], family=binomial(link = "logit"))

library(Hmisc)
somers2(model1$fitted.values, model1$y)

separationplot(pred=model1$fitted.values, actual=model1$y, type="rect",
line=TRUE, show.expected=TRUE, heading="Example 1")
```

| sp.categorical | *Separation plots for variables with more than two outcome levels* |

Description

This function generates separation plots for polytomous dependent variables.

Usage

```
sp.categorical(pred, actual, file = NULL, cex = 1.5, ...)
```

Arguments

pred	A matrix of fitted values. Each row represents one observation, and each column represents the probability of obtaining that outcome. The column names correspond to the outcome categories.
actual	A vector containing the actual outcomes corresponding to each observation.
file	The name and file path of where the pdf output should be written, if desired. If file=NULL the output will be written to the screen.
cex	Character expansion factor used for the outcome category labels.
...	Additional arguments passed to separationplot.

Details

This function is a wrapper for separationplot that generates a series of separation plots for each outcome category for a variable with more than two outcomes.

Please see the paper by Greenhill, Ward and Sacks for more information on the features of the separation plot.

Value

None. This function is used for its side effects only.

Author(s)

Brian Greenhill <<bgreenhill@albany.edu>>

References

Greenhill, Brian, Michael D. Ward, and Audrey Sacks. "The separation plot: A new visual method for evaluating the fit of binary models." American Journal of Political Science 55.4 (2011): 991-1002.

See Also

See separationplot for a description of the core function for generating separation plots.

Examples

```
# This example borrows code from the example given in the documentation for the polr() function
# that uses the "housing" dataset:
options(contrasts = c("contr.treatment", "contr.poly"))
house.plr <- polr(Sat ~ Infl + Type + Cont, weights = Freq, data = housing)

sp.categorical(pred=house.plr$fitted.values,
actual=as.character(house.plr$model[,1]), type="rect", lwd2=2)
 # not a very good fit!
```

Index