

Package ‘IPWboxplot’

January 2, 2019

Type Package

Title Adapted Boxplot to Missing Observations

Version 0.1.0

Author Ana Maria Bianco [aut], Graciela Boente [aut], Ana Perez-Gonzalez [aut]

Maintainer Ana Perez-Gonzalez <anapg@uvigo.es>

Description Boxplots adapted to the happenstance of missing observations where drop-out probabilities can be given by the practitioner or modelled using auxiliary covariates. The paper of “Zhang, Z., Chen, Z., Troendle, J. F. and Zhang, J.(2012) <doi:10.1111/j.1541-0420.2011.01712.x>”, proposes estimators of marginal quantiles based on the Inverse Probability Weighting method.

Imports isotone

Suggests mice, knitr, rmarkdown

License GPL (>= 2)

NeedsCompilation no

Encoding UTF-8

Repository CRAN

VignetteBuilder knitr

Date/Publication 2019-01-02 13:30:10 UTC

R topics documented:

IPW.ASYM.boxplot	2
IPW.boxplot	5
IPW.quantile	8

Index	10
--------------	-----------

 IPW.ASYM.boxplot

Boxplot adapted to skewness and missing values

Description

The function draws a modified boxplot adapted to missing data and skewness. The drop-out probabilities can be given by the practitioner or fitted through a logistic model using auxiliary covariates. The plots are adapted to asymmetric distributions by correcting the whiskers through a measure of the skewness.

Usage

```
IPW.ASYM.boxplot(y,px=NULL,x=NULL,graph=c("IPW","both"),names=c("IPW Asymmetric Boxplot",
  "NAIVE Asymmetric Boxplot"), size.letter=1.2,
  method=c("quartile","octile"), ctea=-4, cteb=3, lim.inf=NULL,lim.sup=NULL,
  main=" ",xlab = " ", ylab = " ",color="black")
```

Arguments

y	Numerical vector of length n with possible missing values codified by NA or NAN.
px	Optional. Numerical vector of drop-out probabilities. If not provided a logistic fit is performed using x as predictive variable. Missing values are not admitted.
x	Optional. The matrix of fully observed variables used to estimate the missing model with dimension nrows=n and ncol=p. Missing values are not admitted. One of the vectors px or x must be supplied.
graph	Optional. Character string indicating if the plot contains two boxplots ("both") or only the boxplot computed with the inverse probability weighted quantiles ("IPW"). The default is "IPW".
names	Optional. Character string to name the boxplots. The default is "IPW Asymmetric Boxplot", when graph="IPW" and c("IPW Asymmetric Boxplot", "NAIVE Asymmetric Boxplot") when graph="both".
size.letter	Optional. The font size of names. Default value is 1.2
method	Optional. Character string indicating if the skewness measure is based on the quartiles ("quartile") or the octiles ("octile"). The default is "quartile".
ctea	Optional. Scaling factors to compute the outlier boundary. The default is -4.
cteb	Optional. Scaling factors to compute the outlier boundary. The default is 3. When ctea=cteb=0 the IPW boxplot for symmetric data is obtained.
lim.inf	Optional. The lower limit of the plot if supplied by the user.
lim.sup	Optional. The upper limit of the plot if supplied by the user.
main	Optional. Character string to title the plot. By default no main title is given.
xlab	Optional. Character string to indicate the label of the horizontal axis.
ylab	Optional. Character string to indicate the label of the vertical axis.
color	Optional. Color for the IPW Boxplot.

Details

The function draws boxplots designed to adjust both for skewness and missingness. The drop-out probabilities can be supplied by the user or estimated through a logistic model from given covariates.

The function plots as default a modified boxplot based on the inverse probability weighted (IPW) quantiles adapting for missing observations as in Zhang et al.(2012), but using a correction factor to adjust for skewness. For that purpose, the function incorporates a skewness measure to compute the whiskers and the outlier cut-off values in a similar way to that considered in Hubert and Vandervieren (2008). The argument `method` selects quartiles (`method="quartile"`) or octiles (`method="octile"`) to calculate the skewness measure *SKEW*, respectively, as

$$SKEW = \frac{(Q_{0.75} - Q_{0.5}) - (Q_{0.5} - Q_{0.25})}{(Q_{0.75} - Q_{0.25})},$$

$$SKEW = \frac{(Q_{0.875} - Q_{0.5}) - (Q_{0.5} - Q_{0.125})}{(Q_{0.875} - Q_{0.125})},$$

where Q_α denotes the α -quantile.

The whiskers and the outlier cut-off values are computed by means of an exponential model in the fashion of Hubert and Vandervieren (2008) taking into account the interval:

$$(Q_{0.25} - 1.5 * \exp(c_i * SKEW) * IQR, Q_{0.75} + 1.5 * \exp(c_s * SKEW) * IQR),$$

where $IQR = Q_{0.75} - Q_{0.25}$ and $c_i = c_{tea}$ and $c_s = c_{teb}$ if *SKEW* is positive, otherwise, $c_i = -c_{teb}$ and $c_s = -c_{tea}$.

The default values for c_{tea} and c_{teb} are -4 and 3 , however, the user may choose other values for these constants.

By specifying `graph = "both"`, the function displays two parallel modified boxplots. The boxplot on the left corresponds to the IPW boxplot adapted for missingness and skewness, while that on the right, to its naive counterpart which is simply based on the observations y at hand without any correction for missingness.

The user can supply a vector of drop-out probabilities px or a set of covariates x to estimate the propensity. When both px and x are supplied, the IPW.ASYM.boxplot is executed using px . When px is not given, it is estimated assuming a logistic model depending on the covariates x . For more details, see Bianco et al. (2018).

Value

The output of the function is a list with components:

<code>px</code>	Numerical vector of drop-out probabilities.
<code>IPW.Quartiles</code>	Numerical vector of inverse probability weighted quartiles.
<code>IPW.whisker</code>	Numerical vector of lower and upper whiskers calculated from IPW quantiles.
<code>out.IPW</code>	Numerical vector of data points detected as atypical by the IPW boxplot adapted to skewness.

SKEW.IPW	Skewness measure based on the IPW quartiles (method="quartile") or IPW octiles (method="octile").
NAIVE.Quartiles	Numerical vector of naive quartiles computed from the subset of non-missing values of y . Returned only when graph="both".
NAIVE.whisker	Numerical vector of lower and upper whiskers obtained from the naive quantiles. Returned only when graph="both".
out.NAIVE	Numerical vector of data points detected as atypical by the naive boxplot. Returned only when graph="both".
SKEW.NAIVE	Skewness measure based on the naive quartiles (method="quartile") or Naive octiles (method="octile"), computed from the subset of non-missing values of y . Returned only when graph="both".

Note

The missing values of y must be codified as NA or NAN.

The numerical vector px and the matrix of covariates x must be fully observed. px or x must be supplied by the user.

The lengths of y , px , and $nrow(x)$ must be equal.

Author(s)

Ana Maria Bianco <abianco@dm.uba.ar>, Graciela Boente <gboente@dm.uba.ar> and Ana Perez-Gonzalez <anapg@uvigo.es>.

References

Bianco, A. M., Boente, G., and Perez-Gonzalez, A. (2018). A boxplot adapted to missing values: an R function when predictive covariates are available. Submitted.

Hubert, M. and Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics & Data Analysis*, 52, 5186-5201.

Zhang, Z., Chen, Z., Troendle, J. F. and Zhang, J. (2012). Causal inference on quantiles with an obstetric application. *Biometrics*, 68, 697-706.

See Also

IPW.quantile, IPW.Boxplot

Examples

```
## A real data example

library(mice)
data(boys)
attach(boys)

# The function plots the IPW boxplot adapted to skewness.
# Some statistical summaries computed using the inverse probability weighting approach
# are also returned.
res1=IPW.ASYM.boxplot(hc,x=age,main="IPW boxplot adjusted for skewness of the head circumference")

# We can compare the naive and IPW approaches. We also can consider the skewness measure computed
# using the quartiles (as default).
res2=IPW.ASYM.boxplot(hc,x=age,method="quartile",graph="both",main=" ")

# The results obtained if the skewness measure is computed with the octiles (method="octile") are:
res3=IPW.ASYM.boxplot(hc,x=age,method="octile",graph="both",main=" ")
```

IPW.boxplot

Boxplot adapted to missing values

Description

The function draws a modified boxplot adapted to missing values. The drop-out probabilities can be given by the practitioner or fitted through a logistic model using auxiliary covariates. The function returns the usual boxplot of the available data as well as a modified plot which takes into account the missing data model and weights the observations using the estimated/given propensity.

Usage

```
IPW.boxplot(y,px=NULL,x=NULL,graph=c("IPW","both"),
names=c("IPW Boxplot","NAIVE Boxplot"),size.letter=1.2,
lim.inf=NULL,lim.sup=NULL,main=" ",xlab=" ",ylab=" ",color="black")
```

Arguments

y	Numerical vector of length n with possible missing values codified by NA or NAN.
px	Optional. Numerical vector of drop-out probabilities. If not provided a logistic fit is performed using x as predictive variable. Missing values are not admitted.
x	Optional. The matrix of fully observed variables used to estimate the missing model with dimension nrow=n and ncol=p. Missing values are not admitted. One of the vectors px or x must be supplied.

graph	Optional. Character string indicating if the plot contains two boxplots ("both") or only the boxplot computed with the inverse probability weighted quantiles("IPW"). The default is "IPW".
names	Optional. Character string to name the boxplots. The default is "IPW Boxplot", when graph="IPW" and c("IPW Boxplot", "NAIVE Boxplot") when graph="both".
size.letter	Optional. The font size of names. Default value is 1.2
lim.inf	Optional. The lower limit of the plot if supplied by the user.
lim.sup	Optional. The upper limit of the plot if supplied by the user.
main	Optional. Character string to title the plot. By default no main title is given.
xlab	Optional. Character string to indicate the label of the horizontal axis.
ylab	Optional. Character string to indicate the label of the vertical axis.
color	Optional. Color for the IPW Boxplot.

Details

The function draws boxplots designed to adjust for missing values. The propensity can be supplied by the user or estimated through a logistic model from given covariates.

The function plots as default a modified boxplot based on the inverse probability weighted (IPW) quantiles adapting for missing observations as in Zhang et al.(2012).

By specifying graph = "both", the function displays two parallel boxplots. The boxplot on the left corresponds to the IPW boxplot adapted for missingness, while on the right, the naive boxplot, i.e., the usual boxplot simply computed with the observations y at hand, is displayed.

The user can supply a vector of probabilities px or a set of covariates x to estimate it. When both px and x are supplied, the IPW.boxplot is executed using px. When px is not supplied, it is estimated assuming a logistic model depending on the covariates x. For more details, see Bianco et al. (2018).

Value

The output of the function is a list with components:

px	Numerical vector of probabilities.
IPW.Quartiles	Numerical vector of inverse probability weighted quartiles.
IPW.whisker	Numerical vector of lower and upper whiskers calculated from IPW quartiles.
out.IPW	Numerical vector of data points detected as atypical by the IPW boxplot.
NAIVE.Quartiles	Numerical vector of naive quartiles computed from the subset of non-missing values of y. Returned only when graph="both".
NAIVE.whisker	Numerical vector of lower and upper whiskers obtained from the naive quartiles. Returned only when graph="both".
out.NAIVE	Numerical vector of data points detected as atypical by the naive boxplot. Returned only when graph="both".

Note

The missing values of y must be codified as NA or NAN.

The numerical vector px and the matrix of covariates x must be fully observed. px or x must be supplied by the user.

The lengths of y , px , and $nrow(x)$ must be equal.

Author(s)

Ana Maria Bianco <abianco@dm.uba.ar>, Graciela Boente <gboente@dm.uba.ar> and Ana Perez-Gonzalez <anapg@uvigo.es>.

References

Bianco, A. M., Boente, G., and Perez-Gonzalez, A. (2018). A boxplot adapted to missing values: an R function when predictive covariates are available. Submitted.

Zhang, Z., Chen, Z., Troendle, J. F. and Zhang, J. (2012). Causal inference on quantiles with an obstetric application. *Biometrics*, 68, 697-706.

See Also

IPW.quantile, IPW.ASYM.Boxplot

Examples

```
## A real data example

library(mice)
data(boys)
attach(boys)

res1=IPW.boxplot(tv,x=age,main="IPW boxplot of the testicular volume")

# We can compare the naive and IPW boxplots
res2=IPW.boxplot(tv,x=age,graph="both",main=" ")
```

IPW.quantile *Computes the IPW quantiles*

Description

The function calculates the inverse probability weighted quantiles of a numeric vector.

Usage

```
IPW.quantile(y, px=NULL,x=NULL,probs = seq(0, 1, 0.25))
```

Arguments

<code>y</code>	Numerical vector of length <code>n</code> with possible missing values codified by NA or NAN.
<code>px</code>	Optional. Numerical vector of drop-out probabilities. If not provided a logistic fit is performed using <code>x</code> as predictive variable. Missing values are not admitted.
<code>x</code>	Optional. The matrix of fully observed variables used to estimate the missing model with dimension <code>nrows=n</code> and <code>ncol=p</code> . Missing values are not admitted. One of the vectors <code>px</code> or <code>x</code> must be supplied.
<code>probs</code>	Required. Numeric vector of probabilities with values in (0,1).

Details

The function computes inverse probability weighted (IPW) quantiles of a numeric vector `y` adapting for missing observations as in Zhang et al.(2012).

The user can supply a vector of drop-out probabilities `px` or a set of covariates `x` to estimate the propensity. When both `px` and `x` are supplied, the `IPW.quantile` is executed using `px`. When `px` is not supplied, the happenstance probabilities are estimated assuming a logistic model depending on the covariates `x`. For more details, see Bianco et al. (2018).

We adapted the function `weighted.fractile` from the **isotone** package to missing values in variable `y`. See **isotone** for more details.

Value

The output of the function is a list with components:

ipw.quantile Numerical vector of length `length(probs)` containing the estimated quantiles.

px Numerical vector of drop-out probabilities.

Note

The missing values of `y` must be codified as NA or NAN.

The numerical vector `px` and the matrix of covariates `x` must be fully observed. `px` or `x` must be supplied by the user.

The lengths of `y`, `px`, and `nrow(x)` must be equal.

Author(s)

Ana Maria Bianco <abianco@dm.uba.ar>, Graciela Boente <gboente@dm.uba.ar> and Ana Perez-Gonzalez <anapg@uvigo.es>.

References

Bianco, A. M., Boente, G. and Perez-Gonzalez, A. (2018). A boxplot adapted to missing values: an R function when predictive covariates are available. Submitted.

Zhang, Z., Chen, Z., Troendle, J. F. and Zhang, J. (2012). Causal inference on quantiles with an obstetric application. *Biometrics*, 68, 697-706.

Examples

```
## A real data example
library(mice)
data(boys)
attach(boys)
# As an illustration, we consider variable testicular volume, tv.
# To compute the inverse probability weighted (IPW) quantiles
# the covariate age is considered as covariate with predictive capability
# to estimate the vector of drop-out probabilities.

res=IPW.quantile(tv,x=age,probs=c(0.25,0.5,0.75))
res$IPW.quantile

# Compute the inverse probability weighted (IPW) quantiles
# corresponding to the fractiles 0.3, 0.8 and 0.9
# using the covariate age to estimate the propensity.

res1=IPW.quantile(tv,x=age,probs=c(0.3,0.8,0.9))
res1$IPW.quantile
```

Index

*Topic **boxplot**

IPW.ASYM.boxplot, 2

IPW.boxplot, 5

*Topic **inverse probability weighted**

IPW.ASYM.boxplot, 2

IPW.boxplot, 5

IPW.quantile, 8

*Topic **missing**

IPW.ASYM.boxplot, 2

IPW.boxplot, 5

IPW.quantile, 8

*Topic **quantile**

IPW.ASYM.boxplot, 2

IPW.boxplot, 5

IPW.quantile, 8

IPW.ASYM.boxplot, 2

IPW.boxplot, 5

IPW.quantile, 8