Package 'gR2'

March 12, 2020

Type Package

Title Generalized R Square Measures for a Mixture of Bivariate Linear
    Dependences

Version 1.2.0

Author Jingyi Jessica Li, Heather J Zhou

Maintainer
    Jingyi Jessica Li <jli@stat.ucla.edu>, Heather J Zhou <heatherjzhou@ucla.edu>

Description This package implements the estimation and inference of generalized R square.

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.0.2

LinkingTo Rcpp, RcppArmadillo

Imports Rcpp, parallel, mvtnorm

R topics documented:

---

gR2		*gR2*

Description

gR2 calculates the sample gR2 under the specified scenario, the unspecified scenario (K
chosen), and the unspecified scenario (K not chosen). It also provides an option to perform
statistical inference on the population gR2.

Usage

```
gR2(
  x,
  y,
  z = NULL,
  K = NULL,
  cand.Ks = 1:4,
  nstart = 30,
  mc.cores = parallel::detectCores() - 1,
  regressionMethod = "MA",
  inference = FALSE,
  conf.level = 0.95,
  method = "general"
)
```

Arguments

x	A numeric vector.
y	A numeric vector of the same length as x.
z	A vector of integers that represents the line membership of all the data points. Must be of the same length as x and y. Default is NULL.
K	Number of lines in the unspecified scenario. Default is NULL.
cand.Ks	A vector of positive integers that represents the candidate K's in the unspecified scenario. Default is 1:4.
nstart	Number of initializations for the K-lines algorithm in the unspecified scenario. Default is 30.
mc.cores	Number of cores to use in the unspecified scenario. Default is the number of CPU cores minus one.
regressionMethod	
	Valid values are "MA" and "LM". Indicates which regression method to use in the K-lines algorithm - major axis regression or linear regression. Default is "MA".
inference	Logical. If TRUE, then a confidence interval for the population gR2 of confidence level conf.level will be calculated. Also will be calculated is a p-value of the hypothesis test where the null hypothesis is that the population gR2 is 0 and the alternative hypothesis is that the population gR2 is greater than 0. Default is FALSE.
conf.level	The confidence level of the confidence interval. See description of inference. Default is 0.95.
method	Valid values are "general" and "binorm". Indicates which asymptotic distribution of the sample gR2 to use for inference. Default is "general".

Details

The arguments that require user input are x and y, which must be numeric vectors of the same length.

There are three broad types of scenarios: the specified scenario, the unspecified scenario (K chosen), and the unspecified scenario (K not chosen). The specified scenario is considered when z is provided; the unspecified scenario (K chosen) is considered when z is not provided but K is provided; and the unspecified scenario (K not chosen) is considered when neither z or K is provided.

In the unspecified scenario (K chosen), we recommend that users set K to be less than or equal to 4 for interpretability.

In the unspecified scenario (K not chosen), the gR2 function will automatically choose a K value from cand.Ks using the Akaike information criterion (AIC). Two plots will be outputted: (1) a scree plot that shows how average squared perpendicular/vertical distance changes with the candidate K, and (2), a plot that shows how AIC changes with the candidate K. Users can decide whether the K value chosen by the gR2 function is reasonable by checking these two plots.

Value

gR2 returns a list consisting of one or more of the following items:

estimate	The sample gR2.
conf.level	The confidence level of the confidence interval (if inference is TRUE).
conf.int	The confidence interval for the population gR2 (if inference is TRUE).
p.val	The p-value of the hypothesis test where the null hypothesis is that the population gR2 is 0 and the alternative hypothesis is that the population gR2 is greater than 0 (if inference is TRUE).
K	The number of lines in the unspecified scenario, either chosen by the user or chosen from cand.Ks by the gR2 function.
membership	The inferred line membership of all the data points in the unspecified scenario.

Author(s)

Jingyi Jessica Li, <jli@stat.ucla.edu>

Heather J Zhou, <heatherjzhou@ucla.edu>

References

Li, J.J., Tong, X., and Bickel, P.J. (2019). Generalized R2 Measures for a Mixture of Bivariate Linear Dependences. arXiv.