Package ‘practicalSigni’

March 29, 2023

Type Package

Title Practical Significance Ranking of Regressors

Version 0.1.1

Date 2023-03-29

Encoding UTF-8

Depends R (>= 4.2.0), np (>= 0.60), xtable (>= 1.8), generalCorr (>= 1.2), NNS (>= 0.9), randomForest (>= 4.7),

Suggests R.rsp

VignetteBuilder R.rsp

Description Consider a possibly nonlinear nonparametric regression with p regressors. We provide evaluations by 13 methods to rank regressors by their practical significance or importance using various methods, including machine learning tools. Comprehensive methods are as follows.
m6=Generalized partial correlation coefficient or
GPCC by Vinod (2021)<doi:10.1007/s10614-021-10190-x> and
Vinod (2022)<https://www.mdpi.com/1911-8074/15/1/32>.
m7= a generalization of psychologists' effect size incorporating nonlinearity and many variables.
m8= local linear partial (dy/dxi) using the 'np' package for kernel regressions.
m9= partial (dy/dxi) using the 'NNS' package.
m10= importance measure using the 'NNS' boost function.
m11= Shapley Value measure of importance (cooperative game theory).
m12 and m13= two versions of the random forest algorithm.

License GPL (>= 2)

RoxygenNote 7.2.3

NeedsCompilation no

Author Hrishikesh Vinod [aut, cre]

Maintainer Hrishikesh Vinod <vinod@fordham.edu>

Repository CRAN

Date/Publication 2023-03-29 17:10:05 UTC

R topics documented:

`effSizCut` *Compute Effect Sizes for continuous or categorical data*

Description

Psychologists' so-called "effect size" reveals the practical significance of only one regressor. This function generalizes their algorithm to two or more regressors (p>2). Generalization first converts the xi regressor into a categorical treatment variable with only two categories. One imagines that observations larger than the median (xit> median(xi)) are "treated," and those below the median are "untreated." The aim is the measure the size of the (treatment) effect of (xi) on y. Denote other variables with postscript "o" as (xo). Since we have p regressors in our multiple regression, we need to remove the nonlinear kernel regression effect of other variables (xo) on y while focusing on the effect of xi. There are two options in treating (xo) (i) letting xo be as they are in the data (ii) converting xo to binary at the median. One chooses the first option (i) by setting the logical argument ane=TRUE in calling the function. ane=TRUE is the default. Set ane=FALSE for the second option.

Usage

```
effSizCut(y, bigx, ane = TRUE)
```

Arguments

`y`	(T x 1) vector of dependent variable data values
`bigx`	(T x p) data matrix of xi regressor variables associated with the regression
`ane`	logical variable controls the treatment of other regressors. If ane=TRUE (default), other regressors are used in kernel regression without forcing them to be binary variables. When ane=FALSE, the kernel regression removes the effect of other regressors when other regressors are also binary type categorical variables,

Value

out vector with p values of t-statistics for p regressors

Note

The aim is to answer the following question. Which regressor has the largest effect on the dependent variable? We assume that the signs of regressors are already adjusted such that a numerically larger effect size suggests that the corresponding regressor is most important, having the largest effect size in explaining y the dependent variable.

Author(s)

Prof. H. D. Vinod, Economics Dept., Fordham University, NY

See Also

pracSig13

Examples

```
set.seed(9)
 y=sample(1:15,replace = TRUE)
 x1=sample(2:16, replace = TRUE)
 x2=sample(3:17, replace = TRUE)
effSizCut(y,bigx=cbind(x1,x2),ane=TRUE)
```

`fncut` *fncut auxiliary converts continuous data into two categories*

Description

This is an internal function of the R package practicalSigni Psychologists use effect size to evaluate the practical importance of a treatment on a dependent variable using a binary [0,1] variable. Assuming numerical data, we can always compute the median and regard values < or = the median as zero and other values as unity.

Usage

```
fncut(x)
```

Arguments

x	numerical vector of data values

Value

x vector of zeros and ones split at the median

Author(s)

Prof. H. D. Vinod, Fordham University, NY pracSig13 *Compute thirteen measures of practical significance*

Description

Thirteen methods are denoted m1 to m13. Each yields p numbers when there are p regressors denoted xi. m1=OLS coefficient slopes. m2= t-stat of each slope. m3= beta coefficients OLS after all variables have mean zero and sd=1. m4= Pearson correlation coefficient between y and xi (only two variables at a time, assuming linearity). Let r*(y|xi) denote the generalized correlation coefficient allowing for nonlinearity from Vinod (2021, 2022). It does not equal analogous r*(xi|y). The larger of the two, max(r*(y|xi), r*(xi|y)), is given by the function depMeas() from the 'generalCorr' package. m5= depMeas, which allows nonlinearity. m5 is not comprehensive because it measures only two variables, y and xi, at a time. m6= generalized partial correlation coefficient or GPCC. This is the first comprehensive measure of practical significance. m7=a generalization of psychologists' "effect size" after incorporating the nonlinear effect of other variables. m8= local linear partial (dy/dxi) using the 'np' package for kernel regressions and local linear derivatives. m9= partial derivative (dy/dxi) using the 'NNS' package. m10=importance measure using NNS.boost() function of 'NNS.' m11=Shapley Value measure of importance (cooperative game theory). m12 and m13= two versions of the random forest algorithm measuring the importance of regressors.

Usage

```
pracSig13(y, bigx, yes13 = rep(1, 13), verbo = FALSE)
```

Arguments

`y`	input dependent variable data as a vector
`bigx`	input matrix of p regressor variables
`yes13`	default vector of ones to compute all 13 measures. e.g., yes13[10]=0 means do not compute the m10 method.
`verbo`	logical to print results along the way default=FALSE

Details

If m6, m10 slow down computations, we recommend setting yes13[6]=0=yes13[10] to turn of computation of m6 and m10 .

Value

output matrix (p x 13) containing m1 to m13 criteria (numerical measures of practical significance) along columns and a row for each regressor (excluding the intercept).

Note

needs the function kern(), which requires package 'np'. also needs 'NNS', 'randomForest', packages.

The machine learning methods are subject to random seeds. For some seed values, m10 values from NNS.boost() become degenerate and are reported as NA or missing. In that case the average ranking output r613 from reportRank() needs manual adjustments.

Author(s)

Prof. H. D. Vinod, Economics Dept., Fordham University, NY

References

Vinod, H. D."Generalized Correlation and Kernel Causality with Applications in Development Economics" in Communications in Statistics -Simulation and Computation, 2015, doi:10.1080/03610918.2015.1122048

Vinod, H. D.", "Generalized Correlations and Instantaneous Causality for Data Pairs Benchmark," (March 8, 2015). `https://www.ssrn.com/abstract=2574891`

Vinod, H. D. "Generalized, Partial and Canonical Correlation Coefficients," Computational Economics (2021) SpringerLink vol. 59, pp.1-28. URL https://link.springer.com/article/10.1007/s10614-021-10190-x

Vinod, H. D. "Kernel regression coefficients for practical significance," Journal of Risk and Financial Management 15(1), 2022 pp.1-13. https://doi.org/10.3390/jrfm15010032

Vinod, H. D.", "Hands-On Intermediate Econometrics Using R" (2022) World Scientific Publishers: Hackensack, NJ. `https://www.worldscientific.com/worldscibooks/10.1142/12831`

See Also

See Also as `effSizCut`,

See Also as `reportRank`

`reportRank` *Function to report ranks of 13 criteria for practical significance*

Description

This function generates a report based on the regression of y on bigx. It acknowledges that some methods for evaluating the importance of regressor in explaining y may give the importance value with a wrong (unrealistic) sign. For example, m2 reports t-values. Imagine that due to collinearity m2 value is negative when the correct sign from prior knowledge of the subject matter is that the coefficient should be positive, and hence the t-stat should be positive. The wrong sign means the importance of regressor in explaining y should be regarded as relatively less important. The larger the absolute size of the t-stat, the less its true importance in measuring y. The ranking of coefficients computed here suitably deprecates the importance of the regressor when its coefficient has the wrong sign (perverse direction).

Usage

```
reportRank(
  y,
  bigx,
  yesLatex = 1,
  yes13 = rep(1, 13),
  bsign = 0,
  dig = 3,
  verbo = FALSE
)
```

Arguments

y	(T x 1) vector of dependent variable data values
bigx	(T x p) data marix of xi regressor variables associated with the regression
yesLatex	default 1 means print Latex-ready Tables
yes13	default vector of ones to compute all 13 measures.
bsign	A (p x 1) vector of right signs of regression coefficients. Default is bsign=0 means the right sign is the same as the sign of the covariance, cov(y, xi)
dig	digits to be printed in latex tables, default, dig=d33
verbo	logical to print results by pracSig13, default=FALSE

Value

v15	practical significance index values (sign adjusted) for m1 to m5 using older linear and /or bivariate methods
v613	practical significance index values for m6 to m13 newer comprehensive and nonlinear methods
r15	ranks and average rank for m1 to m5 using older linear and /or bivariate methods
r613	ranks and average rank for m6 to m13 newer comprehensive and nonlinear methods

Note

The machine learning methods are subject to random seeds. For some seed values, m10 values from NNS.boost() rarely become degenerate and are reported as NA or missing. In that case the average ranking output r613 here needs adjustment.

Author(s)

Prof. H. D. Vinod, Economics Dept., Fordham University, NY

See Also

pracSig13

Examples

```
set.seed(9)
y=sample(1:15,replace = TRUE)
x1=sample(2:16, replace = TRUE)
x2=sample(3:17, replace = TRUE)
x3=sample(4:18,replace = TRUE)
options(np.messages=FALSE)
yes13=rep(1,13)
yes13[10]=0
reportRank(y,bigx=cbind(x1,x2,x3),yes13=yes13)
```

`shapleyvalue` *Shapley Value Regression importance of attributes in linear regression*

Description

'ShapleyValue' package function shapleyvalue() calculates the relative importance of independent variables in linear regression while avoiding collinearity. This function is a minor modification of the function in the package ShapleyValue. The original function fails for me when a variable name is x1. Since Jingyi Liang, the maintainer of 'ShapleyValue,' has ignored several requests to fix the error, this function is reluctantly included here as a separate function.

Usage

```
shapleyvalue(y, x)
```

Arguments

y	input dependent variable data as a vector
x	input matrix of p regressor variables must be a data frame

Value

The structure of the output is a datatable, with two rows: the unstandardized and standardized relative importance of each attributes using shapley value regression method.

Author(s)

H. D. Vinod, Fordham University, NY

References

Liang J (2021). ShapleyValue: Shapley Value Regression for Relative Importance of Attributes. R package version 0.2.0, https://CRAN.R-project.org/package=ShapleyValue.

Index