

Package ‘practicalSigni’

December 1, 2023

Type Package

Title Practical Significance Ranking of Regressors and Exact t Density

Version 0.1.2

Date 2023-11-29

Encoding UTF-8

Depends R (>= 4.3.0), np (>= 0.60), generalCorr (>= 1.2),

Imports xtable (>= 1.8.4), ShapleyValue (>= 0.2.0), NNS (>= 0.9),
randomForest (>= 4.7), hypergeo (>= 1.2.13),

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](https://doi.org/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.

Taraldsen's exact density for sampling distribution of correlations added.

License GPL (>= 2)

RoxygenNote 7.2.3

NeedsCompilation no

Author Hrishikesh Vinod [aut, cre]

Maintainer Hrishikesh Vinod <vinod@fordham.edu>

Repository CRAN

Date/Publication 2023-12-01 09:50:02 UTC

R topics documented:

effSizCut	2
fncut	3
pracSig13	4
pvTarald	5
qTarald	6
reportRank	7

Index	9
--------------	----------

effSizCut	<i>Compute Effect Sizes for continuous or categorical data</i>
-----------	--

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 x_i regressor into a categorical treatment variable with only two categories. One imagines that observations larger than the median ($x_i > \text{median}(x_i)$) are "treated," and those below the median are "untreated." The aim is the measure the size of the (treatment) effect of (x_i) on y . Denote other variables with postscript "o" as (x_o). Since we have p regressors in our multiple regression, we need to remove the nonlinear kernel regression effect of other variables (x_o) on y while focusing on the effect of x_i . There are two options in treating (x_o) (i) letting x_o be as they are in the data (ii) converting x_o 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

<code>y</code>	A ($T \times 1$) vector of dependent variable data values.
<code>bigx</code>	A ($T \times p$) data matrix of x_i regressor variables associated with the regression.
<code>ane</code>	logical variable controls the treatment of other regressors. If <code>ane=TRUE</code> (default), other regressors are used in kernel regression without forcing them to be binary variables. When <code>ane=FALSE</code> , 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

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 $r^*(xily)$. The larger of the two, $\max(r^*(y|xi), r^*(xily))$, 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	vector of ones to compute respective 13 measures m1 to m13. Default is all ones to compute all 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 off slowcomputation of m6 and m10 at least initially to get quick answers for other m's.

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](#)

pvTarald

Compute the p-value for exact correlation significance test using Taraldsen's exact methods.

Description

Compute the p-value for exact correlation significance test using Taraldsen's exact methods.

Usage

pvTarald(n, rho = 0, obsr)

Arguments

n	number of observations, n-1 is degrees of freedom
rho	True unknown population correlation coefficient in the r-interval [-1, 1], default=0
obsr	observed r or correlation coefficient

Value

ans is the p-value or probability from sampling distribution of observing a correlation as extreme or more extreme than the input obsr or observed r.

Note

needs function hypergeo from the package of that name.

Author(s)

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

References

Taraldsen, G. "The Confidence Density for Correlation" Sankhya: The Indian Journal of Statistics 2023, Volume 85-A, Part 1, pp. 600-616.

See Also

See Also as [qTarald](#),

qTarald	<i>Compute the quantile for exact t test density using Taraldsen's methods</i>
---------	--

Description

Compute the quantile for exact t test density using Taraldsen's methods

Usage

```
qTarald(n, rho = 0, cum)
```

Arguments

n	number of observations, n-1 is degrees of freedom
rho	True unknown population correlation coefficient, default=0
cum	cumulative probability for which quantile is needed

Value

r quantile of Taraldsen's density for correlation coefficient.

Note

needs function `hypergeo::hypergeo()`. The quantiles are rounded to 3 places and computed by numerical methods.

Author(s)

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

References

Taraldsen, G. "The Confidence Density for Correlation" *Sankhya: The Indian Journal of Statistics* 2023, Volume 85-A, Part 1, pp. 600-616.

See Also

See Also as [pvTarald](#),

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

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

Value

<code>v15</code>	practical significance index values (sign adjusted) for <code>m1</code> to <code>m5</code> using older linear and/or bivariate methods
<code>v613</code>	practical significance index values for <code>m6</code> to <code>m13</code> newer comprehensive and nonlinear methods
<code>r15</code>	ranks and average rank for <code>m1</code> to <code>m5</code> using older linear and/or bivariate methods
<code>r613</code>	ranks and average rank for <code>m6</code> to <code>m13</code> 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)
x0=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(x0,x2,x3),yes13=yes13)
```


Index

- * **Forest**
 - pracSig13, 4
- * **GPCC**
 - pracSig13, 4
- * **Random**
 - pracSig13, 4
- * **Shapley**
 - pracSig13, 4
- * **Value**
 - pracSig13, 4

effSizCut, 2, 5

fncut, 3

pracSig13, 3, 4, 8

pvTarald, 5, 7

qTarald, 6, 6

reportRank, 5, 7