Package ‘depcoeff’

January 23, 2020

Title Dependency Coefficients
Type Package
Version 0.0.1
Date 2019-12-23
Author Eckhard Liebscher
Maintainer Eckhard Liebscher <eckhard.liebscher@hs-merseburg.de>

Description
Functions to compute coefficients measuring the dependence of two or more than two variables. The functions can be deployed to gain information about functional dependencies of the variables with emphasis on monotone functions. The statistics describe how well one response variable can be approximated by a monotone function of other variables. In regression analysis the variable selection is an important issue. In this framework the functions could be useful tools in modeling the regression function. Detailed explanations on the subject can be found in papers Liebscher (2014) <doi:10.2478/demo-2014-0004>; Liebscher (2017) <doi:10.1515/demo-2017-0012>; Liebscher (2019, submitted).

Depends R (>= 3.5.0)
Suggests MASS
License GPL-2
Encoding UTF-8
LazyData true
LinkingTo Rcpp
Imports Rcpp, copula
RoxygenNote 7.0.2
NeedsCompilation yes
Repository CRAN
Date/Publication 2020-01-23 16:20:05 UTC
R topics documented:

kendr . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
kendrs . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
spearr . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
spearrs . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
zetac . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
zetaci . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
zetapm . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

Index 10

kendr Kendall regression coefficient

Description

The function kendr evaluates the multivariate Kendall regression coefficient. It describes how well the target variable \( y \) can be fit by a function of regressor variables which is increasing w.r.t. some regressors and decreasing w.r.t. the other regressors.

Usage

kendr(x, y, direction=NULL, out=0)

Arguments

x data matrix of regressor variables
y data vector of the target variable
direction vector of length \( d \) (\( d \) is number of regressors), value 1 refers to regressors leading to increasing \( y \) whenever this regressor increases, value -1 refers to regressors leading to decreasing \( y \) whenever this regressor increases. If direction=NULL, then all coefficients are computed.
out value 1: full output, value 0: reduced output, only coefficients that are largest in absolute value

Value

list of Kendall regression coefficients for several directions

References

Eckhard Liebscher (2019). Kendall regression Coefficient. submitted

Examples

library(MASS)
data <- gilgaiskendr(data[,1:3],data[,4],out=1)
kendrs  

Kendall regression coefficient for split domains

Description

The function kendrs evaluates the multivariate Kendall regression coefficient for two regressors and split regressor region. It describes how well the target variable can be fit in each split region by a function which is increasing w.r.t. some regressors and decreasing w.r.t. the other regressors.

Usage

kendrs(x,y,splitp=NULL)

Arguments

x  
datamatrix of regressor variables with two columns,
y  
data vector of the target variable
splitp  
vector of length 2 of the splitting points, If p1 is the first component of this vector, then the point splits the domain of the first regressor into a left region of fraction p1 of data items and a right region of the remaining data items. The same is done for the second regressor. As the result we obtain 4 subregions of the regressor domain. default=c(0.5,0.5)

Value

list of Kendall regression coefficients for the 4 split regions and the total coefficient together with the corresponding optimal directions. direction ++ means that y increases whenever both regressors increases direction +- means that y increases whenever the first regressor increases and the other regressor decreases..etc.

References

Eckhard Liebscher (2019). Kendall regression coefficient. submitted

Examples

library(MASS)
data<- gilgais
kendrs(data[,1:2],data[,3],splitp=c(0.4,0.6))
spearr  
*Spearman regression coefficient*

**Description**

The function `spearr` evaluates the multivariate Spearman regression coefficient. It describes how well the target variable `y` can be fit by a function of regressor variables which is increasing w.r.t. some regressors and decreasing w.r.t. the other regressors.

**Usage**

```
spearr(x, y, direction=NULL, out=0)
```

**Arguments**

- `x` data matrix of regressor variables
- `y` data vector of the target variable
- `direction` vector of length `d` (d is number of regressors), value 1 refers to regressors leading to increasing `y` whenever this regressor increases, value -1 refers to regressors leading to decreasing `y` whenever this regressor increases. If direction=NULL, then all coefficients are computed.
- `out` value 1: full output, value 0: reduced output, only coefficients that are largest in absolute value

**Value**

list of Spearman regression coefficients for several directions

**References**


**Examples**

```
library(MASS)
data <- gilgais
spearr(data[,1:3], data[,4], out=1)
```
**spearrs**  
*Spearman regression coefficient for split domains*

**Description**

The function `spearrs` evaluates the multivariate Spearman regression coefficient for two regressors and split regressor region. It describes how well the target variable can be fit in each split region by a function which is increasing w.r.t. some regressors and decreasing w.r.t. the other regressors.

**Usage**

`spearrs(x, y, splitp=NULL)`

**Arguments**

- `x`: datamatrix of regressor variables with two columns,
- `y`: data vector of the target variable
- `splitp`: vector of length 2 of the splitting points. If p1 is the first component of this vector, then the point splits the domain of the first regressor into a left region of fraction p1 of data items and a right region of the remaining data items. The same is done for the second regressor. As the result we obtain 4 subregions of the regressor domain. default=c(0.5,0.5)

**Value**

list of Kendall regression coefficients for the 4 split regions and the total coefficient together with the corresponding optimal directions. direction ++ means that y increases whenever both regressors increases direction +- means that y increases whenever the first regressor increases and the other regressor decreases..etc.

**References**


**Examples**

```r
library(MASS)
data<- gilgais
spearrs(data[,1:2],data[,3],splitp=c(0.4,0.6))
```
zetac | Zeta dependence coefficient

Description

zetac is a function to evaluate the zeta dependence coefficient (one interval) of two random variables x and y which is based on the copula. Four specific coefficients are available: the Spearman coefficient, Spearman’s footrule, the power coefficient and the Huber function coefficient.

Usage

zetac(x, y, method = "Spearman", methodF = 1, parH = 0.5, parp = 1.5)

Arguments

- **x, y**: data vectors of the two variables whose dependence is analysed.
- **method**: list of names of the coefficients: "Spearman" stands for the Spearman coefficient, "footrule" means Spearman’s footrule, "power" stands for the power function coefficient, "Huber" means the Huber function coefficient. If "all" is assigned to method then all methods are used.
- **methodF**: value 1, 2 or 3 refers to several methods for computation of the distribution function values, 1 is the default value.
- **parH**: parameter of the Huber function (default 0.5). Valid values for parH are between 0 and 1.
- **parp**: parameter of the power function (default 1.5). The parameter has to be positive.

Details

Let $X_1, \ldots, X_n$ be the sample of the $X$ variable. Formulas for the estimators of values $F(X_i)$ of the distribution function: methodF = 1 $\rightarrow \hat{F}(X_i) = \frac{1}{n} \text{rank}(X_i)$ methodF = 2 $\rightarrow \hat{F}^1(X_i) = \frac{1}{n+1} \text{rank}(X_i)$ methodF = 3 $\rightarrow \hat{F}^2(X_i) = \frac{1}{\sqrt{n^2-1}} \text{rank}(X_i)$ The values of the distribution function of $Y$ are treated analogously.

Value

zeta dependence coefficient of two random variables. This coefficient is bounded by 1. The higher the value the stronger is the dependence.

References

Examples

```r
library(MASS)
data<- gilgais
zetac(data[,1],data[,2])
```

---

**Description**

The function `zetaci` evaluates the coefficient of piecewise monotonicity of variables `x` and `y` where the `x`-domain is split into a fixed number of intervals.

**Usage**

```r
zetaci(x,y,a,method="Spearman",methodF=1,parH=0.5,parp=1.5)
```

**Arguments**

- `x, y` data vectors of the two variables whose dependence is analysed.
- `a` vector of fractions \(a_i, 0 < a_i < a_{i+1} < 1\) for the splitting. A fraction of \(a_1, a_2 - a_1, a_3 - a_2\ldots\) of data points are in the corresponding split region. The number of split regions is equal to the length of `a` plus 1.
- `method` value (default "Spearman")
- `methodF` value 1, 2 or 3 refers to several methods for computation of the distribution function values, 1 is the default value.
- `parH` parameter of the Huber function (default 0.5). Valid values for `parH` are between 0 and 1.
- `parp` parameter of the power function (default 1.5). The parameter has to be positive.

**Details**

Let \(X_1, \ldots, X_n\) be the sample of the \(X\) variable. Formulas for the estimators of values \(\hat{F}(X_i)\) of the distribution function:

- `methodF = 1` \(\hat{F}(X_i) = \frac{\text{rank}(X_i)}{n}\)
- `methodF = 2` \(\hat{F}^2(X_i) = \frac{1}{\sqrt{n}^2 - 1} \text{rank}(X_i)\)

The values of the distribution function of \(Y\) are treated analogously.

**Value**

list of zeta dependence coefficients of piecewise monotonicity of two random variables containing the following elements: Spearman...Spearman coefficient footrule...Spearman’s footrule power...power coefficient Huber...Huber function coefficient

**References**

Examples

```r
library(MASS)
data<- gilgais
zetaci(data[, 1], data[, 2], a=c(0.25, 0.5, 0.75))
```

zetapm

Zeta dependence coefficient of piecewise monotonicity

Description

zetapm is a function to evaluate the zeta dependence coefficients of piecewise monotonicity of two random variables x and y which is based on the copula. The regressor domain (domain of x) is split into two parts. The function searches for the optimal splitting point to obtain maximum dependence. The main part of the function is coded as C++ procedure.

Usage

```r
zetapm(x,y,amin=0.25,method="all",methodF=1,parp=1.5,parH=0.5)
```

Arguments

- `x, y`: data vectors of the two variables whose dependence is analysed.
- `amin`: minimum fraction of sample items to be used for one split region.
- `method`: vector of chosen special coefficients: Spearman...Spearman coefficient footrule...Spearman’s footrule power...power coefficient Huber...Huber function coefficient, "all" refers to all coefficients.
- `methodF`: value 1,2 or 3 refers to several methods for computation of the distribution function values, 1 is the default value.
- `parp`: parameter of the power function (default 1.5). The parameter has to be positive.
- `parH`: parameter of the Huber function (default 0.5). Valid values for parH are between 0 and 1.

Details

Let $X_1, \ldots, X_n$ be the sample of the $X$ variable. Formulas for the estimators of values $\hat{F}(X_i)$ of the distribution function: methodF = 1 $\rightarrow \hat{F}(X_i) = \frac{1}{n} \text{rank}(X_i)$ methodF = 2 $\rightarrow \hat{F}^2(X_i) = \frac{1}{n} \text{rank}(X_i)$ methodF = 3 $\rightarrow \hat{F}^3(X_i) = \frac{1}{\sqrt{n-1}} \text{rank}(X_i)$ The values of the distribution function of $Y$ are treated analogously.

Value

list of zeta dependence coefficients (plusminus coefficient and minusplus one) of piecewise monotonicity of two random variables containing the following elements or a subset of it in this order: Spearman coefficient, footrule, power coefficient, Huber function coefficient. position1 and position2 indicate the number of the sample items where the optimized split point is located.
References


Examples

library(MASS)
data<- gilgais
zetapm(data[,1],data[,2])
Index

kendr, 2
kendrs, 3

spearr, 4
spearrs, 5

zetac, 6
zetaci, 7
zetapm, 8