Package 'HellCor'

October 12, 2022

Type Package

Title The Hellinger Correlation

Version 1.3

Date 2020-05-16

Author Gery Geenens [aut],
Pierre Lafaye De Micheaux [aut, cre]

Maintainer Pierre Lafaye De Micheaux <lafaye@unsw.edu.au>

Description Empirical value of the Hellinger correlation, a measure of dependence between two continuous random variables. More details can be found in Geenens and Lafaye De Micheaux (2019) <arXiv:1810.10276v4>.

License GPL (>= 2)

LazyLoad yes

LazyData true

Depends R (>= 2.10.0), energy, FNN, orthopolynom

Imports stats

NeedsCompilation yes

Repository CRAN

Date/Publication 2020-05-17 07:20:02 UTC

R topics documented:

`Chagos` *Coral fish, seabirds and reef productivity study by Graham et al (2018)*

Description

Data set investigated in Graham et al (2018). It contains several variables describing fish and seabirds populations on twelve islands of the Chagos Archipelago (British Indian Ocean Territory).

Usage

```
data(Chagos)
```

Format

A data frame with 12 observations on the following 7 variables.

- `Atoll` Factor with 3 levels. Atoll containing several islands.
- `Island` Factor with 12 levels. Name of the island.
- `Treatment` Factor with 2 levels. Presence or absence of rats on the island.
- `Seabirds_ha` Numeric vector of the number of birds per hectare of island.
- `kg_N_ha_yr` Numeric vector of nitrogen input (in kg) by seabirds per hectare of island over a one year period.
- `Number_of_fishes` Integer vector of the number of fishes recorded during underwater visual surveys along the reef crest of each island on the lagoonal side of each atoll.
- `biomass` numeric vector. Fish counts were converted into biomass using published length-weight relationships from FishBase.

Author(s)

Geenens G., Lafaye de Micheaux P.

References

Graham N. A. J., Wilson S. K., Carr P., Hoey A. S., Jennings S., MacNeil M. A. (2018). Seabirds enhance coral reef productivity and functioning in the absence of invasive rats. *Nature* 559, 250–253.

Examples

```
data(Chagos)
```

HellCor	*The Hellinger Correlation*

Description

Empirical value of the Hellinger correlation between two continuous random variables X and Y.

Usage

```
HellCor(x, y, Kmax = 20L, Lmax = 20L, K = 0L, L = 0L,
        alpha = 6.0, pval.comp = FALSE, conf.level = NULL,
        B1 = 200, B2 = 200, C.version = TRUE)
```

Arguments

`x`	Sample of X-values.
`y`	Sample of Y-values.
`Kmax`	Maximal number of terms to consider in the expansion of the square root of the copula density in the basis of Legendre polynomials.
`Lmax`	Maximal number of terms to consider in the expansion of the square root of the copula density in the basis of Legendre polynomials.
`K`	If K and L are not equal to zero, and Kmax = Lmax = 0, the number of terms is set to K in the expansion of the square root of the copula density in the basis of Legendre polynomials.
`L`	If K and L are not equal to zero, and Kmax = Lmax = 0, the number of terms is set to L in the expansion of the square root of the copula density in the basis of Legendre polynomials.
`alpha`	Parameter of the Beta(alpha,alpha) distribution used to fix boundary issues through transformation when estimating the Hellinger correlation.
`pval.comp`	Logical. Should we compute the p-value?
`conf.level`	If not `NULL`, a number in (0, 1) that sets the confidence level for the computed confidence interval.
`B1`	Numeric. Number of Bootstrap replicates for the outer loop in the double bootstrap procedure used to obtain the confidence interval. Only used if `conf.level` is not `NULL`. For some examples, one might need to increase this value to 1,000.
`B2`	Numeric. Number of Bootstrap replicates for the inner loop in the double bootstrap procedure used to obtain the confidence interval. Only used if `conf.level` is not `NULL`.
`C.version`	Logical. If `FALSE`, the R version is used instead of the C version. Not really useful to change to `FALSE` unless one wants to understand the algorithm used.

Details

When `Kmax = Lmax = K = L = 0`, the value returned in `Hcor` is the unnormalized version of the empirical Hellinger correlation.

Value

List with the following components:

Hcor	Value of the empirical Hellinger correlation.
p.value	The p-value associated to the null hypothesis that the Hellinger correlation is equal to 0 (computed using a Monte-Carlo simulation). Will be NA if pval.comp is not set to TRUE.
conf.int	Confidence interval for the population Hellinger correlation (computed using a double Bootstrap). Will be NA if conf.level is not set to a value in (0,1).
Khat	Value of K selected by cross-validation, if computed, otherwise NA.
Lhat	Value of L selected by cross-validation, if computed, otherwise NA.

Author(s)

Geenens G., Lafaye de Micheaux P.

References

Geenens G., Lafaye de Micheaux P., (2018). The Hellinger correlation. (), –.

Examples

```
# We illustrate the application of our new measure using data extracted
#  from a study on the coral reef published in
#  [Graham, N.A.J. and Wilson, S.K. and Carr, P. and Hoey, A.S. and
#   Jennings, S. and MacNeil, M.A. (2018), Seabirds enhance coral reef
#   productivity and functioning in the absence of invasive rats, Nature,
#   559, 250--253.].
# The two variables we consider are the number of fishes and the nitrogen
# input by seabirds per hectare recorded on n = 12 islands in the Chagos
# Archipelago. Nitrogen input is an indirect measure of the abundance of
# seabirds. It is worthwhile to notice that nitrogen is absorbed by algae
# and that herbivorous fishes eat these algae. Fishes and birds living in
# two different worlds, it might seem odd to suspect a dependence between
# these two variables. Since our measure is tailored to capture dependence
# induced by a lurking variable, we can suspect the existence of such a
# hidden variable. This is indeed what researchers in (Graham et al., 2018)
# were able to show by finding that the presence and abundance of rats on
# an island had dramatic effects on the number of seabirds (the rats eating
# their eggs) and consequently on the input of nitrogen in seabirds' guano.
# In turn, this would diminish the abundance of algae and fishes eating
# these algae.

data(Chagos)
n <- nrow(Chagos)

par.save <- par()$mfrow
par(mfrow = c(1, 2))
plot(Chagos$Seabirds_ha, Chagos$Number_of_fishes, main = "Original data",
```

```
     xlab = "Density of seabirds", ylab = "Density of fishes")
plot(rank(Chagos$Seabirds_ha) / (n + 1), rank(Chagos$Number_of_fishes) /
     (n + 1), main = "Rank-Copula transformed data",
     xlab = "Density of seabirds", ylab = "Density of fishes")
par(mfrow = par.save)

set.seed(1)
# Empirical Hellinger correlation
HellCor(Chagos$Seabirds_ha, Chagos$Number_of_fishes, pval.comp = TRUE)
# Pearson correlation
cor.test(Chagos$Seabirds_ha, Chagos$Number_of_fishes)
# Distance correlation
dcor.test(Chagos$Seabirds_ha, Chagos$Number_of_fishes, R = 200)


set.seed(1)
# Empirical Hellinger correlation
HellCor(Chagos$kg_N_ha_yr, Chagos$Seabirds_ha, pval.comp = TRUE)
# Pearson correlation
cor.test(Chagos$kg_N_ha_yr, Chagos$Seabirds_ha)
# Distance correlation
dcor.test(Chagos$kg_N_ha_yr, Chagos$Seabirds_ha, R = 200)

set.seed(1)
# Empirical Hellinger correlation
HellCor(Chagos$kg_N_ha_yr, Chagos$Number_of_fishes, pval.comp = TRUE)
# Pearson correlation
cor.test(Chagos$kg_N_ha_yr, Chagos$Number_of_fishes)
# Distance correlation
dcor.test(Chagos$kg_N_ha_yr, Chagos$Number_of_fishes, R = 200)

t.test(Chagos$kg_N_ha_yr ~ Chagos$Treatment)

######################################################################
# Geenens G., Lafaye de Micheaux P., (2020). The Hellinger correlation
######################################################################
# Figure 5.2

## Not run:
n <- 500

set.seed(1)
par(mfrow = c(3, 5))

XX <- .datagenW.corrected(n)
plot(XX, xlab = expression(X[1]), ylab = expression(X[2]))
etahat <- HellCor(XX[,1], XX[,2])$Hcor
title(main = substitute(W~~-~~hat(eta)==etahat, list(etahat =
round(etahat, 3))))

XX <- .datagenDiamond(n)
plot(XX, xlab = expression(X[1]), ylab = expression(X[2]))
etahat <- HellCor(XX[,1], XX[,2])$Hcor
```

```
title(main = substitute(Diamond~~-~~hat(eta)==etahat, list(etahat =
round(etahat, 3))))

XX <- .datagenParabola.corrected(n)
plot(XX, xlab = expression(X[1]), ylab = expression(X[2]))
etahat <- HellCor(XX[,1], XX[,2])$Hcor
title(main = substitute(Parabola~~-~~hat(eta)==etahat, list(etahat =
round(etahat, 3))))

XX <- .datagen2Parabolas.corrected(n)
plot(XX, xlab = expression(X[1]), ylab = expression(X[2]))
etahat<-HellCor(XX[,1], XX[,2])$Hcor
title(main = substitute(Two~~parabolae~~-~~hat(eta)==etahat, list(etahat
= round(etahat, 3))))

XX <- .datagenCircle.corrected(n)
plot(XX, xlab = expression(X[1]), ylab = expression(X[2]))
etahat <- HellCor(XX[,1], XX[,2])$Hcor
title(main = substitute(Circle~~-~~hat(eta)==etahat, list(etahat =
round(etahat, 3))))

XX <- .datagen4indclouds(n)
plot(XX, xlab = expression(X[1]), ylab = expression(X[2]))
etahat <- HellCor(XX[,1], XX[,2])$Hcor
title(main = substitute(4~~clouds~~-~~hat(eta)==etahat, list(etahat =
round(etahat, 3))))

XX <- .datagenCubic(n)
plot(XX, xlab = expression(X[1]), ylab = expression(X[2]))
etahat <- HellCor(XX[,1], XX[,2])$Hcor
title(main = substitute(Cubic~~-~~hat(eta)==etahat, list(etahat =
round(etahat, 3))))

XX <- .datagenSine(n)
plot(XX, xlab = expression(X[1]), ylab = expression(X[2]))
etahat <- HellCor(XX[,1], XX[,2])$Hcor
title(main = substitute(Sine~~-~~hat(eta)==etahat, list(etahat =
round(etahat, 3))))

XX <- .datagenWedge(n)
plot(XX, xlab = expression(X[1]), ylab = expression(X[2]))
etahat <- HellCor(XX[,1], XX[,2])$Hcor
title(main = substitute(Wedge~~-~~hat(eta)==etahat, list(etahat =
round(etahat, 3))))

XX <- .datagenCross(n)
plot(XX, xlab = expression(X[1]), ylab = expression(X[2]))
etahat <- HellCor(XX[,1], XX[,2])$Hcor
title(main = substitute(Cross~~-~~hat(eta)==etahat, list(etahat =
round(etahat, 3))))

XX <- .datagenSpiral(n, sigma = 0.05)
plot(XX, xlab = expression(X[1]), ylab = expression(X[2]))
```

```
etahat <- HellCor(XX[,1], XX[,2])$Hcor
title(main = substitute(Spiral~~-~~hat(eta)==etahat, list(etahat =
round(etahat, 3))))

XX <- .datagen4Circles(n)
plot(XX, xlab = expression(X[1]), ylab = expression(X[2]))
etahat <- HellCor(XX[,1], XX[,2])$Hcor
title(main = substitute(Circles~~-~~hat(eta)==etahat, list(etahat =
round(etahat, 3))))

XX <- .datagenHeavisine(n)
plot(XX, xlab = expression(X[1]), ylab = expression(X[2]))
etahat <- HellCor(XX[,1], XX[,2])$Hcor
title(main = substitute(Heavisine~~-~~hat(eta)==etahat, list(etahat =
round(etahat, 3))))

XX <- .datagenDoppler(n)
plot(XX, xlab = expression(X[1]), ylab = expression(X[2]))
etahat<-HellCor(XX[,1], XX[,2])$Hcor
title(main = substitute(Doppler~~-~~hat(eta)==etahat, list(etahat =
round(etahat, 3))))

XX <- .datagen5Clouds(n)
plot(XX, xlab = expression(X[1]), ylab = expression(X[2]))
etahat <- HellCor(XX[,1], XX[,2])$Hcor
title(main = substitute(5~~clouds~~-~~hat(eta)==etahat, list(etahat =
round(etahat, 3))))

## End(Not run)

############################
# Birth rate vs Death rate #
############################

data(worlddemographics)
x <- wdemographics$Death.Rate.Pop
y <- wdemographics$Birth.Rate.Pop
plot(x, y, xlab = "DEATHS/1,000 POPULATION", ylab =
"BIRTHS/1,000 POPULATION", main =
"Birth rate vs Death rate for 229 countries and territories (in 2020)", col = "orangered", pch = 20)
text(x, y, labels = wdemographics$Country, pos = 3, cex = 0.4, offset = 0.2)
## Not run:
 HellCor(x, y, pval.comp = TRUE)
 cor.test(x, y) # Pearson
 dcor.test(x, y, R = 200)

## End(Not run)
```

wdemographics *Birth rates and death rates for 229 countries in 2020*

Description

This data set contains the birth rates and death rates of 229 countries in 2020 ranked in decreasing order.

Usage

```
wdemographics
```

Format

Data frame with 229 observations and 3 variables.

`Country` Character. Name of the country.

`Birth.Rate.Pop` Numeric. Birth rate per 1,000 habitants.

`Death.Rate.Pop` Numeric. Death rate per 1,000 habitants.

Author(s)

Geenens G., Lafaye de Micheaux P.

References

Source: The World Factbook of the Central Intelligence Agency of the United States (https://www.cia.gov/library/publications/the-world-factbook/geos/xx.html).

See also Figure 1.1 in "A consistent test of independence between random vectors", G. Boglioni Beaulieu, Master thesis, Universite de Montreal, (2016).

Examples

```
data(wdemographics)
```

`worlddemographics` *Birth rates and death rates for 229 countries in 2020*

Description

This data set contains the birth rates and death rates of 229 countries in 2020 ranked in decreasing order.

Usage

```
worlddemographics
```

Format

Data frame with 229 observations and 3 variables.

`Country` Character. Name of the country.

`Birth.Rate.Pop` Numeric. Birth rate per 1,000 habitants.

`Death.Rate.Pop` Numeric. Death rate per 1,000 habitants.

Author(s)

Geenens G., Lafaye de Micheaux P.

References

Source: The World Factbook of the Central Intelligence Agency of the United States (https://www.cia.gov/library/publications/the-world-factbook/geos/xx.html).

See also Figure 1.1 in "A consistent test of independence between random vectors", G. Boglioni Beaulieu, Master thesis, Universite de Montreal, (2016).

Examples

```
data(worlddemographics)
```

Index