Package 'SpatialPack'

March 20, 2018

Version 0.3

Date 2018-03-20

Title Tools for Assessment the Association Between Two Spatial Processes

Description Tools to assess the association between two spatial processes. Currently, four methodologies are implemented: A modified t-test to perform hypothesis testing about the independence between the processes, a suitable nonparametric correlation coefficient, the codispersion coefficient, and an F test for assessing the multiple correlation between one spatial process and several others. SpatialPack gives methods to complement methodologies that are available in geoR for one spatial process.

Author Felipe Osorio [aut, cre],
Ronny Vallejos [aut],
Francisco Cuevas [ctb],
Diego Mancilla [ctb]

Maintainer Felipe Osorio <felipe.osorios@usm.cl>

Imports stats, graphics

License GPL (>= 2.0)

URL http://spatialpack.mat.utfsm.cl

LazyLoad yes

LazyData yes

NeedsCompilation yes

Repository CRAN

Date/Publication 2018-03-20 22:18:00 UTC

R topics documented:

`codisp` *Codispersion Coefficient*

Description

Computes the codispersion coefficient between two spatial variables for a given number of classes for the lag distance.

Usage

```
codisp(x, y, coords, nclass = 13)
```

Arguments

`x`	an `n`-dimensional vector of data values.
`y`	an `n`-dimensional vector of data values.
`coords`	an `n-by-2` matrix containing coordinates of the `n` data locations in each row.
`nclass`	a single number giving the number of cells for the codispersion coefficient. The default is 13. If this argument is `NULL` Sturges' formula is used.

Details

The procedure computes the codispersion coefficient for two spatial sequences defined on general (non-rectangular) grids. First, a given number of bins are constructed for the lag distance. Then the codispersion is computed for each bin.

Value

A list with class `"codisp"` containing the following components:

`coef`	a vector of size `nclass` containing the values of the codispersion coefficient.
`upper.bounds`	upper bounds of the intervals constructed to compute the codispersion coefficient.
`card`	number of elements in each interval generated to compute the codispersion coefficient.

The function `plot` can be used to obtain a graph of the codispersion coefficient versus the lag distance.

References

Matheron, G. (1965), *Les Variables Regionalisees et leur Estimation*. Masson, Paris.

Rukhin, A., Vallejos, R. (2008), Codispersion coefficient for spatial and temporal series. *Statistics and Probability Letters* **78**, 1290–1300.

Vallejos, R. (2008). Assessing the association between two spatial or temporal sequences. *Journal of Applied Statistics* **35**, 1323–1343.

Examples

```
# Murray Smelter site dataset
data(murray)

# defining the arsenic (As) and lead (Pb) variables from the murray dataset
x <- murray$As
y <- murray$Pb

# extracting the coordinates from Murray dataset
coords <- murray[c("xpos","ypos")]

# computing the codispersion coefficient
z <- codisp(x, y, coords)
z

## plotting the codispersion coefficient vs. the lag distance
plot(z)

# Comovement between two time series representing the monthly deaths
# from bronchitis, emphysema and asthma in the UK for 1974-1979
x <- mdeaths
y <- fdeaths
coords <- cbind(1:72, rep(1,72))
z <- codisp(x, y, coords)

# plotting codispersion and cross-correlation functions
par(mfrow = c(1,2))
ccf(x, y, ylab = "cross-correlation", max.lag = 20)
plot(z)
```

---

cor.spatial *Tjostheim's Coefficient*

Description

Computes Tjostheim's coefficient for two spatial sequences observed over the same locations on the plane.

Usage

```
cor.spatial(x, y, coords)
```

Arguments

x	an n-dimensional vector of data values.
y	an n-dimensional vector of data values.
coords	an n-by-2 matrix containing coordinates of the n data locations in each row.

Details

The implemented technique is a nonparametric coefficient that summarizes the association between two spatial variables. This coefficient was first introduced by Tjostheim (1978) and later generalized by Hubert and Golledge (1992). The computation of the coefficient is based on the construction of ranks associated to suitable modifications of the coordinates. Tjostheim's coefficient is a variant of the correlation coefficient (`cor`) to be used in a spatial statistics context.

Value

Tjostheim's coefficient. The variance is returned as the attribute `"variance"`.

References

Tjostheim, D., (1978), A measure of association for spatial variables. *Biometrika* **65**, 109–114.

Hubert, L., Golledge, R.G., (1982), Measuring association between spatially defined variables: Tjostheim's coefficient index and some extensions. *Geographical Analysis* **14**, 273–278.

Examples

```
# Murray Smelter site dataset
data(murray)

# defining the arsenic (As) and lead (Pb) variables from the murray dataset
x <- murray$As
y <- murray$Pb

# extracting the coordinates from Murray dataset
coords <- murray[c("xpos","ypos")]

# computing Tjostheim's coefficient
z <- cor.spatial(x, y, coords)
z
```

---

modified.Ftest *Modified F test*

Description

Performs a modified version of the F test to assess the multiple correlation between one spatial processes and several others.

Usage

```
modified.Ftest(x, y, coords, nclass = 13)
```

Arguments

x	an n-by-q matrix of data values.
y	an n-dimensional vector of data values.
coords	an n-by-2 matrix containing coordinates of the n data locations in each row.
nclass	a single number giving the number of cells for Moran's index. The default is 13. If this argument is NULL Sturges' formula us used.

Details

The methodology implemented is a modified F test for assessing the multiple correlation between one spatial process and several others. The test is based on corrections of the multiple correlation coefficient between the two spatially correlated sequences and required the estimation of an effective sample size. This factor takes into account the spatial association of both processes.

Value

A list with class "mod.Ftest" containing the following components:

corr	the sample correlation coefficient.
ESS	the estimated effective sample size.
Fstat	the value of the (unscaled) F-statistic.
df1,df2	degrees of freedom for the F-statistic.
p.value	the p-value for the test.
upper.bounds	upper bounds of the intervals constructed to compute Moran's I.
card	number of elements in each interval generated to compute Moran's I.
imoran	a matrix containing Moran's index for each interval associated with the response and predicted variables.

The generic functions print and summary are used to obtain and print additional details about the modified F test.

References

Dutilleul, P., Pelletier, B., Alpargu, G. (2008). Modified F tests for assessing the multiple correlation between one spatial process and several others. *Journal of Statistical Planning and Inference* **138**, 1402–1415.

Examples

```
# The Pinus Radiata data set
data(radiata)

# defining the response and predictor variables from the radiata data set
y <- radiata$height
x <- radiata[c("basal","altitude","slope")]

# extracting the coordinates from the radiata data set
```

```
coords <- radiata[c("xpos","ypos")]

# computing the modified F-test of spatial association
z <- modified.Ftest(x, y, coords)
z

# display the upper bounds, cardinality and the computed Moran's index
summary(z)
```

`modified.ttest` *Modified t test*

Description

Performs a modified version of the *t* test to assess the correlation between two spatial processes.

Usage

```
modified.ttest(x, y, coords, nclass = 13)
```

Arguments

`x`	an `n`-dimensional vector of data values.
`y`	an `n`-dimensional vector of data values.
`coords`	an `n-by-2` matrix containing coordinates of the `n` data locations in each row.
`nclass`	a single number giving the number of cells for Moran's index. The default is 13. If this argument is `NULL` Sturges' formula us used.

Details

The methodology implemented is a modified *t* test of spatial association based on the work of Clifford and Richardson (1989). The test is based on corrections of the sample correlation coefficient between the two spatially correlated sequences and required the estimation of an effective sample size. This factor takes into account the spatial association of both processes.

Value

A list with class `"mod.ttest"` containing the following components:

`corr`	the sample correlation coefficient.
`ESS`	the estimated effective sample size.
`Fstat`	the value of the (unscaled) *F*-statistic.
`dof`	the estimated degrees of freedom for the *F*-statistic.
`p.value`	the *p*-value for the test.
`upper.bounds`	upper bounds of the intervals constructed to compute Moran's *I*.
`card`	number of elements in each interval generated to compute Moran's *I*.

`imoran` a matrix containing Moran's index for each interval associated with both variables.

The generic functions `print` and `summary` are used to obtain and print additional details about the modified *t* test.

References

Clifford, P., Richardson, S., Hemon, D. (1989). Assessing the significance of the correlation between two spatial processes. *Biometrics* **45**, 123–134.

Dutilleul, P. (1993). Modifying the *t* test for assessing the correlation between two spatial processes. *Biometrics* **49**, 305–314.

Examples

```
# Murray Smelter site dataset
data(murray)

# defining the arsenic (As) and lead (Pb) variables from the murray dataset
x <- murray$As
y <- murray$Pb

# extracting the coordinates from Murray dataset
coords <- murray[c("xpos","ypos")]

# computing the modified t-test of spatial association
z <- modified.ttest(x, y, coords)
z

# display the upper bounds, cardinality and the computed Moran's index
summary(z)
```

---

murray *The Murray smelter site dataset*

Description

The dataset consists of soil samples collected in and around the vacant, industrially contaminated, Murray smelter site (Utah, USA). This area was polluted by airborne emissions and the disposal of waste slag from the smelting process. A total of 253 locations were included in the study, and soil samples were taken from each location. Each georeferenced sample point is a pool composite of four closely adjacent soil samples in which the concentration of the heavy metals arsenic (As) and lead (Pb) was determined.

Usage

```
data(murray)
```

Format

A data frame with 253 observations on the following 5 variables.

As arsenic concentrations measurements.

Pb lead concentrations measurements.

xpos x-coordinates.

ypos y-coordinates.

quad a factor where numbers indicate different sub-regions within the area.

Source

Griffith, D., Paelinck, J.H.P. (2011). *Non-Standard Spatial Statistics*. Springer, New York.

---

radiata *The Pinus Radiata dataset*

Description

Pinus radiata is one of the mostly widely planted species in Chile and is planted in a wide array of soil types and regional climates. The plots were located in the Escuadron sector, south of Concepcion, in the southern portion of Chile and has an area of 1244.43 hectares.

Usage

```
data(radiata)
```

Format

A data frame with 468 observations on the following 6 variables.

xpos x-coordinates.

ypos y-coordinates.

basal basal area measurements.

height dominant tree height.

altitude altitude in meters.

slope slope of the terrain plot.

Source

Cuevas, F., Porcu, E., Vallejos, R. (2013). Study of spatial relationships between two sets of variables: A nonparametric approach. *Journal of Nonparametric Statistics* **25**, 695–714.

Vallejos, R., Osorio, F., Bevilacqua, M. (2018+). *Spatial Relationships Between Two Georeferenced Variables: with Applications in R*. Springer, New York.

Index