

Paula Maria Montoya Pfeiffer

Doutoranda em ecologia, laboratório de abelhas, orientada por Isabel Alves dos Santos. Tema de projeto de tese: fluxo de polen entre fragmentos e áreas restauradas de Mata Atlântica.

Exercícios

exec

Propostas trabalho final

PROPOSTA 1

Fazer uma função com ciclos de regressões lineares simples para identificar as variáveis preditoras de uma variável 'y' de vários níveis (ou classificada por um fator). O produto da função é uma tabela com o resultados das regressões (R2 e valor de p) entre cada nível da variável dependente e as variáveis preditoras.

Exemplo:

Pergunta: Qual o quais são as variáveis que melhor explicam a abundância das espécies em diferentes locais (fragmentos de vegetação).

- Variável dependente: abundancia de cada espécie.
- Variáveis preditoras: associadas as características dos locais (área do fragmento, forma do fragmento, grau de isolamento, etc.)

Resultado:

Nome da especie	Área fragmento	Forma fragmento	Grau de isolamento
Sp1	R2 p valor	R2 p valor	R2 p valor
Sp2	R2 p valor	R2 p valor	R2 p valor
Sp3	R2 p valor	R2 p valor	R2 p valor

PROPOSTA 2

Uma função para fazer uma regressão entre os coeficientes de similaridade de pares de amostras (i.e. variável dependente) e as diferencias entre os valores das variáveis preditoras dos pares de amostras. A função calculará os coeficientes de similaridade e as diferencias entre pares de amostras e depois fará a regressão.

Exemplo: O mesmo anterior.

1. calcular matrizes de coeficientes de similaridade segundo o numero de espécies coexistentes entre locais
2. calcular matrizes de diferencias entre os valores das variáveis preditoras e logo fazer o seguinte

data.frame:

Coef. Similaridade fragmentos	Diferencia tamanho fragmento	Diferencia Forma fragmento	Diferencia grau de isolamento
ij	area 'i'- área 'j'	...i - ...j	...i - ...j
ik	...i - ...k	...i - ...k	...i - ...k
i...n	...	...	...

3. Fazer uma regressão com os dados do data.frame

Ola, Paula! A proposta A me parece muito simples e não haverá muitos desafios para implementá-la. A proposta B é mais adequada para o trabalho final, mas ainda não está muito clara. Senti falta de uma contextualização da importância da função para outras pessoas e da sua utilidade. Poderia citar trabalhos que usaram essa abordagem. —[Leo](#)

Pelo que entendi a ideia do Plano B é fazer uma função que verifique a correlação entre duas matrizes de similaridade ou distância (similaridade entre fragmentos e similaridade entre a comunidade;. Isso basicamente é uma análise de Mantel, já que uma regressão não é adequada dado a dependência entre as observações das similaridades. Sugiro que faça a proposta B a partir de entrada de dois conjuntos de dados: 1. das espécies por fragmentos e 2. dos fragmentos com seus dados de tamanho, isolamento etc. A partir dessa entrada de dados, a função deve: 1. calcular duas matriz de similaridade: 1. para os fragmentos (tamanho, isolamento etc) e 2. a comunidade em cada fragmento (similaridade de espécies). Em seguida fazer o teste o mantel por permutação (fizemos isso em aula). Para tornar o desafio mais interessante já que fizemos parte disso em aula, possibilite ao usuário a escolha de diferentes índices de “distância” ou “similaridade”.

— [Alexandre Adalardo de Oliveira](#) 2015/03/27 14:43

Trabalho final

```
require(vegan)
```

####Input two data sets

```
sophisticated.mantel <- function(datasp, datavar, sample.dist = "bray",
var.dist = "euclidean", cor.method = "pearson", ...)
{
  if(is.numeric(datasp) | is.numeric(datavar))
    stop("data must be a data frame or a matrix") #data set class must be
data frame or matrix
  x <- vegdist(datasp, method = sample.dist, ...) #sample dissimilarity
matrix

##Tables of results
  result <- data.frame(Variables = colnames(datavar), Mantel_statistic_r =
rep(NA, ncol(datavar)), Significance = rep(NA, ncol(datavar))) #Table of
Mantel test results
  result.sp <- matrix(NA, ncol(datasp), ncol(datavar)+3) #Table of Species
analysis results
  colnames(result.sp) <- c("Species", "Abundance", "Frequency",
colnames(datavar))
  result.sp[,1] <- colnames(datasp) #Species names in the tables's first
column
##Mantel test for each independent variable
  for(i in 1:ncol(datavar))
  {
    y <- dist(datavar[,i], method = var.dist, ...) #distance matrix for each
independent variable
    z <- mantel(x,y, method = cor.method) #Mantel test between distance
matrixes of samples and independent variables
    result$Mantel_statistic_r[i] <- round(z$statistic, 3) #R2-values from
the Mantel test
    if(z$signif <= 0.05){
      result$Significance[i] <- paste(z$signif, "*") #significant p-values
from Mantel test
    }
    else{
      result$Significance[i] <- paste(z$signif, " ") #non-significant p-
values
    }
  }

#Species analysis
  for(k in 1:ncol(datasp))
  {
    result.sp[k,2] <- sum(datasp[,k]) #total abundance of each species
    result.sp[k,3] <- sum(datasp[,k] != 0) #total frequency of each species
    for(b in 1:ncol(datavar)){
      d <- anova(lm(datasp[,k]~1), lm(datasp[,k]~datavar[,b])) #linnear
regression between each species abundances and each independent variable
      p <- round(d[2,6], 3) #p-values
      if(d[2,6] <= 0.05){
        result.sp[k,b+3] <- paste(p, "*") #significant p-values
      }
    }
  }
}
```

```
    }
    else{
      result.sp[k,b+3] <- paste(p, " ") #non-significant p-values
    }
  }
}
return(list(Mantel_Test_Results = result, Species_Analysis =
data.frame(result.sp))) #list of results
}
##Examples

data(varespec)
data(varechem)

sophisticated.mantel(varespec, varechem)
```

Help da função

sophisticated.mantel package:unknown R

Documentation

A sophisticated Mantel test for species community analysis

Description:

This function creates distance matrices from a first data set of species abundances in samples and a second data set of explanatory variables from samples. Afterwards runs the mantel test for each explanatory variable. In addition, the function runs simple linear regressions between each species and each explanatory variable.

Usage:

```
sophisticated.mantel(dataspc, datavar, sample.dist = "bray",
var.dist = "euclidean", cor.method = "pearson", ...)
```

Arguments:

dataspc: a data frame or matrix of species abundances (columns) distributed in samples (rows).

datavar: a data frame or matrix of numeric independent variable(s) associated to samples. Row length must be the same as dataspc.

sample.dist: the distance measure to be used for computing the sample/species distance matrix (from dataspc) using 'vegdist' (from vegan package). This must be one of "manhattan", "euclidean", "canberra", "bray" (default), "kulczynski", "jaccard", "gower", "altGower", "morisita", "horn", "mountford", "raup", "binomial", "chao" or "cao".

var.dist: the distance measure to be used for computing distance matrices of the explanatory variables (from datavar), using 'dist'. This must be one of "euclidean" (default), "maximum", "manhattan", "canberra",

"binary" or "minkowski".

`cor.method`: the correlation coefficient to be computed by 'mantel' (from vegan package) as accepted by 'cor'. This can be "pearson"(default), "spearman" or "kendall".

... additional arguments from the functions 'vegdist', 'dist' and 'mantel'.

Details:

'datasp' is converted to a distance matrix which is calculated with the distances between samples.

'datavar' is converted to one or many distance matrices according to the number of variables (columns). Each matrix is calculated with the differences (distances) between the variable's values on every pair of samples. Factor variables must be arranged as binary data on 'datavar', with as many columns as the number of factor levels.

The function returns the coefficients of the correlations between the distance matrix of each variable and the distance matrix of the sample/species matrix. Significance values are calculated by permutation analysis (see 'mantel' in vegan package documentation).

The second part of the function returns a data frame with the species's total abundances and frequencies, as well as the significance values of simple linear regressions between species abundances and explanatory variables using 'lm'.

Value:

The function returns a list with the following components:

`Mantel_Test_Results`: R-square values ("Mantel_statistic_r") and p-values ("Significance") of the correlation with each explanatory variable.

`Species_Analysis`: p-values of simple linear regressions between species abundances and explanatory variables. The asterisk '*' indicates statistically significant relations.

Warning:

Data sets must be data frames or matrices with the same row length and must not have NA values.

Note:

Legendre and Legendre (2012) explain distance methods and give suggestions for using them depending on data characteristics. Distance method explanations are also given at 'vegdist' vegan package documentation.

Author(s):

Paula Maria Montoya-Pfeiffer

References:

Legendre, P. and Legendre, L. (2012) Numerical Ecology. 3rd English Edition. Elsevier.

Jari Oksanen, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre,

Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Helene Wagner (2015). Community Ecology Package 'Vegan'. 280pp. Available in: <http://cran.r-project.org>, <https://github.com/vegandevs/vegan>.

See Also:

'vegdist' and 'mantel' in vegan package and 'dist' for additional arguments and further information.

Examples:

```
require(vegan)
data(varespec)
data(varechem)
sophisticated.mantel(varespec, varechem) ##with default arguments and data frames
matrix.sp <- as.matrix(varespec)
sophisticated.mantel(matrix.sp, varechem) ##with a matrix data set
sophisticated.mantel(varespec, varechem, sample.dist = "morisita",
var.dist="canberra", cor.method = "spearman") ##with different distance methods
```

[sophisticated_mantel.r](#)

From:

<http://ecor.ib.usp.br/> - **ecoR**

Permanent link:

http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2015:alunos:trabalho_final:pmmontoyap:start 

Last update: **2020/08/12 06:04**