

Joice lamara Nogueira



Mestranda em Ecologia pelo PPGE da Universidade de São Paulo. Desenvolvo meu trabalho com *Sinningia allagophylla*, uma Gesneriaceae herbácea, com grande distribuição geográfica (Sudeste e Sul) e que apresenta grande variação fenotípica (como cor e tamanho da flor e quantidade de pelo) dependendo do local em que ocorre. Meu objetivo é entender qual processo evolutivo explica tamanha variação. Em outras palavras, procuro saber se trata-se de um processo de adaptação, ou seja, implicando modificações genéticas, se é um processo de plasticidade fenotípica, atuando no fenótipo e não sendo transmitidas aos descendentes ou se ainda podem ser espécies diferentes.

exec

TRABALHO FINAL

PRINCIPAL

Construir uma função que procure por padrões morfológicos e fenológicos de flores de espécies de plantas que possuem ampla distribuição geográfica e que apresentam fenótipos e fenologia diferenciada.

O input da função será feito por um `data.frame` com dados de fisionomia onde as flores são encontradas (Cerrado Campo Sujo, Cerrado Campo Limpo e Mata Atlântica ou quaisquer outras fisionomias de interesse), cores de corola (ou outra característica morfológica) representados por um vetor numérico e início da floração, também com vetores numéricos.

Esses dados serão aleatorizados por simulações de modelos nulos, onde a hipótese nula é de que não existe um padrão entre a morfologia e/ou a fenologia floral por região amostrada.

Não tenho muita certeza de como será o output dessa função, mas provavelmente algum índice de sobreposição das variáveis, como, por exemplo, o índice de Czekanowski

PLANO B

A mesma ideia presente no trabalho principal, mas o input será uma matriz que utilize somente as fisionomias e os meses de início de floração. O objetivo é procurar padrões de início de floração em diferentes fisionomias.

Nesse caso, o output poderia ser índices de sobreposições como o de Czekanowski (citado acima), que são amplamente utilizados para dados fenológicos. Esse índice equivale, graficamente, à área de intersecção de dois histogramas fenológicos.

Comentário Leandro

Joice, não ficou muito claro como seria a comparação do modelo nulo com seus dados. Acho a função interessante, mas a comparação entre diferentes fatores ficará complexa para ser feita em apenas uma semana.

Outro fator complicador é não saber exatamente o que será produzido como output. Acho que isso precisa estar bem claro antes de começar a elaborar a função, senão você terá muitos problemas

para escrevê-la.

Reformulado a proposta do projeto final

Considerando os comentários e as sugestões, irei reformular meu projeto final.

Como sugerido pelo professor Ale, minha função irá calcular uma matriz de semelhança entre objetos de estudo utilizando o índice de Gower.

O índice de Gower é um coeficiente genérico de similaridade que combina diferentes tipos de características descritora dos objetos de estudo e as processa de acordo com seu tipo matemático próprio. Sendo assim, essa função terá três condições: se as características entre os objetos são binários ("bin"), se são qualitativos ("quali") ou se são quantitativos ("quanti").

O input será uma matriz contendo os objetos de estudo (que pode ser, no meu caso, flores em um determinado local ou mesmo de locais distintos, ou pode ser objetos de estudo generalizados, como populações diferente, etc) e seus descritores; o output será uma matriz de semelhança entre os objetos. Como desafio, tentarei incorporar na função particularidades do índice de Gower, como a função delta Kronecker, que é um jeito especial de tratar de NAs.

Comentário ALe

Concordo com o Leandro: faltou definir melhor o que será a estatística de interesse para conseguir criar um modelo nulo. Como se trata de dados multivariados, poderia usar algum índice para reduzir a variação a um valor... Por exemplo: usar um índice de distância entre cada flor amostrada usando uma medida que permita a entrada de dados de diferentes naturezas (já que cor é nominal e tamanho é contínuo p.ex) A distância de Gower parece muito usada nesses casos. A partir desses dados poderia fazer um teste de hipótese por permutação para verificar se as distâncias médias dentro de cada região são maiores que as distâncias médias entre regiões (parece um modelo de ANOVA) com isso responde a pergunta! Só acho que é um desafio muito grande para o pouco tempo! Que tal ficar só com uma função que cria a matriz de distância entre amostras usando o índice de Gower... o primeiro passo!

Comentário Paulo

Concordo que pode ficar muita coisa para fazer em pouco tempo. Com Alê sugeriu, vc pode usar este trabalho para fazer uma parte, que depois você poderá incorporar na função completa (lembre-se que pode-se chamar uma função dentro de outra!). Neste sentido, outra sugestão é implementar a randomização para apenas uma localidade. A hipótese nula seria então que não há associação entre as características florais. Se vc chegar até aí e sobrar tempo, aí parta para incluir região.

Após a entrega

Se entendi sua função, ela aceita apenas um tipo de caracter, mas a vantagem do índice de Gower é justamente aceitar diferentes tipos de dados. Você pode incorporar isto facilmente na sua função, permitindo que o argumento `caracter` aceite um vetor de comprimento igual ao número de caracteres no objeto de entrada.

Página de ajuda

`dist.gower`

BIE5782

RDocumentation

Índice de similaridade de Gower

Description

A função calcula uma matriz de semelhança para diferentes objetos. O índice de Gower é um coeficiente genérico de similaridade que combina diferentes tipos de características e as processa de acordo com seu tipo matemático próprio. Calcula-se primeiro a similaridade parcial de cada descritor entre os objetos de interesse.

Usage

```
dist.gower(x, caracter="bin")
```

Arguments

`x` Matrix de dados

`caracter` Natureza matemática das características que descrevem os objetos ("bin", "quali" e "quanti").

Details

As naturezas matemáticas das características podem ser:

"bin" matriz binária, onde os objetos são descritos apenas por 1 e 0 (presença e ausência das características, respectivamente).

"quali" matriz com dados qualitativos, descritos por fatores numéricos.

"quanti" características quantitativas.

Quando as características forem categorizadas como qualitativas, a matriz de dados será transformada em matriz binária, ocorrendo a perda de algumas informações valiosas. Para dados quantitativos, no entanto, o índice diminuiu essa perda por considerar as diferenças entre os valores quantitativos.

Value:

É gerada uma matriz de semelhança entre os objetos de estudo presente na matriz de dados.

Warnings

O índice de Gower não realiza nenhuma comparação quando há dados faltantes, portanto, NA's são automaticamente retirados na função, não sendo facultativo a permanência ou retirada dos mesmos.

Authors

Joice Iamara Nogueira
joyce.iamara@usp.br

References:

Legendre, P. & Legendre, L. (1998) Numerical Ecology. Elsevier Science, Amsterdã.

See also

'vegdist', no pacote vegan

Example

```
source("dist.gower.r")

# qual a semelhança entre o comprimento da pétala e o comprimento da sépala
em iridaceae?
iris
iris.2<-matrix(c(iris[1:50,3],iris[1:50,1]), ncol=2, nrow=50)
colnames(iris.2)<-c("comp.sépala","compr.petala")
rownames(iris.2)<-c(1:50)
dist.gower(iris.2, caract="quanti")
```

Código da função

```
dist.gower<- function (x, character="bin")
{
  x1<-(na.omit(x))

  nobj<-dim(x1)[2] # numero de objetos de estudo

  ncar<-dim(x1)[1] # numero de descritores

  similar=matrix(1,ncol=nobj,nrow=nobj) # cria o esqueleto de uma matriz para
  comparar os
  objetos entre si posteriormente

  rownames(similar)=paste("obj",c(1:nobj))
```

```
colnames(similar)=paste("obj",c(1:nobj))

if (caracter=="bin") # se os descritores são binários (ausencia e presença,
por exemplo)

{

for (i in 1:nobj-1) # criando o ciclo

{

j=i+1

for (j in j:nobj)

{

comp<-sum(x1[,i]==x1[,j]) # soma das colunas de i que é igual ao de j, ou
seja se em i tiver 1 e em j tb, vai somar. A mesma coisa com doble-zeros.

total<-sum(comp)/ncar # a distância é igual a soma anterior dividido pelo
numero de descritores

similar[i,j]<-total

similar[j,i]<-total

diag(similar)<-1

}

}

return(similar)

}

if (caracter=="quali") # se as caracterísitcas forem qualitativas, a matriz
de dados será transfromada em matriz binária

{

x1[x1>0]=1 # transformando em matriz binária

# e repetindo o mesmo passo anterior

for (i in 1:nobj-1) # criando o ciclo

{

j=i+1
```

```
for (j in j:nobj)

{

comp<-sum(x1[,i]==x1[,j]) # soma das colunas de i que é igual ao de j, ou
seja se em i tiver
1 e em j tb, vai somar. A mesma coisa com zeros.

total<-sum(comp)/ncar # a distância é igual a soma anterior dividido pelo
numero de descritores.

similar[i,j]<-total

similar[j,i]<-total

diag(similar)<-1

}

}

return(similar)

}

if (caracter=="quanti") # se os descritores são quantitativos

matriz.ref<-matrix(NA,ncol=nobj, nrow=nobj)

difere<-as.matrix(dist (t(x1),method="maximum")) #calculando a difere de um
objeto com os outros

intervalo<-range(x1) # calcula-se o intervalo entre as características

inter<-abs(intervalo[1]-intervalo[2]) # e depois o R, que é a diferença dos
intervalos

divi<-matrix(NA,nrow=nrow(difere),ncol=ncol(difere))

for (i in 1:ncol(difere)) # agora preciso dividir as colunas da diferença
entre os descitores...

{

div<-difere[,i]/inter

divi[,i]<-div # lindo!!!

}
```

```
dif.total<-matrix(NA,nrow=nrow(difere),ncol=ncol(difere)) # novamente uma
matriz para colocar o resultado

for (i in 1:ncol(divi))

{

dif.2<-abs(divi[,i]-1) # transformando um dado que é distância em
semelhança, ao diminuir por 1

dif.total[,i]<-dif.2

}

similar<-(dif.total[,])/ncar <- calculando a similaridade

diag(similar)<-1 # as diagonais devem ser igual a 1, pois os objetos das
diagonais são os mesmos e, portanto, a semelhança é total.

return(similar)

}
```

Arquivo da função

[dist.gower.r](#)

[help_dist.gower.r](#)

From:

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

Permanent link:

http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2011:alunos:trabalho_final:joice:start



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