

Estrutura de Comunidades

Três aspectos essenciais da estrutura das comunidades são a riqueza de espécies, suas abundâncias e a composição de espécies. Um dos grandes desafios em ecologia é avaliar o grau de similaridade entre comunidades em relação a esses três aspectos e quais os possíveis fatores que determinam os padrões observados.



Introdução

Comunidades vegetais podem apresentar diferentes conjuntos de espécies mesmo em localidades espacialmente próximas e compreender os processos que determinam essas diferenças é essencial para entender a estruturação e o funcionamento dessas comunidades. O primeiro passo para essa compreensão é identificar o quão similar são as diferentes localidades em relação às espécies presentes em cada localidade.


Considerando que cada espécie representa uma variável para a definição das similaridades entre as localidades e que serão analisadas várias espécies simultaneamente, é necessário o uso de técnicas de **Análises Multivariadas**. Existem muitos procedimentos analíticos diferentes e, em função do pouco tempo, teremos que focar em alguns procedimentos específicos. Entretanto, alguns textos e links para páginas importantes estão disponíveis no final desse roteiro para quem se interessar mais pelo assunto.

Na primeira etapa do nosso roteiro vamos utilizar os chamados **Métodos de Ordenação Irrestrita** para identificarmos locais mais similares e as espécies que determinam essas similaridades. Após identificarmos as similaridades entre locais, vamos propor alguns processos que poderiam explicar esses padrões e vamos avaliar a relação entre esses processos e os locais, usando os **Métodos de Ordenação Restrita**. Por fim, vamos quantificar a importância relativa de cada processo usando técnicas de **Partição da Variação**.

Preparação para o exercício

Para começarmos nosso exercício prático de hoje, crie uma pasta para você na área de trabalho (desktop) do seu computador. Copie para essa pasta os três arquivos abaixo:

- [especies.csv](#)
- [ambientais.csv](#)
- [coordenadas.csv](#)

Em seguida, abra o programa R, clicando no ícone  que está na área de trabalho do seu computador.

Se tudo deu certo até aqui, abrirá uma janela do R como essa:



Já com a janela do programa R aberto, o próximo passo será mudar o diretório de trabalho para aquela pasta que você acabou de criar. Com isso será mais fácil importar os dados dos arquivos “.csv” para dentro do ambiente R.

A mudança de diretório deve ser feita da seguinte forma:

- Abra o Menu “Arquivo” (ou “File”);
- Selecione “Mudar dir” (ou “Change dir”);
- Escolha a sua pasta na janela que abrir.

[Obs. Para Mac, essa opção está no Menu “Misc” e a opção é “Change working dir”]

Para checar se você está na pasta correta, copie e cole o comando abaixo na janela (console) do R. Atenção: O comando deve ser colado na frente do símbolo “>” (circundado em azul na imagem anterior):

```
getwd()
```

Após colar, aperte a tecla “enter” e veja se aparece o nome da sua pasta. Se sim, ótimo. Se não, chame um monitor ou professor.

Para fazer as análises vamos usar um pacote muito legal para quem quer trabalhar com análises de vegetação. O pacote se chama *vegan* e para instalá-lo, use o comando:

```
install.packages("vegan")
```

Se o programa perguntar se você quer instalar as dependências, selecione **sim/yes!**

E agora, carregue o pacote no R, usando o comando library:

```
library(vegan)
```

Se você for usar o mesmo computador outras vezes, não precisará instalar novamente o pacote *vegan* (a não ser quando lançarem uma nova versão), mas toda vez que for fazer essas análises usando esse pacote você precisará carregar (com o código “library(vegan)” ou “require(vegan)”), ok?

Antes de aplicar os métodos de ordenação, vamos entender um pouquinho algumas das informações que serão analisadas no nosso exercício.

Conhecendo os dados das espécies de plantas

Imagine que amostramos uma área de 1km x 1km, dividindo-a em 100 parcelas de 100 x 100m. Amostramos todas as espécies arbóreas acima de 20 cm de DAP e anotamos o número de indivíduos de cada espécie (ou seja, a abundância de cada espécie) em cada parcela.

Tabulamos os dados em uma planilha cujas linhas são as amostras (parcelas ou *sites*) e as colunas são as abundâncias das espécies ¹⁾. Se quiser ver a planilha completa de espécies por parcela abra o

arquivo “especies.csv” diretamente em um programa de planilha eletrônica (p. ex. Excel).

Agora vamos importar a planilha para o R e visualizá-la dentro do próprio R, como explicaremos abaixo:

O primeiro passo é importar a planilha de dados para o R e criar um objeto para essa planilha, dentro do R:

```
sp<- read.csv2 ("especies.csv")
```

O que esse comando está “dizendo” é: crie um objeto chamado “sp” dentro do R e coloque nele o arquivo “especies.csv” usando a função de leitura de arquivo “read.csv2”. Existem outras funções de leitura, mas essa é mais adequada para esse formato de arquivo que temos.

Se quiser ter certeza que a planilha foi copiada corretamente, visualize as primeiras 5 linhas do objeto “sp” (que é a mesma planilha, agora importada para o R) usando a função *head*:

```
head(sp)
```

Se quiser ver a planilha inteira, basta escrever o nome do objeto (nesse caso, “sp”).

```
sp
```

Como essa planilha não é enorme, não é tão ruim de visualizar, mas quando as planilhas são muito grandes, vale a pena ver o resumo das variáveis. Para isso use a função *summary*:

```
summary(sp)
```

Você entende o que está apresentado nesse objeto? Se não, converse com seus colegas, monitores e professores. Não prossiga se não entendeu.

Visualizando Correlações

Uma forma de tentar entender como as espécies estão distribuídas nas parcelas seria olhar para as correlações entre as abundâncias das espécies (par-a-par) em cada parcela. Isso nos ajuda a entender quais espécies tendem a ocorrer com abundâncias mais similares nas parcelas. Quanto maior for a correlação, mais associada será a distribuição de um dado par de espécies.

Então, vamos fazer uma matriz de correlação entre todos os pares de espécies da nossa tabela e observar quais espécies estão mais associadas e qual a forma dessa associação.

Para facilitar a nossa vida, vamos antes importar uma função gráfica (“panel.cor”) que será utilizada simplesmente para melhorar a visualização dessas correlações. Copie tudo que está dentro desse quadro, cole no R e aperte “enter”:

```
## Função para colocar coeficientes de correlação em uma das diagonais
panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  rs <- cor(x, y)
  r <- abs(rs)
  txt <- format(c(rs, 0.123456789), digits = digits)[1]
  txt <- paste0(prefix, txt)
  if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
  text(0.5, 0.5, txt, cex = cex.cor * r)
}
```

Não aconteceu nada?? Que ótimo, não precisava acontecer mesmo. A função simplesmente foi salva e armazenada no R para ser usada depois...😁.

Agora vamos aplicar ao nosso objeto “sp” uma função chamada “pairs” que vai calcular a correlação par-a-par para as 10 espécies. Vamos adicionar a função “panel.cor” dentro da função “pairs” para melhorar a visualização.

```
pairs (sp, upper.panel=panel.cor)
```

Salve essa figura em pdf na sua pasta para que possa abrir com facilidade sempre que precisar consultá-la

Para salvar como pdf, copie e cole esses comandos:

```
pdf(file="correlacoes.pdf")
pairs (sp, upper.panel=panel.cor)
dev.off()
```

Como entender essa figura?



Na diagonal principal da figura temos os códigos das espécies.

Na parte inferior e à esquerda da diagonal principal da figura, temos os gráficos que mostram as relações entre cada par de espécies. Os pontos nessas figuras são as parcelas. Para entender qual espécie está representada no eixo X e qual está no eixo Y de cada um dos gráficos, basta encontrar os códigos das espécies nas linhas e colunas correspondentes. Por exemplo, o gráfico localizado na linha 4, coluna 3, tem o gráfico de dispersão das abundâncias da espécie 3 (Sp3) no eixo X, e da espécie 4 (Sp4) no eixo Y.

Na parte superior e à direita da diagonal principal da figura temos os valores das correlações entre as abundâncias das espécies nas parcelas. Esses valores são referentes aos gráficos que se encontram na posição espelhada na parte inferior da diagonal principal. Note que os valores de correlação variam entre -1,0 e +1,0 e que os maiores valores indicam os pares de espécies que estão mais fortemente associadas.

Para facilitar a visualização, quanto maior for o valor da correlação, maior será a letra do número. Correlações muito baixas sequer aparecem.

Anote os pares de espécies com correlações mais fortes, avalie se as correlações entre elas são positivas ou negativas e anote um par para o qual não há nenhuma evidência de correlação.

Repare que como são apenas 10 espécies não é difícil visualizar as correlações par-a-par desta maneira, pois são gerados apenas 45 gráficos. Mas imagine se tivéssemos as mais de 1000 espécies da parcela permanente de Manaus? Só para terem uma ideia, para 100 espécies, teríamos 4.950 gráficos...😬

Entendendo os princípios de um Método de Ordenação

Os métodos de ordenação são aplicados quando desejamos observar padrões entre objetos (parcelas, no nosso caso), mas temos um conjunto muito grande de variáveis (espécies, no nosso caso) que definem esses objetos e dificultam a nossa compreensão. Os objetivos dos métodos de ordenação consistem em condensar a informação de todas as variáveis em um número menor de eixos que consigam representar a maior parte da variação nos dados. Com um número menor de eixos é possível observar padrões que seriam dificilmente observados analisando cada variável inicial separadamente.

Se tivéssemos apenas duas espécies e duas parcelas, seria bem fácil saber o quanto as parcelas são similares entre si. Bastaria plotarmos as abundâncias de cada espécie em um gráfico e medir a distância entre as duas parcelas. Veja esse exemplo hipotético de duas espécies em duas parcelas:



Se tivéssemos três espécies (uma em cada eixo da figura) e 10 parcelas (pontos com cores diferentes), ainda seria possível visualizarmos parcialmente a semelhança entre as parcelas, mas com muito mais dificuldade.



Porém, acima de três espécies, precisamos da ajuda de métodos de ordenação! Inicialmente vamos aprender um dos métodos de ordenação conhecido como Análise de Componentes Principais (*Principal Component Analysis* - sigla **PCA**).

Como funciona uma PCA (Análise de Componentes Principais)?

Se quiséssemos visualizar a similaridade entre as nossas 100 parcelas usando diretamente os dados das nossas 10 espécies, precisaríamos de um gráfico com 10 dimensões (eixos), mas não teríamos a capacidade de visualizá-lo, pois só conseguimos visualizar até 3 dimensões (sim, é triste admitir, mas nós humanos somos muito limitados...). Então, quando temos mais de 3 espécies, torna-se impossível identificar visualmente quais parcelas são mais similares entre si, levando em consideração individualmente cada uma das espécies e suas abundâncias. Entretanto, se existem correlações fortes entre algumas dessas espécies, podemos reduzir as dimensões a um número menor de eixos derivados (também chamados de “componentes”) nos quais a maior parte das relações entre as parcelas estará representada. Matematicamente, isso será obtido por meio de álgebra matricial.

Antes de aplicarmos os procedimentos matemáticos, vamos visualizar um exemplo do que acontece quando temos três dimensões e queremos reduzir para dois eixos derivados (componentes), usando o método de Análise de Componentes Principais:

Nessa figura abaixo temos as abundâncias de três espécies (1, 2 e 3) representadas nos eixos X1, X2 e X3 para um conjunto de 26 parcelas, cujos nomes são indicados por letras de a até z. Esses são os dados brutos:



A Análise de Componentes Principais (PCA) vai “procurar”, a partir da nuvem de pontos formada pelas abundâncias das 3 espécies nas 26 parcelas, um eixo imaginário que capta a maior parte da variação nos dados. Esse eixo, conhecido como **Componente Principal 1** (ou pela sigla **PCA1**), expressará as similaridades e as diferenças mais importantes entre as parcelas.

Em um extremo desse novo eixo derivado estarão projetadas as parcelas mais parecidas entre si, enquanto no extremo oposto estarão as parcelas mais diferentes dessas (porém similares entre si). Além disso, dentre as três espécies, aquelas que estiverem mais correlacionadas com esse novo eixo derivado, serão as que foram mais importantes para defini-lo.

Você consegue visualizar quais são as parcelas mais similares e quais as principais espécies que determinam o eixo **PCA1** na figura abaixo?



Depois da definição do eixo 1, em geral, ainda haverá uma certa quantidade de variação nos dados. A Análise de Componentes Principais vai então procurar o segundo eixo ²⁾, que será ortogonal ao eixo 1 (e por isso, independente do eixo 1) e vai explicar o máximo possível da variação remanescente nesse novo eixo (veja o eixo **PCA2** na figura acima).

Agora podemos projetar os dois novos eixos obtidos e conseguiremos avaliar com maior facilidade as parcelas que são mais similares entre si:



Quando temos 3 variáveis (espécies, no caso acima) conseguimos visualizar de forma aproximada esse processo de redução de dimensões, mas quando temos 4 variáveis ou mais, como já explicamos, não temos a capacidade de visualizar o processo. Porém, é possível, matematicamente, fazer a redução e identificar os principais eixos derivados e as variáveis que mais contribuíram para a formação desses eixos derivados.

Por meio de álgebra matricial, esses eixos derivados serão definidos a partir da matriz de correlações entre as espécies (ou seja, aqueles valores que estavam na parte direita e acima da diagonal da matriz de correlação que visualizamos anteriormente). Dessa matriz serão obtidos os autovalores e os autovetores ³⁾ que definirão esses novos eixos derivados.

Agora, vamos aplicar o procedimento de análise na nossa comunidade de 10 espécies e 100 parcelas!

Aplicando PCA para ordenar as parcelas em relação às espécies

Para realizar a Análise de Componentes Principais para a nossa comunidade vamos utilizar uma função existente no pacote *vegan* do R.

Vamos aplicar a função “rda” ao nosso objeto “sp” e vamos gerar um novo objeto “pca.sp”, que vai conter todos os resultados dessa análise. O comando vai ficar assim:

```
pca.sp <- rda(sp, scale = TRUE)
```

Mas, o que é **scale=TRUE**?

Isso é um requisito do pacote “vegan” para a realização de uma PCA baseada em correlações. Com esse comando os valores que serão utilizados na análise não serão os valores reais das abundâncias das espécies nas parcelas, mas sim os valores corrigidos (ou melhor, escalonados) pelo desvio padrão das abundâncias das espécies nas parcelas. Não se preocupe em entender isso detalhadamente agora. É simplesmente uma forma de evitar problemas com valores muito discrepantes de abundâncias entre as espécies.

Bom, agora vamos visualizar somente o primeiro eixo gerado pela PCA e olhar quais parcelas são mais parecidas e quais são mais diferentes. Para isso vamos aplicar uma função chamada “plot” no objeto que contém os resultados da análise (pca.sp) e vamos pedir para essa função nos mostrar apenas o eixo 1 (usando o termo “choices = c(1)”). ⁴⁾ O comando completo fica assim:

```
plot(pca.sp, choices= c(1), display= c("sp","wa"), scaling = "species",  
cex=0.5)  
## Para gravar um pdf com o grafico  
pdf(file="pca1.pdf", height=9)  
plot(pca.sp, choices= c(1), display= c("sp","wa"), scaling = "species",  
cex=0.5)  
dev.off()
```

Esse gráfico parece estranho, mas não é complicado. Vamos entendê-lo:



A linha representa o primeiro eixo da PCA que, nesse caso, está definido por um intervalo de aproximadamente -1,5 (extremo à esquerda) a +1,5 (extremo à direita). A PCA atribui um valor para cada parcela nesse eixo, que chamamos de **escores das parcelas**. De maneira análoga, a PCA também cria os **escores das espécies**, nesse mesmo eixo.

Identifique quais são as espécies (sp1, sp2,...sp3?) que estão nos extremos desse eixo e as parcelas (sit1, sit2,...,sit10,...,sit100?) mais próximas dessas espécies. Identifique também alguns conjuntos de parcelas que estão mais próximas entre si (em qualquer região do eixo).

Até aqui, tudo bem? Tomara que sim! Porém uma PCA não produz apenas um eixo de explicação da variação dos dados. O primeiro eixo (esse que acabamos de visualizar) é sempre o eixo que explica a maior parte da variação dos dados. Como explicado anteriormente, o segundo eixo produzido pela PCA será um eixo ortogonal ao primeiro eixo, pois assim, fica garantido que ele é independente do primeiro. Ele será o eixo ortogonal que explica a maior parte da variação restante. E assim por diante com os outros eixos...Os últimos eixos geralmente explicam muito pouco sobre a ordenação das parcelas.

Ainda está difícil de visualizar?



Veja estes gráficos interativos dos casos com duas e três variáveis:

- <http://setosa.io/ev/principal-component-analysis/>

Entendendo os resultados da PCA

Então, agora, vamos ver os resultados numéricos da PCA para a nossa comunidade. Para isso, basta colar no R o nome do objeto que foi criado:

```
pca.sp
```

Vamos entender alguns pontos importantes desse resultado:

Inércia (*Inertia*) é o termo usado no pacote *vegan* para indicar a variação nos dados. O valor total aqui nesse caso que estamos calculando é 10, pois estamos usando a matriz de correlação escalonada com valores de 0 a 1 para cada variável (espécies) e temos 10 variáveis. Então, o valor máximo que a inércia total pode assumir é 10.

Logo abaixo, a gente tem os valores de variação explicada por cada um dos eixos derivados. Para sabermos a porcentagem de explicação de cada eixo derivado, basta dividirmos esse valor pela inércia total.

E esse tal de “*Unconstrained*”??? Calma, mais abaixo a gente explica... por agora, não se preocupe

com isso.

Mas, ainda temos um resultado muito simplista da PCA. Vamos usar a função “summary” para vermos um resultado mais completo e entendermos melhor o que são e como são compostos esses diferentes eixos derivados:

```
summary (pca.sp)
```

Agora já podemos ver a proporção da variação total que cada eixo explica (“*Proportion Explained*”) e também a proporção acumulada à medida que adicionamos eixos (“*Cumulative Proportion*”).

Além disso, mais embaixo, você verá a tabela com os “*Species scores*”. Nas linhas estão as espécies e nas colunas os diferentes eixos derivados da PCA.

Compare as espécies que você identificou no gráfico anterior como aquelas presentes nos extremos do eixo 1 com os valores apresentados na primeira coluna (que representa o primeiro eixo - PC1) dessa tabela.

É isso aí. Se tudo deu certo, esses valores devem ser relativamente parecidos.

Então, os escores das espécies (“*Species scores*”) representam a ordem de importância das espécies na definição de cada um dos eixos derivados calculados pela PCA. Maiores valores, sejam eles positivos ou negativos, indicam as espécies que mais contribuíram para a formação desse eixo.

Olhando agora para o eixo 2 (PC2), identifique quais são as espécies que têm os maiores valores positivos. São as mesmas do eixo 1? E as espécies que tinham maiores valores no eixo 1, quais valores têm nesse segundo eixo, altos? baixos? intermediários?

Você consegue visualizar o que está acontecendo com a ordem de importância das espécies? Difícil fazer isso olhando os números apenas, né?

Antes de mostrarmos o gráfico com esses eixos plotados, vamos olhar agora para a segunda tabela, dos “*Sites scores*”. Nessa tabela, são as parcelas (sit1, sit2, sit3, etc) que estão representadas nas linhas.

Procure as parcelas e siga o mesmo raciocínio que usamos acima para as espécies.

Ok, agora chega de imaginar. Vamos ver como ficam os dois primeiros eixos da PCA para a nossa comunidade em um gráfico e observar como as parcelas estão organizadas nesse espaço bidimensional.

Visualizando os gráficos de PCA e interpretando um plano de ordenação

Vamos aplicar uma função especial (*biplot*) sobre o nosso objeto “pca.sp” para fazermos esse tipo de gráfico⁵⁾:

```
biplot(pca.sp, scaling=2, type = c("text", "text"))
```

Salve essa figura em sua pasta, da mesma forma já realizada anteriormente. Faça o mesmo com as próximas figuras que produzir. Lembre sempre de colar o comando **dev.off()** no final do comando para salvar o gráfico.



Obs.: Note que aquele eixo 1 que a gente já visualizou sozinho, anteriormente, é o mesmo que está nesse gráfico, ok? Para ter certeza disso, projete a localização de algumas parcelas ou espécies diretamente no eixo 1 dessa figura e veja que estão exatamente na mesma posição que estavam no gráfico anterior

Neste tipo de gráfico que foi produzido, os vetores (setas) indicam em que direção aumenta a abundância de cada espécie e o comprimento dos vetores indica a importância da espécie para cada eixo. Os pontos representam as parcelas.

Olhando para o gráfico, a qual conclusão você chega em relação à ordenação das parcelas em função das espécies e suas abundâncias utilizando os dois primeiros eixos?

Dicas

- Quais espécies têm abundâncias correlacionadas, ou seja, quais são as associações de espécies?
- Quais parcelas têm cada associação?
- Qual a sua interpretação de cada eixo?

Precisamos olhar também a relação com o terceiro eixo?

Para esse nosso exercício prático não vamos olhar, pois os dois primeiros eixos (PC1 e PC2) já explicam cerca de 65% da variação e consideramos que isso é suficiente para os nossos propósitos. Porém, em alguns casos o terceiro eixo também pode trazer informações importantes⁶⁾.

Muito bem, depois de entendermos a lógica básica de funcionamento de uma PCA e interpretarmos a ordenação produzida pelos dois primeiros eixos, a pergunta que surge naturalmente é:

“Por que algumas espécies ocorrem predominantemente e em maior abundância em algumas parcelas e outras não?”

Tempo para pensar!



Uma das respostas mais rápidas que vêm à mente de um(a) ecólogo(a) é que parcelas com maior similaridade de espécies podem ter variáveis ambientais mais parecidas e que as diferenciam do restante.

Como podemos avaliar isso?

Conhecendo as variáveis ambientais

Podemos ir a campo e medir variáveis ambientais em todas as parcelas, como características físicas do solo (granulometria, porosidade, etc), características químicas do solo (N, P, K, Ca, Mg, pH, Matéria Orgânica, Capacidade de Troca Catiônica, etc), topografia e luminosidade, etc. Com esses dados podemos construir uma tabela em que as linhas são as parcelas e as colunas são as variáveis ambientais. Com essa tabela podemos fazer uma PCA. Nesse caso, ao invés das espécies, as variáveis que definirão a ordenação das parcelas serão as variáveis ambientais.

Vamos conhecer então as variáveis ambientais desse nosso exercício prático:

```
amb<-read.csv2 ("ambientais.csv")
head (amb)
summary(amb)
```

Nessa planilha temos as 5 variáveis ambientais hipotéticas e os valores de cada uma delas em cada parcela de amostragem.

Vamos ver como essas variáveis estão correlacionadas par-a-par:

```
pairs(amb, upper.panel=panel.cor)
```



Quais variáveis estão mais correlacionadas?

Calculando PCA para variáveis ambientais

Agora, vamos ver como fica a ordenação das parcelas de acordo com essas variáveis ambientais, aplicando a PCA nesse conjunto de dados somente com as variáveis ambientais.

```
pca.amb<-rda(amb, scale = TRUE)
pca.amb
summary (pca.amb)
```

```
biplot (pca.amb, scaling=2,type = c("text", "text"))
```



Será que essa ordenação é parecida com a ordenação obtida para as espécies?

Podemos colocar os gráficos lado a lado e comparar. Porém, fazendo dessa forma, teremos apenas uma comparação visual e não podemos avaliar o quanto essas variáveis ambientais explicam a distribuição das espécies nas parcelas. Felizmente, existem métodos analíticos para nos ajudar com isso.

Análise de ordenação restrita (Constrained ordination analysis)

Podemos fazer uma análise que junta essas duas matrizes de informações sobre as parcelas (abundâncias das espécies e variáveis ambientais), de modo que a ordenação das parcelas pelas espécies seja ajustada pelas variáveis ambientais ("*constrained ordination*"). A partir das diferenças entre a ordenação irrestrita ("*unconstrained*") e da ordenação restrita ("*constrained*") é possível calcular o quanto as variáveis ambientais explicam a distribuição das espécies nas parcelas.

Uma das análises que faz esses cálculos é a chamada de Análise de Redundância (*Redundancy Analysis - RDA*), que é feita a partir da mesma lógica que usamos antes para fazer as PCAs, porém, reordenando as parcelas em função das variáveis ambientais.

Relações entre espécies e variáveis ambientais

Antes de fazermos a análise é importante visualizarmos as relações entre as abundâncias das espécies e os valores das variáveis ambientais em cada parcela.

Vamos olhar como ficam as relações no caso das nossas amostras. Ou seja, vamos ver quais das 10 espécies estão relacionadas linearmente com quais variáveis ambientais. Para isso, copie e cole os comandos abaixo:

```
par(mfrow=c(10,5), mar=c(0.1,0.1,0.1,0.1))
for(i in 1:10){
  for(j in 1:5){
    plot(sp[,i] ~ amb[,j], axes=FALSE)
    box()
  }
}
```

Nessa figura cada gráfico apresenta a relação entre os valores de abundância de uma espécie (linhas) e os valores de uma variável ambiental (colunas) para todas as parcelas. Assim o primeiro gráfico em cima e à esquerda apresenta a relação entre a espécie 1 (sp1) e a variável ambiental 1 (x1). O gráfico logo abaixo apresenta a relação entre a espécie 2 (sp2) e a variável 1 (x1), e assim por diante.



Veja quais são as variáveis ambientais que apresentam relações mais lineares e mais fortes com as espécies.

Calculando RDA para espécies x variáveis ambientais

Não vamos entrar em muitos detalhes técnicos dessa análise, mas vale a pena lembrar que agora as variáveis ambientais é que condicionam a ordenação das parcelas. Essa nova ordenação não será feita diretamente com os valores reais das abundâncias das espécies (aqueles que estavam no nosso objeto “sp”). Os valores usados na RDA serão os valores de abundância ajustados em função das variáveis ambientais (que estão no objeto “amb”).

Esse tipo de análise é chamado de **Análise de Ordenação Restrita** (ou *constrained ordination analysis*) exatamente porque um dos conjuntos de dados (espécies, no nosso caso) vai ser condicionado (*constrained*) pelo outro conjunto de dados (variáveis ambientais, no nosso caso).

Olhando para aquela figura que mostra a relação entre cada espécie e cada variável ambiental, podemos notar que algumas espécies estão fortemente relacionadas a algumas variáveis ambientais, enquanto outras espécies estão fracamente relacionadas às variáveis ambientais. Em uma análise de RDA, quanto maior o número de espécies relacionadas às variáveis ambientais e quanto mais forte forem essas relações, maior será a importância das variáveis ambientais na definição da similaridade entre as parcelas.

Vamos então fazer a RDA para os nossos dados, ajustando os dados das abundâncias das espécies (“sp”) em função das variáveis ambientais (“amb”):

```
rda.sp.amb<- rda(sp, amb)
```

OBS.: note que estamos usando a mesma função “rda”, mas agora temos duas planilhas dentro dos parênteses (“sp” e “amb”). Além disso, não estamos usando “**scale=TRUE**” como fizemos na análise de PCA, pois isso não é necessário para uma RDA.

Vamos olhar e interpretar os resultados obtidos:

```
summary(rda.sp.amb)
```

A primeira informação importante desses resultados é a tabela de Partição da Variância (*Partitioning of variance*).

No item “*Constrained*”, o valor apresentado representa a proporção da variação das abundâncias das espécies nas parcelas que é explicada pelas variáveis ambientais que adicionamos na análise de ordenação restrita (*Constrained*). Esse valor tem um significado “similar” ao do R^2 de uma análise de regressão linear

A proporção apresentada no item “*Unconstrained*” representa a variação das abundâncias das espécies nas parcelas que não é explicada pelas variáveis ambientais escolhidas.



E então, você acha que nossas variáveis ambientais explicam pouco ou muito da distribuição das espécies?

Continuando a analisar os resultados obtidos, temos abaixo os valores de proporção de explicação dos eixos de ordenação criados pela RDA. Note que são 5 eixos restritos (RDs) e 10 eixos não restritos (PCs). Se você somar a proporção explicada pelos 5 eixos RDs verá que o valor é igual ao valor total do item “*Constrained*” na tabela anterior. Já na soma dos eixos PCs o valor é igual ao do item “*Unconstrained*”

Mais abaixo estão as tabelas com os valores:

- dos escores das espécies (“sp” que será usado no gráfico abaixo)
- dos escores das parcelas (sites) ponderados pelas abundâncias das espécies (não serão usados no gráfico abaixo)
- das escores das parcelas (sites) mas já com os ajustes impostos pelas variáveis ambientais (“lc” que será usado no gráfico abaixo)
- dos escores das variáveis ambientais (“bp” que será usado no gráfico abaixo)

A partir dos dados dessas tabelas, nós podemos fazer um gráfico que mostre a ordenação das parcelas, das espécies e das variáveis ambientais nos dois primeiros eixos (RDA1 e RDA2) para entender quais são as variáveis ambientais que determinam mais fortemente a distribuição das espécies, que por sua vez determinam a ordenação das parcelas.

Esse gráfico é um pouco mais complicado, então, vamos explicar passo-a-passo

Primeiro vamos montar um **gráfico vazio**... estranho, hein???? Mas acredite, vai ser legal!

```
plot(rda.sp.amb, type="none", choices=c(1,2))
```

Depois, vamos adicionando os elementos aos poucos. Vá copiando e colando linha por linha e veja como o gráfico vai sendo montado.

```
points(rda.sp.amb, display="lc", cex=0.01) ##Adicionando os pontos
referentes às parcelas

text(rda.sp.amb, display="lc", cex=0.5) ##Adicionando os nomes das parcelas

text(rda.sp.amb, display="bp", col="blue") ##Adicionando as setas das
variáveis ambientais

spamb2.scores<-scores(rda.sp.amb, choices=1:2, display="sp") ##selecionando
os escores das espécies (foram selecionados apenas os eixos 1 e 2)

arrows(0,0,spamb2.scores[,1],spamb2.scores[,2], length=0, lty=1, col="red")
##adicionando linhas para as espécies

text(rda.sp.amb, display = "species", cex=0.8, adj=-0.3, col="red")
```

##Adicionando nomes às linhas das espécies

Neste gráfico as variáveis ambientais estão representadas pelas setas azuis, as espécies pelas linhas em vermelho e as parcelas pelos pontos pretos.



Ok, até aqui, então, temos uma explicação bem razoável da distribuição das espécies nas parcelas, pois as variáveis ambientais parecem ser importantes na determinação da distribuição das espécies.

Porém, existe ainda uma outra explicação que os(as) ecólogos(as) sempre precisam avaliar antes de afirmar que são as condições ambientais que determinam a coocorrência de algumas espécies em determinadas parcelas.

Tempo para pensar



Ok, vamos lá, você já deve estar começando a ficar cansado(a), então vamos ajudar.

As espécies podem estar distribuídas de forma **espacialmente agregada** em algumas parcelas e isso pode ser simplesmente um reflexo da limitação da dispersão das espécies.⁷⁾ Na agregação de espécies são as populações de várias espécies que estão agregadas em locais próximos. Se isso estiver acontecendo, parcelas mais próximas devem apresentar maior similaridade de composição e abundância de espécies. Como parcelas mais próximas também podem ter condições ambientais mais semelhantes, a associação entre espécies e variáveis ambientais pode ser apenas uma correlação espúria, devido à agregação das espécies. Precisamos avaliar se isso acontece!

Então, é importante que tenhamos os dados de localização espacial de cada uma das parcelas estudadas para podermos incluir na nossa próxima análise.

Dados das coordenadas das parcelas

É exatamente isso que temos na nossa planilha “coordenadas.csv”.

Vamos importá-la para o R:

```
coords<-read.csv2 ("coordenadas.csv")
head (coords)
summary(coords)
```

Se usássemos essa planilha diretamente na nossa próxima análise, apenas os ângulos retos e as diagonais de distância geográfica entre as parcelas seriam consideradas nas análises que visam entender como estão distribuídas espacialmente aquelas parcelas que apresentam maior similaridade em composição de espécies.

Porém, a estruturação espacial das parcelas mais similares pode ser bem mais complicada que as descritas por distâncias em linhas retas e diagonais. Na imagem abaixo são mostrados alguns exemplos de diferentes estruturas espaciais possíveis:



Para construir essa figura ⁸⁾ foram simuladas várias estruturas espaciais hipotéticas para serem comparadas com os dados reais.

Vamos imaginar que os dados reais que queremos analisar são a composição e abundância das espécies em parcelas que estão ordenadas ao longo do eixo 1 de uma PCA. Então, nessa figura os quadrados maiores representam os valores mais positivos (pretos) e mais negativos (brancos) dos escores das parcelas para esse eixo. Quanto menor o tamanho do quadrado, mais próxima do centro ("zero") do eixo 1 estará a parcela. Apesar das duas primeiras figuras (PCNM1 e PCNM2) serem facilmente descritas por distâncias em linha reta, note como seria complicado descrever alguns desses padrões (por exemplo, PCNM15 e PCNM21) usando sistemas simples de coordenadas.

Diante disso, para podermos incorporar outras estruturas espaciais possíveis teremos que aplicar dois procedimentos - *PCNM* - *Principal Coordinates of Neighbour Matrices* e *forward selection*. Esses procedimentos são muito usados em trabalhos científicos cujos objetivos envolvem a análise da importância do espaço na estruturação de comunidades. Vale a pena conhecer como funcionam, mas neste exercício não vamos entrar em detalhes sobre esses métodos e vamos usar uma tabela com as variáveis espaciais já trabalhadas e selecionadas para o nosso conjunto de dados.

Então, vamos pedir que acredite em nós!



Se você acreditar, copie essa tabela para a sua pasta:

[tab_coords_sel.csv](#)

Agora, importe a tabela para o R:

```
tab.coords.sel<-read.csv2("tab_coords_sel.csv")
```

Essa tabela importada ("tab.coords.sel") contém as informações que serão usadas para representar a estrutura espacial dos dados na última etapa do nosso exercício.

Nessa tabela estão apenas as coordenadas PCNMs que foram previamente selecionadas porque melhor explicam a distribuição espacial das abundâncias das espécies nas parcelas

Partição da Variação

Então, agora estamos interessados em calcular qual a porcentagem de variação da distribuição das espécies nas parcelas que é explicada:

- exclusivamente pelas variáveis ambientais
- exclusivamente pela estrutura espacial
- pelos dois fatores em conjunto

O valor dos fatores em conjunto das variáveis ambientais e da estrutura espacial revela a relação entre essas duas matrizes de dados. Valores altos para os “efeitos em conjunto” indicam que parcelas mais próximas espacialmente têm também variáveis ambientais mais similares, ou, em outras palavras, isso significa que **as variáveis ambientais estão espacialmente estruturadas**.

Poderíamos fazer diferentes RDAs separadamente para cada situação e analisar as mudanças nos valores de **porcentagem explicada**. Mas temos uma análise que faz isso para nós de uma só vez. Vamos então aplicar essa análise conhecida como **Partição da Variação** (*Variation Partitioning*) e analisar as porcentagens explicadas pelas variáveis ambientais, pela estrutura espacial e por ambas. Ah! E também poderemos avaliar qual a **porcentagem não explicada** por esses fatores que estamos analisando. Para isso, vamos aplicar a função “varpart” para os nossos três objetos criados anteriormente nesse roteiro (“sp”, “amb” e “tab.coords.sel”) e vamos criar um objeto (“particao”) que contém os resultados dessa análise:

```
(particao <- varpart(sp, amb , tab.coords.sel))
```

A tabela resultante mostra a variação total, decomposta em:

- $a+b+c$ = \$ variação total
- $a+b$ = \$ variação total devido ao ambiente (\$X1\$)
- $b+c$ = \$ variação total devido ao espaço (\$X2\$)

Note que:

- a = \$ variação exclusivamente devido ao ambiente (\$X1|X2\$)⁹⁾
- c = \$ variação exclusivamente devido à estrutura espacial (\$X2|X1\$)¹⁰⁾

Além disso, temos também:

- d = \$ variação não explicada (resíduo)

Veja como ficaram os resultados para a nossa comunidade:



Um gráfico (Diagrama de Venn) que ajuda a entender essa partição:

```
plot(particao)
```



Obs.: Não leve em consideração o tamanho dos círculos, mas sim os valores indicados neles e fora deles.

Você entende o que é o componente b da partição?

E então, como você interpreta esses resultados obtidos, no contexto de processos determinísticos baseados em nicho e processos neutros baseados em limitação de dispersão?

Se tiver um tempinho ainda, tente entender a diferença entre os resultados obtidos para comunidades de samambaias em dois transectos na Bolívia

Conhecendo melhor os dados analisados

Você tem curiosidade de saber como os valores (abundâncias das espécies e variáveis ambientais) estão distribuídos nas nossas parcelas hipotéticas?

Veja abaixo os mapas com as distribuições dos valores das 5 variáveis ambientais e dos valores das abundâncias das espécies (sp1, sp2, sp3,..., sp10) na nossa grade de 10 x 10 parcelas.

As parcelas com cores mais escuras representam valores mais altos, seja de abundância das espécies ou das variáveis ambientais.

Em ambas as figuras, a parcela número 1 (sit1) é a primeira à esquerda e embaixo, a parcela 10 (sit10) é a última à direita e embaixo, a parcela 91 (sit91) é a primeira à esquerda e em cima e a parcela 100 (sit100) é a última à direita e em cima.

NOSSAS 5 VARIÁVEIS AMBIENTAIS:



NOSSAS 10 ESPÉCIES:



Se você gostar de programação e quiser entender como os valores foram produzidos, clique [aqui](#) e veja o roteiro utilizado.

Quer saber mais?

- Borcard et al. (2011). Numerical Ecology with R.
- Legendre & Legendre (2003). Numerical Ecology.
- Sobre os PCNMs: Borcard, D. and P. Legendre. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecological Modelling 153(1-2): 51-68.
- Ótima página de Michael Palmer
- Ótima página de David Zeleny
- Tutorial do vegan
- Introdução a análise multivariada do vegan

1)

ou seja, cada célula da planilha é a abundância de uma espécie em uma parcela

2)

o número de eixos que podem ser obtidos é igual ao número de variáveis originais, porém, esperamos que uma PCA consiga explicar a maior parte da variação em poucos eixos

3)

Cuidado! A matriz da qual os autovalores são obtidos aqui na PCA é muito diferente daquelas matrizes da Análise de Projeção Matricial que aprendemos no módulo de Dinâmica de Populações. Apesar da matemática por trás do cálculo dos autovalores ser a mesma, as interpretações são diferentes!

4)

Algumas outras características do gráfico são definidas pelos termos “display”, “scaling” e “cex”, mas não vamos entrar em detalhes sobre isso aqui, pois são peculiaridades da função plot e do programa

R

5)

Novamente, não se preocupe com os outros termos que estão dentro dos parênteses, pois são peculiaridades do R

6)

Nessas situações, o ideal é fazer também um gráfico relacionando os eixos PC1 e PC3 e um outro gráfico relacionando os eixos PC2 e PC3. Existem técnicas para determinar quantos eixos devem ser inspecionados para entender a maior parte da variação. Se tiver interesse, veja nos materiais suplementares algumas dessas técnicas

7)

Cuidado! Aqui estamos falando de agregação de espécies, ou seja, um conjunto formado pelas mesmas espécies e com altas abundâncias ocorrendo em parcelas espacialmente próximas. Isso é um pouco diferente da agregação dos indivíduos de uma população de apenas uma espécie, que abordamos no módulo sobre Estrutura de Populações

8)

obtida na página do *California Soil Resource Lab* da University of California, Davis

9)

ou seja, variação que pode ser atribuída ao ambiente, descontado o efeito da estrutura espacial.

10)

ou seja, variação que pode ser atribuída à estrutura espacial, descontado o efeito das variáveis ambientais.

From:

<http://labtrop.ib.usp.br/> - **Laboratório de Ecologia de Florestas Tropicais**

Permanent link:

<http://labtrop.ib.usp.br/doku.php?id=cursos:popcom:2017:roteiros:ec>



Last update: **2021/07/20 12:43**