

# 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

Hoje em dia sabemos que as comunidades não são homogêneas e, portanto, diferentes conjuntos de espécies são encontrados em diferentes localidades. Para entender essas diferenças, o primeiro passo é saber quais localidades são mais similares entre si e quais são mais dissimilares. Uma forma de fazer isso é utilizando os chamados métodos de ordenação. Existem vários métodos com características diferentes e adequados a conjuntos de dados específicos. Alguns textos importantes estão disponíveis no final desse roteiro para quem se interessar mais pelo assunto.

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.

Mude o diretório de trabalho do programa R para a sua pasta. Isso pode ser feito de duas formas:

- Digitando `setwd("C:/desktop/nome da sua pasta")` no R
- Abrindo o Menu "Arquivo" (ou "File") e selecionando "Mudar dir" (ou "Change dir"). Nesse caso, escolha a sua pasta na janela que abrir. [Obs. Para Mac, essa opção está no Menu "Misc" e a opção é "Change working dir", ou usando a tecla `command+D`]

Para checar se você está na pasta correta, copie e cole o comando abaixo no console do R:

```
getwd()
```

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 as funções já prontas de 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!**

E agora, carregue o pacote no R

```
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 eles utilizam.

## Conhecendo os dados das espécies de plantas

Imagine que amostramos uma área de 1km x 1km, dividindo-a em 100 parcelas de 100x100m. 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 <sup>1)</sup>. 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:

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

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

```
head(sp)
```

Se quiser ver um resumo das variáveis existentes na planilha, use a função `summary`:

```
summary(sp)
```

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

Se tivéssemos apenas duas espécies, seria bem fácil saber quais parcelas são mais parecidas entre si. Bastaria plotarmos as abundâncias de cada espécie em um gráfico de dispersão e medir a distância em linha reta entre as duas parcelas (essa distância é chamada de **Distância Euclidiana**).

Veja esse exemplo hipotético de duas espécies em duas parcelas:



Se existissem mais parcelas, bastaria medir as distâncias entre cada par de parcelas <sup>2)</sup>.

Se tivéssemos três espécies, ainda seria possível visualizar a semelhança entre as parcelas, mas com mais dificuldade. Note que nesse exemplo abaixo temos 10 parcelas.



Porém, acima de três espécies fica praticamente impossível visualizar as relações entre as parcelas.

## Visualizando Correlações

Uma outra forma de tentar entender como as espécies estão distribuídas nas parcelas seria olhar para as correlações entre as abundâncias de cada par de espécies nas parcelas. Isso nos ajudaria a entender quais pares de espécies tendem a ocorrer com abundâncias similares nas mesmas parcelas.

Então, vamos fazer a matriz de correlação entre todos os pares de espécies da nossa tabela e olhar como elas se relacionam!

Antes, precisamos importar uma função para facilitar a visualização dessas correlações. Copie tudo que está dentro desse quadro e cole no R:

```
## 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))
  r <- abs(cor(x, y))
  txt <- format(c(r, 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 uma função ("pairs") que vai calcular a correlação par-a-par na nossa tabela de espécies por parcelas (que se chama "sp") e vamos utilizar a função "panel.cor" para melhorar a visualização.

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

**Salve essa figura em pdf na sua pasta**

Como entender essa figura?

Essa figura é uma representação da matriz de correlação entre as abundâncias das espécies nas parcelas. O nome/código das espécies está na diagonal principal da matriz.

Para entender qual espécie está representada no eixo x e qual está no eixo y de cada um dos gráficos (mostrados na parte de baixo e à esquerda da diagonal), basta encontrar os nomes 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 no eixo x, e da espécie 4 no eixo y.

Os valores que estão acima e à direita da diagonal principal são os valores da correlação referentes aos gráficos que se encontram na posição espelhada abaixo da diagonal. Note que os valores de correlação variam entre -1,0 e +1,0. 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. Mas imagine se tivéssemos 144 espécies (como naquele exemplo lá no sul da Bahia) ou as mais de 1000 espécies da parcela permanente de Manaus? Só para terem uma ideia, para as 144 espécies, teríamos 10.296 gráficos...😬

## Análise de Ordenação

Inicialmente vamos aprender um dos métodos de ordenação, conhecido como Análise de Componentes Principais (*Principal Component Analysis - PCA*).

### Como funciona uma PCA?

Se conseguíssemos visualizar as correlações entre as 10 espécies em um mesmo gráfico seria lindo, mas infelizmente não temos essa capacidade...<sup>3)</sup>

Quando temos mais de 3 espécies, fica difícil identificar quais parcelas são mais parecidas com as outras, levando em consideração todas as espécies e suas abundâncias. Entretanto, matematicamente isso é possível, por meio de álgebra matricial.

A Análise de Componentes Principais (PCA) procura, na nuvem de pontos formada pelas abundâncias das 10 espécies nas 100 parcelas, um eixo imaginário que melhor expressa a similaridade e as diferenças entre as parcelas. Matematicamente, esse eixo será definido pelo cálculo dos autovalores a partir dos valores de correlação apresentados na matriz de correlação entre as espécies (ou seja, aqueles valores que estavam na parte direita da matriz de correlação que vimos anteriormente).<sup>4)</sup>

Esse primeiro eixo é conhecido como Componente Principal 1 (ou pela sigla PC1) e ele vai usar as correlações entre as espécies para colocar em um extremo do eixo as parcelas mais parecidas entre si e no extremo oposto, as parcelas mais diferentes dessas (porém similares entre si). As parcelas que não são parecidas com nenhum desses extremos ficam na região intermediária do eixo.

Então, vamos ver na nossa comunidade hipotética, como fica esse primeiro eixo.

## Calculando PCA para espécies

Vamos realizar a Análise de Componentes Principais, utilizando uma função já pronta no pacote *vegan* do R. Vamos colocar um nome no objeto que vai ser gerado pela análise, pois queremos ver detalhadamente os resultados encontrados.

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

A função `rda` está sendo aplicada na planilha “sp” e vai gerar o objeto “pca.sp”.

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 o primeiro eixo gerado pela PCA e olhar quais parcelas são mais parecidas e quais são mais diferentes.

```
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:

Identifique quais são as parcelas (sit1, sit2,...,sit10,...,sit100?) mais parecidas entre si nesse eixo e

quais são as espécies(sp1, sp2,...sp3?) que estão mais próximas dessas parcelas nesse eixo. Essas são as espécies que estão definindo a similaridade entre as parcelas, de acordo com esse eixo.

A linha vertical representa o primeiro eixo do PCA, que no caso é o intervalo de -1,5 a 1,5. O PCA atribui um valor para cada parcela neste eixo, que chamamos de **escores das parcelas**. De maneira análoga, o PCA também cria os **escores das espécies**, neste mesmo eixo.

Agora que você sabe disso, identifique as espécies com os cinco maiores escores. Anote o valor aproximado do escore de cada espécie, por inspeção visual da posição delas. Faça o mesmo para as parcelas com os 5 maiores escores.

Ok, mas 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. O segundo eixo produzido pela PCA será um eixo ortogonal ao primeiro eixo, pois assim, fica garantido que o segundo eixo 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.

## Entendendo os resultados da PCA

Então, agora, vamos ver os resultados da PCA para a nossa comunidade hipotética:

```
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. Para sabermos a porcentagem de explicação de cada eixo, 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. Para ver um resultado mais completo e entender melhor o que são esses diferentes eixos, use:

```
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 eixos da PCA.

**Compare agora os valores que você anotou antes para os “escores das parcelas no 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 projeção da espécie em cada um dos eixos calculados pela PCA.

**Olhando agora para o eixo 2 (PC2), identifique quais são as espécies que têm os maiores valores positivos. São as mesmas que as 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 as espécies?**

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 mais parecidas e veja os valores que elas têm no eixo 1 (PC1). Veja se são similares aos valores que você estimou (tomara que sim...). Agora olhe a localização dessas parcelas no eixo 2 (PC2) 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 em um gráfico e como as parcelas estão organizadas nesse espaço bidimensional.

### **Visualizando os gráficos de PCA**

Vamos usar uma função especial (*biplot*) para fazer esse tipo de gráfico:

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

Salve essa figura em sua pasta. Faça o mesmo com as próximas figuras de ordenação que produzir.

**Obs.: Note que aquele eixo 1 que a gente já visualizou sozinho na posição vertical, agora está na horizontal, ok?**

## Interpretando um plano de ordenação

Os vetores indicam em que direção aumenta a abundância de cada espécie. Os pontos neste gráfico 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 importantes<sup>5)</sup>.

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 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"))
```

**Salve essa figura em sua pasta.**

## 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 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 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 qual(is) das 10 espécies está(ão) relacionada(s) linearmente com qual(is) variável(is) ambiental(is). 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()
  }
}
par(mfrow=c(1,1))
```

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 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 ambiente

Não vamos entrar nos detalhes técnicos da análise, mas vale a pena lembrar que agora as variáveis ambientais é que condicionam a ordenação das parcelas. Essa nova ordenação será feita a partir dos valores das abundâncias das espécies ajustados às variáveis ambientais. É por esse motivo que esse tipo de análise é chamado de análise de ordenação restrita (ou *constrained ordination analysis*).

```
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 <sup>6</sup>.

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
- dos escores das parcelas (sites) ponderados pelas abundâncias das espécies
- dos escores das parcelas (sites) mas já com os ajustes impostos pelas variáveis ambientais
- dos escores das variáveis ambientais

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 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 é montado.

```
points(rda.sp.amb, display="sites", cex=0.01) ##Adicionando os pontos
```

referentes às parcelas

```
text(rda.sp.amb, display="sites", 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
```

Como vocês já perceberam, 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 a distribuição das espécies é definida pelas condições ambientais.**

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 <sup>7)</sup>. 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, a estrutura espacial das parcelas que são mais similares em termos de abundâncias das espécies seria avaliada considerando-se apenas os ângulos retos e as diagonais de distância entre parcelas.

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



Para construir essa figura <sup>8)</sup> 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 “zero” (centro) 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.

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 previamente selecionadas que melhor explicam a distribuição espacial das abundâncias das espécies nas parcelas.

## Partição da Variação

Por fim, vamos calcular a porcentagem de variação da distribuição das abundâncias das espécies que é explicada:

- exclusivamente pelas variáveis ambientais
- exclusivamente pela estrutura espacial

- pelos dois efeitos em conjunto

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

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.

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

A tabela mostra a variação total das abundâncias das espécies, decomposta em:

- $a+b+c =$  variação total
- $a+b =$  variação total devido ao ambiente ( $X_1$ )
- $b+c =$  variação total devido ao espaço ( $X_2$ )

Note que:

- $a =$  variação exclusivamente devido ao ambiente ( $X_1|X_2$ )<sup>9)</sup>
- $c =$  variação exclusivamente devido à estrutura espacial ( $X_2|X_1$ )<sup>10)</sup>

Além disso, temos também:

- $d =$  variação não explicada (resíduo)

Um diagrama que ajuda a entender essa partição:

```
plot(particao)
```

### E então, como você interpreta esse resultado?

Dica: você precisa entender o que é o componente  $b$  da partição.

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

Para visualizar os mapas com as distribuições das abundâncias das espécies (sp1, sp2, sp3...) na grade de 10 x 10 parcelas, clique [aqui](#) e para a distribuição dos valores das variáveis ambientais (var1, var2,...) clique [aqui](#).

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.

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

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)

As distâncias, ou qualquer outra medida de diferença, podem ser organizadas em uma matriz parcelas x parcelas, em que cada célula é a distância entre um par. Esta seria uma **matriz de distâncias euclidianas**

3)

Quando tiver um tempinho (por. ex. em casa quando estiver revisando o exercício) veja: [TedTalk Rogerio Martins](#). Agora sim, tá explicada a diferença entre bege e marfim. Entenderam?

4)

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 1, de Dinâmica de Populações. Apesar da matemática por trás do cálculo dos autovalores ser a mesma, as interpretações são diferentes!

5)

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

6)

Na verdade, é necessário fazer uma correção do  $R^2$  para interpretá-lo corretamente. Essa correção foi proposta por Peres-Neto et al (2006) e tem uma função (RsquareAdj) que já calcula o  $R^2$  ajustado. Geralmente o valor de  $R^2$  ajustado é menor

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 foi o que abordamos no módulo sobre Estrutura de Populações - ok? Na agregação de espécies são as populações de várias espécies que estão agregadas em locais próximos

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:2016:roteiros:ec>



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