

# Introdução ao R com aplicações em biodiversidade e conservação

2020-11-03



# Contents

<b>1</b>	<b>Pré-requisitos</b>	<b>5</b>
<b>2</b>	<b>Introdução</b>	<b>7</b>
2.1	Histórico deste livro . . . . .	7
2.2	Objetivo deste livro . . . . .	8
2.3	O que você não encontrará neste livro . . . . .	8
2.4	Por que usar o R? . . . . .	9
2.5	Como usar este livro . . . . .	9
<b>3</b>	<b>Estatística básica</b>	<b>11</b>
3.1	Teste T (de Student) para duas amostras independentes . . . . .	11
3.2	Teste T para amostras pareadas . . . . .	19
3.3	Correlação de Pearson . . . . .	22
3.4	Regressão Linear Simples . . . . .	25
3.5	Regressão Linear Múltipla . . . . .	29
3.6	Análises de Variância (ANOVA) . . . . .	34
3.7	ANOVA de um fator . . . . .	35
3.8	ANOVA de dois fatores ou Anova fatorial . . . . .	39
3.9	ANOVA em blocos aleatorizados . . . . .	48
3.10	Análise de covariância (ANCOVA) . . . . .	53
<b>4</b>	<b>Introdução à Análises Multidimensionais</b>	<b>59</b>
4.1	Backgorund da análise . . . . .	60
4.2	Exemplo 1: . . . . .	60
4.3	Exemplo 2: . . . . .	64
<b>5</b>	<b>K-means e agrupamentos não-hierarquicos</b>	<b>67</b>
5.1	Backgorund da análise . . . . .	67
5.2	Exemplo 1: . . . . .	67
5.3	Backgorund da análise . . . . .	70
5.4	Exemplo 1: . . . . .	71
<b>6</b>	<b>Rarefação</b>	<b>75</b>
6.1	Background da análise . . . . .	75

6.2	Exemplo prático 1 - Morcegos . . . . .	76
6.3	Exemplo prático 2 - Rarefação . . . . .	78
6.4	Para se aprofundar . . . . .	80
<b>7</b>	<b>Estimadores de Riqueza</b>	<b>81</b>
7.1	Backgorund da análise . . . . .	81
7.2	Estimadores baseados na abundância das espécies . . . . .	82
7.3	Estimadores baseados na incidência das espécies . . . . .	88
<b>8</b>	<b>Diversidade Funcional</b>	<b>103</b>
8.1	2.1 Aspectos teóricos . . . . .	103
8.2	2.2. Definindo a dis(similaridade) entre espécies . . . . .	104
8.3	2.3 Métricas de diversidade funcional (alpha) . . . . .	110
8.4	2.3 Métricas de diversidade funcional (beta) . . . . .	126

## Chapter 1

# Pré-requisitos



## Chapter 2

# Introdução

### 2.1 Histórico deste livro

Este livro foi estruturado a partir da apostila elaborada pelos pesquisadores Diogo B. Provete, Fernando R. da Silva e Thiago Gonçalves-Souza para ministrar o curso *Estatística aplicada à ecologia usando o R* no PPG em Biologia Animal da UNESP de S.J. Rio Preto em abril de 2011. Os três pesquisadores eram alunos do PPG em Biologia Animal quando elaboraram o material disponibilizado na apostila. A proposta de transformar a apostila em livro era um tópico recorrente desde 2011, e concretizada 10 anos depois. Neste período, Diogo, Fernando e Thiago foram contratados pela Universidade Federal de Mato Grosso do Sul, Universidade Federal de São Carlos, campus Sorocaba e Universidade Federal Rural de Pernambuco, respectivamente. Nestes anos eles ofertaram diferentes versões do curso *Estatística aplicada à ecologia usando o R* para alunos de graduação e pós graduação em diferentes instituições do Brasil. A possibilidade da oferta destes novos cursos fortaleceu a ideia de transformar a apostila em um livro com base nas experiências em sala de aula. Considerando que novas abordagens ecológicas vem sendo descritas a uma taxa elevada nos últimos anos, era de se esperar que as informações disponíveis na apostila estivessem defasadas após 10 anos. Por este motivo, Diogo, Fernando e Thiago resolveram convidar outros dois pesquisadores, Gustavo B. Paterno da Universität München e Maurício H. Vancine do PPG em Zoologia da UNESP, campus Rio Claro, que são referências no uso de estatística em ecologia usando o R. Com o time completo, passamos mais de um ano realizando reuniões, compartilhando scripts e pagando cerveja para os co-autores por capítulo atrasado até chegarmos neste primeira versão do livro.

## 2.2 Objetivo deste livro

Nossa proposta com este livro é de traçar o melhor caminho (pelo menos em nosso ponto de vista) entre questões ecológicas e os métodos estatísticos mais robustos para testá-las. Guiar seus passos nesse caminho (nem sempre linear) necessita que você utilize um requisito básico: o de utilizar seu esforço para caminhar. O nosso esforço, em contrapartida, será o de indicar as melhores direções para que adquira certa independência em análises ecológicas. Um dos nossos objetivos é mostrar que o conhecimento de teorias ecológicas e a utilização de questões apropriadas são o primeiro passo na caminhada rumo à compreensão da lógica estatística. Não deixe que a estatística se torne a “pedra no seu caminho”. Em nossa opinião, programas com ambiente de programação favorecem o entendimento da lógica estatística, uma vez que cada passo (lembre-se de que você está caminhando em uma estrada desconhecida) precisa ser coordenado, ou seja, as linhas de comando (detalhes abaixo) precisam ser compreendidas para que você teste suas hipóteses.

A primeira parte deste livro pretende utilizar uma estratégia que facilita a escolha do teste estatístico apropriado, por meio da seleção de questões/hipóteses claras e da ligação dessas hipóteses com a teoria e o método (veja esquema no Capítulo 2). Enfatizamos que é sempre bom ter em mente que é muito importante saber aonde se quer chegar para poder escolher o que fazer. Posteriormente à escolha de suas questões é necessário transferir o contexto ecológico para um contexto meramente estatístico (hipótese nula/alternativa). A partir da definição de sua hipótese nula partiremos para a aplicação de cada teste estatístico (de modelos lineares generalizados à análises multivariadas) utilizando como plataforma o programa R. Antes de detalhar cada análise estatística, apresentaremos os comandos básicos para a utilização do R e os tipos de distribuição estatística que são essenciais para a compreensão dos testes estatísticos. Para isso, organizamos um esquema que chamamos de “estrutura lógica” que facilita a compreensão dos passos necessários para testar suas hipóteses (veja esquema no Capítulo 2).

## 2.3 O que você não encontrará neste livro

Aprofundamento teórico, detalhes matemáticos, e explicação dos algoritmos são informações que infelizmente não serão abordadas neste livro. O foco do livro é a explicação de como cada teste funciona (teoria e procedimentos matemáticos básicos) e sua aplicação em testes ecológicos usando o programa R. Para tanto, o livro dos irmãos Pierre e Louis Legendre (Legendre & Legendre 2012) é uma leitura que permite o aprofundamento de cada uma das análises propostas aqui. Além disso, são de fundamental importância para o amadurecimento em análises ecológicas as seguintes leituras: Manly (1991), Pinheiro & Bates (2000), Scheiner & Gurevitch (2001), Quinn & Keough (2002), Venables & Ripley (2002), Magurran (2011), Gotelli & Ellison (2010) e Zar (2009).



## 2.4 Por que usar o R?

Os criadores do R o chamam de uma linguagem e ambiente de programação estatística e gráfica. A linguagem R também é chamada de programação “orientada ao objeto” (*object oriented programming*), o que significa que utilizar o R envolve basicamente a criação e manipulação de objetos em uma tela branca em que o usuário tem de dizer exatamente o que deseja que o programa execute ao invés de simplesmente pressionar um botão. E vem daí uma das grandes vantagens em se usar o R: o usuário tem total controle sobre o que está acontecendo e também tem de compreender o que deseja antes de executar uma análise. Além disso, o R permite integração com outros programas escritos em C++, Python e Java permitindo que os usuários possam aplicar novas metodologias sem ter que aprender novas linguagens.

Na página pessoal do Prof. Nicolas J. Gotelli (<http://www.uvm.edu/~ngotelli/homepage.html>) existem vários conselhos para um estudante iniciante de ecologia. Dentre esses conselhos, o Prof. Gotelli menciona que o domínio de uma linguagem de programação é uma das habilidades mais importantes, porque dá liberdade ao ecólogo para executar tarefas que vão além daquelas disponíveis em pacotes comerciais. Além disso, a maioria das novas análises propostas nos mais reconhecidos periódicos em ecologia normalmente são implementadas em linguagem R, e os autores incluem normalmente o código fonte no material suplementar dos artigos, tornando a análise acessível. A partir do momento que essas análises ficam disponíveis (seja por código fornecido pelo autor ou por implementação em pacotes pré-existentes), é mais simples entendermos a lógica de análises complexas, especialmente as multivariadas, com nossos próprios dados realizando-as passo a passo. Sem a utilização do R, normalmente temos que contatar os autores que nem sempre são acessíveis.

Uma última vantagem é que por ser um software livre, a citação do R em artigos é permitida e até aconselhável. Para saber como citar o R, digite `citation()` na linha de comando. Para citar um pacote específico, digite `citation()` com o nome do pacote entre aspas dentro dos parênteses. Neste ponto, esperamos ter convencido você leitor de que aprender a utilizar o R tem inúmeras vantagens, vai ser difícil no começo mas continue e perceberá que o investimento vai valer à pena no futuro.

## 2.5 Como usar este livro

Os conteúdos apresentados em cada capítulo são independentes entre si. Portanto, você pode utilizar este livro de duas formas. A primeira é seguir uma ordem sequencial (capítulos 1, 2, 3, ...) que recomendamos, principalmente, para as pessoas que não possuem familiaridade com a programação em R. A segunda é selecionar o capítulo que contém a análise de seu interesse e mudar de um capítulo para outro sem seguir a sequência apresentada no livro.

Com exceção dos capítulos 2 e 3, os outros capítulos foram elaborados seguindo

a mesma estrutura contendo uma descrição da análise estatística (*background*) e exemplos relacionados com questões ecológicas que podem ser respondidas por esta análise. Todos os exemplos são compostos por: i) uma descrição dos dados utilizados, ii) pergunta e predição do trabalho, iii) descrição das variáveis resposta(s) e preditora(s), e iv) descrição e explicação das linhas de comando do R necessárias para realização da análise. Os exemplos utilizados são baseados em dados reais que já foram publicados em artigos científicos ou são dados coletados por um dos autores deste livro. Nós recomendamos que primeiro você utilize estes exemplos para se familiarizar com análises e a formatação das linhas e colunas das planilhas. Em seguida, utilize os seus dados para realizar as análises. Esta é a melhor maneira para se familiarizar com as linhas de comando do R. Realçamos que não estamos abordando todas as possibilidades disponíveis e existem outros pacotes e funções no R que realizam as mesmas análises. Contudo, esperamos que o conteúdo apresentado permita que os leitores adquiram independência e possam caminhar sozinhos na exploração de novos pacotes e funções para responderem suas perguntas biológicas.

## Chapter 3

# Estatística básica

### 3.1 Teste T (de Student) para duas amostras independentes

#### 3.1.1 Background da análise

Uma das perguntas mais comum em estatística é saber se há diferença entre as médias de dois grupos ou tratamentos. Para responder esta pergunta, William Sealy Gosset, químico da cervejaria Guinness, em 1908 desenvolveu o Teste T que é uma estatística que segue uma distribuição t de Student para rejeitar ou não uma hipótese nula de médias iguais entre os grupos.

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{2S_p^2}{n}}}$$

Onde:

- $\bar{X}_1 - \bar{X}_2$  = diferença entre as médias das duas amostras,
- $S_p^2$  = desvio padrão das amostras,
- $n$  = tamanho das amostras.

##### 3.1.1.1 Premissas do Teste t :

- As amostras devem ser independentes;
- As unidades amostrais são selecionadas aleatoriamente;
- Distribuição normal (gaussiana) dos resíduos. **Observação:** Zar (2010, p. 136) indica que o Test T é robusto mesmo com moderada violação da normalidade, principalmente se o tamanho amostral for alto.

- Homogeneidade da variância. **Observação.** Caso as variâncias não sejam homogêneas, isso deve ser informado na linha de comando, pois o denominador da fórmula acima será corrigido.

### 3.1.1.2 Exemplo prático 1 - Teste T para duas amostras com variâncias iguais

**3.1.1.2.1 Explicação dos dados** Neste exemplo avaliaremos o comprimento rostro-cloacal (CRC em milímetros) de machos de *Physalaemus nattereri* (Anura:Leptodactylidae) amostrados em diferentes estações do ano com armadilhas de interceptação e queda na região noroeste do estado de São Paulo (da Silva & Rossa-Feres 2010).

#### Pergunta:

O CRC dos machos de *P. nattereri* é maior na estação chuvosa do que na estação seca?

#### Predições

O CRC dos machos será maior na estação chuvosa porque há uma vantagem seletiva para os indivíduos maiores durante a atividade reprodutiva.

#### Variáveis

- Variáveis preditoras
  - Dataframe com os indivíduos (unidade amostral) nas linhas e CRC (mm - variável resposta contínua) e estação (variável preditora categórica) como colunas.

#### Checklist

- Verificar se o seu dataframe está com as unidades amostrais nas linhas e variáveis preditores nas colunas

## 3.1.2 Análise

Cálculo do Teste T para duas amostras independentes com variâncias iguais

```
## IMPORTANDO OS DADOS
#####
CRC_PN_macho <- ecodados::teste_t_var_igual
# verificar se o dataframe foi lido corretamente e se não há erros
# Esses comandos são úteis para planilhas grandes
head(CRC_PN_macho) # mostra as seis primeiras linhas da planilha
```

```
##      CRC Estacao
## 1 3.82 Chuvosa
```

### 3.1. TESTE T (DE STUDENT) PARA DUAS AMOSTRAS INDEPENDENTES13

```
## 2 3.57 Chuvosa
## 3 3.67 Chuvosa
## 4 3.72 Chuvosa
## 5 3.75 Chuvosa
## 6 3.83 Chuvosa
```

```
tail(CRC_PN_macho) # mostra as seis últimas linhas da planilha
```

```
##      CRC Estacao
## 46 3.16      Seca
## 47 3.48      Seca
## 48 3.48      Seca
## 49 3.49      Seca
## 50 3.51      Seca
## 51 3.30      Seca
```

```
# TESTE NORMALIDADE
```

```
*****
```

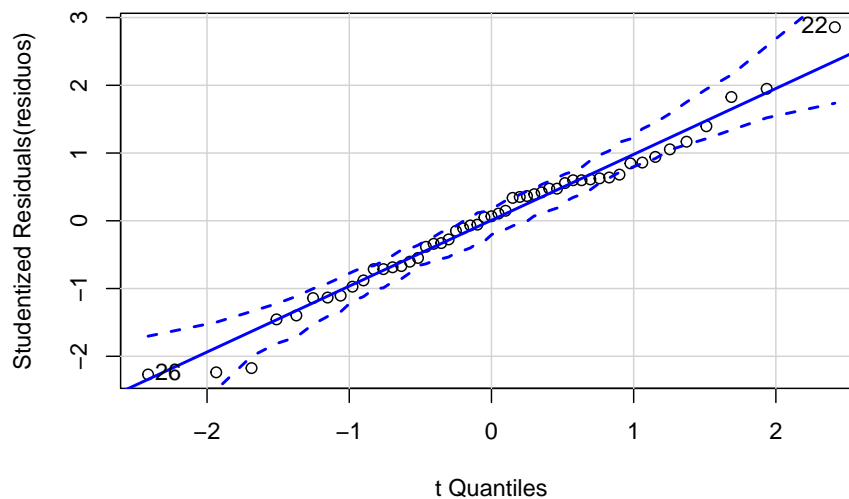
```
## Verificando normalidade usando QQ-plot
```

```
## Os pontos não podem fugir da reta criando formas como U
```

```
residuos <- lm(CRC ~ Estacao, data = CRC_PN_macho)
```

```
library("car")
```

```
qqPlot(residuos)
```



```
## [1] 22 26
```

```
## Outra possibilidade é usar o teste de Shapiro-Wilk para verificar normalidade
## Hipótese nula que a distribuição é normal
## valor de  $p < 0.05$  significa que os dados não apresentam distribuição normal
## valor de  $p > 0.05$  significa que os dados apresentam distribuição normal
shapiro.test(CRC_PN_macho$CRC)
```

```
##
## Shapiro-Wilk normality test
##
## data: CRC_PN_macho$CRC
## W = 0.95559, p-value = 0.05417
```

```
# TESTE DE HOMOGENEIDADE DA VARIÂNCIA
#####
## Hipótese nula que a variância é homogênea
## valor de  $p < 0.05$  significa que os dados não apresentam homogeneidade
## valor de  $p > 0.05$  significa que os dados apresentam homogeneidade
library(car)
leveneTest(CRC ~ Estacao, data = CRC_PN_macho)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 1  1.1677 0.2852
##      49
```

```
# TESTE T AMOSTRAS INDEPENDENTES E VARIÂNCIAS IGUAIS
#####888
t.test(CRC ~ Estacao, data = CRC_PN_macho, var.equal = TRUE)
```

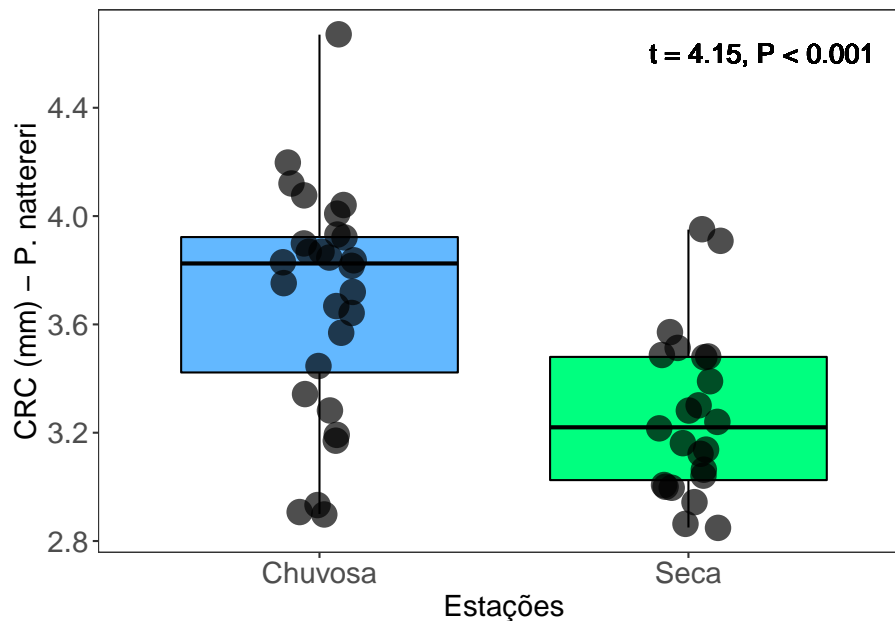
```
##
## Two Sample t-test
##
## data: CRC by Estacao
## t = 4.1524, df = 49, p-value = 0.000131
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.2242132 0.6447619
## sample estimates:
## mean in group Chuvosa      mean in group Seca
##           3.695357           3.260870
```

Visualizar os resultados em gráfico

```
library(ggplot2)
ggplot(data = CRC_PN_macho, aes(x= Estacao, y= CRC, color = Estacao)) +
  labs(x = "Estações", y = "CRC (mm) - P. nattereri", size = 15) +
  geom_boxplot(fill=c("steelblue1", "springgreen1"), color="black", outlier.shape = NA) +
  geom_jitter(shape = 16, position=position_jitter(0.1), cex = 6, alpha = 0.7) +
```

### 3.1. TESTE T (DE STUDENT) PARA DUAS AMOSTRAS INDEPENDENTES<sup>15</sup>

```
scale_color_manual(values = c("black", "black")) +  
geom_text(x = 2.2, y = 4.6, label = "t = 4.15, P < 0.001", color = "black", size = 5) +  
theme_bw() +  
theme(axis.text.y = element_text(size = 15), axis.text.x = element_text(size = 15)) +  
theme(axis.title.y = element_text(size = 15), axis.title.x = element_text(size = 15)) +  
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank()) +  
theme(legend.position = "none")
```



#### Interpretação dos resultados

Neste exemplo, rejeitamos a hipótese nula que as médias do CRC dos machos entre as estações seca e chuvosa são iguais ( $t = 4,15$ ,  $P < 0,001$ ). Os resultados mostram que os machos de *P. nattereri* coletados na estação chuvosa foram em média 0,43mm maiores do que os coletados na estação seca.

#### 3.1.2.1 Exemplo prático 2 - Teste T para duas amostras independentes com variâncias diferentes

**3.1.2.1.1 Explicação dos dados** Neste exemplo, avaliaremos o comprimento rostro-cloacal (CRC - milímetros) de fêmeas de *Leptodactylus podicipinus* amostradas em diferentes estações do ano com armadilhas de interceptação e queda na região noroeste do estado de São Paulo (da Silva & Rossa-Feres 2010). **Observação:** Os dados foram alterados em relação a publicação original para se enquadrarem no exemplo de amostras com variâncias diferentes.

**Pergunta:**

O CRC das fêmeas de *L. podicipinus* é maior na estação chuvosa do que na estação seca?

**Predições**

O CRC das fêmeas será maior na estação chuvosa porque há uma vantagem seletiva para os indivíduos maiores durante a atividade reprodutiva.

**Variáveis**

- Variáveis preditoras
  - Dataframe com os indivíduos (unidade amostral) nas linhas e CRC (mm - variável resposta contínua) e estação (variável preditora categórica) como colunas.

**Checklist**

- Verificar se o seu dataframe está com as unidades amostrais nas linhas e variáveis preditores nas colunas

**3.1.3 Análise**

Calculo do Teste T para duas amostras com variâncias diferentes

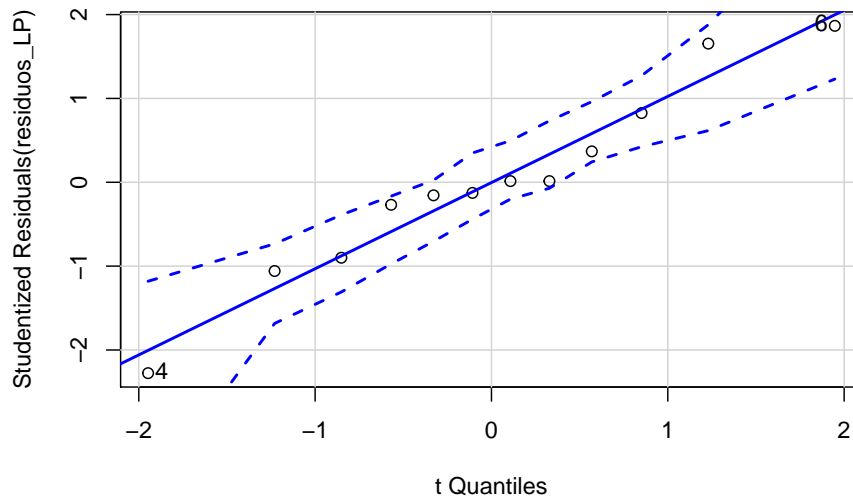
```
## IMPORTANDO OS DADOS
#####
CRC_LP_femea <- ecodados::teste_t_var_diferente
head(CRC_LP_femea) # verificar se o dataframe foi lido corretamente

##      CRC Estacao
## 1 2.72 Chuvosa
## 2 2.10 Chuvosa
## 3 3.42 Chuvosa
## 4 1.50 Chuvosa
## 5 3.90 Chuvosa
## 6 4.00 Chuvosa

# TESTE NORMALIDADE
#####
## Verificando normalidade usando QQ-plot
## Os pontos não podem fugir da reta criando formas como U
residuos_LP <- lm(CRC ~ Estacao, data = CRC_LP_femea)
library("car")
qqPlot(residuos_LP)
```



### 3.1. TESTE T (DE STUDENT) PARA DUAS AMOSTRAS INDEPENDENTES<sup>17</sup>



```
## [1] 4 6
```

```
## Outra possibilidade é usar o teste de Shapiro-Wilk para verificar normalidade
## Hipótese nula que a distribuição é normal
## valor de  $p < 0.05$  significa que os dados não apresentam distribuição normal
## valor de  $p > 0.05$  significa que os dados apresentam distribuição normal
shapiro.test(CRC_LP_femea$CRC)
```

```
##
## Shapiro-Wilk normality test
##
## data: CRC_LP_femea$CRC
## W = 0.88195, p-value = 0.09284
```

```
# TESTE DE HOMOGENEIDADE DA VARIÂNCIA
#*****
## Hipótese nula que a variância é homogênea
## valor de  $p < 0.05$  significa que os dados não apresentam homogeneidade
## valor de  $p > 0.05$  significa que os dados apresentam homogeneidade
library(car)
leveneTest(CRC ~ Estacao, data = CRC_LP_femea)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 1  9.8527 0.01053 *
##      10
```

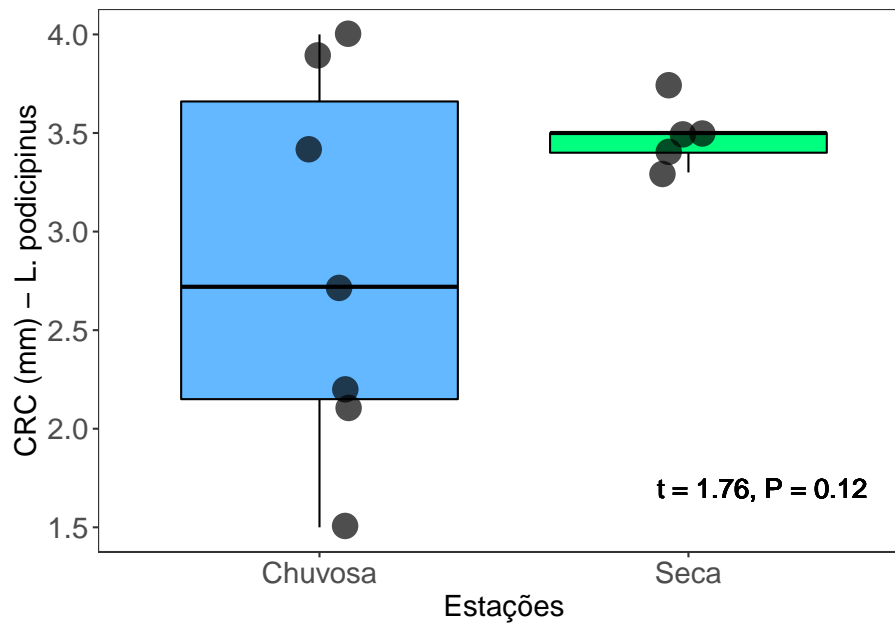
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# TESTE T COM AMOSTRAS INDEPENDENTES E VARIÂNCIS DIFERENTES
#####
## Com base no teste de Levene, avise na linha de comando que as variâncias
## não são iguais (var.equal = FALSE).
t.test(CRC ~ Estacao, data = CRC_LP_femea, var.equal = FALSE)

##
## Welch Two Sample t-test
##
## data: CRC by Estacao
## t = -1.7633, df = 6.4998, p-value = 0.1245
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.5489301  0.2375016
## sample estimates:
## mean in group Chuvosa      mean in group Seca
##           2.834286           3.490000
```

Visualizar os resultados em gráfico

```
library(ggplot2)
ggplot(data = CRC_LP_femea, aes(x= Estacao, y= CRC, color = Estacao)) +
  labs(x = "Estações", y = "CRC (mm) - L. podicipinus", size = 15) +
  geom_boxplot(fill=c("steelblue1", "springgreen1"), color="black", outlier.shape = NA) +
  geom_jitter(shape = 16, position=position_jitter(0.1), cex = 6, alpha = 0.7) +
  scale_color_manual(values = c("black", "black")) +
  geom_text(x = 2.2, y = 1.7, label = "t = 1.76, P = 0.12", color = "black", size = 5) +
  theme_bw() +
  theme(axis.text.y = element_text(size = 15), axis.text.x = element_text(size = 15)) +
  theme(axis.title.y = element_text(size = 15), axis.title.x = element_text(size = 15)) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank()) +
  theme(legend.position = "none")
```



### Interpretação dos resultados

Neste exemplo, não rejeitamos a hipótese nula e consideramos que as médias do CRC das fêmeas entre as estações seca e chuvosa são iguais ( $t = 1,76$ ,  $P = 0,12$ ). Os resultados mostram que as fêmeas de *L. podicipinus* coletadas na estação chuvosa não são maiores do que as fêmeas coletadas na estação seca.

## 3.2 Teste T para amostras pareadas

### 3.2.1 Background da análise

O Teste T Pareado é uma estatística que usa dados medidos duas vezes na mesma unidade amostral, resultando em pares de observações para cada amostra (amostras pareadas). Ele determina se a diferença da média entre duas observações é zero.

$$t = \frac{\bar{d}}{S_{\bar{d}}}$$

Onde:

- $\bar{d}$  = média da diferença das medidas pareadas. Observe que o teste não usa as medidas originais, e sim, a diferença para cada par,

- $S\bar{d}$  = erro padrão da diferença das medidas pareadas.

### 3.2.1.1 Premissas do Teste t para amostras pareadas:

- As unidades amostrais são selecionadas aleatoriamente;
- Distribuição normal (gaussiana) dos valores da diferença para cada par;

### 3.2.1.2 Exemplo prático 1 - Teste T para amostras pareadas

**3.2.1.2.1 Explicação dos dados** Neste exemplo avaliaremos a diferença na riqueza de espécies de artrópodes registradas em 27 localidades. Todas as localidades foram amostradas duas vezes. A primeira amostragem foi realizada com na localidade antes da perturbação e a segunda amostragem foi realizada após a localidade ter sofrido uma queimada.

#### Pergunta:

A riqueza de espécies de artrópodes é prejudicada pelas queimadas?

#### Predições

A riqueza de espécies de artrópodes será maior antes da queimada devido a extinção local das espécies.

#### Variáveis

- Variáveis preditoras
  - Dataframe com as localidades nas linhas e riqueza de espécies (variável preditora contínua) e estado (Pre-queimada ou Pós-queimada - variável categórica) da localidade nas colunas.

#### Checklist

- Verificar se o seu dataframe está com as unidades amostrais nas linhas e variáveis preditores nas colunas

## 3.2.2 Análise

Calculo do Teste T com amostras pareadas

```
## IMPORTANDO OS DADOS
#####
Pareado <- ecodados::teste_t_pareado
head(Pareado) # verificar se o dataframe foi lido corretamente

##   Areas Riqueza      Estado
## 1     1      92 Pre-Queimada
## 2     2      74 Pre-Queimada
## 3     3      96 Pre-Queimada
```

```
## 4      4      89 Pre-Queimada
## 5      5      76 Pre-Queimada
## 6      6      80 Pre-Queimada
```

```
tail(Pareado)
```

```
##      Areas Riqueza      Estado
## 49      22      37 Pos-Queimada
## 50      23      20 Pos-Queimada
## 51      24      12 Pos-Queimada
## 52      25      22 Pos-Queimada
## 53      26      27 Pos-Queimada
## 54      27      28 Pos-Queimada
```

```
# TESTE T PAREADO
```

```
*****
```

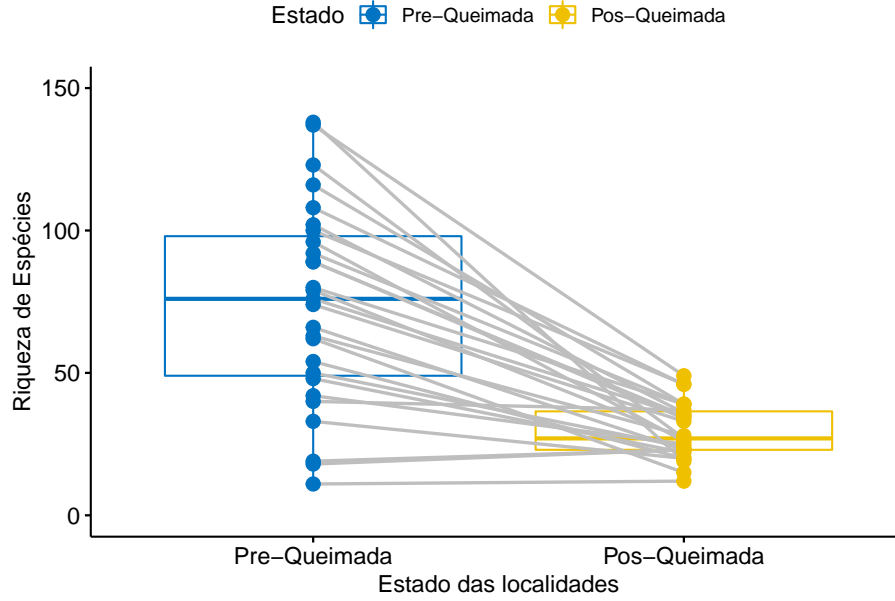
```
# O uso do [] é para selecionar dentro do vetor/coluna *Riqueza* os 27 primeiros números [1:27] q
```

```
t.test(Pareado$Riqueza[1:27], Pareado$Riqueza[28:54], paired = TRUE)
```

```
##
## Paired t-test
##
## data: Pareado$Riqueza[1:27] and Pareado$Riqueza[28:54]
## t = 7.5788, df = 26, p-value = 4.803e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 32.47117 56.63994
## sample estimates:
## mean of the differences
## 44.55556
```

Visualizar os resultados em gráfico

```
library("ggpubr")
ggpaired(Pareado, x = "Estado", y = "Riqueza",
          color = "Estado", line.color = "gray", line.size = 0.8, palette = "jco", width = 0.8,
          point.size = 3, xlab = "Estado das localidades", ylab = "Riqueza de Espécies") +
  expand_limits(y=c(0,150))
```



### Interpretação dos resultados

Neste exemplo, rejeitamos a hipótese nula que a riqueza de espécies de artrópodes é igual antes e depois da queimada ( $t = 7,57$ ,  $P < 0,001$ ). Os resultados mostram que as localidades após as queimadas apresentam em média 44,5 espécies de artrópodes a menos do que antes das queimadas.

## 3.3 Correlação de Pearson

### 3.3.1 Background da análise

É um teste que mede a o grau de associação entre duas variáveis contínuas (X e Y). Importante ressaltar que a análise de correlação não assume que a variável X influencie a variável Y ou que exista uma relação de causa e efeito entre elas (Zar 2016). A análise é definida em termos da variância de X, a variância de Y, e a covariância de X e Y (i.e. como elas variam juntas).

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left(\sum X^2 - \frac{\sum X^2}{n}\right) \left(\sum Y^2 - \frac{\sum Y^2}{n}\right)}}$$

Onde:

- $r$  = coeficiente de correlação que indica a força da relação entre as duas variáveis. Seu range de valores está entre  $-1 \leq r \leq 1$ . A correlação positiva indica que o aumento no valor de uma das variáveis é acompanhado pelo aumento no valor da outra variável. A correlação negativa indica que um aumento no valor de uma das variáveis é acompanhado pela diminuição no valor da outra variável. Se  $r$  é igual a zero, não existe correlação entre as variáveis.

#### 3.3.1.1 Premissas da Correlação de Person:

- As amostras devem ser independentes e pareadas (i.e. as duas variáveis devem ser medidas na mesma unidade amostral);
- As unidades amostrais são selecionadas aleatoriamente;
- A relação entre as variáveis tem que ser linear.

#### 3.3.1.2 Exemplo prático 1 - Correlação de Pearson

**3.3.1.2.1 Explicação dos dados** Neste exemplo avaliaremos a correlação entre a altura do tronco e o tamanho da raiz medidos em 35 indivíduos de uma espécie vegetal arbustiva.

##### Pergunta:

Existe correlação entre a altura do tronco e o tamanho da raiz dos arbustos?

##### Predições

A altura do tronco é positivamente correlacionado com o tamanho da raiz.

##### Variáveis

- Variáveis preditoras
  - Dataframe com os indivíduos (unidade amostral) nas linhas e altura do tronco e tamanho da raiz (duas variáveis tem que ser contínuas) como colunas.

##### Checklist

- Verificar se o seu dataframe está com as unidades amostrais nas linhas e variáveis preditores nas colunas

#### 3.3.2 Análise

Calculo do Teste de Correlação de Pearson

```
## IMPORTANDO OS DADOS
#*****
```

```

correlacao_arbustos <- ecodados::correlacao
head(correlacao_arbustos) # verificar se o dataframe foi lido corretamente

##      Tamanho_raiz Tamanho_tronco
## 1      10.177049      19.54383
## 2       6.622634      17.13558
## 3       7.773629      19.50681
## 4      11.055257      21.57085
## 5       4.487274      13.22763
## 6      11.190216      21.62902

# Teste de Correlação de Pearson
#####
# Para outros testes de correlação como Kendall ou Spearman é só alterar na linha de c
cor.test(correlacao_arbustos$Tamanho_raiz, correlacao_arbustos$Tamanho_tronco, method =

##
## Pearson's product-moment correlation
##
## data: correlacao_arbustos$Tamanho_raiz and correlacao_arbustos$Tamanho_tronco
## t = 11.49, df = 33, p-value = 4.474e-13
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7995083 0.9457816
## sample estimates:
##      cor
## 0.8944449

```

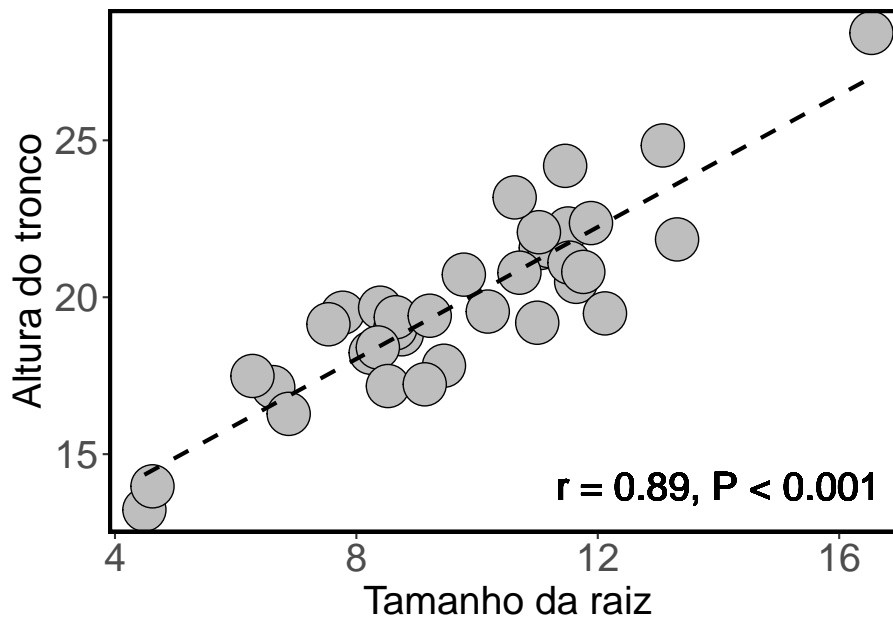
Visualizar os resultados em gráfico

```

library(ggplot2)
ggplot(data = correlacao_arbustos, aes(x= Tamanho_raiz, y= Tamanho_tronco)) +
  labs(x = "Tamanho da raiz", y = "Altura do tronco", size = 20) +
  geom_point(size = 10, shape = 21, fill = "gray") +
  geom_text(x = 14, y = 14, label = "r = 0.89, P < 0.001", color = "black", size = 7) +
  theme_bw() +
  theme(axis.title.y = element_text(size = 20), axis.title.x = element_text(size = 20),
        axis.text.y = element_text(size = 20), axis.text.x = element_text(size = 20)) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size = 2)) +
  geom_smooth(method = lm, se = FALSE, color = "black", linetype="dashed")

```





#### Interpretação dos resultados

Neste exemplo, rejeitamos a hipótese nula que as variáveis não são correlacionadas ( $r = 0.89$ ,  $P < 0.001$ ). Os resultados mostram que o aumento na altura dos arbustos é acompanhado pelo aumento no tamanho da raiz.

### 3.4 Regressão Linear Simples

#### 3.4.1 Background da análise

A regressão simples é usada para analisar a relação entre uma variável preditora (plotada no eixo-X) e uma variável resposta (plotada no eixo-Y). As duas variáveis devem ser contínuas. Diferente das correlações, a regressão assume uma relação de causa e efeito entre as variáveis. O valor da variável preditora (X) causa, direta ou indiretamente, o valor da variável resposta (Y). Assim, Y é uma função linear de X:

$$Y = \beta_0 + \beta_1 X_i + \epsilon_i$$

Onde:

- $\beta_0$  = intercepto que representa o valor da função quando  $X = 0$ ,

- $\beta_1$  = inclinação (*slope*) que mede a mudança na variável Y para cada mudança de unidade da variável X. Este parâmetro é usado para testar a hipótese nula da regressão que assume que  $\beta_1 = 0$ .
- $\epsilon_1$  = erro aleatório referente a variável Y que não pode ser explicado pela variável X.

#### 3.4.1.1 Premissas da Regressão Linear Simples:

- As amostras devem ser independentes;
- As unidades amostrais são selecionadas aleatoriamente;
- Distribuição normal (gaussiana) dos resíduos;
- Homogeneidade da variância.

#### 3.4.1.2 Exemplo prático 1 - Regressão linear simples

**3.4.1.2.1 Explicação dos dados** Neste exemplo, avaliaremos a relação entre o gradiente de temperatura média anual (°C) e o tamanho médio do comprimento rostro-cloacal (CRC em mm) de populações de *Dendropsophus minutus* (Anura:Hylidae) amostradas em 109 localidades no Brasil (Boaratti & da Silva 2015).

##### Pergunta:

Há relação entre o tamanho do CRC das populações e a temperatura das localidades onde os indivíduos ocorrem?

##### Predições

O CRC das populações serão menores em localidades mais quentes do que em localidades mais frias de acordo com a Hipótese do balanço de calor.

##### Variáveis

- Variáveis preditoras
  - Dataframe com as populações (unidade amostral) nas linhas e CRC médio (mm) e temperatura média anual como colunas.

##### Checklist

- Verificar se o seu dataframe está com as unidades amostrais nas linhas e variáveis preditores nas colunas

#### 3.4.2 Análise

Calculo da regressão linear simples

```
## IMPORTANDO DADOS
#####
dados_regressao <- ecodados::regressoes
head(dados_regressao) # verificar se o dataframe foi lido corretamente
```

```
##      Municipio      CRC Temperatura Precipitacao
## 1   Acorizal 22.98816    24.13000    1228.2
## 2  Alpinopolis 22.91788    20.09417    1487.6
## 3 Alto_Paraiso 21.97629    21.86167    1812.4
## 4   Americana 23.32453    20.28333    1266.2
## 5     Apiacas 22.83651    25.47333    2154.0
## 6  Arianopolis 20.86989    20.12167    1269.2
```

```
# ANALISE DA REGRESSÃO
```

```
#####
```

```
modelo_regressao <- lm(CRC ~ Temperatura, data = dados_regressao)
```

```
# PRIMEIRO VAMOS VERIFICAR A NORMALIDADE E HOMOGENEIDADE DAS VARIÂNCIAS
```

```
#####
```

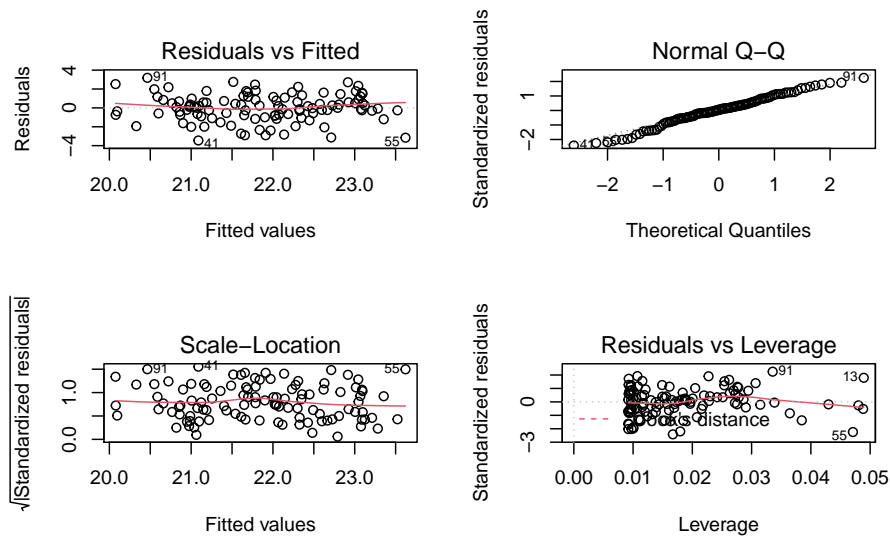
```
# Os gráficos *Residuals vs Fitted*, *Scale-Location*, e *Residual vs Leverage* estão relacionados
```

```
# O gráfico *Normal Q-Q* está relacionado com a distribuição normal dos resíduos. Neste gráfico,
```

```
par(mfrow = c(2, 2), oma = c(0, 0, 2, 0))
```

```
plot(modelo_regressao)
```

lm(CRC ~ Temperatura)



```
dev.off()
```

```
## null device
##      1
```

```
# VERIFICANDO OS RESULTADOS DA REGRESSÃO
*****
anova(modelo_regressao)
```

```
## Analysis of Variance Table
##
## Response: CRC
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Temperatura  1  80.931   80.931   38.92 9.011e-09 ***
## Residuals 107 222.500    2.079
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

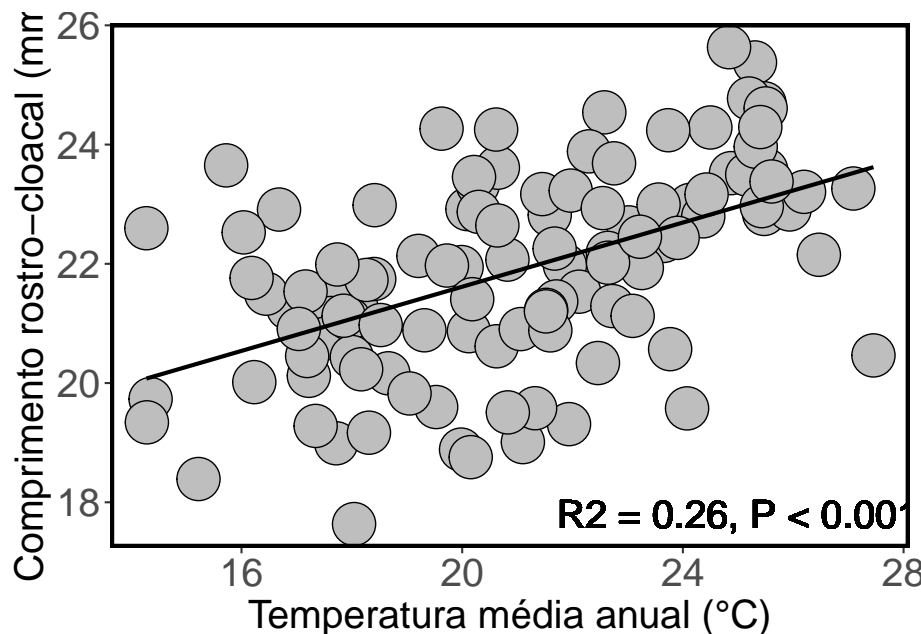
```
# ou
# esta função apresenta os resultados mais detalhados com a estimativa do intercepto,
summary(modelo_regressao)
```

```
##
## Call:
## lm(formula = CRC ~ Temperatura, data = dados_regressao)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4535 -0.7784  0.0888  0.9168  3.1868
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 16.23467    0.91368   17.768 < 2e-16 ***
## Temperatura  0.26905    0.04313    6.239 9.01e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.442 on 107 degrees of freedom
## Multiple R-squared:  0.2667, Adjusted R-squared:  0.2599
## F-statistic: 38.92 on 1 and 107 DF,  p-value: 9.011e-09
```

Visualizar os resultados em gráfico

```
library(ggplot2)
ggplot(data = dados_regressao, aes(x= Temperatura, y= CRC)) +
  labs(x = "Temperatura média anual (°C)", y = "Comprimento rostro-cloacal (mm)", size = 10) +
  geom_point(size = 10, shape = 21, fill = "gray") +
  geom_text(x = 25, y = 17.8, label = "R2 = 0.26, P < 0.001", color = "black", size = 10)
```

```
theme_bw() +
  theme(axis.title.y = element_text(size = 20), axis.title.x = element_text(size = 20)) +
  theme(axis.text.y = element_text(size = 20), axis.text.x = element_text(size = 20)) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size = 2)) +
  geom_smooth(method = lm, se = FALSE, color = "black")
```



#### Interpretação dos resultados

Neste exemplo, rejeitamos a hipótese nula que não existe relação entre o tamanho do CRC das populações de *D. minutus* e a temperatura da localidade onde elas ocorrem ( $F_{1,107} = 38,92$ ,  $P < 0,001$ ). Os resultados mostram que o tamanho do CRC das populações tem uma relação positiva com a temperatura das localidades. Assim, populações de *D. minutus* em localidades mais quentes apresentam maior CRC do que as populações em localidades mais frias.

## 3.5 Regressão Linear Múltipla

### 3.5.1 Background da análise

A regressão múltipla é uma extensão da regressão simples. Ela é usada quando queremos determinar o valor da variável resposta (Y) com base nos valores de duas ou mais variáveis preditoras ( $X_1$ ,  $X_2$ ,  $X_n$ ).

$$Y = \beta_0 + \beta_1 X_1 + \beta_n X_n + \epsilon_i$$

Onde:

- $\beta_0$  = intercepto que representa o valor da função quando  $X = 0$ ;
- $\beta_n$  = inclinação (*slope*) que mede a mudança na variável  $Y$  para cada mudança de unidade das variáveis  $X_n$ ;
- $\epsilon_1$  = erro aleatório referente a variável  $Y$  que não pode ser explicado pelas variáveis preditoras.

#### 3.5.1.1 Premissas da Regressão Linear Múltipla:

- As amostras devem ser independentes;
- As unidades amostrais são selecionadas aleatoriamente;
- Distribuição normal (gaussiana) dos resíduos;
- Homogeneidade da variância.

#### 3.5.1.2 Exemplo prático 1 - Regressão linear múltipla

**3.5.1.2.1 Explicação dos dados** Utilizaremos o mesmo exemplo da regressão simples. Contudo, além do gradiente de temperatura média anual (°C) incluiremos o gradiente de precipitação anual (mm) como outra variável preditora do tamanho médio do comprimento rostro-cloacal (CRC em mm) de populações de *Dendropsophus minutus* (Anura:Hylidae) amostradas em 109 localidades no Brasil (Boaratti & da Silva 2015).

##### Pergunta:

O tamanho do CRC das populações de *D. minutus* é influenciado pela temperatura e precipitação das localidades onde os indivíduos ocorrem?

##### Predições

O CRC das populações serão menores em localidades com clima quente e chuvoso do que em localidades com clima frio e seco.

##### Variáveis

- Variáveis preditoras
  - Dataframe com as populações (unidade amostral) nas linhas e CRC médio (mm) e temperatura e precipitação como colunas.

##### Checklist

- Verificar se o seu dataframe está com as unidades amostrais nas linhas e variáveis preditores nas colunas

### 3.5.2 Análise

Cálculo da regressão linear múltipla

```
## IMPORTANDO DADOS
#####
dados_regressao_mul <- ecodados::regressoes
head(dados_regressao_mul) # verificar se o dataframe foi lido corretamente
```

```
##      Municipio      CRC Temperatura Precipitacao
## 1    Acorizal 22.98816    24.13000    1228.2
## 2  Alpinopolis 22.91788    20.09417    1487.6
## 3 Alto_Paraiso 21.97629    21.86167    1812.4
## 4    Americana 23.32453    20.28333    1266.2
## 5      Apiacas 22.83651    25.47333    2154.0
## 6  Arianopolis 20.86989    20.12167    1269.2
```

```
# ANÁLISE DA REGRESSÃO
#####
modelo_regressao_mul <- lm(CRC ~ Temperatura + Precipitacao, data = dados_regressao_mul)
```

```
# MULTICOLINEARIDADE
#####
```

```
# Multicolinearidade ocorre quando as variáveis preditoras são correlacionadas. Essa correlação é
library(car)
vif(modelo_regressao_mul)
```

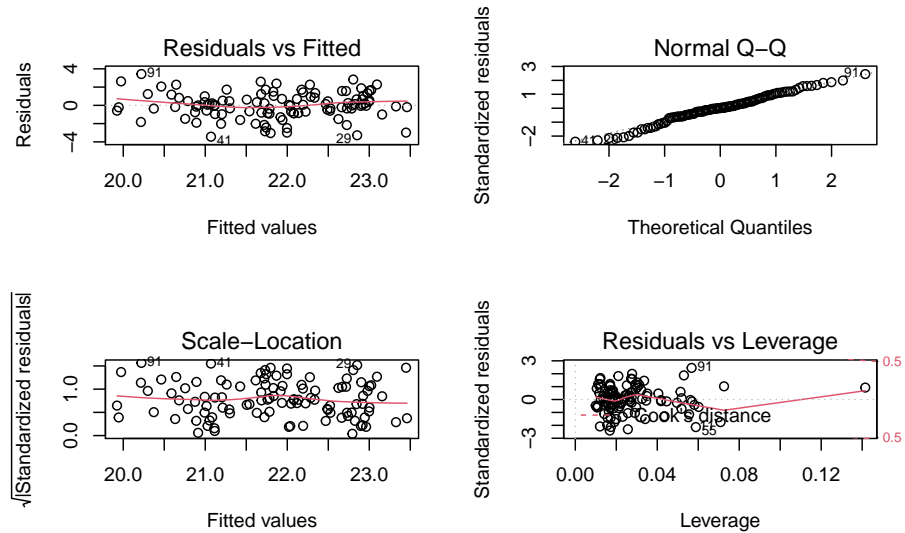
```
##      Temperatura Precipitacao
##      1.041265      1.041265
```

```
# VERIFICANDO A NORMALIDADE E HOMOGENEIDADE DAS VARIÂNCIAS
```

```
#####
```

```
# Os gráficos *Residuals vs Fitted*, *Scale-Location*, e *Residual vs Leverage* estão relacionados.
# O gráfico *Normal Q-Q* está relacionado com a distribuição normal dos resíduos. Neste gráfico,
par(mfrow = c(2, 2), oma = c(0, 0, 2, 0))
plot(modelo_regressao_mul)
```

lm(CRC ~ Temperatura + Precipitacao)



```
dev.off()
```

```
## null device
##      1
```

```
# VERIFICANDO OS RESULTADOS DA REGRESSÃO
#*****
anova(modelo_regressao_mul)
```

```
## Analysis of Variance Table
##
## Response: CRC
##      Df Sum Sq Mean Sq F value    Pr(>F)
## Temperatura  1  80.931   80.931  39.0028 8.94e-09 ***
## Precipitacao  1   2.549    2.549   1.2283  0.2702
## Residuals    106 219.951    2.075
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# ou
summary(modelo_regressao_mul)
```

```
##
## Call:
## lm(formula = CRC ~ Temperatura + Precipitacao, data = dados_regressao_mul)
##
## Residuals:
```



```
##      Min      1Q  Median      3Q      Max
## -3.4351 -0.8026  0.0140  0.9420  3.4300
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  16.7162571  1.0108674  16.537 < 2e-16 ***
## Temperatura  0.2787445  0.0439601   6.341 5.71e-09 ***
## Precipitacao -0.0004270  0.0003852  -1.108    0.27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.44 on 106 degrees of freedom
## Multiple R-squared:  0.2751, Adjusted R-squared:  0.2614
## F-statistic: 20.12 on 2 and 106 DF,  p-value: 3.927e-08
```

Percebam que a temperatura tem uma relação significativa com o tamanho do CRC das populações ( $P < 0.001$ ), enquanto que a precipitação não apresenta relação com o CRC ( $P = 0.27$ ). Neste caso, é interessante saber se um modelo mais simples (e.g. contendo apenas temperatura) explicaria a distribuição tão bem ou melhor do que este modelo mais complexo considerando dois parâmetros (temperatura e precipitação).

Para isso, podemos utilizar a *LIKELIHOOD RATIO TEST (LRT)* para comparar modelos. A LRT compara dois modelos aninhados, testando se os parâmetros do modelo mais complexo diferem significativamente do modelo mais simples. Em outras palavras, ele testa se há necessidade de se incluir um parâmetro extra no modelo para explicar os dados.

```
## CRIANDO OS MODELOS ANINHADOS
## *****
modelo_regressao_mul <- lm(CRC ~ Temperatura + Precipitacao, data = dados_regressao_mul)
modelo_regressao <- lm(CRC ~ Temperatura, data = dados_regressao_mul)

# LIKELIHOOD RATIO TEST (LRT)
# *****
# A hipótese nula é que o modelo mais simples é melhor
# Valores de  $p < 0.05$  rejeita a hipótese nula e o modelo mais complexo é o melhor
# Valores de  $p > 0.05$  não rejeita a hipótese nula e o modelo mais simples é o melhor
library(lmtest)
lrtest(modelo_regressao_mul, modelo_regressao)

## Likelihood ratio test
##
## Model 1: CRC ~ Temperatura + Precipitacao
## Model 2: CRC ~ Temperatura
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1    4 -192.93
```

```
## 2    3 -193.55 -1 1.2558    0.2624
# COMPARANDO COM O MODELO SOMENTE COM O INTERCEPTO
#*****
# criando um modelo sem parâmetros, só o intercepto
modelo_intercepto <- lm(CRC ~ 1, data = dados_regressao_mul)
lrtest(modelo_regressao, modelo_intercepto)

## Likelihood ratio test
##
## Model 1: CRC ~ Temperatura
## Model 2: CRC ~ 1
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    3 -193.55
## 2    2 -210.46 -1 33.815  6.061e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Interpretação dos resultados

Neste exemplo, a precipitação não está associada com a variação no tamanho do CRC das populações de *D. minutus*. Por outro lado, temperatura explicou 26% da variação do tamanho do CRC das populações.

## 3.6 Análises de Variância (ANOVA)

### 3.6.1 Background da análise

Anova refere-se a uma variedade de delineamentos experimentais nos quais a variável preditora é categórica e a variável resposta é contínua (Gotelli & Ellison 2013). Exemplos desses delineamentos experimentais são: Anova de um fator, Anova de dois fatores, Anova em blocos aleatorizados, Anova de medidas repetidas e Anova *split-plot*. De forma geral, a Anova é um teste estatístico usado para comparar a média entre grupos amostrados independentemente. Para isso, o teste leva em conta, além das médias dos grupos, a variação dos dados dentro e entre os grupos. Neste capítulo, iremos demonstrar as linhas de comandos para alguns dos principais delineamentos experimentais.

#### 3.6.1.1 Premissas da Anova:

- As amostras devem ser independentes;
- As unidades amostrais são selecionadas aleatoriamente;
- Distribuição normal (gaussiana) dos resíduos;
- Homogeneidade da variância.

## 3.7 ANOVA de um fator

Este teste considera delineamentos experimentais com apenas um fator (ou tratamento) que pode ser composto por três ou mais grupos (ou níveis).

### 3.7.0.1 Exemplo prático 1 - Anova de um fator

**3.7.0.1.1 Explicação dos dados** Neste exemplo, avaliaremos se o adubo X-2020 disponibilizado recentemente no mercado melhora o crescimento dos indivíduos de *Coffea arabica* como divulgado pela empresa responsável pela venda do produto. Para isso, foi realizado um experimento com indivíduos de *C. arabica* cultivados em três grupos: i) grupo controle onde os indivíduos não receberam adubação, ii) grupo onde os indivíduos receberam a adição do adubo tradicional mais utilizado pelos produtores de *C. arabica*, e iii) grupo onde os indivíduos receberam a adição do adubo X-2020.

#### Pergunta:

O crescimento dos indivíduos de *C. arabica* é melhorado pela adição do adubo X-2020?

#### Predições

O crescimento dos indivíduos de *C. arabica* será maior no grupo que recebeu o adubo X-2020.

#### Variáveis

- Variáveis preditoras
  - Dataframe com as plantas (unidade amostral) nas linhas e o tratamento na coluna.

#### Checklist

- Verificar se o seu dataframe está com as unidades amostrais nas linhas e variável preditora na coluna.

### 3.7.1 Análise

Cálculo da Anova de um fator

```
## IMPORTANDO DADOS
#####
dados_anova_simples <- ecodados::anova_simples
head(dados_anova_simples) # verificar se o dataframe foi lido corretamente

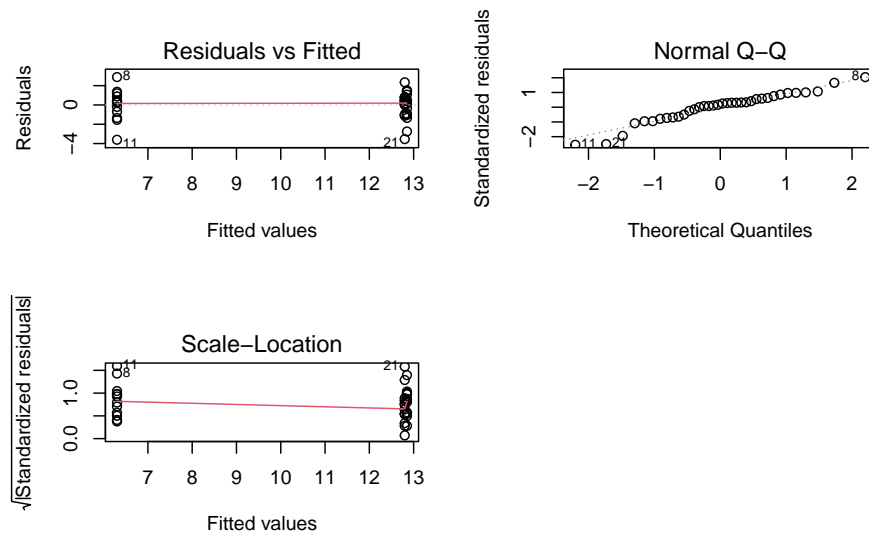
##   Crescimento Tratamento
## 1         7.190    Controle
## 2         6.758    Controle
## 3         6.101    Controle
## 4         4.758    Controle
```

```
## 5      6.542  Controle
## 6      7.667  Controle

# ANÁLISE ANOVA de um fator
#*****
Modelo_anova <- aov(Crescimento ~ Tratamento, data = dados_anova_simples)

# VERIFICANDO A NORMALIDADE E HOMOGENEIDADE DAS VARIÂNCIAS
#*****
# Os gráficos *Residuals vs Fitted* e *Scale-Location* estão relacionados com a homogeneidade das variâncias.
# O gráfico *Normal Q-Q* está relacionado com a distribuição normal dos resíduos. Nestes casos, os pontos devem estar próximos da linha de referência.
par(mfrow = c(2, 2), oma = c(0, 0, 2, 0))
plot(Modelo_anova)
```

aov(Crescimento ~ Tratamento)



```
dev.off()
```

```
## null device
##      1

# Se preferir, você pode utilizar testes estatísticos
# Teste de Shapiro-Wilk para normalidade separadamente para cada grupo
shapiro.test(dados_anova_simples$Crescimento[1:12])

##
## Shapiro-Wilk normality test
##
## data:  dados_anova_simples$Crescimento[1:12]
```

```
## W = 0.96731, p-value = 0.8806
shapiro.test(dados_anova_simples$Crescimento[13:24])

##
## Shapiro-Wilk normality test
##
## data:  dados_anova_simples$Crescimento[13:24]
## W = 0.87324, p-value = 0.07184
shapiro.test(dados_anova_simples$Crescimento[25:36])

##
## Shapiro-Wilk normality test
##
## data:  dados_anova_simples$Crescimento[25:36]
## W = 0.9294, p-value = 0.3738
# Teste de Bartlett para homogeneidade da variância
bartlett.test(Crescimento ~ Tratamento, data = dados_anova_simples)

##
## Bartlett test of homogeneity of variances
##
## data:  Crescimento by Tratamento
## Bartlett's K-squared = 0.61835, df = 2, p-value = 0.7341
# VERIFICANDO OS RESULTADOS DA ANOVA
#*****
anova(Modelo_anova)

## Analysis of Variance Table
##
## Response: Crescimento
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Tratamento  2 340.32  170.160   77.989 3.124e-13 ***
## Residuals   33   72.00    2.182
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Percebam que o resultado da Anova ( $\text{Pr}(>F) < 0.001$ ) indica que devemos rejeitar a hipótese nula que não há diferença entre as médias dos grupos. Contudo, os resultados não mostram quais são os grupos que apresentam diferenças. Para isso, temos que realizar testes de comparações múltiplas *post-hoc* para detectar os grupos que apresentam diferenças significativas entre as médias. **Observação** Os testes *post-hoc* só devem ser utilizados quando rejeitamos a hipótese nula ( $P < 0.05$ ) no teste da Anova.

```
# Diferenças entre os tratamentos
#*****
# Teste de Tuckey's honest significant difference
TukeyHSD(Modelo_anova)
```

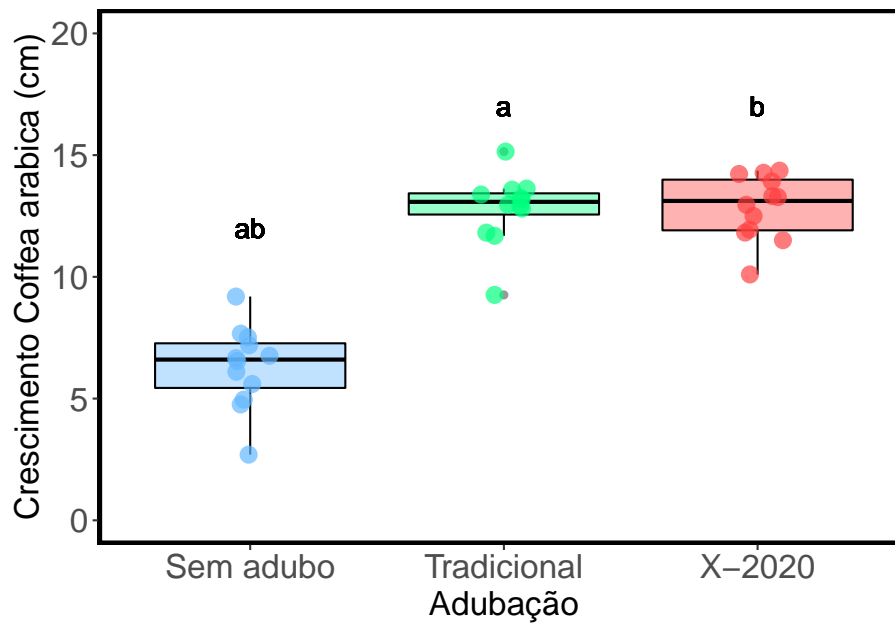
```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Crescimento ~ Tratamento, data = dados_anova_simples)
##
## $Tratamento
##
```

	diff	lwr	upr	p adj
Adubo_X-2020-Adubo_Tradicional	0.04991667	-1.429784	1.529617	0.9962299
Controle-Adubo_Tradicional	-6.49716667	-7.976867	-5.017466	0.0000000
Controle-Adubo_X-2020	-6.54708333	-8.026784	-5.067383	0.0000000

Visualizar os resultados em gráfico

```
# Reordenando a ordem que os grupos irão aparecer no gráfico
dados_anova_simples$Tratamento <- factor(dados_anova_simples$Tratamento ,
                                           levels=c("Controle", "Adubo_Tradicional", "Adubo_X-2020"))

# Gráfico
library(ggplot2)
ggplot(data = dados_anova_simples, aes(x= Tratamento, y= Crescimento, color = Tratamento)) +
  labs(x = "Adubação", y = "Crescimento Coffea arabica (cm)", size = 20) +
  geom_boxplot(fill=c("steelblue1", "springgreen1", "brown1"), color="black", show.legend = TRUE,
               alpha = 0.4) +
  geom_jitter(shape = 16, position=position_jitter(0.1), cex = 4, alpha = 0.7) +
  scale_color_manual(values = c("steelblue1", "springgreen1", "brown1")) +
  scale_y_continuous(limits = c(0, 20), breaks = c(0, 5, 10, 15, 20)) +
  geom_text(x = 1, y = 12, label = "ab", color = "black", size = 5) +
  geom_text(x = 2, y = 17, label = "a", color = "black", size = 5) +
  geom_text(x = 3, y = 17, label = "b", color = "black", size = 5) +
  scale_x_discrete(labels=c("Sem adubo", "Tradicional", "X-2020")) +
  theme_bw() +
  theme(axis.title.y = element_text(size = 17), axis.title.x = element_text(size = 17),
        axis.text.y = element_text(size = 17), axis.text.x = element_text(size = 17)) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size = 2)) +
  theme(legend.position = "none")
```



#### Interpretação dos resultados

Neste exemplo, os indivíduos de *C. arabica* que receberam adubação (tradicional e X-2020) apresentaram maior crescimento do que os indivíduos que não receberam adubação. Contudo, diferente do que foi divulgado pela empresa, o adubo X-2020 não apresentou melhor desempenho que o adubo tradicional já utilizado pelos produtores.

## 3.8 ANOVA de dois fatores ou Anova fatorial

Este teste considera delineamentos amostrais com dois fatores (ou tratamento) que podem ser compostos por dois ou mais grupos (ou níveis). Esta análise tem uma vantagem, pois permite avaliar o efeito da interação entre os fatores na variável resposta. Quando a interação está presente, o impacto de um fator depende do nível (ou grupo) do outro fator.

### 3.8.0.1 Exemplo prático 1 - Anova de dois fatores

**3.8.0.1.1 Explicação dos dados** Neste exemplo, avaliaremos se o tempo que o corpo leva para eliminar uma droga utilizada em exames de ressonância magnética está relacionado com o sistema XY de determinação do sexo e/ou com a idade dos pacientes. Para isso, foi realizado um experimento com 40 pacientes distribuídos da seguinte maneira: i) 10 indivíduos XX - jovens, ii) 10

indivíduos XX - idosas, iii) 10 indivíduos XY - jovens, e iv) 10 indivíduos XY - idosos.

#### Pergunta:

O tempo de eliminação da droga é dependente do sistema XY de determinação do sexo e idade dos pacientes?

#### Predições

O tempo de eliminação da droga vai ser mais rápido nas pacientes XX e jovens.

#### Variáveis

- Variáveis preditoras
  - Dataframe com os pacientes (unidade amostral) nas linhas e os tratamentos nas colunas.

#### Checklist

- Verificar se o seu dataframe está com as unidades amostrais nas linhas e as variáveis preditoras nas colunas.

### 3.8.1 Análise

Cálculo da Anova de dois fator

```
## IMPORTANDO DADOS
#####
dados_dois_fatores <- ecodados::anova_dois_fatores
head(dados_dois_fatores)

##      Tempo Pessoas Idade
## 1 18.952      XX Jovem
## 2 16.513      XX Jovem
## 3 17.981      XX Jovem
## 4 21.371      XX Jovem
## 5 14.470      XX Jovem
## 6 19.130      XX Jovem

# Análise anova de dois fatores
#####
# A interação entre os fatores é representada por *
Modelo1 <- aov(Tempo ~ Pessoas * Idade, data = dados_dois_fatores)

# Olhando os resultados
anova(Modelo1)

## Analysis of Variance Table
##
```



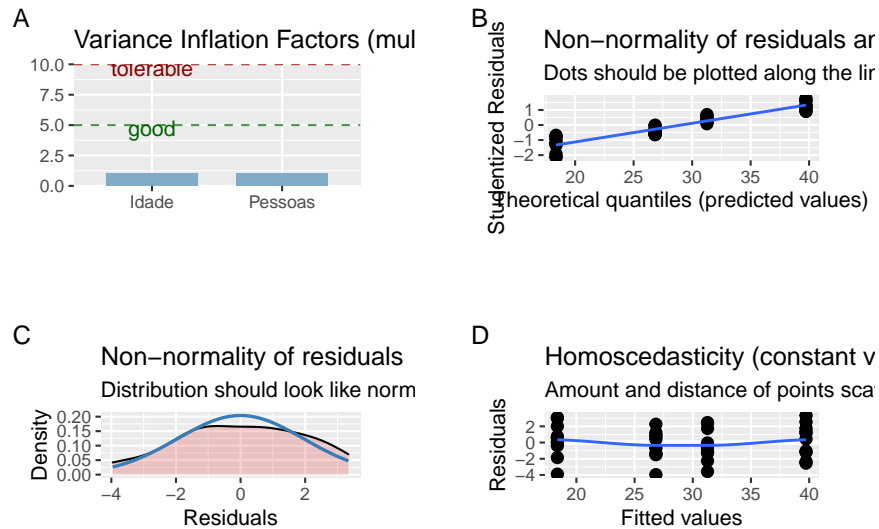
```
## Response: Tempo
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Pessoas      1  716.72   716.72 178.8538 1.56e-15 ***
## Idade         1 1663.73 1663.73 415.1724 < 2.2e-16 ***
## Pessoas:Idade 1    4.77    4.77  1.1903  0.2825
## Residuals    36  144.26    4.01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Percebam que a interação não apresenta um efeito significativo ( $P > 0.05$ ). Assim, iremos retirar a interação
Modelo2 <- aov(Tempo ~ Pessoas + Idade, data = dados_dois_fatores)

# A hipótese nula é que o modelo mais simples é melhor
# Valores de  $p < 0.05$  rejeita a hipótese nula e o modelo mais complexo é o melhor
# Valores de  $p > 0.05$  não rejeita a hipótese nula e o modelo mais simples é o melhor
library(lmtest)
lrtest(Modelo1, Modelo2)
```

```
## Likelihood ratio test
##
## Model 1: Tempo ~ Pessoas * Idade
## Model 2: Tempo ~ Pessoas + Idade
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1    5 -82.413
## 2    4 -83.063 -1  1.3012    0.254

# VERIFICANDO A NORMALIDADE E HOMOGENEIDADE DAS VARIÂNCIAS
#*****
# Esta função mostra os resultados para multicolinearidade (a), dois gráficos avaliando a normalidade dos resíduos
library(sjPlot)
plot_grid(plot_model(Modelo2, type = "diag"))
```



```
# VERIFICANDO OS RESULTADOS DA ANOVA
#*****
anova(Modelo2)
```

```
## Analysis of Variance Table
##
## Response: Tempo
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Pessoas   1  716.72  716.72    177.94 1.041e-15 ***
## Idade     1 1663.73 1663.73    413.05 < 2.2e-16 ***
## Residuals 37  149.03    4.03
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Percebam que o resultado da Anova ( $\text{Pr}(>F) < 0.001$ ) indica que devemos rejeitar a hipótese nula que não há diferença entre as médias dos sistema XY e idade dos pacientes. Neste caso, não precisamos realizar testes de comparações múltiplas *post-hoc* porque os fatores apresentam apenas dois níveis. Contudo, se no seu delineamento experimental um dos fatores apresentar três ou mais níveis, você deverá utilizar os testes de comparações *post-hoc* para determinar as diferenças entre os grupos. **Observação** Os testes *post-hoc* só devem ser utilizados quando rejeitamos a hipótese nula ( $P < 0.05$ ) no teste da Anova.

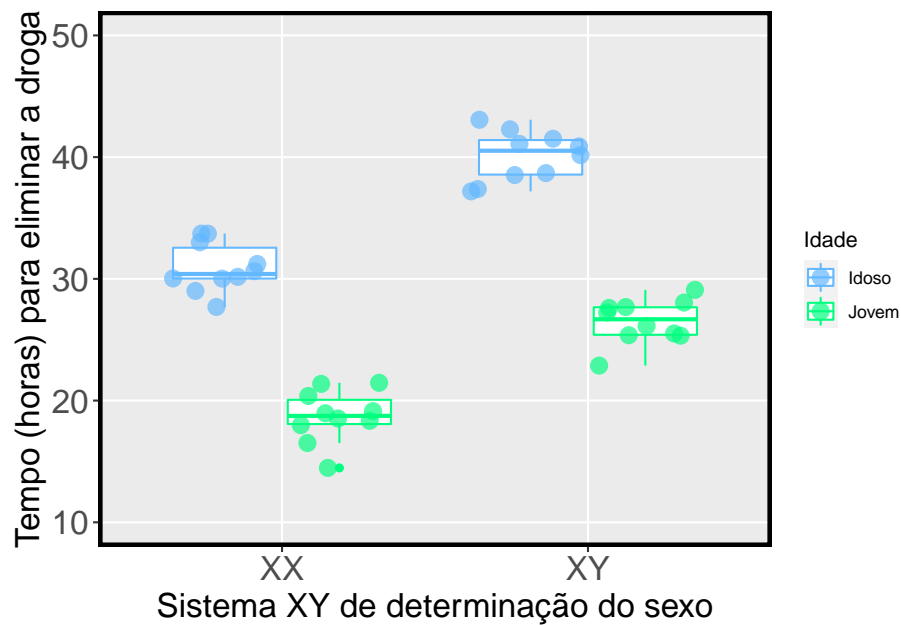
```
# Diferenças entre os tratamentos
#*****
# Teste de Tuckey's honest significant difference
```

```
TukeyHSD(Modelo2)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Tempo ~ Pessoas + Idade, data = dados_dois_fatores)
##
## $Pessoas
##          diff          lwr          upr p adj
## XY-XX 8.46595 7.180008 9.751892      0
##
## $Idade
##          diff          lwr          upr p adj
## Jovem-Idoso -12.89855 -14.18449 -11.61261      0
```

Visualizar os resultados em gráfico

```
# Gráfico
library(ggplot2)
ggplot(data = dados_dois_fatores, aes(y= Tempo, x= Pessoas, color = Idade)) +
  geom_boxplot() +
  labs(x = "Sistema XY de determinação do sexo", y = "Tempo (horas) para eliminar a droga",
       size = 20) +
  geom_jitter(shape = 16, position=position_jitterdodge(), cex = 4, alpha = 0.7) +
  scale_color_manual(values = c("steelblue1", "springgreen1")) +
  scale_y_continuous(limits = c(10, 50), breaks = c(10, 20, 30, 40, 50)) +
  theme(axis.title.y = element_text(size = 17), axis.title.x = element_text(size = 17)) +
  theme(axis.text.y = element_text(size = 17), axis.text.x = element_text(size = 17)) +
  theme(panel.grid.minor = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size = 2))
```



### Interpretação dos resultados

Neste exemplo, O sistema XY de determinação do sexo e a idade dos pacientes tem um efeito no tempo de eliminação da droga do organismo. Os pacientes XX e jovens apresentam eliminação mais rápida da droga do que pacientes XY e idosos.

#### 3.8.1.1 Exemplo prático 2 - Anova de dois fatores com efeito da interação

**3.8.1.1.1 Explicação dos dados** Neste exemplo usaremos os mesmos dados do exemplo anterior. Neste caso, alteremos os dados para que a interação seja significativa.

```
## IMPORTANDO DADOS
#####
dados_dois_fatores_interacao <- ecodados::anova_dois_fatores_interacao1
head(dados_dois_fatores_interacao)
```

```
##      Tempo Pessoas Idade
## 1 31.092      XX Jovem
## 2 27.331      XX Jovem
## 3 24.874      XX Jovem
## 4 27.369      XX Jovem
## 5 28.125      XX Jovem
```

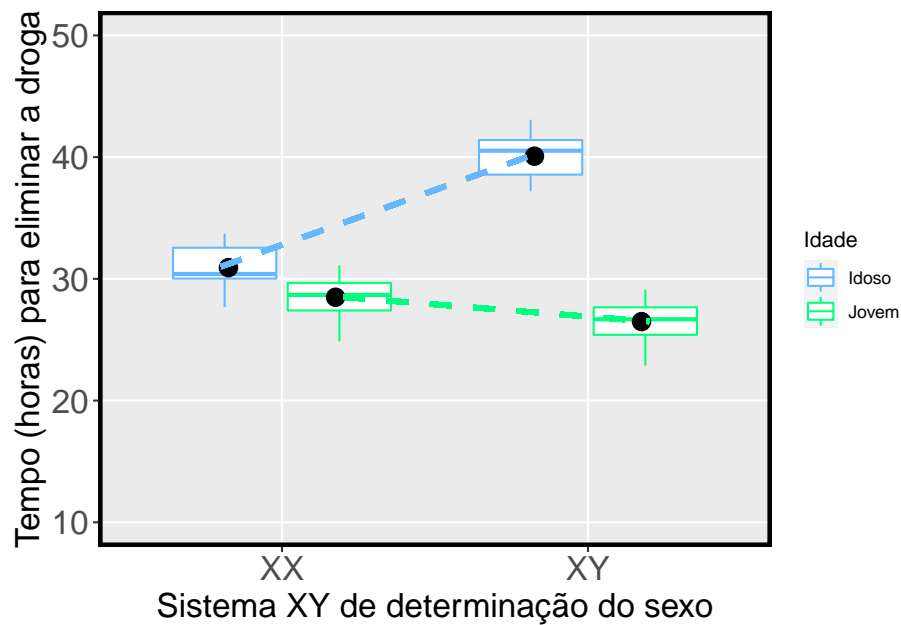
```
## 6 29.607      XX Jovem
# Análise anova de dois fatores
#####
# A interação entre os fatores é representada por *
Modelo_interacao1 <- aov(Tempo ~ Pessoas * Idade, data = dados_dois_fatores_interacao)

# Olhando os resultados
anova(Modelo_interacao1)
```

```
## Analysis of Variance Table
##
## Response: Tempo
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Pessoas      1 128.04  128.04   34.841 9.377e-07 ***
## Idade         1 641.75  641.75  174.623 2.236e-15 ***
## Pessoas:Idade 1 311.17  311.17   84.672 5.463e-11 ***
## Residuals    36 132.30    3.68
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Percebam que a interação é significativa ( $P < 0.05$ ). Agora nossa interpretação precisa ser baseada na interação entre os fatores. Vamos visualizar os resultados em gráfico.

```
# Gráfico
library(ggplot2)
library(ggforce)
ggplot(data = dados_dois_fatores_interacao, aes(y= Tempo, x= Pessoas, color = Idade)) +
  geom_boxplot() +
  stat_summary(fun = mean, geom="point", aes(group=Idade, x = Pessoas), color = "black",
              position = position_dodge(0.7), size = 4) +
  geom_link(aes(x = 0.8, y = 31, xend = 1.8, yend = 40), color = "steelblue1",
            lwd = 1.3, linetype = 2) +
  geom_link(aes(x = 1.2, y = 28.5, xend = 2.2, yend = 26.5), color = "springgreen1",
            lwd = 1.3, linetype = 2) +
  labs(x = "Sistema XY de determinação do sexo", y = "Tempo (horas) para eliminar a droga",
       size = 20) +
  scale_color_manual(values = c("steelblue1", "springgreen1", "steelblue1",
                                "springgreen1")) +
  scale_y_continuous(limits = c(10, 50), breaks = c(10, 20, 30, 40, 50)) +
  theme(axis.title.y = element_text(size = 17), axis.title.x = element_text(size = 17)) +
  theme(axis.text.y = element_text(size = 17), axis.text.x = element_text(size = 17)) +
  theme(panel.grid.minor = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size = 2))
```



### Interpretação dos resultados

Percebam que para saber a resposta do fator idade (jovem ou idoso) na eliminação da droga, você precisa saber com qual pessoa (XX ou XY) ele está associado. Isso porque a resposta de um fator depende do outro fator. Jovens eliminam a droga do corpo mais rápido nas pessoas XY enquanto os idosos eliminam a droga mais rápido nas pessoas XX.

#### 3.8.1.2 Exemplo prático 3 - Anova de dois fatores com efeito da interação

**3.8.1.2.1 Explicação dos dados** Neste exemplo usaremos os mesmos dados do exemplo anterior. Neste caso, alteremos os dados para que a interação seja significativa.

```
## IMPORTANDO DADOS
#####
dados_dois_fatores_interacao2 <- ecodados::anova_dois_fatores_interacao2
head(dados_dois_fatores_interacao2)
```

```
##      Tempo Pessoas Idade
## 1 18.952      XX Jovem
## 2 16.513      XX Jovem
## 3 17.981      XX Jovem
## 4 21.371      XX Jovem
```

```
## 5 14.470      XX Jovem
## 6 19.130      XX Jovem

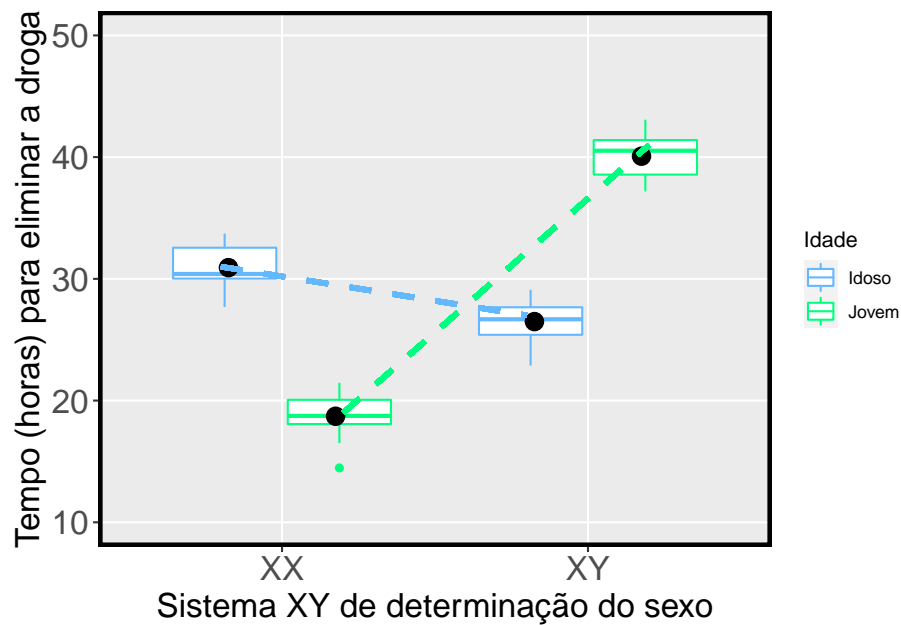
# Análise anova de dois fatores
#####
# A interação entre os fatores é representada por *
Modelo_interacao2 <- aov(Tempo ~ Pessoas * Idade, data = dados_dois_fatores_interacao2)

# Olhando os resultados
anova(Modelo_interacao2)
```

```
## Analysis of Variance Table
##
## Response: Tempo
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Pessoas    1  716.72   716.72 178.8538 1.56e-15 ***
## Idade       1    4.77     4.77   1.1903  0.2825
## Pessoas:Idade 1 1663.73 1663.73 415.1724 < 2.2e-16 ***
## Residuals   36  144.26     4.01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Percebam que a interação é significativa ( $P < 0.05$ ), mas a idade não é significativa. Nossa interpretação precisa ser baseada na interação entre os fatores. Vamos visualizar os resultados em gráfico.

```
# Gráfico
library(ggplot2)
library(ggforce)
ggplot(data = dados_dois_fatores_interacao2, aes(y= Tempo, x= Pessoas, color = Idade)) +
  geom_boxplot() +
  stat_summary(fun = mean, geom="point", aes(group=Idade, x = Pessoas), color = "black",
              position = position_dodge(0.7), size = 4) +
  geom_link(aes(x = 0.8, y = 31, xend = 1.8, yend = 27), color = "steelblue1",
            lwd = 1.3, linetype = 2) +
  geom_link(aes(x = 1.2, y = 19, xend = 2.2, yend = 41), color = "springgreen1",
            lwd = 1.3, linetype = 2) +
  labs(x = "Sistema XY de determinação do sexo", y = "Tempo (horas) para eliminar a droga",
       size = 20) +
  scale_color_manual(values = c("steelblue1", "springgreen1", "steelblue1", "springgreen1")) +
  scale_y_continuous(limits = c(10, 50), breaks = c(10, 20, 30, 40, 50)) +
  theme(axis.title.y = element_text(size = 17), axis.title.x = element_text(size = 17)) +
  theme(axis.text.y = element_text(size = 17), axis.text.x = element_text(size = 17)) +
  theme(panel.grid.minor = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size = 2))
```



#### Interpretação dos resultados

Percebam que as linhas se cruzam. Esse é exemplo clássico de interação. Novamente, para saber a resposta do fator idade (jovem ou idoso), você precisa saber com qual pessoa (XX ou XY) ele está associado. Jovens são mais rápidos para eliminar a droga em pessoas XX enquanto os idosos são mais rápidos para eliminar a droga nas pessoas XY.

### 3.9 ANOVA em blocos aleatorizados

No delineamento experimental com blocos aleatorizados, cada fator é agrupado em blocos, com réplicas de cada nível do fator representado em cada bloco (Gotelli & Elisson 2013). O bloco é uma área ou período de tempo dentro do qual as condições ambientais são relativamente homogêneas. O objetivo do uso dos blocos é controlar fontes de variações indesejadas na variável dependente que não são de interesse do pesquisador. Desta maneira, podemos retirar dos resíduos, os efeitos das variações indesejadas que não são do nosso interesse, e testar com maior poder estatístico os efeitos dos tratamentos de interesse. Importante, os blocos devem ser arranjados de forma que as condições ambientais sejam mais similares dentro dos blocos do que entre os blocos.



### 3.9.0.1 Exemplo prático 1 - Anova em blocos aleatorizados

**3.9.0.1.1 Explicação dos dados** Neste exemplo, avaliaremos a riqueza de espécies de anuros amostradas em poças artificiais instaladas a diferentes distâncias de seis fragmentos florestais no sudeste do Brasil (da Silva et al. 2020). Os fragmentos florestais apresentam diferenças entre si que não do interesse do pesquisador. Por isso, eles foram incluídos como blocos nas análises. As poças artificiais foram instaladas em todos os fragmentos florestais baseado no seguinte delineamento experimental (da Silva et al. 2012): i) 4 poças no interior do fragmento a 100m de distância da borda do fragmento; ii) 4 poças no interior no fragmento a 50m de distância da borda do fragmento; iii) 4 poças na borda do fragmento; iv) 4 poças na matriz de pastagem a 50m de distância da borda do fragmento; e v) 4 poças na matriz de pastagem a 100m de distância da borda do fragmento. Percebam que todos os tratamentos foram instalados em todos os blocos.

#### Pergunta:

A distância da poça artificial ao fragmento florestal influencia a riqueza de espécies anuros?

#### Predições

Poças na borda do fragmento florestal apresentarão maior riqueza de espécies do que poças distantes da borda.

#### Variáveis

- Variáveis preditoras
  - Dataframe com as poças (unidade amostral) nas linhas e o tratamento e bloco nas colunas.

#### Checklist

- Verificar se o seu dataframe está com as unidades amostrais nas linhas e variáveis preditoras nas colunas.

## 3.9.1 Análise

Cálculo da Anova em blocos aleatorizados

```
## IMPORTANDO DADOS
#####
dados_bloco <- ecodados::anova_bloco
str(dados_bloco) # verificar se o dataframe foi lido corretamente

## 'data.frame': 30 obs. of 3 variables:
## $ Riqueza: int  90 95 107 92 89 92 81 92 93 80 ...
## $ Blocos : chr  "A" "A" "A" "A" ...
## $ Pocas : chr  "Int-50m" "Int-100m" "Borda" "Mat-50m" ...
```

```

# ANALISE ANOVA em blocos aleatorizados
#####
# Há duas maneiras para incluir os efeitos dos blocos
model_bloco1 <- aov(Riqueza ~ Pocas + Blocos, data = dados_bloco)

model_bloco2 <- aov(Riqueza ~ Pocas + Error(Blocos), data = dados_bloco)

# Percebam que as duas formas apresentam os mesmo resultados para o efeito da distância
anova(model_bloco1)

## Analysis of Variance Table
##
## Response: Riqueza
##           Df Sum Sq Mean Sq F value Pr(>F)
## Pocas      4 1504.5   376.12   2.9071 0.0478 *
## Blocos      5 1089.0   217.79   1.6834 0.1846
## Residuals 20 2587.5   129.38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(model_bloco2)

##
## Error: Blocos
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  5   1089    217.8
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## Pocas      4   1504    376.1   2.907 0.0478 *
## Residuals 20   2588    129.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# O que não pode acontecer é ignorar o efeito do bloco que é incorporado pelos resíduos.

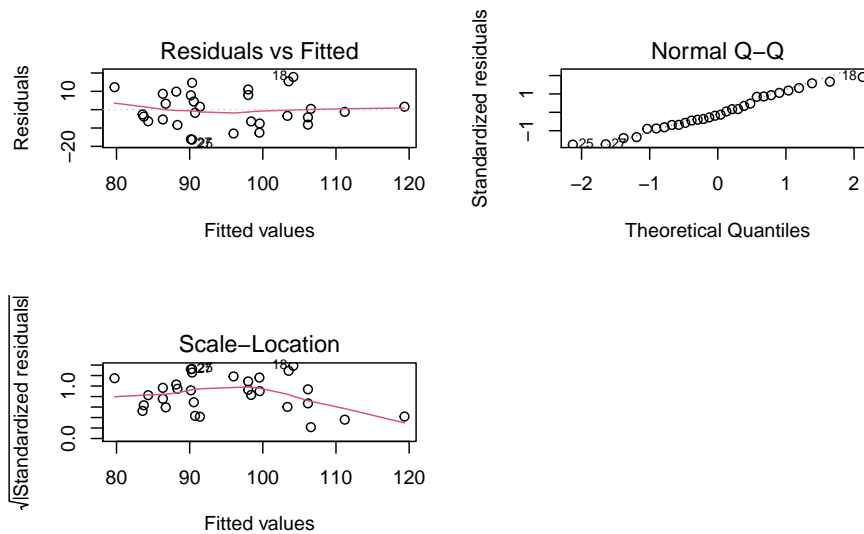
modelo_errado <- aov(Riqueza ~ Pocas, data = dados_bloco)
anova(modelo_errado)

## Analysis of Variance Table
##
## Response: Riqueza
##           Df Sum Sq Mean Sq F value Pr(>F)
## Pocas      4 1504.5   376.12   2.5576 0.06359 .
## Residuals 25 3676.5   147.06
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
# VERIFICANDO A NORMALIDADE E HOMOGENEIDADE DAS VARIÂNCIAS
#*****
# Os gráficos *Residuals vs Fitted* e *Scale-Location* estão relacionados com a homogeneidade da
# O gráfico *Normal Q-Q* está relacionado com a distribuição normal dos resíduos. Neste gráfico,
par(mfrow = c(2, 2), oma = c(0, 0, 2, 0))
plot(model_bloco1)
```

aov(Riqueza ~ Pocas + Blocos)



```
dev.off()
```

```
## null device
##          1
```

Percebam que o resultado da Anova ( $\Pr(>F) < 0.001$ ) indica que devemos rejeitar a hipótese nula que não há diferença entre as médias dos grupos. Contudo, os resultados não mostram quais são os grupos que apresentam diferenças. Para isso, temos que realizar testes de comparações múltiplas *post-hoc* para detectar os grupos que apresentam diferenças significativas entre as médias. **Observação** Os testes *post-hoc* só devem ser utilizados quando rejeitamos a hipótese nula ( $P < 0.05$ ) no teste da Anova.

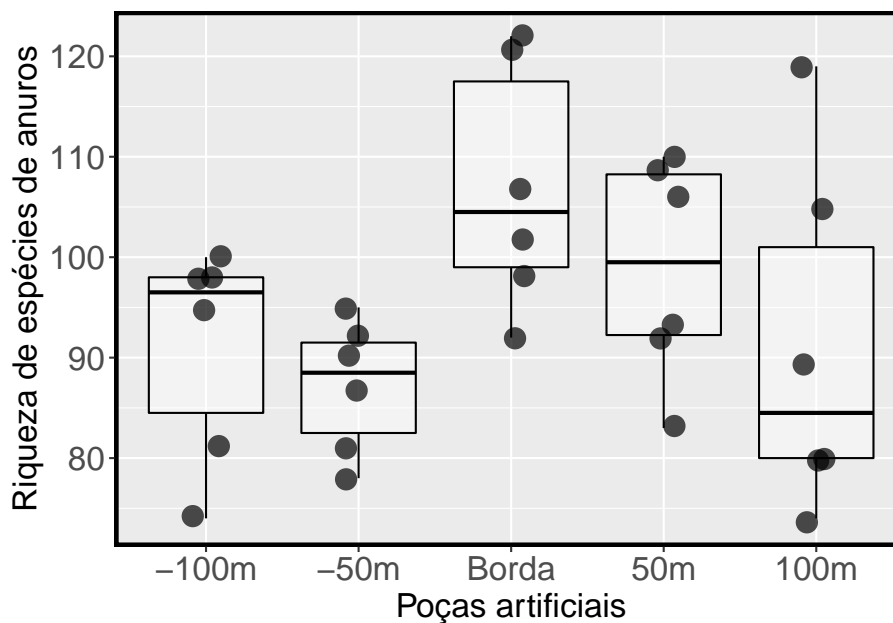
```
# Diferenças entre os tratamentos
#*****
# Teste de Tukey's honest significant difference
library(emmeans)
pairs(lsmeans(model_bloco1, "Pocas"), adjust = "tukey")
```

```
## contrast          estimate    SE df t.ratio p.value
## Borda - (Int-100m)      16.000 6.57 20   2.436  0.1463
## Borda - (Int-50m)      19.833 6.57 20   3.020  0.0472
## Borda - (Mat-100m)     15.833 6.57 20   2.411  0.1531
## Borda - (Mat-50m)       8.167 6.57 20   1.244  0.7269
## (Int-100m) - (Int-50m)   3.833 6.57 20   0.584  0.9760
## (Int-100m) - (Mat-100m) -0.167 6.57 20  -0.025  1.0000
## (Int-100m) - (Mat-50m)  -7.833 6.57 20  -1.193  0.7553
## (Int-50m) - (Mat-100m)  -4.000 6.57 20  -0.609  0.9720
## (Int-50m) - (Mat-50m) -11.667 6.57 20  -1.777  0.4135
## (Mat-100m) - (Mat-50m)  -7.667 6.57 20  -1.167  0.7692
##
## Results are averaged over the levels of: Blocos
## P value adjustment: tukey method for comparing a family of 5 estimates
```

Visualizar os resultados em gráfico

```
# Reordenando a ordem que os grupos irão aparecer no gráfico
dados_bloco$Pocas <- factor(dados_bloco$Pocas,
                             levels=c("Int-100m", "Int-50m", "Borda", "Mat-50m", "Mat-100m"))

# Gráfico
library(ggplot2)
ggplot(data = dados_bloco, aes(x= Pocas, y= Riqueza)) +
  labs(x = "Poças artificiais", y = "Riqueza de espécies de anuros", size = 20) +
  geom_boxplot(color="black", show.legend = FALSE,
               alpha = 0.4) +
  geom_jitter(shape = 16, position=position_jitter(0.1), cex = 5, alpha = 0.7) +
  scale_x_discrete(labels=c("-100m", "-50m", "Borda", "50m", "100m")) +
  theme(axis.title.y = element_text(size = 17), axis.title.x = element_text(size = 17)) +
  theme(axis.text.y = element_text(size = 17), axis.text.x = element_text(size = 17)) +
  theme(panel.border = element_rect(colour = "black", fill=NA, size = 2))
```



#### Interpretação dos resultados

Neste exemplo, rejeitamos a hipótese nula que a distância da poças artificiais até as bordas dos fragmentos florestais não influencia a riqueza de espécies de anuros. As poças artificiais instaladas nas bordas dos fragmentos florestais apresentaram maior riqueza de espécies do que as poças distantes.

### 3.10 Análise de covariância (ANCOVA)

A ANCOVA pode ser compreendida como uma extensão da Anova com a adição de variável contínua (covariável) medida em todas as unidades amostrais (Gotelli & Ellison 2013). A ideia é que a covariável também afete os valores da variável resposta. Não incluir a covariável irá fazer com que a variação não explicada pelo modelo concentre-se nos resíduos. Incluindo a covariável, o tamanho do resíduo é menor, e o teste para avaliar as diferenças nos tratamentos terá mais poder estatístico.

#### 3.10.0.1 Exemplo prático 1 - ANCOVA

**3.10.0.1.1 Explicação dos dados** Neste exemplo, avaliaremos o efeito da herbivoria na biomassa dos frutos de uma espécie de árvore na Mata Atlântica. O delineamento experimental permitiu que alguns indivíduos sofressem

herbivoria e outros não. Os pesquisadores também mediram o tamanho da raiz dos indivíduos para inseri-la como uma covariável no modelo.

#### Pergunta:

A herbivoria diminuiu a biomassa dos frutos?

#### Predições

Os indivíduos que sofreram herbivoria irão produzir frutos com menor biomassa do que os indivíduos sem herbivoria.

#### Variáveis

- Variáveis preditoras
  - Dataframe com as indivíduos da espécie de planta (unidade amostral) nas linhas e o tratamento e a covariável nas colunas.

#### Checklist

- Verificar se o seu dataframe está com as unidades amostrais nas linhas e variáveis preditoras nas colunas.

### 3.10.1 Análise

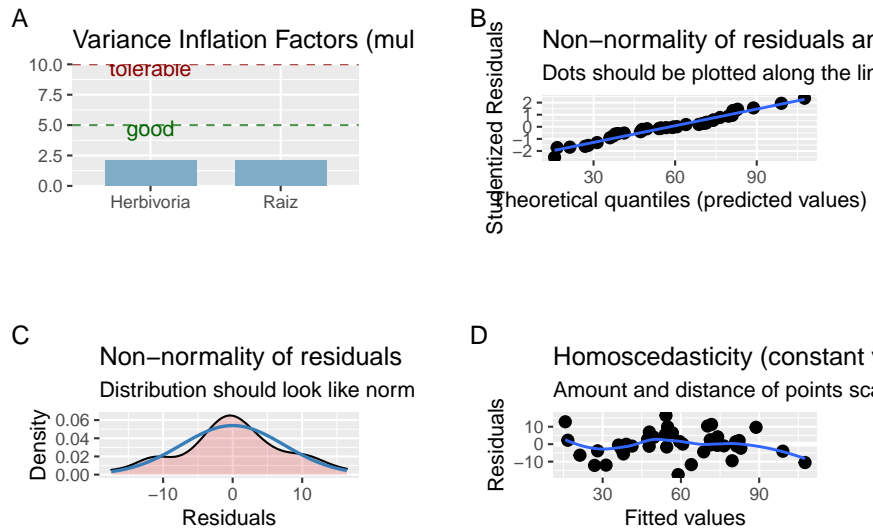
#### Cálculo da ANCOVA

```
## IMPORTANDO DADOS
#####
dados_ancova <- ecodados::ancova
str(dados_ancova) # verificar se o dataframe foi lido corretamente

## 'data.frame': 40 obs. of 3 variables:
## $ Raiz : num 6.22 6.49 4.92 5.13 5.42 ...
## $ Biomassa : num 59.8 61 14.7 19.3 34.2 ...
## $ Herbivoria: chr "Sem_herb" "Sem_herb" "Sem_herb" "Sem_herb" ...

# ANALISE ANCOVA
#####
modelo_ancova <- lm(Biomassa ~ Herbivoria + Raiz, data = dados_ancova)

# VERIFICANDO A NORMALIDADE E HOMOGENEIDADE DAS VARIÂNCIAS
#####
# Esta função mostra os resultados para multicolinearidade (a), dois gráficos avaliand
library(sjPlot)
plot_grid(plot_model(modelo_ancova, type = "diag"))
```



```
# OLHANDO OS RESULTADOS
```

```
#####
```

```
anova(modelo_ancova)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: Biomassa
```

```
##          Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## Herbivoria  1  1941.9  1941.9   33.759 1.135e-06 ***
```

```
## Raiz        1 17434.1 17434.1  303.075 < 2.2e-16 ***
```

```
## Residuals  37  2128.4    57.5
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Percebam que o resultado da ANCOVA ( $\text{Pr}(>F) < 0.001$ ) indica que tanto a herbivoria como a o tamanho da raiz (covariável) tem efeitos significativos na biomassa dos frutos. Contudo, a interação entre as variáveis não foi significativa. Vamos usar o Likelihood ratio test (LRT) para ver se podemos seguir com um modelo mais simples (sem interação).

```
modelo_ancova <- lm(Biomassa ~ Herbivoria * Raiz, data = dados_ancova)
```

```
modelo_ancova2 <- lm(Biomassa ~ Herbivoria + Raiz, data = dados_ancova)
```

```
# LRT
```

```
library(lmtest)
```

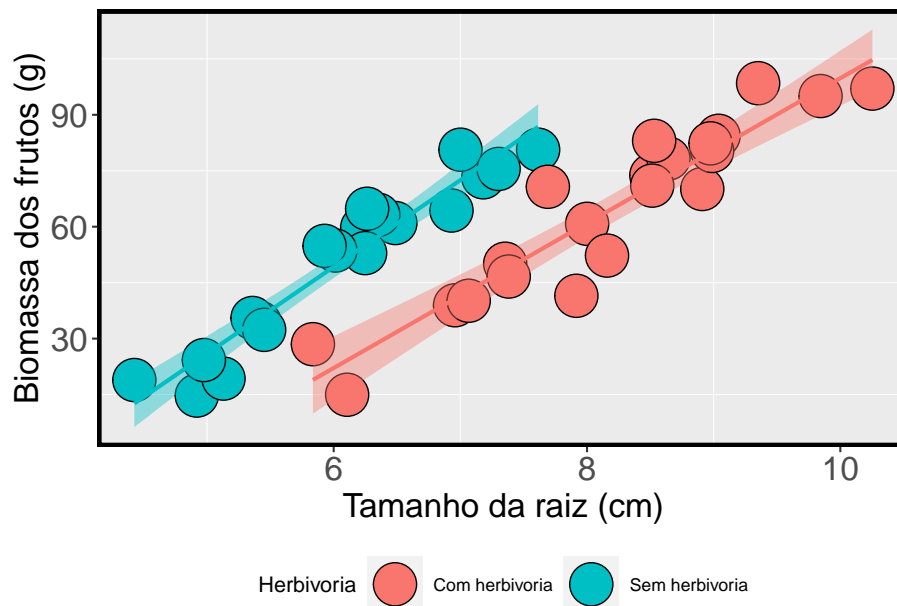
```
# A hipótese nula é que o modelo mais simples é melhor
```

```
# Valores de  $p < 0.05$  rejeita a hipótese nula e o modelo mais complexo é o melhor
# Valores de  $p > 0.05$  não rejeita a hipótese nula e o modelo mais simples é o melhor
lrtest(modelo_ancova, modelo_ancova2)
```

```
## Likelihood ratio test
##
## Model 1: Biomassa ~ Herbivoria * Raiz
## Model 2: Biomassa ~ Herbivoria + Raiz
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1    5 -134.91
## 2    4 -136.24 -1  2.6554    0.1032
```

Visualizar os resultados em gráfico

```
# Gráfico
library(ggplot2)
ggplot(data = dados_ancova, aes(x= Raiz, y= Biomassa, fill = Herbivoria)) +
  labs(x = "Tamanho da raiz (cm)", y = "Biomassa dos frutos (g)", size = 20) +
  geom_point(size = 10, shape = 21) +
  theme(axis.title.y = element_text(size = 17), axis.title.x = element_text(size = 17)) +
  theme(axis.text.y = element_text(size = 17), axis.text.x = element_text(size = 17)) +
  theme(panel.grid.major = element_blank(),
        panel.border = element_rect(colour = "black", fill=NA, size = 2)) +
  theme(legend.position="bottom") +
  scale_fill_discrete(name = "Herbivoria", labels = c("Com herbivoria", "Sem herbivoria")) +
  geom_smooth(aes(color=Herbivoria), method="lm", show.legend = FALSE)
```





**Interpretação dos resultados**

Neste exemplo, o tamanho da raiz (covariável) tem uma relação positiva com a biomassa dos frutos. Quanto maior o tamanho da raiz, maior a biomassa dos frutos. Usando a ANCOVA e controlando o efeito da covariável, percebemos que a herbivoria também afeta a biomassa dos frutos. Os indivíduos que não sofreram herbivoria produziram frutos com maior biomassa do que os indivíduos com herbivoria.

**3.10.2 Para se aprofundar**

- Recomendamos aos interessados os livros: i) Zar (2010) Biostatistical analysis; ii) Gotelli & Ellison (2013) A primer of ecological statistics; e iii) Quinn & Keough (2002) Experimental design and data analysis for biologists.



## Chapter 4

# Introdução à Análises Multidimensionais

Neste módulo iremos aprender como implementar no R as análises multivariadas mais comumente utilizadas em ecologia de comunidades. Para isso precisaremos dos pacotes `vegan`, `labdsv` e `ade4`. Procuraremos explicar brevemente a lógica por trás de cada teste, a sua aplicação em problemas comumente encontrados em estudos ecológicos, mas não destrinchar detalhadamente como cada método funciona e o seu componente matemático.

Em geral, análises multivariadas têm três principais utilidades: reduzir a dimensionalidade dos dados e encontrar a principal direção de variação dos dados, testar relações entre matrizes, ou ainda encontrar diferenças entre grupos. Apesar dessas análises também serem utilizadas como análises exploratórias e para descrever padrões em estudos ecológicos, a necessidade de se ter hipóteses, ou ao menos expectativas a priori, não pode ser ignorada. Se quiser saber mais sobre aspectos teóricos e filosóficos das análises, sugerimos consultar James & McCulloch (1990). Antes de entrar de cabeça nas análises multivariadas também, sugerimos fortemente o estudo de métodos de amostragem e como fazer boas perguntas. Não vamos nos estender muito nesses tópicos porque eles foram abordados nas aulas disponíveis no YouTube.

Análises multivariadas podem ser divididas, grosseiramente, em dois tipos: agrupamento e ordenação. Análises de agrupamento em geral tentam agrupar objetos (observações) em grupos de maneira que objetos do mesmo grupo sejam mais semelhantes entre si do que objetos de outros grupos. Mais formalmente, o agrupamento de objetos (ou descritores) é uma operação pela qual um conjunto de objetos (ou descritores) é particionado em dois ou mais subconjuntos, usando regras pré-estabelecidas de aglomeração ou divisão (Legendre & Legendre, 2012). Por outro lado, a análise de ordenação é uma operação pela qual os objetos (ou descritores) são posicionados num espaço que contém menos dimen-

sões que o conjunto de dados original; a posição dos objetos ou descritores em relação aos outros também podem ser usadas para agrupá-los.

Vamos começar com análises de agrupamento. Aqui vamos exemplificar dois métodos: uma técnica de agrupamento hierárquica (dendrograma) e outra não-hierárquica (k-means).

## 4.1 Background da análise

O objetivo da análise de agrupamento é agrupar objetos admitindo que haja um grau de similaridade entre eles. Esta análise pode ser utilizada ainda para classificar uma população em grupos homogêneos de acordo com uma característica de interesse. A grosso modo, uma análise de agrupamento tenta resumir uma grande quantidade de dados e apresentá-la de maneira fácil de visualizar e entender (em geral, na forma de um dendrograma). No entanto, os resultados da análise podem não refletir necessariamente toda a informação originalmente contida na matriz de dados. Para avaliar o quão bem uma análise de agrupamento representa os dados originais existe uma métrica — o coeficiente de correlação cofenético — o qual discutiremos em detalhes mais adiante.

Antes de considerar algum método de agrupamento, pense porque você esperaria que houvesse uma descontinuidade nos dados; ou ainda, considere se existe algum ganho prático em dividir uma nuvem de objetos contínuos em grupos. O padrão apresentado pelo dendrograma depende do protocolo utilizado (método de agrupamento e índice de dissimilaridade); os grupos formados dependem do nível de corte escolhido. O leitor interessado é remetido à duas referências: Legendre & Legendre (2012) e Borcard et al. (2018).

## 4.2 Exemplo 1:

### Pergunta:

Existem grupos de espécies de anfíbios anuros com padrões de ocorrência similar ao longo de poças?

### Predições

- 1: Iremos encontrar ao menos dois grupos de espécies: aquelas que ocorrem em poças dentro de floresta vs. aquelas que ocorrem em poças de áreas abertas.

### Variáveis

- Variáveis preditoras
  - 1. A nossa matriz de dados contém a abundância das espécies nas linhas e locais (poças) nas colunas.

### 4.2.1 Explicação da análise

A matriz deve conter os objetos a serem agrupados (e.g., espécies) nas linhas e as variáveis (e.g., locais de coleta ou medidas morfológicas) nas colunas. A escolha do método de agrupamento é crítico para a escolha de um coeficiente de associação. É importante compreender as propriedades dos métodos de agrupamento para interpretar corretamente a estrutura ecológica que eles evidenciam (Legendre & Legendre, 2012). De acordo com a classificação de Sneath & Sokal (1973) existem cinco tipos de métodos: 1) seqüenciais ou simultâneos; 2) aglomerativo ou divisivo ;3) monotéticos ou politéticos; 4) hierárquico ou não hierárquicos e 5) probabilístico.

Métodos hierárquicos podem ser divididos naqueles que consideram o centróide ou amédia aritmética entre os grupos. O principal método hierárquico que utiliza a média aritmética é o UPGMA (Agrupamento pelas médias aritméticas não ponderadas), e o principal método que utiliza centróides é a Distância mínima de Ward.

O UPGMA funciona da seguinte forma: a maior similaridade (ou menor distância) identifica os próximos agrupamentos a serem formados. Após esse evento, o método calcula a média aritmética das similaridades ou distâncias entre um objeto e cada um dos membros do grupo ou, no caso de um grupo previamente formado, entre todos os membros dos dois grupos. Todos os objetos recebem pesos iguais no cálculo.

O método de Ward é baseado no critério de quadrados mínimos (OLS), o mesmo utilizado para ajustar um modelo linear. O objetivo é definir os grupos de maneira que a soma de quadrados (i.e. similar ao erro quadrado da ANOVA) dentro dos grupos seja minimizada (Borcard et al. 2018).

#### Checklist

- Verifique se não há espaço nos nomes das colunas e linhas
- Se os dados forem de abundância, recomenda-se realizar a transformação de Hellinger (Legendre & Gallagher, 2001).
- Se a matriz original contiver muitos valores discrepantes (e.g., uma espécie muito mais ou muito menos abundante que outras) é necessário transformar os dados usando `log1p`.
- Se as variáveis forem medidas tomadas em diferentes escalas (metros, graus celcius etc), é necessário padronizar cada variável para ter a média 0 e desvio padrão 1. Isso pode ser feito utilizando a função `decostand` do pacote `vegan`.

### 4.2.2 Análise

Para começar, vamos primeiro importar os dados e depois calcular a matriz de distância que seja adequada para o tipo de dado que temos (abundância de espécies - dados de contagem)

```
library(ecodados) # Carrega o arquivo multivar_bocaina
library(vegan)
```

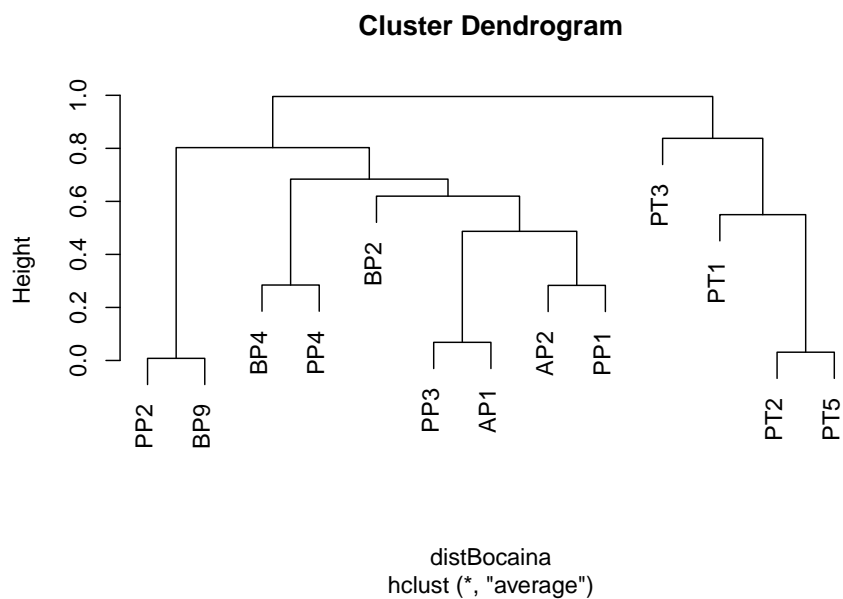
```
#sp_compos <- read.table("bocaina.txt", h=TRUE)
sp_compos <- multivar_bocaina
head(sp_compos)
```

```
##          BP4  PP4 PP3  AP1 AP2 PP1 PP2 BP9  PT1 PT2 PT3 BP2 PT5
## Aper         0   3   0    0  2   0   0   0    0  0   0 181   0
## Bahe      859  14  14    0 87 312 624 641    0  0   0  14   0
## Rict     1772 1517 207   573 796   0   0   0    0  0   0   0   0
## Cleuco      0   0   0    0   0   0   0   0    0 29 369   0  84
## Dmic        0   0   6   60   4   0   0   0 2758 319  25   0 329
## Dmin        0  84 344 1045  90   0   0   0    8   0   0   0   0
```

```
bocaina <- t(sp_compos) #transpondo a matriz para obter a classificação por linhas
distBocaina <- vegdist(bocaina, method="horn") #produz uma matriz de similaridade com o
dendro <- hclust(distBocaina, method="average") #produz um agrupamento com a função hclust
```

Visualizar os resultados

```
plot(dendro)
```



#### 4.2.2.1 Interpretação dos resultados

Antes de começarmos a interpretar os resultados precisamos verificar que o agrupamento reduziu a dimensionalidade da matriz de forma eficiente, de maneira a não distorcer a informação. Fazemos isso calculando o **Coefficiente de correlação cofenética (CCC)**

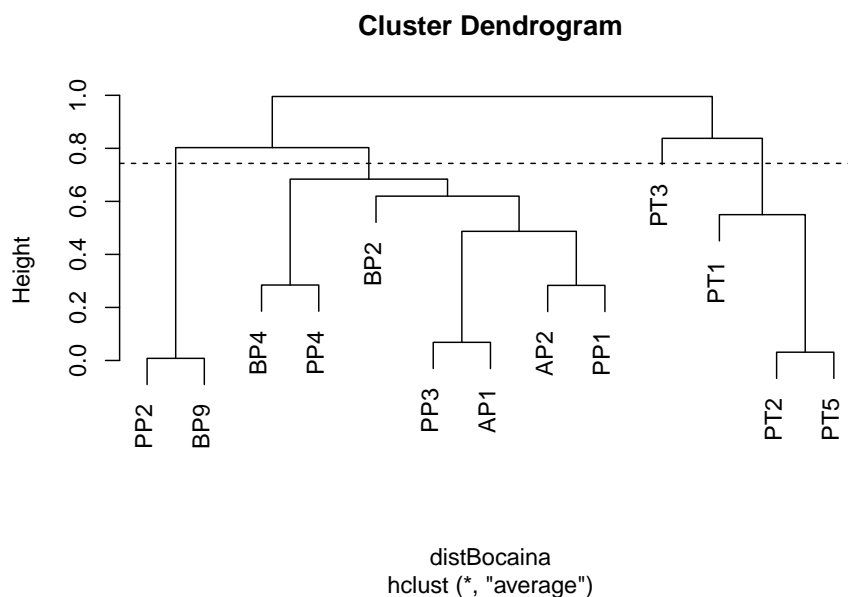
```
cofresult <- cophenetic(dendro)
cor(cofresult, distBocaina)
```

```
## [1] 0.8819701
```

Um CCC > .7 indica uma boa representação. Portanto, o nosso resultado de 0.8819701 é bastante alto, garantindo que o dendrograma é adequado.

No entanto, para interpretar os resultados precisamos antes definir um nível de corte, que vai nos dizer quantos grupos existem. Há vários métodos para definir grupos, desde os heurísticos aos que utilizam bootstrap. Se quisermos interpretar este dendrograma, podemos por exemplo estabelecer um nível de corte de 50% de distância (ou seja, grupos cujos objetos tenham ao menos 50% de similaridade entre si).

```
plot(dendro)
k = 4
n = nrow(bocaina)
MidPoint = (dendro$height[n-k] + dendro$height[n-k+1]) / 2
abline(h = MidPoint, lty=2)
```



Nesse caso teremos a formação de cinco grupos, representados pelos nós que estão abaixo da linha de corte.

### 4.3 Exemplo 2:

A seguir, vamos utilizar o pacote `pvclust` que calcula automaticamente o nível de corte de similaridade baseado no Bootstrap de cada nó. Uma desvantagem deste método é que ele somente aceita índices de similaridade da função `dist` que possui apenas a distância Euclidiana, Manhattan e Canberra. Uma maneira de contornarmos essa limitação é utilizar transformações dos dados disponíveis na função `disttransform` no pacote `BiodiversityR` ou o `decostand` do pacote `vegan`. Também é possível utilizar a transformação de Box-Cox para dados multivariados, disponível no material suplementar de Legendre & Borcard (2018) aqui

```
library(pvclust)
library(BiodiversityR)
```

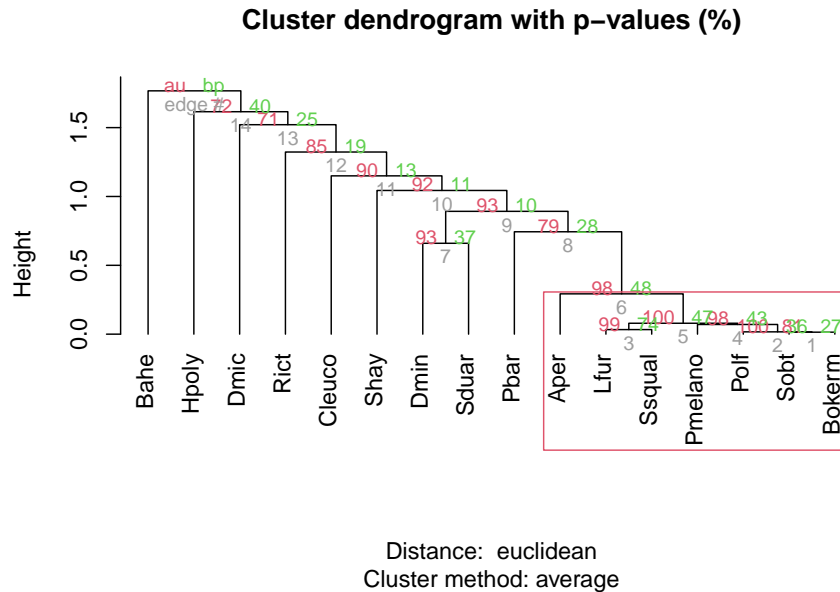
Aqui vamos utilizar a distância de Chord para calcular a matriz de distância. Se transformarmos uma matriz usando a transformação Chord e depois calcularmos a distância Euclidiana, isso equivale à calcular diretamente a distância de Chord:

```
bocaina_transf <- disttransform(bocaina, "chord")
analise <- pvclust(bocaina_transf, method.hclust="average", method.dist="euclidean")
```

```
## Bootstrap (r = 0.46)... Done.
## Bootstrap (r = 0.54)... Done.
## Bootstrap (r = 0.69)... Done.
## Bootstrap (r = 0.77)... Done.
## Bootstrap (r = 0.85)... Done.
## Bootstrap (r = 1.0)... Done.
## Bootstrap (r = 1.08)... Done.
## Bootstrap (r = 1.15)... Done.
## Bootstrap (r = 1.23)... Done.
## Bootstrap (r = 1.38)... Done.
```

```
plot(analise, hang=-1)
pvrect(analise)
```



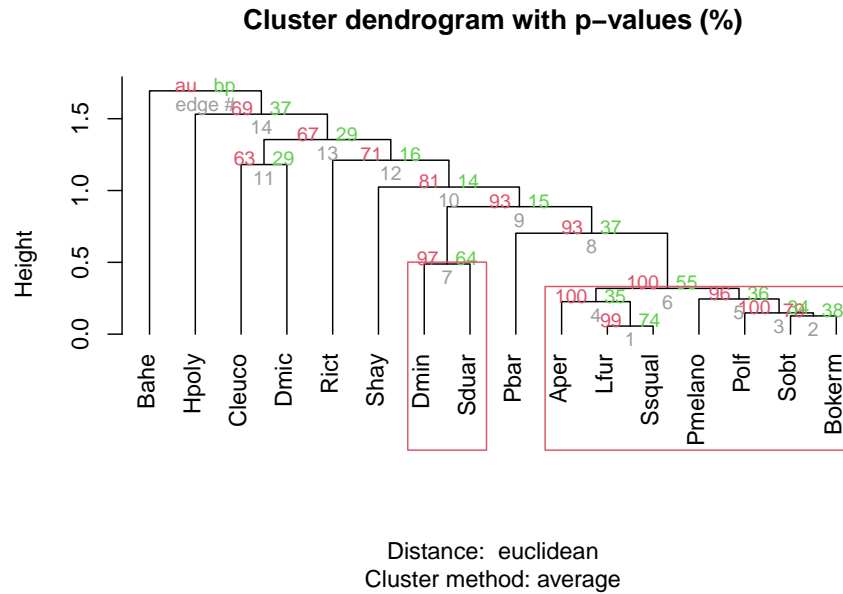


É possível notar que existe um único grupo com BS > 95%. Agora vamos tentar usar a distância de Hellinger:

```
bocaina_transf2 <- disttransform(bocaina, "hellinger")
analise2 <- pvclust(bocaina_transf2, method.hclust="average", method.dist="euclidean")
```

```
## Bootstrap (r = 0.46)... Done.
## Bootstrap (r = 0.54)... Done.
## Bootstrap (r = 0.69)... Done.
## Bootstrap (r = 0.77)... Done.
## Bootstrap (r = 0.85)... Done.
## Bootstrap (r = 1.0)... Done.
## Bootstrap (r = 1.08)... Done.
## Bootstrap (r = 1.15)... Done.
## Bootstrap (r = 1.23)... Done.
## Bootstrap (r = 1.38)... Done.
```

```
plot(analise2, hang=-1)
pvrect(analise2)
```



Notem que se mudarmos o coeficiente de associação, o resultado também muda. Agora temos 1 grupo a mais, composto por *Dendropsophus minutus* e *Scinax duartei* que não apareciam antes. Isso se deve ao fato de que a distância de Hellinger dá menos peso para espécies raras do que a Chord.

## Chapter 5

# K-means e agrupamentos não-hierárquicos

### 5.1 Backgorund da análise

K-means é um tipo de agrupamento não hierárquico porque não busca obter grupos menores que por sua vez pertencem a grupos maiores. Resumidamente, podemos calcular o K-means apartir de uma matriz quadrada ou de distância. Essa técnica procura particionar os objetos em  $k$  grupos de maneira a minimizar a soma de quadrados entre grupos e maximizá-la dentro dos grupos. Um critério similar ao de uma ANOVA.

### 5.2 Exemplo 1:

#### Pergunta:

Qual é o número de grupos que melhor sumariza o padrão de ocorrência de espécies de peixes ao longo de um riacho?

#### Predições

- 1: O agrupamento ideal para explicar a variância no padrão de ocorrência de espécies é 4.

#### Variáveis

- Variáveis resposta
  - 1. Para este exemplo iremos utilizar um conjunto de dados disponível no pacote `ade4` que contém dados de 27 espécies de peixes coletados em 30 pontos ao longo do Rio Doubs, na fronteira entre a França e Suíça.

### 5.2.1 Explicação da análise

Um diferencial do K-means em relação aos agrupamentos hierarquicos (=clusters) é que o usuário pode escolher antecipadamente o número de grupos que quer formar.

#### Checklist

- Vamos normalizar os dados de abundância antes de entrar na análise propriamente, já que existem muitos zeros na matriz.

### 5.2.2 Análise

```
library(ade4)
data(doubs)
head(doubs$fish)
```

```
##      Cogo Satr Phph Neba Thth Teso Chna Chto Lele Lece Baba Spbi Gogo Eslu Pefl
## 1      0   3   0   0   0   0   0   0   0   0   0   0   0   0   0   0
## 2      0   5   4   3   0   0   0   0   0   0   0   0   0   0   0   0
## 3      0   5   5   5   0   0   0   0   0   0   0   0   0   0   1   0
## 4      0   4   5   5   0   0   0   0   0   0   1   0   0   1   2   2
## 5      0   2   3   2   0   0   0   0   0   5   2   0   0   2   4   4
## 6      0   3   4   5   0   0   0   0   0   1   2   0   0   1   1   1
##      Rham Legi Scer Cyca Titi Abbr Icme Acce Ruru Blbj Alal Anan
## 1      0   0   0   0   0   0   0   0   0   0   0   0   0
## 2      0   0   0   0   0   0   0   0   0   0   0   0   0
## 3      0   0   0   0   0   0   0   0   0   0   0   0   0
## 4      0   0   0   0   1   0   0   0   0   0   0   0   0
## 5      0   0   2   0   3   0   0   0   0   5   0   0   0
## 6      0   0   0   0   2   0   0   0   0   1   0   0   0
```

```
spe <- doubs$fish[-8,]# retiro a linha 8, pois não há dados
```

```
spe.norm <- decostand(spe, "normalize") # função do pacote vegan, ela faz várias padronizações
```

O argumento `centers` na função abaixo indica o número de grupos que se quer formar. Neste exemplo estamos utilizando `centers=4`.

```
spe.kmeans <- kmeans(spe.norm, centers=4, nstart=100)
spe.kmeans
```

```
## K-means clustering with 4 clusters of sizes 12, 3, 6, 8
```

```
##
```

```
## Cluster means:
```

```
##           Cogo           Satr           Phph           Neba           Thth           Teso
## 1 0.10380209 0.542300691 0.50086515 0.43325916 0.114024105 0.075651573
## 2 0.00000000 0.000000000 0.00000000 0.00000000 0.000000000 0.000000000
## 3 0.06167791 0.122088022 0.26993915 0.35942538 0.032664966 0.135403325
```

```
## 4 0.00000000 0.006691097 0.02506109 0.06987391 0.006691097 0.006691097
##      Chna      Chto      Lele      Lece      Baba      Spbi      Gogo
## 1 0.00000000 0.00000000 0.06983991 0.1237394 0.02385019 0.00000000 0.05670453
## 2 0.05205792 0.00000000 0.07647191 0.3166705 0.00000000 0.00000000 0.20500174
## 3 0.06212775 0.21568957 0.25887226 0.2722562 0.15647062 0.1574388 0.16822286
## 4 0.10687104 0.09377516 0.14194394 0.2011411 0.24327992 0.1326062 0.28386032
##      Eslu      Pefl      Rham      Legi      Scer      Cyca      Titi
## 1 0.04722294 0.02949244 0.00000000 0.00000000 0.00000000 0.00000000 0.03833408
## 2 0.07647191 0.00000000 0.00000000 0.05205792 0.07647191 0.00000000 0.00000000
## 3 0.12276089 0.17261621 0.0793181 0.06190283 0.04516042 0.06190283 0.14539027
## 4 0.20630360 0.16920496 0.2214275 0.19066542 0.13171275 0.16019126 0.26230024
##      Abbr      Icme      Acce      Ruru      Blbj      Alal      Anan
## 1 0.00000000 0.00000000 0.00000000 0.01049901 0.00000000 0.00000000 0.00000000
## 2 0.00000000 0.00000000 0.18058775 0.31667052 0.05205792 0.7618709 0.00000000
## 3 0.01473139 0.00000000 0.03192175 0.32201597 0.01473139 0.1095241 0.04739636
## 4 0.19561641 0.1331835 0.26713081 0.32103755 0.22883055 0.3326939 0.18873077
##
## Clustering vector:
## 1 2 3 4 5 6 7 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
## 1 1 1 1 3 1 1 3 1 1 1 1 1 1 3 3 3 3 4 4 4 2 2 2 4 4
## 28 29 30
## 4 4 4
##
## Within cluster sum of squares by cluster:
## [1] 2.5101386 0.3560423 1.7361453 0.4696535
## (between_SS / total_SS = 66.7 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

O objeto que fornece o resultado contém: 1) o tamanho (número de objetos) em cada um dos 4 grupos; 2) o centroid de cada grupo e o pertencimento de cada espécie a cada grupo; e 3) o quando da Soma de Quadrados dos dados é explicada por esta conformação de grupos.

No entanto, não é possível saber a priori qual o número *ideal* de grupos. Para descobrir isso repetimos o k-means com uma série de valores de **K**. Isso pode ser feito na função `cascadeKM`.

```
spe.KM.cascade <- cascadeKM(spe.norm, inf.gr=2, sup.gr=10, iter=100, criterion="ssi")
```

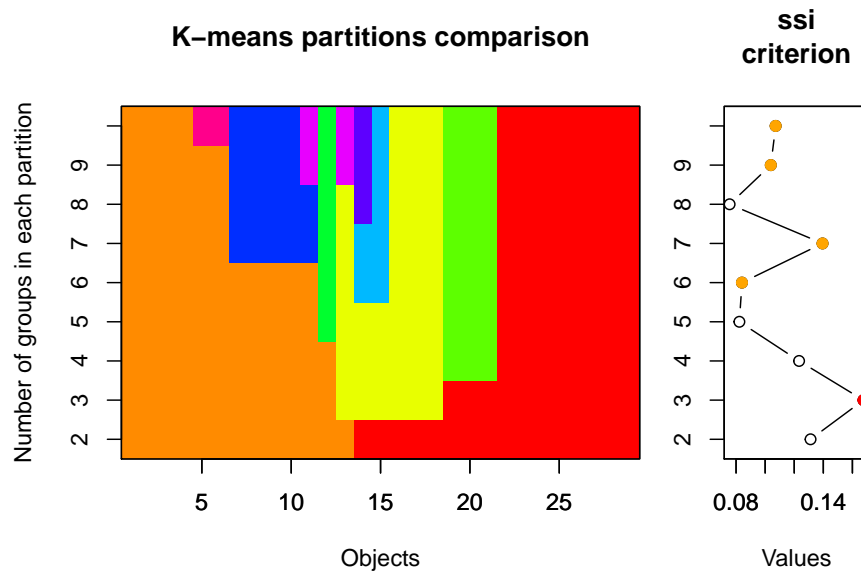
Tanto `calinski` quando `ssi` são bons critérios para encontrar o número ideal de grupos. Quanto maior o valor de `ssi` melhor (veja `?cascadeKM` mais detalhes).

```
# summary
spe.KM.cascade$results
```

```
##      2 groups 3 groups 4 groups 5 groups 6 groups 7 groups 8 groups
## SSE 8.2149405 6.4768108 5.0719796 4.30155732 3.58561200 2.9523667 2.48405487
## ssi 0.1312111 0.1675852 0.1233266 0.08226682 0.08405525 0.1394601 0.07555728
##      9 groups 10 groups
## SSE 2.0521888 1.7599292
## ssi 0.1039046 0.1072322
```

SSE: critério utilizado pelo algoritmo para achar o agrupamento ótimo dos objetos.

```
plot(spe.KM.cascade, sortg=TRUE)
```



Este resultado nos mostra que o número ideal de grupos é 3, vejam que o SSI máximo é alcançado neste número de grupos 0.1675852 (também indicado pela bola vermelha no plot).

#Espécies indicadoras

### 5.3 Backgorund da análise

Uma pergunta normalmente feita por ecólogos é: qual espécie pode ser indicadora de uma determinada condição ambiental?

O índice IndVal mede dois aspectos das espécies: Especificidade e fidelidade. Uma alta fidelidade significa que espécies ocorrem em todos os locais do grupo e uma alta especificidade significa que as espécies ocorrem somente naquele grupo. Uma boa espécie indicadora é aquela na qual todos os indivíduos ocorrem em

todas as amostras referentes a um grupo específico. A Especificidade é dada pela divisão da abundância média da espécie no grupo pela somatória das abundâncias médias dos grupos. Fidelidade é igual ao número de lugares no grupo onde a espécie está presente dividido pelo número total de lugares do grupo (Dufrêne & Legendre, 1997).

Espécies raras podem receber o mesmo valor de IndVal das espécies indicadoras e são chamadas de indicadoras assimétricas, i.e., contribuem com a especificidade do habitat mas não servem para prever grupos. Ao contrário, as espécies indicadoras são verdadeiros indicadores simétricos e podem ser usadas para prever grupos.

## 5.4 Exemplo 1:

### Pergunta:

Qual espécie de anfíbio anuro na fase larval pode ser indicadora da fitofisionomia onde é encontrada?

### Predições

- 1: Espécies terrestres serão indicadoras de área aberta, enquanto espécies arborícolas serão indicadoras de áreas florestais.

### Variáveis

- Variáveis resposta
  - 1. Mesma matriz já utilizada contendo a abundância de girinos ao longo de poças na Serra da Bocaina.

### 5.4.1 Explicação da análise

A análise procede da seguinte forma:

- 1. Uma matriz de distância é construída e as unidades amostrais são classificadas com alguma análise de agrupamento, hierárquico ou não;
- 2. A variável ambiental para a qual se deseja classificar os grupos é inserida;
- 3. As espécies indicadoras de cada grupo são formadas através do cálculo da especificidade e fidelidade, obtendo-se o valor de IndVal para cada espécie;
- 4. Por fim, o conjunto de dados originais é comparado para ver se a análise faz sentido.

O cálculo da significância do índice de IndVal é feito por aleatorização de Monte Carlo. Assim, o valor do índice é aleatorizado 999 vezes (ou o número de vezes que você optar) dentro dos tratamentos e o valor de  $P$  é dado pelo número de vezes em que o índice observado foi igual ou maior que os valores aleatorizados.

### 5.4.2 Análise

O IndVal está disponível tanto no pacote `indicspecies` quando no `labdsv`. Para este exemplo iremos usar o `labdsv`.

```
library(labdsv)
```

Primeiro vamos agrupar as unidades amostrais (poças) que informa os grupos de fitofisionomias onde as poças se localizam e para os quais deseja-se encontrar espécies indicadoras:

```
fitofis <- c(rep(1,4), rep(2,4), rep(3,4), rep(4,4), rep(5,4))

resultado <- indval(bocaina, fitofis)
summary(resultado) #só exibe o resultado para as espécies indicadoras

##          cluster indicator_value probability
## Rict          1          0.8364          0.008
## Sduar          1          0.7475          0.041
##
## Sum of probabilities                = 8.068
##
## Sum of Indicator Values              = 7.3
##
## Sum of Significant Indicator Values  = 1.58
##
## Number of Significant Indicators     = 2
##
## Significant Indicator Distribution
##
## 1
## 2
```

Para apresentar uma tabela dos resultados para todas as espécies temos de processar os dados:

```
resultado$maxcls

##      Aper      Bahe      Rict      Cleuco      Dmic      Dmin      Hpoly      Lfur      Pbar      Polf
##         3         2         1         3         3         1         2         3         1         3
## Pmelano      Sduar      Shay      Sibt      Ssqual      Bokerm
##         2         1         3         2         3         2

resultado$indcls

##      Aper      Bahe      Rict      Cleuco      Dmic      Dmin      Hpoly      Lfur
## 0.2432796 0.6487329 0.8363823 0.4128631 0.6645244 0.7032145 0.6208711 0.2279412
##      Pbar      Polf      Pmelano      Sduar      Shay      Sibt      Ssqual      Bokerm
## 0.2813725 0.2437500 0.2500000 0.7474527 0.4930269 0.2222222 0.2500000 0.4583333
```



```
resultado$pval
```

```
##      Aper      Bahe      Rict      Cleuco      Dmic      Dmin      Hpoly      Lfur      Pbar      Polf
##      1.000      0.053      0.008      0.428      0.206      0.097      0.273      1.000      0.608      1.000
## Pmelano      Sduar      Shay      Sibt      Ssqual      Bokerm
##      1.000      0.041      0.399      0.721      1.000      0.234
```

```
tab.resultado=cbind(resultado$maxcls,resultado$indcls,resultado$pval)
colnames(tab.resultado)<-c("maxgrp", "ind. value", "P")
tab.resultado
```

```
##      maxgrp ind. value      P
## Aper      3  0.2432796 1.000
## Bahe      2  0.6487329 0.053
## Rict      1  0.8363823 0.008
## Cleuco     3  0.4128631 0.428
## Dmic       3  0.6645244 0.206
## Dmin       1  0.7032145 0.097
## Hpoly      2  0.6208711 0.273
## Lfur       3  0.2279412 1.000
## Pbar       1  0.2813725 0.608
## Polf       3  0.2437500 1.000
## Pmelano    2  0.2500000 1.000
## Sduar      1  0.7474527 0.041
## Shay       3  0.4930269 0.399
## Sibt       2  0.2222222 0.721
## Ssqual     3  0.2500000 1.000
## Bokerm     2  0.4583333 0.234
```

No resultado podemos ver que temos duas espécies indicadoras da fitofisionomia 1: *Rhinella icterica* (Rict) e *Scinax duartei* (Sduar). Nenhuma espécie foi indicadora dos outros grupos neste exemplo.

### 5.4.3 Para se aprofundar

- Agrupamento de espécies e locais baseado em modelos
- Numerical Ecology with R



## Chapter 6

# Rarefação

### 6.1 Background da análise

Uma das dificuldades na comparação da riqueza de espécies entre comunidades é decorrente da diferença no esforço amostral (e.g. diferença no número de indivíduos, discrepância na quantidade de unidades amostrais ou área amostrada) que inevitavelmente influenciará no número de espécies observadas (Gotelli & Chao 2013). O método de rarefação nos permite comparar o número de espécies entre comunidades quando o tamanho da amostra ou a abundância de indivíduos não são iguais. A rarefação calcula o número esperado de espécies em cada comunidade tendo como base comparativa um valor em que todas as amostras atinjam um tamanho padrão, ou comparações baseadas na comunidade com menor número de amostragens ou com menos indivíduos. O teste foi formulado considerando seguinte pergunta: Se considerarmos  $n$  indivíduos ou amostras ( $n < N$ ) para cada comunidade, quantas espécies registraríamos nas comunidades considerando o mesmo número de indivíduos ou amostras?

$$E(S) = \sum 1 - \frac{(N - N_1)/n}{N/n}$$

Onde:

- $E(S)$  = Número de espécies esperado,
- $N$  = Número total de indivíduos na amostra,
- $N_i$  = Número de indivíduos da  $i$ ésima espécie,
- $n$  = tamanho da amostra padronizada (menor amostra).

Gotelli & Collwel (2001) descrevem este método e discutem em detalhes as restrições sobre seu uso na ecologia:

- As amostras a serem comparados devem ser consistentes do ponto de vista taxonômico, ou seja, todos os indivíduos devem pertencer ao mesmo grupo taxonômico;
- As comparações devem ser realizadas somente entre amostras com as mesmas técnicas de coleta;
- Os tipos de hábitat onde as amostras são obtidas devem ser semelhantes;
- É um método para estimar a riqueza de espécies em uma amostra menor – não pode ser usado para extrapolar e estimar riqueza.

Contudo, é importante ressaltar que esta última restrição foi superada por Colwell et al. (2012) e Chao & Jost (2012) que desenvolveram uma nova abordagem onde os dados podem ser interpolados (rarefeito) para amostras menores e extrapolados para amostras maiores.

## 6.2 Exemplo prático 1 - Morcegos

### 6.2.1 Explicação

#### Explicação dos dados

Neste exemplo usaremos os dados de espécies de morcegos amostradas em três fragmentos florestais (Breviglieri 2008): i) Mata Ciliar do Córrego Talhadinho com 12 hectares inserida em uma matriz de pastagem; ii) Mata Ciliar do Córrego dos Tenentes com 10 hectares inserida em uma matriz de cultivo de cana-de-açúcar e pastagem; e iii) Fazenda Experimental de Pindorama com 128 hectares inserida uma matriz de cana-de-açúcar e pastagem.

#### Pergunta:

A riqueza de espécies de morcegos é maior na Fazenda Experimental do que nos fragmentos florestais menores?

#### Predições

O número de espécies será maior em fragmentos florestais maiores.

#### Variáveis

- Variáveis preditoras
  - matriz ou dataframe com as abundâncias das espécies de morcegos registradas nos três fragmentos florestais

#### Checklist

- Verificar se a sua matriz ou dataframe estão com as espécies nas linhas e os fragmentos florestais nas colunas

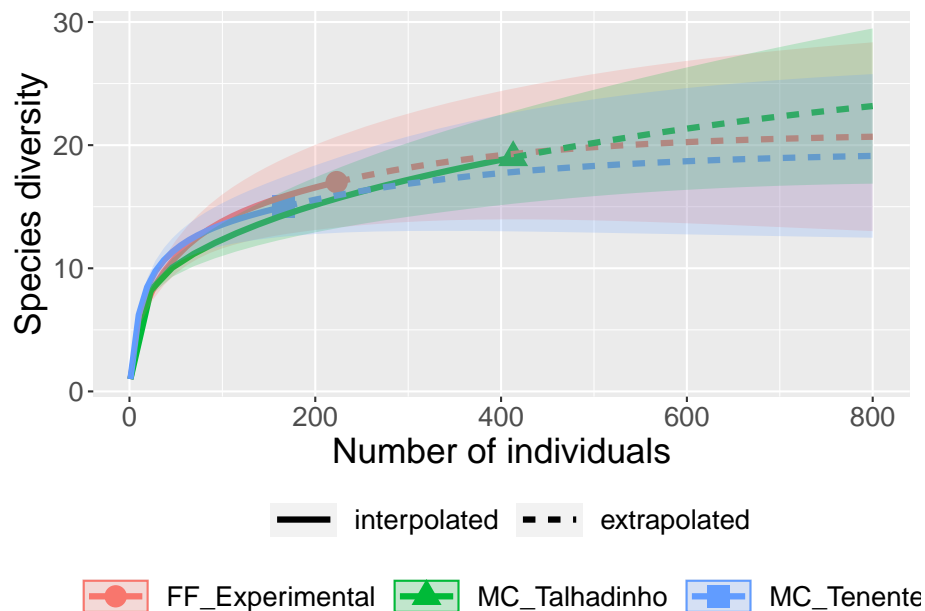
### 6.2.2 Análise

Calculo da rarefação

```
library(iNEXT)
library(devtools)
devtools::install_github("paternogbc/ecodados")
library(ecodados)

dados_rarefacao <- rarefacao_morcegos
resultados_morcegos <- iNEXT(dados_rarefacao, q = 0, datatype = "abundance", endpoint = 800)
# q refere-se a família *Hill-numbers* (Hill 1973) onde 0 = riqueza de espécies, 1 = diversidade
# Veja mais detalhes sobre os números de Hill no Capítulo 7 onde tratamos de extrapolações.
# datatype refere-se ao tipo de dados que você vai analisar (e.g. abundância, incidência).
# endpoint refere-se ao valor de referência que você determina para a extrapolação.

# Visualizar os resultados
ggiNEXT(resultados_morcegos, type = 1)
```



### 6.2.3 Interpretação dos resultados

Neste exemplo, foram registrados 166 indivíduos na MC\_Tenentes, 413 na MC\_Talhadinho e 223 na FF\_Experimental. Lembrando, você não pode comparar a riqueza de espécies observada diretamente: 15 espécies na MC\_Tenentes, 17 espécies na MC\_Talhadinho, e 13 espécies na FF\_Experimental. A comparação da riqueza de espécies entre as comunidades

deve ser feita com base na riqueza de espécies estimada que é calculada com base no número de indivíduos da comunidade com menor abundância (166 indivíduos). Olhando o gráfico é possível perceber que a riqueza de espécies de morcegos estimada não é diferente entre os três fragmentos florestais quando corrigimos o problema da abundância pela rarefação. A interpretação é feita com base no intervalo de confiança de 95%. As curvas serão diferentes quando os intervalos de confiança não se sobreporem (Chao et al. 2014). Percebam que está abordagem, além da interpolação (rarefação), também realiza extrapolações que podem ser usadas para estimar o número de espécies caso o esforço de coleta fosse maior. Este é o assunto do nosso próximo capítulo.

## 6.3 Exemplo prático 2 - Rarefação

### 6.3.1 Explicação

#### Explicação dos dados

Neste exemplo iremos comparar o número de espécies de anuros e répteis (serpentes e lagartos) usando informações dos indivíduos depositados em coleções científicas e coletas de campo (da Silva et al. 2017).

#### Pergunta:

A riqueza de espécies de anuros e répteis é maior em coleções científicas do que nas coletas de campo?

#### Predições

O número de espécies será maior em coleções científicas devido ao maior esforço amostral (i.e. maior variação temporal para depositar os indivíduos e maior número de pessoas contribuindo com as informações de diferentes estudos e/ou coletas esporádicas).

#### Variáveis

- Variáveis preditoras
  - matriz ou dataframe com as abundâncias das espécies de anuros e répteis (planilhas separadas) registradas em coleções científicas e coletas de campo.

#### Checklist

- Verificar se a sua matriz ou dataframe estão com as espécies nas linhas e a fonte dos dados nas colunas.

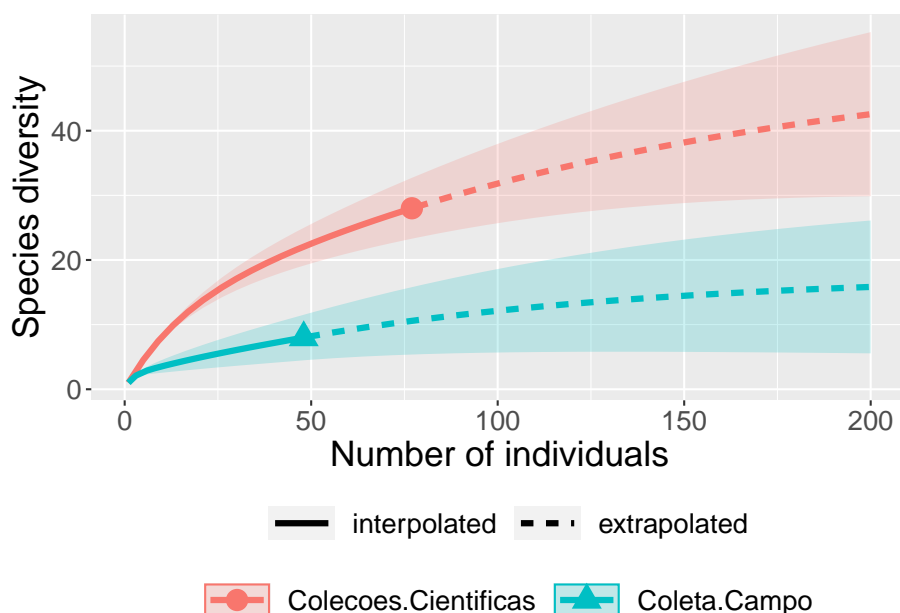
### 6.3.2 Análise

Calculo da rarefação para os dados de répteis

```
library(iNEXT)

rarefacao_repteis <- rarefacao_repteis
resultados_repteis <- iNEXT(rarefacao_repteis, q = 0, datatype = "abundance", endpoint = 200)

# Visualizar os resultados
ggiNEXT(resultados_repteis, type = 1)
```



### 6.3.3 Interpretação dos resultados

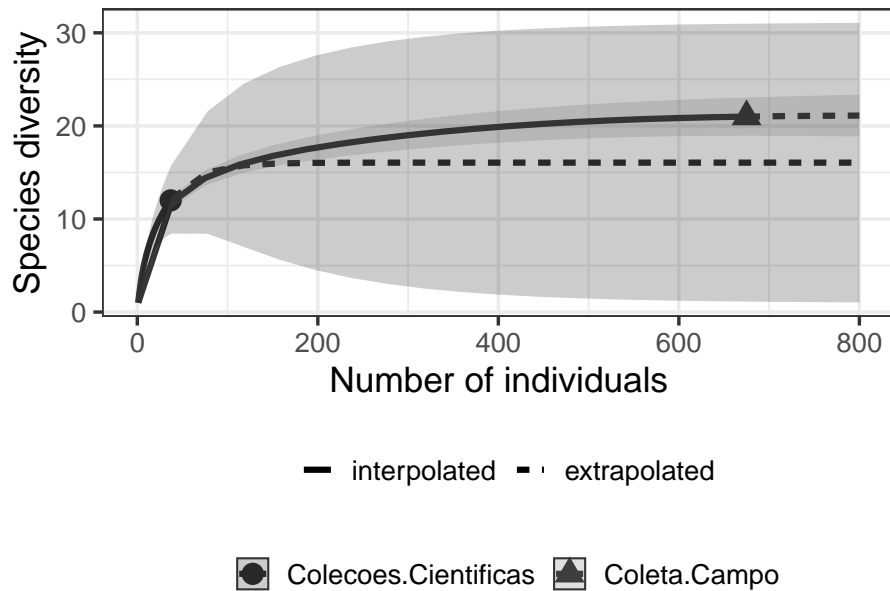
Neste exemplo, foram registradas oito espécies de répteis nas coletas de campo (40 indivíduos) e 28 espécies nas coleções científicas (77 indivíduos). Com base na rarefação, concluímos que a riqueza de espécies de répteis obtida nas coleções científicas é 2,5 vezes maior do que a obtida em coletas de campo.

Calculo da rarefação para os dados dos anuros

```
library(iNEXT)

rarefacao_anuros <- rarefacao_anuros
resultados_anuros <- iNEXT(rarefacao_anuros, q = 0, datatype = "abundance", endpoint = 800)

# Visualizar os resultados
ggiNEXT(resultados_anuros, type = 1, grey = TRUE)
```



#### Interpretação dos resultados

Neste exemplo, foram registradas 21 espécies de anuros nas coletas de campo (709 indivíduos) e 12 espécies nas coleções científicas (37 indivíduos). Com base na rarefação, concluímos que não há diferença entre a riqueza de espécies de anuros obtida em coletas de campo e coleções científicas.

### 6.4 Para se aprofundar

- Recomendamos aos interessados que olhem a página do EstimateS software e baixem o manual do usuário que contém informações detalhadas sobre os índices de rarefação. Este site foi criado e é mantido pelo Dr. Robert K. Colwell, um dos maiores especialistas do mundo em estimativas da biodiversidade
- Recomendamos também o livro Magurran & McGill (2010) - Biological Diversity Frontiers in Measurement and Assessment.



## Chapter 7

# Estimadores de Riqueza

### 7.1 Backgorund da análise

Uma vez que determinar o número total de espécies numa área é praticamente impossível, principalmente em regiões com alta riqueza de espécies, os estimadores são úteis para extrapolar a riqueza observada e tentar estimar a riqueza total através de uma amostra incompleta de uma comunidade biológica (Walther & Moore 2005). Neste capítulo serão considerados os estimadores não paramétricos que usam informações da frequência de espécies raras na comunidade (Gotelli & Chao 2013). Isto porque tanto os testes paramétricos que tentam determinar os parâmetros de uma curva usando o formato da curva de acumulação de espécies (e.g. equação logística, Michaelis-Menten) quanto os testes que usam a frequência do número de indivíduos para enquadrá-las em uma das distribuições de abundância das espécies (e.g. distribuições log-séries, log-normal) não funcionam muito bem com dados empíricos (Gotelli & Chao 2013). Para mais detalhes sobre os testes paramétricos veja Magurran (2004) e Colwell (2019).

#### **7.1.0.1 Quatro características para um bom estimador de riqueza (Chazdon et al. 1998; Horter et al. 2006):**

- Independência do tamanho da amostra (quantidade de esforço amostral realizado);
- Insensibilidade a diferentes padrões de distribuições (diferentes equitabilidades);
- Insensibilidade em relação à ordem das amostragens;
- Insensibilidade à heterogeneidade entre as amostras usadas entre estudos.

## 7.2 Estimadores baseados na abundância das espécies

### 7.2.1 CHAO 1 - (Chao 1984, 1987):

Estimador simples do número absoluto de espécies em uma comunidade. É baseado no número de espécies raras dentro de uma amostra.

$$Chao_1 = S_{obs} + \left( \frac{n-1}{n} \right) \frac{F_1(F_1-1)}{2(F_2+1)}$$

onde:

- Sobs = o número de espécies na comunidade,
- $n$  = número de amostras,
- $F_1$  = número de espécies observadas com abundância de um indivíduo (espécies *singleton*),
- $F_2$  = número de espécies observadas com abundância de dois indivíduos (espécies *doubletons*).

O valor de Chao 1 é máximo quando todas as espécies menos uma são únicas (*singleton*). Neste caso, a riqueza estimada é aproximadamente o dobro da riqueza observada.

#### 7.2.1.1 Exemplo prático - Chao 1

**7.2.1.1.1 Explicação dos dados** Neste exemplo usaremos os dados de 17 espécies de anuros amostradas em 14 dias de coletas de campo em um habitat reprodutivo localizado na região noroeste do estado de São Paulo, Brasil.

#### Pergunta:

Quantas espécies a mais poderiam ser amostradas caso aumentássemos o esforço amostral?

#### Predições

- O número de espécies estimadas é similar ao número de espécies observada;
- O número de espécies estimadas é maior do que o número de espécies observada.

#### Variáveis

- Variáveis preditoras
  - matriz ou vetor com as abundâncias das espécies de anuros registradas em um habitat reprodutivo

### Checklist

- Verificar se a sua matriz está com as espécies nas colunas e as amostragens nas linhas
- Verificar se os dados são de abundância e não presença e ausência

### 7.2.2 Análise

Calculo do estimador de riqueza - Chao 1

```
library(ecodados)
library(vegan)
dados_coleta <- poca_anuros
est_chao1 <- estaccumR(dados_coleta, permutations = 100)
summary(est_chao1, display = "chao")
```

```
## $chao
##      N      Chao  2.5%   97.5% Std.Dev
## Dia_10  1  6.939167  3.00 12.33333 2.703161
## Dia_2   2  9.476262  4.95 14.64375 2.674539
## Dia_12  3 11.140333  7.00 17.05000 2.607404
## Dia_7   4 12.385833  9.00 19.00000 2.531193
## Dia_9   5 13.020000  9.00 18.00000 2.381187
## Dia_14  6 13.738333  9.00 20.00000 2.510277
## Dia_4   7 14.491667 11.00 22.00000 2.515538
## Dia_6   8 15.288333 12.00 22.00000 2.540011
## Dia_5   9 16.256667 12.00 22.00000 2.601413
## Dia_13 10 16.718333 12.00 22.00000 2.384562
## Dia_1   11 17.608333 13.00 22.00000 2.367401
## Dia_8   12 18.335000 14.50 22.00000 2.128516
## Dia_11  13 19.160000 15.50 22.00000 1.689062
## Dia_3   14 20.000000 20.00 20.00000 0.000000
##
## attr(,"class")
## [1] "summary.poolaccum"
```

Visualizar os resultados com intervalo de confiança de 95%.

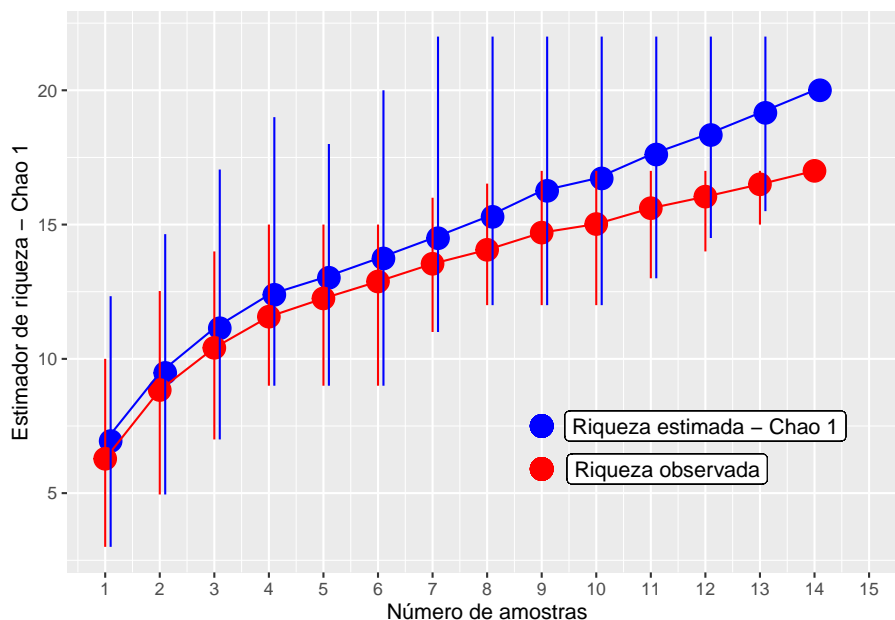
```
library(ggplot2)
# preparando os dados para fazer o gráfico
resultados <- summary(est_chao1, display = c("S", "chao"))
res_chao <- cbind(resultados$chao[,1:4], resultados$S[,2:4])
res_chao <- as.data.frame(res_chao)
colnames(res_chao) <- c("Amostras", "Chao", "C_inferior", "C_superior", "Riqueza",
                        "R_inferior", "R_superior")

# comando para o gráfico
ggplot(res_chao, aes(y = Riqueza, x = Amostras)) +
```

```

geom_point(aes(y = Chao, x = Amostras + 0.1), size = 5, color = "blue", alpha = 1) +
geom_point(aes(y = Riqueza, x = Amostras), size = 5, color = "red", alpha = 1) +
geom_line(aes(y = Chao, x = Amostras), color = "blue") +
geom_line(aes(y = Riqueza, x = Amostras), color = "red") +
geom_linerange(aes(ymin = C_inferior, ymax = C_superior, x = Amostras + 0.1),
color = "blue") +
geom_linerange(aes(ymin = R_inferior, ymax = R_superior, x = Amostras), color = "red") +
ylab("Estimador de riqueza - Chao 1") +
xlab("Número de amostras") +
scale_x_continuous(limits = c(1,15), breaks=seq(1,15,1)) +
geom_point(y= 7.5, x = 9, size = 5, color = "blue", alpha = 1) +
geom_point(y= 5.9, x = 9, size = 5, color = "red", alpha = 1) +
geom_label(y = 7.5, x = 12, label = "Riqueza estimada - Chao 1") +
geom_label(y = 5.9, x = 11.3, label = "Riqueza observada")

```



### 7.2.2.1 Interpretação dos resultados

Com base no número de espécies raras (*singletons* e *doubletons*), o estimador Chao 1 indica a possibilidade de encontrarmos mais três espécies caso o esforço amostral fosse maior e não mostra tendência de estabilização da curva em uma assíntota.

### 7.2.3 ACE - *Abundance-based Coverage Estimator* (Chao & Lee 1992, Chao et al. 2000):

Este método trabalha com a abundância das espécies raras (i.e. abundância baixa). Entretanto, diferente do estimador anterior, esse método permite ao pesquisador determinar os limites para os quais uma espécie seja considerada rara. Em geral, são consideradas raras espécies com abundância entre 1 e 10 indivíduos. A riqueza estimada pode variar conforme se aumente ou diminua o limiar de abundância, e infelizmente não existem critérios biológicos definidos para a escolha do melhor intervalo.

$$ACE = S_{abund} + \frac{S_{rare}}{C_{ace}} + \frac{F_1}{C_{ace}} Y_{ace}^2$$

onde:

$$Y_{ace}^2 = \max \left[ \frac{S_{rare}}{C_{ace}} \frac{\sum_{i=1}^{10} i(i-1)F_i}{(N_{rare})(N_{rare}-1)} - 1, 0 \right]$$

$$C_{ace} = 1 - \frac{F_1}{N_{rare}}$$

$$N_{rare} = \sum_{i=1}^{10} iF_i$$

Não precisa fazer cara feia, é óbvio que iremos usar o programa para fazer esses cálculos.

#### 7.2.3.1 Exemplo prático - ACE

**7.2.3.1.1 Explicação dos dados** Usaremos os mesmos dados de 17 espécies de anuros amostradas em 14 dias de coletas de campo em um habitat reprodutivo localizado na região noroeste do estado de São Paulo, Brasil.

#### Pergunta:

Quantas espécies a mais poderiam ser amostradas caso aumentássemos o esforço amostral?

#### Predições

- O número de espécies estimadas é similar ao número de espécies observada;
- O número de espécies estimadas é maior do que o número de espécies observada.

#### Variáveis

- Variáveis preditoras
  - matriz ou vetor com as abundâncias das espécies de anuros registradas em um habitat reprodutivo

### Checklist

- Verificar se a sua matriz está com as espécies nas colunas e as amostragens nas linhas
- Verificar se os dados são de abundância e não presença e ausência

### 7.2.4 Análise

Calculo do estimador de riqueza - ACE

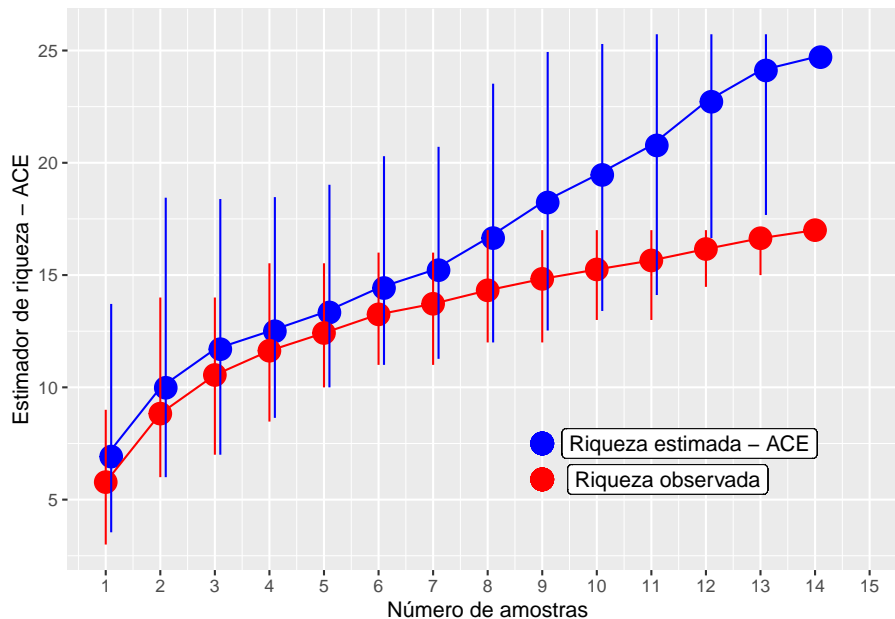
```
library(vegan)
dados_coleta <- poca_anuros
est_ace <- estaccumR(dados_coleta, permutations = 100)
summary(est_ace, display = "ace")
```

```
## $ace
##      N      ACE      2.5%      97.5% Std.Dev
## Dia_12 1  6.915877  3.545190 13.71429 2.702247
## Dia_3   2  9.980312  6.000000 18.44439 3.427675
## Dia_2   3 11.698589  7.000000 18.38713 3.152887
## Dia_7   4 12.497742  8.641454 18.47101 2.563305
## Dia_8   5 13.334706 10.000000 19.02155 2.382534
## Dia_13  6 14.422381 11.000000 20.29373 2.385637
## Dia_9   7 15.217600 11.269677 20.71163 2.490499
## Dia_5   8 16.646320 12.000000 23.52215 3.207545
## Dia_6   9 18.230771 12.531112 24.93470 3.649519
## Dia_4  10 19.459136 13.400722 25.28994 3.846804
## Dia_14 11 20.773557 14.111345 25.72368 3.894644
## Dia_10 12 22.716886 16.647490 25.72368 3.170156
## Dia_11 13 24.111305 17.676471 25.72368 2.080821
## Dia_1  14 24.703704 24.703704 24.70370 0.000000
##
## attr(,"class")
## [1] "summary.poolaccum"
```

Visualizar os resultados com intervalo de confiança de 95%

```
library(ggplot2)
# preparando os dados para fazer o gráfico
resultados_ace <- summary(est_ace, display = c("S", "ace"))
res_ace <- cbind(resultados_ace$ace[,1:4], resultados_ace$S[,2:4])
res_ace <- as.data.frame(res_ace)
colnames(res_ace) <- c("Amostras", "ACE", "ACE_inferior", "ACE_superior", "Riqueza",
                      "R_inferior", "R_superior")
```

```
# comando para o gráfico
ggplot(res_ace, aes(y = Riqueza, x = Amostras)) +
  geom_point(aes(y = ACE, x = Amostras + 0.1), size = 5, color = "blue", alpha = 1) +
  geom_point(aes(y = Riqueza, x = Amostras), size = 5, color = "red", alpha = 1) +
  geom_line(aes(y = ACE, x = Amostras), color = "blue") +
  geom_line(aes(y = Riqueza, x = Amostras), color = "red") +
  geom_linerange(aes(ymin = ACE_inferior, ymax = ACE_superior, x = Amostras + 0.1),
  color = "blue") +
  geom_linerange(aes(ymin = R_inferior, ymax = R_superior, x = Amostras), color = "red") +
  ylab("Estimador de riqueza - ACE") +
  xlab("Número de amostras") +
  scale_x_continuous(limits = c(1,15), breaks=seq(1,15,1)) +
  geom_point(y = 7.5, x = 9, size = 5, color = "blue", alpha = 1) +
  geom_point(y = 5.9, x = 9, size = 5, color = "red", alpha = 1) +
  geom_label(y = 7.5, x = 11.7, label = "Riqueza estimada - ACE") +
  geom_label(y = 5.9, x = 11.3, label = "Riqueza observada")
```



#### 7.2.4.1 Interpretação dos resultados

Com base no número de espécies raras (abundância menor que 10 indivíduos - *default*), o estimador ACE indica a possibilidade de encontrarmos mais sete espécies caso o esforço amostral fosse maior e não mostrou tendência de estabilização da curva em uma assíntota.

## 7.3 Estimadores baseados na incidência das espécies

### 7.3.1 CHAO 2 - (Chao 1987):

De acordo com Anne Chao, o estimador Chao 1 pode ser modificado para uso com dados de presença/ausência levando em conta a distribuição das espécies entre amostras. Neste caso é necessário conhecer o número de espécies encontradas em somente uma amostra e o número de espécies encontradas exatamente em duas amostras. Essa variação ficou denominada como Chao 2:

$$Chao_2 = S_{obs} + \left( \frac{m-1}{m} \right) \left( \frac{Q_1(Q_1-1)}{2(Q_2+1)} \right)$$

onde:

- $S_{obs}$  = o número de espécies na comunidade,
- $m$  = número de amostragens,
- $Q_1$  = número de espécies observadas em uma amostragem (espécies *uniques*),
- $Q_2$  = número de espécies observadas em duas amostragens (espécies *duplicates*).

O valor de Chao2 é máximo quando as espécies menos uma são únicas (*uniques*). Neste caso, a riqueza estimada é aproximadamente o dobro da riqueza observada. Colwell & Coddington (1994) encontraram que o valor de Chao 2 mostrou ser o estimador menos enviesado para amostras com tamanho pequeno.

#### 7.3.1.1 Exemplo prático - Chao 2

**7.3.1.1.1 Explicação dos dados** Usaremos os mesmos dados de 17 espécies de anuros amostradas em 14 dias de coletas de campo em um habitat reprodutivo localizado na região noroeste do estado de São Paulo, Brasil.

##### Pergunta:

Quantas espécies a mais poderiam ser amostradas caso aumentássemos o esforço amostral?

##### Predições

- O número de espécies estimadas é similar ao número de espécies observada;
- O número de espécies estimadas é maior do que o número de espécies observada.



**Variáveis**

- Variáveis preditoras
  - matriz ou vetor com a incidência das espécies de anuros registradas em um habitat reprodutivo

**Checklist**

- Verificar se a sua matriz está com as espécies nas colunas e as amostragens nas linhas

**7.3.2 Análise**

Calculo do estimador de riqueza - Chao 2

```
library(vegan)
dados_coleta <- poca_anuros
est_chao2 <- poolaccum(dados_coleta, permutations = 100)
summary(est_chao2, display = "chao")
```

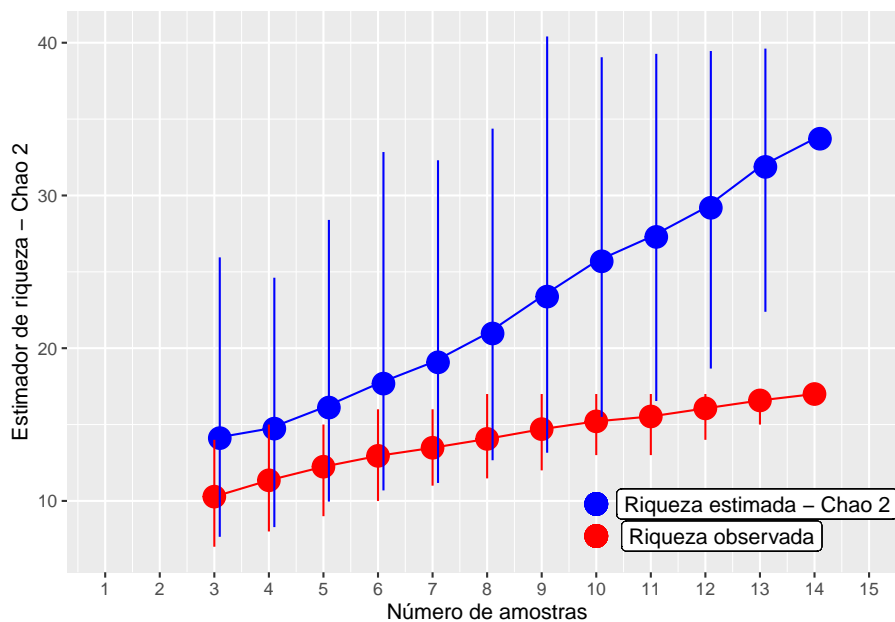
```
## $chao
##      N      Chao      2.5%    97.5% Std.Dev
## [1,]  3 14.11136  7.650000 25.94583 5.915790
## [2,]  4 14.73408  8.286719 24.61406 4.222528
## [3,]  5 16.11380  9.958333 28.40000 4.600280
## [4,]  6 17.68049 10.687500 32.84375 5.850734
## [5,]  7 19.07486 11.176786 32.30357 5.917198
## [6,]  8 20.96375 12.662500 34.37500 6.075325
## [7,]  9 23.37778 13.155556 40.41111 7.066263
## [8,] 10 25.68825 15.500000 39.05000 6.750631
## [9,] 11 27.28682 16.542045 39.27273 6.478553
## [10,] 12 29.19354 18.666667 39.45833 6.340934
## [11,] 13 31.87077 22.384615 39.61538 5.152470
## [12,] 14 33.71429 33.714286 33.71429 0.000000
##
## attr(,"class")
## [1] "summary.poolaccum"
```

Visualizar os resultados com intervalo de confiança de 95%

```
library(ggplot2)
# preparando os dados para fazer o gráfico
resultados_chao2 <- summary(est_chao2, display = c("S", "chao"))
res_chao2 <- cbind(resultados_chao2$chao[,1:4], resultados_chao2$S[,2:4])
res_chao2 <- as.data.frame(res_chao2)
colnames(res_chao2) <- c("Amostras", "Chao2", "C_inferior", "C_superior", "Riqueza",
                        "R_inferior", "R_superior")

# comando para o gráfico
```

```
ggplot(res_chao2, aes(y = Riqueza, x = Amostras)) +
  geom_point(aes(y = Chao2, x = Amostras + 0.1), size = 5, color = "blue", alpha = 1) +
  geom_point(aes(y = Riqueza, x = Amostras), size = 5, color = "red", alpha = 1) +
  geom_line(aes(y = Chao2, x = Amostras), color = "blue") +
  geom_line(aes(y = Riqueza, x = Amostras), color = "red") +
  geom_linerange(aes(ymin = C_inferior, ymax = C_superior, x = Amostras + 0.1),
    color = "blue") +
  geom_linerange(aes(ymin = R_inferior, ymax = R_superior, x = Amostras), color = "red") +
  ylab("Estimador de riqueza - Chao 2") +
  xlab("Número de amostras") +
  scale_x_continuous(limits = c(1,15), breaks=seq(1,15,1)) +
  geom_point(y= 9.8, x = 10, size = 5, color = "blue", alpha = 1) +
  geom_point(y= 7.7, x = 10, size = 5, color = "red", alpha = 1) +
  geom_label(y = 9.8, x = 12.95, label = "Riqueza estimada - Chao 2") +
  geom_label(y = 7.7, x = 12.3, label = "Riqueza observada")
```



### 7.3.2.1 Interpretação dos resultados

Com base no número de espécies raras (*uniques* e *duplicates*), Chao 2 estimou a possibilidade de encontrarmos mais dezesseis espécies caso o esforço amostral fosse maior e não mostrou tendência de estabilização da curva em uma assíntota.

**7.3.3 JACKKNIFE 1 (Burnham & Overton 1978, 1979):**

Este estimador baseia-se no número de espécies que ocorrem em somente uma amostra (Q1).

$$S_{jack1} = S_{obs} + Q1 \left( \frac{m-1}{m} \right)$$

onde:

- Sobs = o número de espécies na comunidade,
- Q1 = número de espécies observadas em uma amostragem (espécies *uniques*),
- $m$  = número de amostragens.

Palmer (1990) verificou que Jackknife 1 foi o estimador mais preciso e menos enviesado comparado a outros métodos de extrapolação.

**7.3.3.1 Exemplo prático - Jackknife 1**

**7.3.3.1.1 Explicação dos dados** Usaremos os mesmos dados de 17 espécies de anuros amostradas em 14 dias de coletas de campo em um habitat reprodutivo localizado na região noroeste do estado de São Paulo, Brasil.

**Pergunta:**

Quantas espécies a mais poderiam ser amostradas caso aumentássemos o esforço amostral?

**Predições**

- O número de espécies estimadas é similar ao número de espécies observada;
- O número de espécies estimadas é maior do que o número de espécies observada.

**Variáveis**

- Variáveis preditoras
  - matriz ou vetor com as abundâncias das espécies de anuros registradas em um habitat reprodutivo

**Checklist**

- Verificar se a sua matriz está com as espécies nas colunas e as amostragens nas linhas

### 7.3.4 Análise

Calculo do estimador de riqueza - Jackknife 1

```
library(vegan)
dados_coleta <- poca_anuros
est_jack1 <- poolaccum(dados_coleta, permutations = 100)
summary(est_jack1, display = "jack1")
```

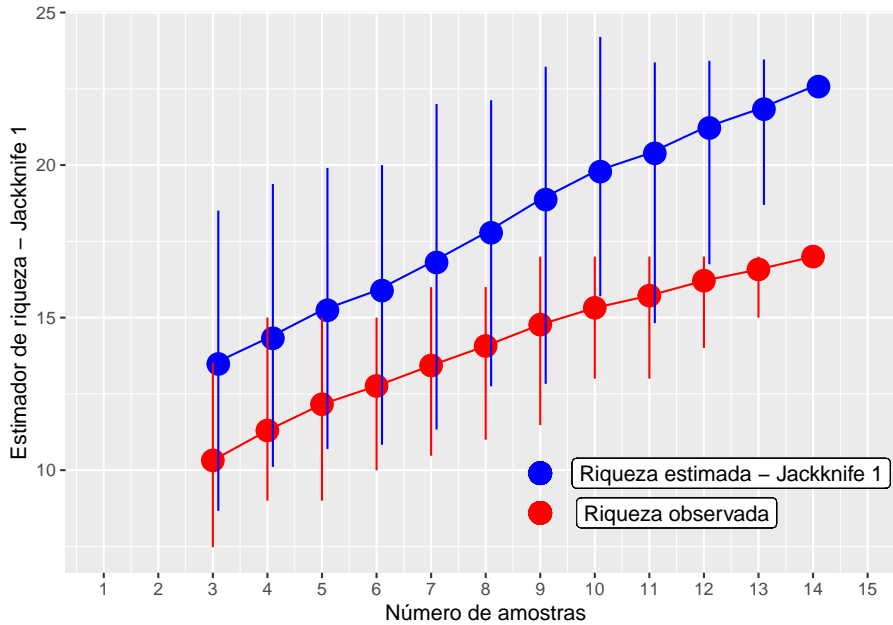
```
## $jack1
##          N Jackknife 1      2.5%    97.5% Std.Dev
## [1,]  3    13.48000  8.666667 18.50833 2.655837
## [2,]  4    14.32250 10.106250 19.38125 2.610269
## [3,]  5    15.24000 10.695000 19.90500 2.603184
## [4,]  6    15.88500 10.833333 20.00000 2.466577
## [5,]  7    16.80714 11.332143 22.00000 2.514799
## [6,]  8    17.78000 12.750000 22.12500 2.450979
## [7,]  9    18.86778 12.830556 23.22222 2.535922
## [8,] 10    19.78400 15.700000 24.20000 2.468866
## [9,] 11    20.38364 14.818182 23.36364 2.447600
## [10,] 12    21.21500 16.750000 23.41667 2.012035
## [11,] 13    21.83231 18.692308 23.46154 1.449584
## [12,] 14    22.57143 22.571429 22.57143 0.000000
##
## attr(,"class")
## [1] "summary.poolaccum"
```

Visualizar os resultados com 95% intervalo de confiança

```
library(ggplot2)
# preparando os dados para fazer o gráfico
resultados_jack1 <- summary(est_jack1, display = c("S", "jack1"))
res_jack1 <- cbind(resultados_jack1$jack1[,1:4], resultados_jack1$S[,2:4])
res_jack1 <- as.data.frame(res_jack1)
colnames(res_jack1) <- c("Amostras", "JACK1", "JACK1_inferior", "JACK1_superior", "Riqueza",
                        "R_inferior", "R_superior")

# comando para o gráfico
ggplot(res_jack1, aes(y = Riqueza, x = Amostras)) +
  geom_point(aes(y = JACK1, x = Amostras + 0.1), size = 5, color = "blue", alpha = 1) +
  geom_point(aes(y = Riqueza, x = Amostras), size = 5, color = "red", alpha = 1) +
  geom_line(aes(y = JACK1, x = Amostras), color = "blue") +
  geom_line(aes(y = Riqueza, x = Amostras), color = "red") +
  geom_linerange(aes(ymin = JACK1_inferior, ymax = JACK1_superior, x = Amostras + 0.1),
  color = "blue") +
  geom_linerange(aes(ymin = R_inferior, ymax = R_superior, x = Amostras), color = "red") +
  ylab("Estimador de riqueza - Jackknife 1") +
  xlab("Número de amostras") +
```

```
scale_x_continuous(limits = c(1,15), breaks=seq(1,15,1)) +
geom_point(y= 9.9, x = 9, size = 5, color = "blue", alpha = 1) +
geom_point(y= 8.6, x = 9, size = 5, color = "red", alpha = 1) +
geom_label( y = 9.9, x = 12.5, label = "Riqueza estimada - Jackknife 1") +
geom_label( y = 8.6, x = 11.5, label = "Riqueza observada")
```



#### 7.3.4.1 Interpretação dos resultados

Com base no número de espécies raras, o estimador Jackknife 1 calculou a possibilidade de encontrarmos mais seis espécies caso o esforço amostral fosse maior e não mostrou tendência de estabilização da curva em uma assíntota.

#### 7.3.5 JACKKNIFE 2 (Burnham & Overton 1978, 1979, Palmer 1991):

Este método basea-se no número de espécies que ocorrem em apenas uma amostra e no número de espécies que ocorrem em exatamente duas amostras.

$$S_{jack2} = S_{obs} + \left[ \frac{Q_1(2m-3)}{m} - \frac{Q_2(m-2)^2}{m(m-1)} \right]$$

onde:

- $S_{obs}$  = o número de espécies na comunidade,

- $m$  = número de amostragens,
- $Q1$  = número de espécies observadas em uma amostragem (espécies *uniques*),
- $Q2$  = número de espécies observadas em duas amostragens (espécies *duplicates*).

### 7.3.5.1 Exemplo prático - Jackknife 2

**7.3.5.1.1 Explicação dos dados** Usaremos os mesmos dados de 17 espécies de anuros amostradas em 14 dias de coletas de campo em um habitat reprodutivo localizado na região noroeste do estado de São Paulo, Brasil.

#### Pergunta:

Quantas espécies a mais poderiam ser amostradas caso aumentássemos o esforço amostral?

#### Predições

- O número de espécies estimadas é similar ao número de espécies observada;
- O número de espécies estimadas é maior do que o número de espécies observada.

#### Variáveis

- Variáveis preditoras
  - matriz ou vetor com as abundâncias das espécies de anuros registradas em um habitat reprodutivo

#### Checklist

- Verificar se a sua matriz está com as espécies nas colunas e as amostragens nas linhas

### 7.3.6 Análise

Cálculo do estimador de riqueza - Jackknife 2

```
library(vegan)
dados_coleta <- poca_anuros
est_jack2 <- poolaccum(dados_coleta, permutations = 100)
summary(est_jack2, display = "jack2")
```

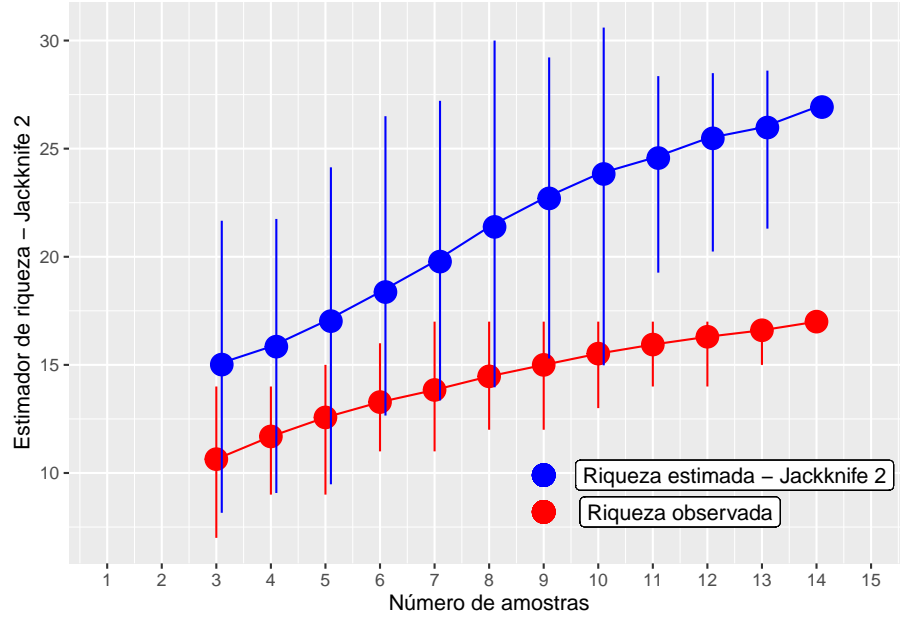
```
## $jack2
##      N Jackknife 2      2.5%    97.5% Std.Dev
## [1,]  3    15.01333  8.158333 21.66667 3.163225
## [2,]  4    15.84333  9.075000 21.75000 3.396421
```

```
## [3,] 5      17.02300  9.477500 24.13625 3.658051
## [4,] 6      18.36800 12.656667 26.50083 3.464244
## [5,] 7      19.77690 13.357143 27.21250 3.918379
## [6,] 8      21.38143 13.964286 30.00000 4.095742
## [7,] 9      22.69972 15.295486 29.21840 3.854432
## [8,] 10     23.82878 14.977778 30.60000 3.742583
## [9,] 11     24.56573 19.265909 28.35455 3.114132
## [10,] 12     25.47720 20.242424 28.49242 2.716484
## [11,] 13     25.97635 21.301282 28.60897 2.100889
## [12,] 14     26.92308 26.923077 26.92308 0.000000
##
## attr(,"class")
## [1] "summary.poolaccum"
```

Visualizar os resultados com intervalo de confiança de 95%

```
library(ggplot2)
# preparando os dados para fazer o gráfico
resultados_jack2 <- summary(est_jack2, display = c("S", "jack2"))
res_jack2 <- cbind(resultados_jack2$jack2[,1:4], resultados_jack2$S[,2:4])
res_jack2 <- as.data.frame(res_jack2)
colnames(res_jack2) <- c("Amostras", "JACK2", "JACK2_inferior", "JACK2_superior", "Riqueza",
                        "R_inferior", "R_superior")

# comando para o gráfico
ggplot(res_jack2, aes(y = Riqueza, x = Amostras)) +
  geom_point(aes(y = JACK2, x = Amostras + 0.1), size = 5, color = "blue", alpha = 1) +
  geom_point(aes(y = Riqueza, x = Amostras), size = 5, color = "red", alpha = 1) +
  geom_line(aes(y = JACK2, x = Amostras), color = "blue") +
  geom_line(aes(y = Riqueza, x = Amostras), color = "red") +
  geom_linerange(aes(ymin = JACK2_inferior, ymax = JACK2_superior, x = Amostras + 0.1),
                color = "blue") +
  geom_linerange(aes(ymin = R_inferior, ymax = R_superior, x = Amostras), color = "red") +
  ylab("Estimador de riqueza - Jackknife 2") +
  xlab("Número de amostras") +
  scale_x_continuous(limits = c(1,15), breaks=seq(1,15,1)) +
  geom_point(y= 9.9, x = 9, size = 5, color = "blue", alpha = 1) +
  geom_point(y= 8.2, x = 9, size = 5, color = "red", alpha = 1) +
  geom_label(y = 9.9, x = 12.5, label = "Riqueza estimada - Jackknife 2") +
  geom_label(y = 8.2, x = 11.5, label = "Riqueza observada")
```



### 7.3.6.1 Interpretação dos resultados

Com base no número de espécies raras, o estimador Jackknife 2 calculou a possibilidade de encontrarmos mais dez espécies caso o esforço amostral fosse maior e não mostrou tendência estabilização da curva em uma assíntota.

### 7.3.7 BOOTSTRAP (Smith & van Belle 1984):

Este método difere dos demais por utilizar dados de todas as espécies coletadas para estimar a riqueza total, não se restringindo às espécies raras. Ele requer somente dados de incidência. A estimativa pelo bootstrap é calculada somando-se a riqueza observada à soma do inverso da proporção de amostras em que cada espécie ocorre.

$$S_{boot} = S_{obs} + \sum_{k=1}^{S_{obs}} (1 - P_k)^m$$

onde:

- $S_{obs}$  = o número de espécies na comunidade,
- $m$  = número de amostragens,
- $P_k$  = proporção do número de amostras em que cada espécie foi registrada.



### 7.3.7.1 Exemplo prático - Bootstrap

**7.3.7.1.1 Explicação dos dados** Usaremos os mesmos dados de 17 espécies de anuros amostradas em 14 dias de coletas de campo em um habitat reprodutivo localizado na região noroeste do estado de São Paulo, Brasil.

#### Pergunta:

Quantas espécies a mais poderiam ser amostradas caso aumentássemos o esforço amostral?

#### Predições

- O número de espécies estimadas é similar ao número de espécies observada;
- O número de espécies estimadas é maior do que o número de espécies observada.

#### Variáveis

- Variáveis preditoras
  - matriz ou vetor com as abundâncias das espécies de anuros registradas em um habitat reprodutivo

#### Checklist

- Verificar se a sua matriz está com as espécies nas colunas e as amostragens nas linhas

## 7.3.8 Análise

Calculo do estimador de riqueza - Bootstrap

```
library(vegan)
dados_coleta <- poca_anuros
est_boot <- poolaccum(dados_coleta, permutations = 100)
summary(est_boot, display = "boot")
```

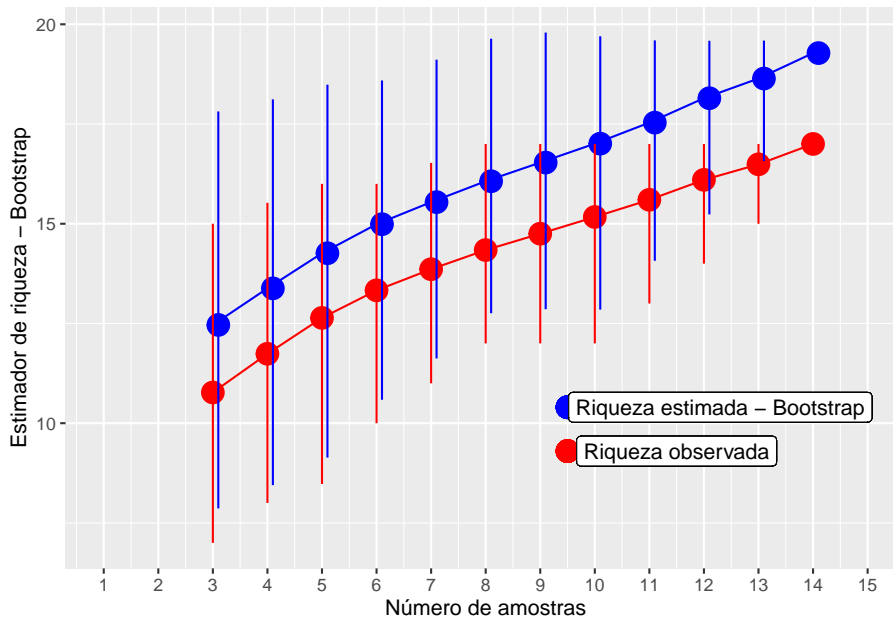
```
## $boot
##      N Bootstrap      2.5%    97.5%   Std.Dev
## [1,]  3  12.46296   7.863889 17.81481  2.5418338
## [2,]  4  13.38020   8.447266 18.11807  2.5853800
## [3,]  5  14.26049   9.137768 18.48586  2.3931326
## [4,]  6  14.98800  10.587626 18.59184  2.3787503
## [5,]  7  15.54129  11.622499 19.11226  2.0485404
## [6,]  8  16.06517  12.756676 19.63881  2.0302798
## [7,]  9  16.53307  12.857386 19.79244  2.0746699
## [8,] 10  17.00749  12.847702 19.69951  1.9199967
## [9,] 11  17.53435  14.071729 19.59628  1.5488150
## [10,] 12  18.14692  15.233594 19.58721  1.2796709
## [11,] 13  18.63976  16.570376 19.59107  0.9825201
```

```
## [12,] 14 19.27832 19.278321 19.27832 0.0000000
##
## attr(,"class")
## [1] "summary.poolaccum"
```

Visualizar os resultados com intervalo de confiança de 95%

```
library(ggplot2)
# preparando os dados para fazer o gráfico
resultados_boot <- summary(est_boot, display = c("S", "boot"))
res_boot <- cbind(resultados_boot$boot[,1:4], resultados_boot$S[,2:4])
res_boot <- as.data.frame(res_boot)
colnames(res_boot) <- c("Amostras", "BOOT", "BOOT_inferior", "BOOT_superior", "Riqueza",
                        "R_inferior", "R_superior")

# comando para o gráfico
ggplot(res_boot, aes(y = Riqueza, x = Amostras)) +
  geom_point(aes(y = BOOT, x = Amostras + 0.1), size = 5, color = "blue", alpha = 1) +
  geom_point(aes(y = Riqueza, x = Amostras), size = 5, color = "red", alpha = 1) +
  geom_line(aes(y = BOOT, x = Amostras), color = "blue") +
  geom_line(aes(y = Riqueza, x = Amostras), color = "red") +
  geom_linerange(aes(ymin = BOOT_inferior, ymax = BOOT_superior, x = Amostras + 0.1),
color = "blue") +
  geom_linerange(aes(ymin = R_inferior, ymax = R_superior, x = Amostras), color = "red") +
  ylab("Estimador de riqueza - Bootstrap") +
  xlab("Número de amostras") +
  scale_x_continuous(limits = c(1,15), breaks=seq(1,15,1)) +
  geom_point(y= 10.4, x = 9.5, size = 5, color = "blue", alpha = 1) +
  geom_point(y= 9.3, x = 9.5, size = 5, color = "red", alpha = 1) +
  geom_label(y = 10.4, x = 12.3, label = "Riqueza estimada - Bootstrap") +
  geom_label(y = 9.3, x = 11.5, label = "Riqueza observada")
```



#### 7.3.8.1 Interpretação dos resultados

Com base na frequência de ocorrência das espécies, o estimador bootstrap calculou a possibilidade de encontrarmos mais duas espécies caso o esforço amostral fosse maior e não mostrou tendência de estabilização da curva em uma assíntota.

### 7.3.9 Interpolação e Extrapolação baseadas em rarefação usando amostragens de incidência ou abundância (Chao & Jost 2012, Colwell et al. 2012):

Este método utiliza teoria de amostragem (e.g. modelos multinomial, Poisson e Bernoulli) para conectar rarefação (interpolação) e predição (extrapolação) com base no tamanho da amostra. Contudo, é importante enfatizar que a extrapolação torna-se altamente incerta quando estendida para o dobro do tamanho da amostragem. Este método utiliza uma abordagem com bootstrap para calcular o intervalo de confiança de 95%. Uma das vantagens desta abordagem é que ela permite além da riqueza de espécies, interpolar e extrapolar os índices de diversidade de Shannon entropy (i.e. quantifica a incerteza da identidade da espécie baseado na amostragem aleatória de um indivíduo da comunidade) e Gini-Simpson (i.e. quantifica a probabilidade que dois indivíduos escolhidos aleatoriamente sejam de diferentes espécies). Contudo, estes índices de diversidades são transformados e apresentados em unidades de riqueza de espécies (*Números de Hill*, Hill 1973). Hill percebeu que poderíamos calcular a riqueza de espécies máxima usando os índices de Shannon entropy e Gini-Simpson quando consideramos que todas as espécies na comunidade possuem abundâncias idên-

ticas (máxima equitabilidade). Então, eles propôs a transformação dos índices de diversidade determinando qual seria o número de espécies equivalente da nossa comunidade (observado) se todas as espécies fossem igualmente abundantes (teórico). Desta maneira, os índices podem ser comparáveis pois estão representados pela mesma unidade - riqueza de espécies. Os números de Hill são representados pelo parâmetro  $q$  que controla a sensibilidade do índice em relação a abundância relativa das espécies (Gotelli & Chao 2013). Quando  $q = 0$  ele não considera a abundância das espécies e é igual a riqueza de espécies. Quando  $q = 1$  ele é o exponencial da diversidade de Shannon, que dá uma peso maior para as espécies raras. Quando  $q = 2$  ele é a diversidade de Simpson, que dá um peso maior para as espécies mais comuns na comunidade.

### 7.3.9.1 Exemplo prático

**7.3.9.1.1 Explicação dos dados** Usaremos os mesmos dados de 17 espécies de anuros amostradas em 14 dias de coletas de campo em um habitat reprodutivo localizado na região noroeste do estado de São Paulo, Brasil.

#### Pergunta:

Quantas espécies a mais poderiam ser amostradas caso aumentássemos o esforço amostral?

#### Predições

- O número de espécies estimadas é similar ao número de espécies observada;
- O número de espécies estimadas é maior do que o número de espécies observada.

#### Variáveis

- Variáveis preditoras
  - matriz ou vetor com as abundâncias das espécies de anuros registradas em um habitat reprodutivo

#### Checklist

- Verificar se a sua matriz está com as espécies nas colunas e as amostragens nas linhas.

### 7.3.10 Análise

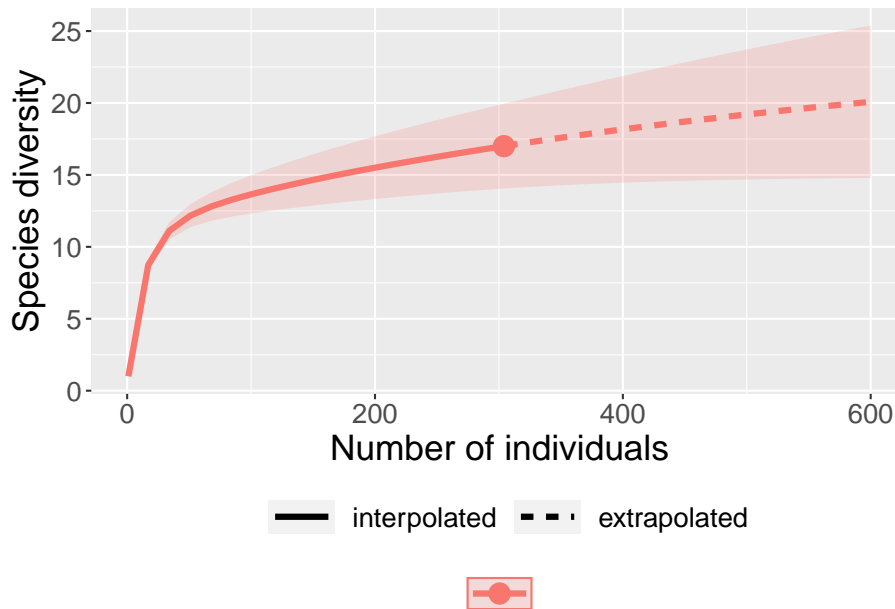
Calculo da extrapolação da riqueza com base no número de indivíduos

```
library(iNEXT)
dados_coleta <- poca_anuros

# preparando os dados para análises considerando a abundância
dados_inext_abu <- colSums(dados_coleta)
```

```
resultados_abundancia <- iNEXT(dados_inext_abu, q = 0, datatype = "abundance",
                              endpoint = 600)

# Visualizar os dados no gráfico
ggiNEXT(resultados_abundancia, type = 1)
```



#### 7.3.10.1 Interpretação dos resultados

Veja que o ponto no final da linha contínua representa as 17 espécies de anuros (eixo Y) observadas entre os 304 indivíduos (eixo X). A extrapolação máxima (600 indivíduos no nosso exemplo), estima um aumento de até oito espécies (intervalo de confiança) caso amostrássemos mais 300 indivíduos.

Calculo da extrapolação da riqueza com base no número de amostras

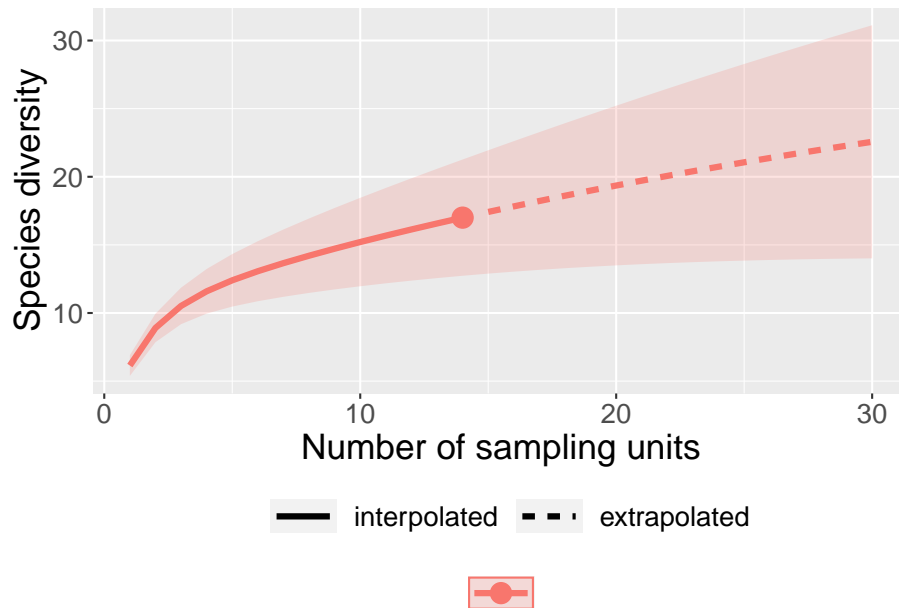
```
library(iNEXT)
dados_coleta <- poca_anuros

# preparando os dados para análises considerando a incidência
dados_inext <- as.incfreq(t(dados_coleta)) # preciso transpor o dataframe

resultados_incidencia <- iNEXT(dados_inext, q = 0, datatype = "incidence_freq",
                              endpoint = 30)

# Visualizar os dados no gráfico
```

```
ggiNEXT(resultados_incidencia, type = 1)
```



### 7.3.10.2 Interpretação dos resultados

Veja que o ponto no final da linha contínua representa as 17 espécies de anuros (eixo Y) observadas nos 14 dias de coleta (eixo X - amostras). A extrapolação máxima (30 dias de coleta no nosso exemplo), estima um aumento de até 13 espécies (intervalo de confiança) caso amostrássemos mais 16 dias.

### 7.3.11 Para se aprofundar

- Recomendamos aos interessados que olhem a página do EstimateS software e baixem o manual do usuário que contém informações detalhadas sobre os índices de rarefação e estimadores de riqueza. Este site foi criado e é mantido pelo Dr. Robert K. Colwell, um dos maiores especialistas do mundo em estimativas da biodiversidade
- Recomendamos também o livro Magurran & McGill (2010) - Biological Diversity Frontiers in Measurement and Assessment.

## Chapter 8

# Diversidade Funcional

### 8.1 2.1 Aspectos teóricos

Até a década de 1990, a teoria ecológica investigava basicamente quais processos determinavam a abundância e riqueza de espécies no espaço e tempo. Esta década foi marcada por intensos debates sobre as regras de montagem de comunidades e como interações e filtros ambientais determinavam a coexistência de espécies (Strong et al. 1984). Porém, a década 2000 foi marcada pelo uso mais explícito das características das espécies como uma variável fundamental tanto para explicar como as espécies seriam afetadas pelo ambiente, quanto para entender como tais espécies afetariam o ecossistema (Diaz & Cabido 2001, McGill et al. 2006). O primeiro estudo que utilizou o termo Diversidade Funcional foi publicado por Williams (1994), que comparou espécies de náuplios filogeneticamente relacionadas e demonstrou que elas possuem alta plasticidade funcional que favorecem ampla variação de comportamentos e, desse modo, permitem que sejam espécies generalistas em ambientes em constante mudança. A unidade básica desses estudos, o atributo funcional (do inglês “functional trait”), é definido como uma propriedade mensurável dos organismos (geralmente em nível individual) que represente características morfológicas, fisiológicas ou fenológicas que afetam a aptidão alterando aspectos do crescimento, reprodução e sobrevivência (Violle et al. 2007). Mais especificamente, o atributo funcional pode ser dividido em atributo efeito (i.e., atributos do organismo que afetam condições ambientais ou propriedades do ecossistema) e resposta (i.e., atributos do organismo que variam em resposta a condições ambientais) (Violle et al. 2007).

As medidas de diversidade, então, passam a ser representadas não somente por diferenças no número e na quantidade de espécies. Ao invés disso, as diferenças/semelhanças dos atributos funcionais das espécies dentro e entre localidades é usada como uma medida de diversidade. Desse modo, a variação no grau de expressão de diferentes atributos funcionais entre diferentes populações,

comunidades ou ecossistemas é definida como *Diversidade Funcional* (**sensu** Garnier 2015). Porém, a diversidade funcional não deve ser usada como medida única, uma vez que tais diferenças entre os atributos funcionais pode ser medida a partir da abundância relativa, riqueza e variação dos atributos funcionais. Desse modo, podemos dividir a diversidade funcional em três diferentes medidas: (1) riqueza funcional, (2) divergência funcional, e (3) regularidade funcional (Villéger et al. 2008). Existem dezenas de métricas que calculam cada uma dessas “dimensões” da diversidade funcional, mas se destacam aquelas baseadas em dendrograma (e.g., FD: Petchey & Gaston 2004) ou em medidas de distância (e.g., Villeger et al. 2008). Assim como a diversidade taxonômica (*Capítulo XXX*), a diversidade funcional pode ser medida em componentes alfa e beta. A seguir, apresentamos diferentes maneiras de calcular a distância entre localidades tendo como base os atributos funcionais das espécies e, além disso, demonstramos como calcular algumas das métricas de diversidade (alfa e beta) funcional mais usadas em ecologia. A parte final deste capítulo apresenta dois exemplos de como podemos testar hipóteses ecológicas comparando a diversidade funcional alfa e beta.

## 8.2 2.2. Definindo a dis(similaridade) entre espécies

Definir o quão diferente ou semelhante são duas espécies que ocorrem em um determinada localidade é a base para calcular a diversidade alfa e beta funcional. Para isso, é fundamental ter em mente que os atributos funcionais podem ser de vários tipos como, por exemplo, contínuo (e.g., tamanho corporal em cm), categórico (e.g., guilda: frugívoro, detritívoro, etc.), ordinal (e.g., 1 para organismo até 5 cm, 2 para organismos entre 5 e 30 cm, e 3 para organismos maiores do que 30 cm), binários (e.g., presença ou ausência de espinho), entre outros (veja Tabela 1 em Gonçalves-Souza et al. 2019). Por este motivo, a decisão do método de distância só será possível após o reconhecimento do tipo de cada um dos atributos funcionais escolhidos. Em linhas gerais, para variáveis contínuas a distância euclidiana é a melhor opção, enquanto para os outros tipos de variáveis ou para matrizes com mais de um tipo de variáveis, a distância de Gower deve ser usada (Pavoine et al. 2009).

```
library(FD)
library(ade4)
library(ecodados)
library(gridExtra)
library(ggplot2)
library(ggrepel)
library(tidyverse)
library(picante)
```



### Baixe todos esses dados, pois serão utilizados em outras partes deste capítulo

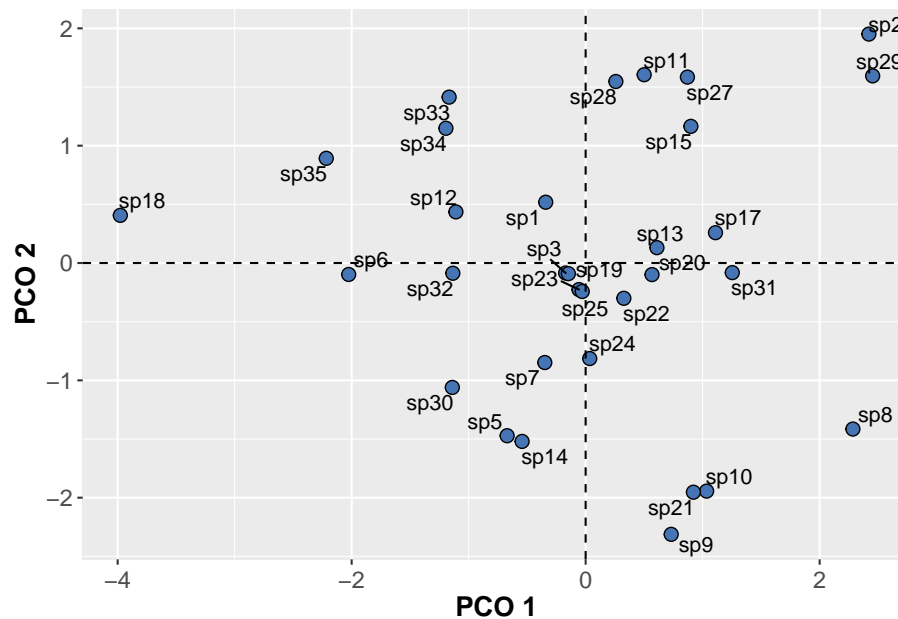
```
comun_fren_dat <- ecodados::fundiv_frenette2012a_comu
ambie_fren_dat <- ecodados::fundiv_frenette2012a_amb
trait_fren_dat <- ecodados::fundiv_frenette2012a_trait
trait_dat      <- ecodados::fundiv_barbaro2009a_trait
comun_dat      <- ecodados::fundiv_barbaro2009a_comu
ambie_dat      <- ecodados::fundiv_barbaro2009a_amb
```

# Medidas de distância com dados contínuos (quantitativos)

```
trait_cont <- trait_fren_dat
trait_pad <- decostand(trait_cont, "standardize")
euclid_dis <- vegdist(trait_pad, "euclidean")
```

```
pcoa_traits_cont <- pcoa(euclid_dis, correction="cailliez") # Resultados são idênticos ao de uma
eixos_cont <- as.data.frame(pcoa_traits_cont$vectors[,1:2]) # Selecionar os dois primeiros eixos
```

```
eixos_cont %>%
  ggplot(aes(x=Axis.1, y=Axis.2)) +
  geom_point(pch=21, size=3, color = "black", fill="#4575b4") +
  geom_text_repel(aes(Axis.1, Axis.2, label = rownames(eixos_cont))) +
  xlab("PCO 1") + ylab("PCO 2") +
  theme(axis.title.x = element_text(face="bold", size=14),
        axis.text.x = element_text(vjust=0.5, size=12)) +
  theme(axis.title.y = element_text(face="bold", size=14),
        axis.text.y = element_text(vjust=0.5, size=12)) +
  geom_hline(yintercept = 0, linetype=2) +
  geom_vline(xintercept = 0, linetype=2) +
  theme(legend.position = "top", legend.title=element_blank()) -> plot_trait_cont
plot_trait_cont
```



```
ggsave("trait_cont.pdf", plot_trait_cont, height = 14, width = 14, dpi = 600, units = "cm")

# Medidas de distância com dados categóricos

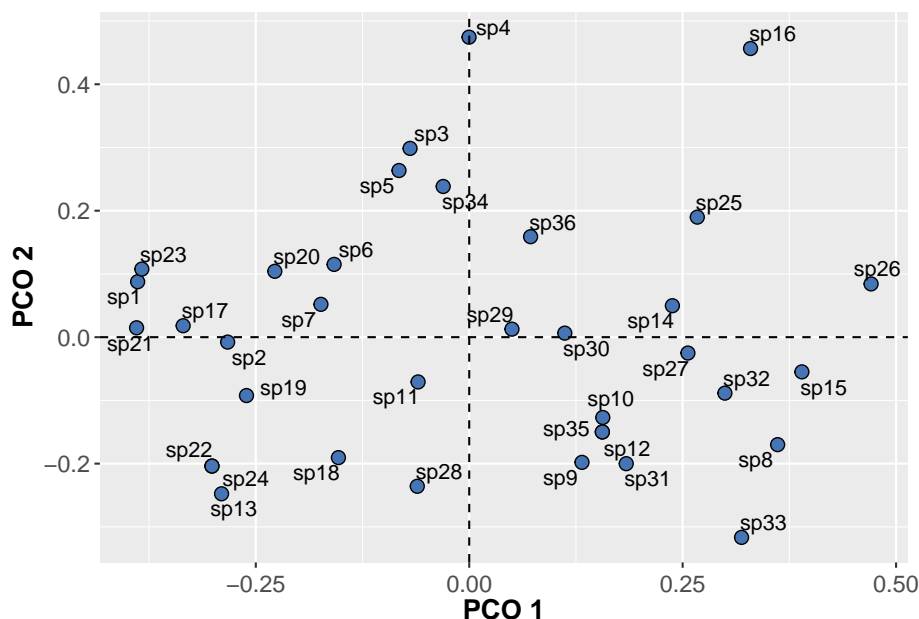
trait_dat %>%
  dplyr::select_if(is.character) -> trait_cat
dist_categ <- gowdis(trait_cat)

# Visualize os dados com uma PCoA

pcoa_traits_cat <- pcoa(dist_categ, correction="cailliez")
eixos_cat <- as.data.frame(pcoa_traits_cat$vectors[,1:2]) # Selecionar os dois primeiros eixos

eixos_cat %>%
  ggplot(aes(x=Axis.1, y=Axis.2)) +
  geom_point(pch=21, size=3, color = "black", fill="#4575b4") +
  geom_text_repel(aes(Axis.1, Axis.2, label = rownames(eixos_cat))) +
  xlab("PCO 1") + ylab("PCO 2") +
  theme(axis.title.x = element_text(face="bold", size=14),
        axis.text.x = element_text(vjust=0.5, size=12)) +
  theme(axis.title.y = element_text(face="bold", size=14),
        axis.text.y = element_text(vjust=0.5, size=12)) +
  geom_hline(yintercept = 0, linetype=2) +
  geom_vline(xintercept = 0, linetype=2) +
  theme(legend.position = "top", legend.title=element_blank()) -> plot_trait_cat
```

```
plot_trait_cat
```



```
ggsave("trait_cat.pdf", plot_trait_cat, height = 14, width = 14, dpi = 600, units = "cm")
```

```
# Medidas de distância com dados mistos: contínuos, categóricos, entre outros
# Primeiro, é necessário organizar cada tipo de dado com a função prep.TYPE
```

```
# Verifique a classe de todos os traits e veja se estão de acordo com sua expectativa
```

```
trait_dat %>%
  dplyr::summarise_all(class) %>%
  tidyr::gather(variable, class)
```

```
##   variable    class
## 1   trend character
## 2  redlist character
## 3   regio  integer
## 4    biog character
## 5  activ character
## 6    diet character
## 7  winter character
## 8   color character
## 9   breed character
## 10  body  integer
## 11  wing character
```

```
## 12    period character
```

```
# Atribuir classe ORDINAL para os traits ordinais, uma vez que foram reconhecidos como
```

```
trait_dat$regio <- as.ordered(trait_dat$regio)
```

```
trait_dat$body <- as.ordered(trait_dat$body)
```

```
## Combinar cada tipo de trait em um data.frame separado
```

```
# Categóricos
```

```
trait_categ <- cbind.data.frame(trend=trait_dat$trend, redlist=trait_dat$redlist, biogeog=trait_dat$biogeog)
```

```
# Ordinais
```

```
trait_ord <- cbind.data.frame(regio=trait_dat$regio, body=trait_dat$body)
```

```
rownames(trait_categ) <- rownames(trait_dat)
```

```
rownames(trait_ord) <- rownames(trait_dat)
```

```
# Agora, combinar os dois data.frames em uma lista chamada "ktab"
```

```
ktab_list <- ktab.list.df(list(trait_categ, trait_ord))
```

```
# Por fim, calcular a distância funcional entre as espécies
```

```
# Em "type", a letra N indica variável categórica (ou nominal), enquanto a letra
```

```
# O indica variável ordinal
```

```
dist_mist <- dist.ktab(ktab_list, type= c("N", "O"))
```

```
# Visualize os dados com uma PCoA (Ver Capítulo XXX)
```

```
pcoa_traits_mist <- pcoa(dist_mist, correction="cailliez")
```

```
eixos_mist <- as.data.frame(pcoa_traits_mist$vectors[,1:2]) # exportar 2 eixos
```

```
eixos_mist %>%
```

```
  ggplot(aes(x=Axis.1, y=Axis.2)) +
```

```
  geom_point(pch=21, size=3, color = "black", fill="#d73027") +
```

```
  geom_text_repel(aes(Axis.1, Axis.2, label = rownames(eixos_mist))) +
```

```
  xlab("PCO 1") + ylab("PCO 2") +
```

```
  theme(axis.title.x = element_text(face="bold", size=14),
```

```
        axis.text.x = element_text(vjust=0.5, size=12)) +
```

```
  theme(axis.title.y = element_text(face="bold", size=14),
```

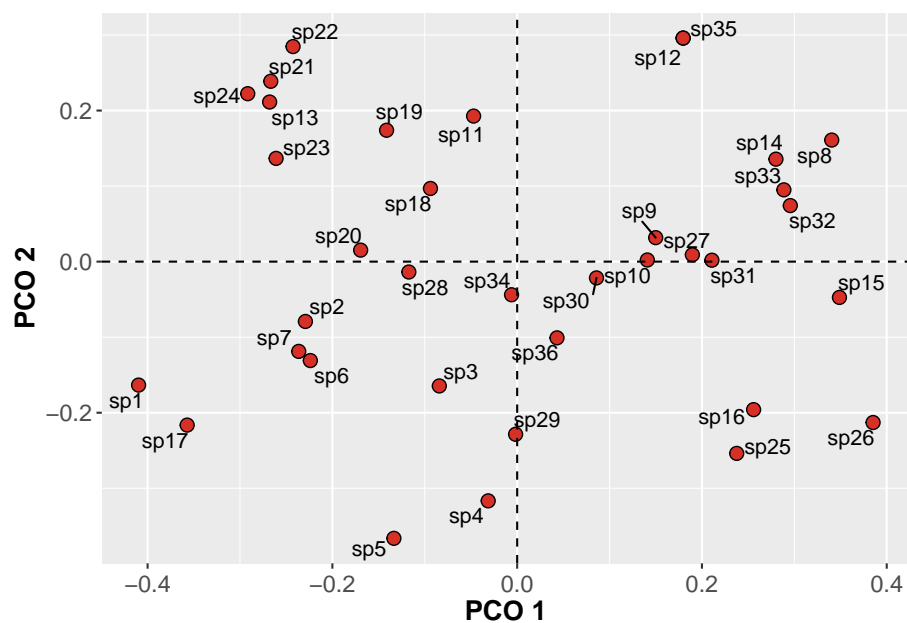
```
        axis.text.y = element_text(vjust=0.5, size=12)) +
```

```
  geom_hline(yintercept = 0, linetype=2) +
```

```
  geom_vline(xintercept = 0, linetype=2) +
```

```
  theme(legend.position = "top", legend.title=element_blank()) -> plot_trait_mist
```

```
plot_trait_mist
```

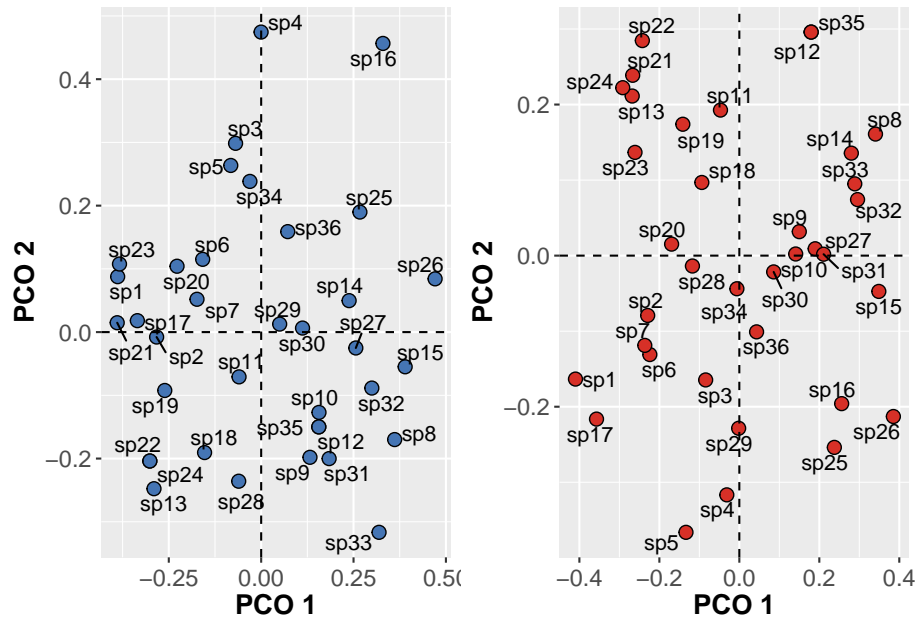


```
## Exportar resultad em pdf. Você pode mudar .pdf por .png ou .svg.
```

```
ggsave("trait_mist.pdf", plot_trait_mist, height = 14, width = 14, dpi = 600, units = "cm")
```

```
# Agora podemos combinar os dois gráficos para comparar as duas medidas de distância, uma somente
```

```
grid.arrange(plot_trait_cat, plot_trait_mist, ncol=2)
```



## 8.3 2.3 Métricas de diversidade funcional (alpha)

### 8.3.1 2.2.1 Riqueza funcional

A riqueza funcional mede a quantidade de espaço funcional preenchido pela espécie de uma comunidade (Mason & Mouillot 2013). A estimativa desse espaço pode ser calculada usando dendrogramas (Petchey & Gaston 2002) ou através do método **Convex Hull** (Cornwell et al. 2006) que dão origem, respectivamente, as duas métricas mais usadas: (1) Diversidade Funcional (FD) e (Riqueza Funcional - FRic). Os índices de riqueza funcional geralmente são usados como indicadores do espaço de nicho que é potencialmente usado ou não (Schleuter et al. 2010).

*# Antes de calcular as métricas de diversidade funcional, vamos calcular a riqueza de*

```
richness <- dbFD(dist_mist, comun_dat)$nbsp # Calcula a riqueza de espécies
```

```
## FRic: Dimensionality reduction was required. The last 17 PCoA axes (out of 19 in total)
```

```
## FRic: Quality of the reduced-space representation = 0.3243851
```

```
## CWM: When 'x' is a distance matrix, CWM cannot be calculated.
```

*# Primeiro passo: definir uma distância apropriada (veja descrição anterior)*

*# Aqui, iremos utilizar a distância de todos os dados trait\_dat (variáveis categóricas)*

dist\_mist

##	sp1	sp2	sp3	sp4	sp5	sp6	sp7
## sp2	0.5000000						
## sp3	0.7107801	0.6808389					
## sp4	0.7771900	0.8538292	0.7179711				
## sp5	0.6107116	0.7345988	0.7381353	0.5106682			
## sp6	0.5041691	0.6487320	0.6527339	0.6522593	0.5177440		
## sp7	0.5041691	0.5809646	0.7137190	0.7132850	0.5177440	0.4082483	
## sp8	0.9574271	0.9128709	0.8447140	0.8106696	0.7689990	0.8466582	0.8466582
## sp9	0.8322170	0.7743619	0.8705685	0.7862568	0.6712758	0.8618286	0.7590006
## sp10	0.8196798	0.8447140	0.8164966	0.7259317	0.6556685	0.7452327	0.8497285
## sp11	0.7381353	0.7658396	0.6755013	0.7674028	0.7671647	0.5971604	0.5971604
## sp12	0.8322170	0.8264198	0.7689102	0.8859268	0.8370252	0.7019605	0.7019605
## sp13	0.6454972	0.7637626	0.8196798	0.7845499	0.6516846	0.6192712	0.6192712
## sp14	0.9128709	0.7637626	0.7395100	0.8605338	0.9172565	0.7959251	0.8466582
## sp15	0.8684391	0.7100609	0.7423352	0.8468228	0.8274814	0.7637626	0.8164966
## sp16	0.8660254	0.8660254	0.6808389	0.6290397	0.7892837	0.7664984	0.8190563
## sp17	0.2886751	0.5773503	0.7671647	0.7771900	0.6107116	0.5809646	0.5809646
## sp18	0.6652707	0.5160455	0.7689102	0.8859268	0.7307151	0.7019605	0.6398556
## sp19	0.7892837	0.6755013	0.8184398	0.7432689	0.8164966	0.7754948	0.7197631
## sp20	0.6149610	0.6479535	0.5381778	0.7674028	0.7671647	0.5971604	0.7233721
## sp21	0.6755013	0.6755013	0.7658396	0.7432689	0.8164966	0.7197631	0.6593372
## sp22	0.6107116	0.6107116	0.7658396	0.8480067	0.7637626	0.5927834	0.5927834
## sp23	0.6573806	0.5906064	0.6921158	0.7395100	0.8199027	0.7096247	0.7096247
## sp24	0.5292494	0.5948301	0.7155949	0.8234825	0.7375481	0.5252793	0.5252793
## sp25	0.8404178	0.7345988	0.7093497	0.7432689	0.6454972	0.7754948	0.7197631
## sp26	0.8886143	0.7892837	0.8184398	0.7973594	0.7071068	0.8274814	0.7754948
## sp27	0.8347284	0.8289488	0.7716277	0.7712263	0.7920294	0.6652707	0.6652707
## sp28	0.6454972	0.7637626	0.7671647	0.8290703	0.7345988	0.5809646	0.5809646
## sp29	0.7137190	0.6839065	0.7100609	0.7660943	0.5971604	0.7107801	0.5818433
## sp30	0.8347284	0.8289488	0.7155949	0.6543114	0.7920294	0.6652707	0.6652707
## sp31	0.8259233	0.6573806	0.8035489	0.8447140	0.8199027	0.7096247	0.7660943
## sp32	0.9213121	0.8259233	0.9013088	0.7938566	0.9159187	0.8680825	0.8186781
## sp33	0.8414675	0.7842952	0.8794158	0.8705685	0.8355590	0.7276576	0.7276576
## sp34	0.7520331	0.7456127	0.6813161	0.6558429	0.7252023	0.6787075	0.5421967
## sp35	0.8322170	0.8264198	0.7689102	0.8859268	0.8370252	0.7019605	0.7019605
## sp36	0.6944209	0.7995447	0.7942658	0.6558429	0.6652707	0.6142561	0.7375481
##	sp8	sp9	sp10	sp11	sp12	sp13	sp14
## sp2							
## sp3							
## sp4							
## sp5							
## sp6							
## sp7							

```

## sp8
## sp9 0.5160455
## sp10 0.5448624 0.4177992
## sp11 0.6867080 0.7182727 0.7127595
## sp12 0.5160455 0.7071068 0.7126637 0.4272966
## sp13 0.7071068 0.6652707 0.6495191 0.5887001 0.6652707
## sp14 0.7637626 0.8264198 0.7938566 0.7449169 0.5913006 0.8660254
## sp15 0.6192712 0.6950778 0.6560273 0.8051504 0.6950778 0.8945186 0.5478718
## sp16 0.7637626 0.8753873 0.8447140 0.8184398 0.7185423 1.0000000 0.6454972
## sp17 0.9574271 0.8322170 0.8196798 0.7925762 0.8808623 0.6454972 0.9574271
## sp18 0.7743619 0.5773503 0.6515798 0.6577099 0.7071068 0.6652707 0.7185423
## sp19 0.8213969 0.7241058 0.7988958 0.7395100 0.7795271 0.6516846 0.7689990
## sp20 0.8971628 0.8261854 0.7127595 0.6454972 0.7182727 0.7164039 0.6230847
## sp21 0.8213969 0.7795271 0.8494515 0.6808389 0.7241058 0.5842598 0.7689990
## sp22 0.7127595 0.7241058 0.7449169 0.5448624 0.5980489 0.4179627 0.7689990
## sp23 0.8814629 0.8425832 0.8605338 0.7432689 0.7915890 0.6660657 0.7259317
## sp24 0.7526397 0.7417301 0.7470295 0.5253079 0.6192712 0.3992684 0.8061016
## sp25 0.7127595 0.6640752 0.7449169 0.7395100 0.7241058 0.8706470 0.7127595
## sp26 0.5842598 0.6640752 0.7449169 0.8447140 0.7241058 0.9172565 0.7127595
## sp27 0.6950778 0.8466582 0.8513047 0.5253079 0.5478718 0.7590006 0.6950778
## sp28 0.8164966 0.7805458 0.7671647 0.6793455 0.7805458 0.5773503 0.8660254
## sp29 0.8248465 0.6256088 0.7417301 0.7197631 0.8008660 0.7452327 0.8248465
## sp30 0.7526397 0.7959251 0.8008660 0.5994011 0.7417301 0.7590006 0.7526397
## sp31 0.7259317 0.6781935 0.6381106 0.7432689 0.7370752 0.7812235 0.6002584
## sp32 0.7259317 0.7370752 0.8106696 0.8480067 0.7370752 0.8814629 0.6002584
## sp33 0.6136882 0.6660657 0.6719621 0.7165635 0.6002584 0.7435692 0.6136882
## sp34 0.8312616 0.7689990 0.8751661 0.5078283 0.6516846 0.7856703 0.6640752
## sp35 0.5160455 0.7071068 0.7126637 0.4272966 0.0000000 0.6652707 0.5913006
## sp36 0.7241058 0.8213969 0.7182727 0.7126637 0.7127595 0.7307151 0.7241058
##          sp15          sp16          sp17          sp18          sp19          sp20          sp21
## sp2
## sp3
## sp4
## sp5
## sp6
## sp7
## sp8
## sp9
## sp10
## sp11
## sp12
## sp13
## sp14
## sp15
## sp16 0.6487320
## sp17 0.9151611 0.9128709

```



```

## sp18 0.6950778 0.9217572 0.7252023
## sp19 0.8763897 0.8404178 0.7345988 0.6640752
## sp20 0.8051504 0.7658396 0.6793455 0.6577099 0.6808389
## sp21 0.8763897 0.7892837 0.7345988 0.6640752 0.4082483 0.6808389
## sp22 0.7754948 0.8886143 0.6755013 0.5980489 0.5773503 0.6808389 0.4082483
## sp23 0.8186781 0.7738320 0.7179711 0.6136882 0.5055430 0.6849200 0.2981729
## sp24 0.7743619 0.8777753 0.6028584 0.6192712 0.6063788 0.6652932 0.4483621
## sp25 0.6593372 0.6755013 0.7892837 0.7241058 0.8164966 0.8447140 0.8660254
## sp26 0.5177440 0.6755013 0.8404178 0.7241058 0.8660254 0.9381942 0.9128709
## sp27 0.7185423 0.7214496 0.7832230 0.8466582 0.7859359 0.7805650 0.8372745
## sp28 0.7664984 0.9128709 0.5773503 0.7252023 0.7345988 0.7381353 0.7345988
## sp29 0.7395100 0.7964891 0.6527339 0.6889991 0.6939744 0.7197631 0.8051504
## sp30 0.7743619 0.7770604 0.7832230 0.7959251 0.6715866 0.7252230 0.7310006
## sp31 0.5035545 0.8259233 0.7738320 0.6136882 0.7110371 0.7432689 0.8199027
## sp32 0.6482545 0.6573806 0.8749186 0.7915890 0.5821572 0.7973594 0.7110371
## sp33 0.5194903 0.7842952 0.7904013 0.7259317 0.7224104 0.7725261 0.8297852
## sp34 0.8372745 0.6874627 0.8055353 0.7127595 0.7185423 0.6515798 0.6580043
## sp35 0.6950778 0.7185423 0.8808623 0.7071068 0.7795271 0.7182727 0.7241058
## sp36 0.6063788 0.6874627 0.7520331 0.8213969 0.8753873 0.7689102 0.7743619
##          sp22          sp23          sp24          sp25          sp26          sp27          sp28
## sp2
## sp3
## sp4
## sp5
## sp6
## sp7
## sp8
## sp9
## sp10
## sp11
## sp12
## sp13
## sp14
## sp15
## sp16
## sp17
## sp18
## sp19
## sp20
## sp21
## sp22
## sp23 0.5055430
## sp24 0.1853695 0.5194903
## sp25 0.8164966 0.8199027 0.8372745
## sp26 0.8660254 0.8692374 0.8856421 0.4082483
## sp27 0.7310006 0.8774225 0.7071068 0.6715866 0.6715866

```

```

## sp28 0.6107116 0.7738320 0.5292494 0.7892837 0.7892837 0.6028584
## sp29 0.7516208 0.8468228 0.7126637 0.5611897 0.6939744 0.6515798 0.5854299
## sp30 0.7310006 0.7766618 0.7071068 0.7310006 0.7310006 0.5000000 0.5292494
## sp31 0.7110371 0.7637626 0.7210202 0.5821572 0.5821572 0.6607094 0.5906064
## sp32 0.8199027 0.7637626 0.8285752 0.7674028 0.6498003 0.6607094 0.7179711
## sp33 0.7224104 0.8753873 0.7096247 0.7224104 0.5959950 0.5035545 0.5398157
## sp34 0.7185423 0.7224104 0.7197631 0.6580043 0.7743619 0.5927834 0.7520331
## sp35 0.5980489 0.7915890 0.6192712 0.7241058 0.7241058 0.5478718 0.7805458
## sp36 0.6580043 0.7779525 0.6593372 0.7743619 0.7185423 0.6593372 0.7520331
##          sp29          sp30          sp31          sp32          sp33          sp34          sp35
## sp2
## sp3
## sp4
## sp5
## sp6
## sp7
## sp8
## sp9
## sp10
## sp11
## sp12
## sp13
## sp14
## sp15
## sp16
## sp17
## sp18
## sp19
## sp20
## sp21
## sp22
## sp23
## sp24
## sp25
## sp26
## sp27
## sp28
## sp29
## sp30 0.7126637
## sp31 0.6834536 0.5943093
## sp32 0.7419179 0.5943093 0.5773503
## sp33 0.6543114 0.5804313 0.4277495 0.4277495
## sp34 0.6652932 0.5177440 0.7779525 0.7224104 0.7674028
## sp35 0.8008660 0.7417301 0.7370752 0.7370752 0.6002584 0.6516846
## sp36 0.7805650 0.7754948 0.7779525 0.8297852 0.7110371 0.7637626 0.7127595

```

```
# O índice "Functional Richness" só funciona para comunidades com 3 ou mais espécies
# Caso você tenha comunidades com 1 ou 2 espécies, o valor será NA
```

```
fric <- dbFD(dist_mist, comun_dat)$FRic
```

```
## FRic: Dimensionality reduction was required. The last 17 PCoA axes (out of 19 in total) were r
## FRic: Quality of the reduced-space representation = 0.3243851
## CWM: When 'x' is a distance matrix, CWM cannot be calculated.
```

```
# Functional Diversity (Petchey & Gaston 2002)
```

```
## Passo 1: análise de agrupamento para criar o dendrograma
```

```
dend <- hclust(dist_mist, "average")
```

```
## Passo 2: transformar o dendrograma em um arquivo da classe phylo
```

```
tree_dend <- as.phylo(dend)
```

```
## Passo 3: calcular o valor da diversidade funcional
```

```
FD <- pd(comun_dat, tree_dend)$PD
```

### 8.3.2 2.2.2. Divergência funcional

```
#!!!! Dúvida — Como inserir uma Figura externa? !!!!!!!
```

A divergência funcional é uma medida que descreve a irregularidade na distribuição dos valores dos atributos no volume do espaço funcional ocupado por todas as espécies de uma certa comunidade (Garnier 2015, Figura ???). Para obter os valores de divergência, o espaço funcional é calculado através do método **Convex Hull** (Functional Divergence) ou do espaço multidimensional calculado com um PCoA (Functional Dispersion). Nos dois casos, o valor da métrica representa a distância média das espécies para o centro de gravidade ou centroide do espaço funcional, ponderado pela abundância relativa das espécies (Villegier et al. 2008, Laliberté & Legendre 2010). Desse modo, a divergência funcional é uma medida que calcula o grau de diferenciação em que a distribuição da abundância maximiza a divergência entre os atributos funcionais (Mason & Mouillot). Em geral, estudos que usam esses índices buscam entender o grau de diferenciação de recursos de espécies que coexistem em uma comunidade (Garnier 2015).

### 8.3.3 Análise

```
library(FD)
library(tidyverse)
```

```
library(ecodados)
library(vegan)
library(SYNCSA)
```

```
# Primeiro passo: definir uma distância apropriada (veja descrição anterior)
```

```
# Aqui, iremos utilizar a distância de todos os dados trait_dat (variáveis categóricas)
```

```
dist_mist
```

```
##           sp1           sp2           sp3           sp4           sp5           sp6           sp7
## sp2  0.5000000
## sp3  0.7107801 0.6808389
## sp4  0.7771900 0.8538292 0.7179711
## sp5  0.6107116 0.7345988 0.7381353 0.5106682
## sp6  0.5041691 0.6487320 0.6527339 0.6522593 0.5177440
## sp7  0.5041691 0.5809646 0.7137190 0.7132850 0.5177440 0.4082483
## sp8  0.9574271 0.9128709 0.8447140 0.8106696 0.7689990 0.8466582 0.8466582
## sp9  0.8322170 0.7743619 0.8705685 0.7862568 0.6712758 0.8618286 0.7590006
## sp10 0.8196798 0.8447140 0.8164966 0.7259317 0.6556685 0.7452327 0.8497285
## sp11 0.7381353 0.7658396 0.6755013 0.7674028 0.7671647 0.5971604 0.5971604
## sp12 0.8322170 0.8264198 0.7689102 0.8859268 0.8370252 0.7019605 0.7019605
## sp13 0.6454972 0.7637626 0.8196798 0.7845499 0.6516846 0.6192712 0.6192712
## sp14 0.9128709 0.7637626 0.7395100 0.8605338 0.9172565 0.7959251 0.8466582
## sp15 0.8684391 0.7100609 0.7423352 0.8468228 0.8274814 0.7637626 0.8164966
## sp16 0.8660254 0.8660254 0.6808389 0.6290397 0.7892837 0.7664984 0.8190563
## sp17 0.2886751 0.5773503 0.7671647 0.7771900 0.6107116 0.5809646 0.5809646
## sp18 0.6652707 0.5160455 0.7689102 0.8859268 0.7307151 0.7019605 0.6398556
## sp19 0.7892837 0.6755013 0.8184398 0.7432689 0.8164966 0.7754948 0.7197631
## sp20 0.6149610 0.6479535 0.5381778 0.7674028 0.7671647 0.5971604 0.7233721
## sp21 0.6755013 0.6755013 0.7658396 0.7432689 0.8164966 0.7197631 0.6593372
## sp22 0.6107116 0.6107116 0.7658396 0.8480067 0.7637626 0.5927834 0.5927834
## sp23 0.6573806 0.5906064 0.6921158 0.7395100 0.8199027 0.7096247 0.7096247
## sp24 0.5292494 0.5948301 0.7155949 0.8234825 0.7375481 0.5252793 0.5252793
## sp25 0.8404178 0.7345988 0.7093497 0.7432689 0.6454972 0.7754948 0.7197631
## sp26 0.8886143 0.7892837 0.8184398 0.7973594 0.7071068 0.8274814 0.7754948
## sp27 0.8347284 0.8289488 0.7716277 0.7712263 0.7920294 0.6652707 0.6652707
## sp28 0.6454972 0.7637626 0.7671647 0.8290703 0.7345988 0.5809646 0.5809646
## sp29 0.7137190 0.6839065 0.7100609 0.7660943 0.5971604 0.7107801 0.5818433
## sp30 0.8347284 0.8289488 0.7155949 0.6543114 0.7920294 0.6652707 0.6652707
## sp31 0.8259233 0.6573806 0.8035489 0.8447140 0.8199027 0.7096247 0.7660943
## sp32 0.9213121 0.8259233 0.9013088 0.7938566 0.9159187 0.8680825 0.8186781
## sp33 0.8414675 0.7842952 0.8794158 0.8705685 0.8355590 0.7276576 0.7276576
## sp34 0.7520331 0.7456127 0.6813161 0.6558429 0.7252023 0.6787075 0.5421967
```

```

## sp35 0.8322170 0.8264198 0.7689102 0.8859268 0.8370252 0.7019605 0.7019605
## sp36 0.6944209 0.7995447 0.7942658 0.6558429 0.6652707 0.6142561 0.7375481
##          sp8          sp9          sp10          sp11          sp12          sp13          sp14
## sp2
## sp3
## sp4
## sp5
## sp6
## sp7
## sp8
## sp9 0.5160455
## sp10 0.5448624 0.4177992
## sp11 0.6867080 0.7182727 0.7127595
## sp12 0.5160455 0.7071068 0.7126637 0.4272966
## sp13 0.7071068 0.6652707 0.6495191 0.5887001 0.6652707
## sp14 0.7637626 0.8264198 0.7938566 0.7449169 0.5913006 0.8660254
## sp15 0.6192712 0.6950778 0.6560273 0.8051504 0.6950778 0.8945186 0.5478718
## sp16 0.7637626 0.8753873 0.8447140 0.8184398 0.7185423 1.0000000 0.6454972
## sp17 0.9574271 0.8322170 0.8196798 0.7925762 0.8808623 0.6454972 0.9574271
## sp18 0.7743619 0.5773503 0.6515798 0.6577099 0.7071068 0.6652707 0.7185423
## sp19 0.8213969 0.7241058 0.7988958 0.7395100 0.7795271 0.6516846 0.7689990
## sp20 0.8971628 0.8261854 0.7127595 0.6454972 0.7182727 0.7164039 0.6230847
## sp21 0.8213969 0.7795271 0.8494515 0.6808389 0.7241058 0.5842598 0.7689990
## sp22 0.7127595 0.7241058 0.7449169 0.5448624 0.5980489 0.4179627 0.7689990
## sp23 0.8814629 0.8425832 0.8605338 0.7432689 0.7915890 0.6660657 0.7259317
## sp24 0.7526397 0.7417301 0.7470295 0.5253079 0.6192712 0.3992684 0.8061016
## sp25 0.7127595 0.6640752 0.7449169 0.7395100 0.7241058 0.8706470 0.7127595
## sp26 0.5842598 0.6640752 0.7449169 0.8447140 0.7241058 0.9172565 0.7127595
## sp27 0.6950778 0.8466582 0.8513047 0.5253079 0.5478718 0.7590006 0.6950778
## sp28 0.8164966 0.7805458 0.7671647 0.6793455 0.7805458 0.5773503 0.8660254
## sp29 0.8248465 0.6256088 0.7417301 0.7197631 0.8008660 0.7452327 0.8248465
## sp30 0.7526397 0.7959251 0.8008660 0.5994011 0.7417301 0.7590006 0.7526397
## sp31 0.7259317 0.6781935 0.6381106 0.7432689 0.7370752 0.7812235 0.6002584
## sp32 0.7259317 0.7370752 0.8106696 0.8480067 0.7370752 0.8814629 0.6002584
## sp33 0.6136882 0.6660657 0.6719621 0.7165635 0.6002584 0.7435692 0.6136882
## sp34 0.8312616 0.7689990 0.8751661 0.5078283 0.6516846 0.7856703 0.6640752
## sp35 0.5160455 0.7071068 0.7126637 0.4272966 0.0000000 0.6652707 0.5913006
## sp36 0.7241058 0.8213969 0.7182727 0.7126637 0.7127595 0.7307151 0.7241058
##          sp15          sp16          sp17          sp18          sp19          sp20          sp21
## sp2
## sp3
## sp4
## sp5
## sp6
## sp7
## sp8

```

```

## sp9
## sp10
## sp11
## sp12
## sp13
## sp14
## sp15
## sp16 0.6487320
## sp17 0.9151611 0.9128709
## sp18 0.6950778 0.9217572 0.7252023
## sp19 0.8763897 0.8404178 0.7345988 0.6640752
## sp20 0.8051504 0.7658396 0.6793455 0.6577099 0.6808389
## sp21 0.8763897 0.7892837 0.7345988 0.6640752 0.4082483 0.6808389
## sp22 0.7754948 0.8886143 0.6755013 0.5980489 0.5773503 0.6808389 0.4082483
## sp23 0.8186781 0.7738320 0.7179711 0.6136882 0.5055430 0.6849200 0.2981729
## sp24 0.7743619 0.8777753 0.6028584 0.6192712 0.6063788 0.6652932 0.4483621
## sp25 0.6593372 0.6755013 0.7892837 0.7241058 0.8164966 0.8447140 0.8660254
## sp26 0.5177440 0.6755013 0.8404178 0.7241058 0.8660254 0.9381942 0.9128709
## sp27 0.7185423 0.7214496 0.7832230 0.8466582 0.7859359 0.7805650 0.8372745
## sp28 0.7664984 0.9128709 0.5773503 0.7252023 0.7345988 0.7381353 0.7345988
## sp29 0.7395100 0.7964891 0.6527339 0.6889991 0.6939744 0.7197631 0.8051504
## sp30 0.7743619 0.7770604 0.7832230 0.7959251 0.6715866 0.7252230 0.7310006
## sp31 0.5035545 0.8259233 0.7738320 0.6136882 0.7110371 0.7432689 0.8199027
## sp32 0.6482545 0.6573806 0.8749186 0.7915890 0.5821572 0.7973594 0.7110371
## sp33 0.5194903 0.7842952 0.7904013 0.7259317 0.7224104 0.7725261 0.8297852
## sp34 0.8372745 0.6874627 0.8055353 0.7127595 0.7185423 0.6515798 0.6580043
## sp35 0.6950778 0.7185423 0.8808623 0.7071068 0.7795271 0.7182727 0.7241058
## sp36 0.6063788 0.6874627 0.7520331 0.8213969 0.8753873 0.7689102 0.7743619
##          sp22          sp23          sp24          sp25          sp26          sp27          sp28
## sp2
## sp3
## sp4
## sp5
## sp6
## sp7
## sp8
## sp9
## sp10
## sp11
## sp12
## sp13
## sp14
## sp15
## sp16
## sp17
## sp18

```

```

## sp19
## sp20
## sp21
## sp22
## sp23 0.5055430
## sp24 0.1853695 0.5194903
## sp25 0.8164966 0.8199027 0.8372745
## sp26 0.8660254 0.8692374 0.8856421 0.4082483
## sp27 0.7310006 0.8774225 0.7071068 0.6715866 0.6715866
## sp28 0.6107116 0.7738320 0.5292494 0.7892837 0.7892837 0.6028584
## sp29 0.7516208 0.8468228 0.7126637 0.5611897 0.6939744 0.6515798 0.5854299
## sp30 0.7310006 0.7766618 0.7071068 0.7310006 0.7310006 0.5000000 0.5292494
## sp31 0.7110371 0.7637626 0.7210202 0.5821572 0.5821572 0.6607094 0.5906064
## sp32 0.8199027 0.7637626 0.8285752 0.7674028 0.6498003 0.6607094 0.7179711
## sp33 0.7224104 0.8753873 0.7096247 0.7224104 0.5959950 0.5035545 0.5398157
## sp34 0.7185423 0.7224104 0.7197631 0.6580043 0.7743619 0.5927834 0.7520331
## sp35 0.5980489 0.7915890 0.6192712 0.7241058 0.7241058 0.5478718 0.7805458
## sp36 0.6580043 0.7779525 0.6593372 0.7743619 0.7185423 0.6593372 0.7520331
##          sp29          sp30          sp31          sp32          sp33          sp34          sp35
## sp2
## sp3
## sp4
## sp5
## sp6
## sp7
## sp8
## sp9
## sp10
## sp11
## sp12
## sp13
## sp14
## sp15
## sp16
## sp17
## sp18
## sp19
## sp20
## sp21
## sp22
## sp23
## sp24
## sp25
## sp26
## sp27
## sp28

```

```
## sp29
## sp30 0.7126637
## sp31 0.6834536 0.5943093
## sp32 0.7419179 0.5943093 0.5773503
## sp33 0.6543114 0.5804313 0.4277495 0.4277495
## sp34 0.6652932 0.5177440 0.7779525 0.7224104 0.7674028
## sp35 0.8008660 0.7417301 0.7370752 0.7370752 0.6002584 0.6516846
## sp36 0.7805650 0.7754948 0.7779525 0.8297852 0.7110371 0.7637626 0.7127595

# O índice "Functional Divergence" só funciona para comunidades com 3 ou mais espécies
# Caso você tenha comunidades com 1 ou 2 espécies, a análise irá retornar o valor NA

fdiv <- dbFD(dist_mist, comun_dat)$FDiv

## FRic: Dimensionality reduction was required. The last 17 PCoA axes (out of 19 in total) were retained.
## FRic: Quality of the reduced-space representation = 0.3243851
## CWM: When 'x' is a distance matrix, CWM cannot be calculated.

# O índice "Functional Dispersion" atribui valor 0 para comunidades com 1 ou 2 espécies
fdis <- dbFD(dist_mist, comun_dat)$FDis# Fdis=0 em locais com somente uma espécie

## FRic: Dimensionality reduction was required. The last 17 PCoA axes (out of 19 in total) were retained.
## FRic: Quality of the reduced-space representation = 0.3243851
## CWM: When 'x' is a distance matrix, CWM cannot be calculated.
```

### 8.3.4 2.2.3. Regularidade funcional

A regularidade funcional (do inglês **Functional Evenness**) mede o quão regular é a distribuição da abundância dos valores dos atributos funcionais no espaço funcional. Diferente dos outros métodos, a versão multidimensional deste índice utiliza um método chamado **Minimum Spanning Tree** (MST) para conectar todas espécies no espaço funcional. A distância par-a-par das espécies na MST é ponderada pela abundância relativa das espécies e, desse modo, o valor final da regularidade funcional (FEve) vai variar de 0 (máxima irregularidade da distribuição da abundância ou distância funcional das espécies) a 1 (máxima regularidade).

### 8.3.5 Análise

```
library(FD)
library(tidyverse)
library(ecodados)
library(vegan)
library(GGally)

# Aqui, iremos utilizar a distância de todos os dados trait_dat (variáveis categóricas)
```



```
dist_mist
```

```
##          sp1          sp2          sp3          sp4          sp5          sp6          sp7
## sp2  0.5000000
## sp3  0.7107801 0.6808389
## sp4  0.7771900 0.8538292 0.7179711
## sp5  0.6107116 0.7345988 0.7381353 0.5106682
## sp6  0.5041691 0.6487320 0.6527339 0.6522593 0.5177440
## sp7  0.5041691 0.5809646 0.7137190 0.7132850 0.5177440 0.4082483
## sp8  0.9574271 0.9128709 0.8447140 0.8106696 0.7689990 0.8466582 0.8466582
## sp9  0.8322170 0.7743619 0.8705685 0.7862568 0.6712758 0.8618286 0.7590006
## sp10 0.8196798 0.8447140 0.8164966 0.7259317 0.6556685 0.7452327 0.8497285
## sp11 0.7381353 0.7658396 0.6755013 0.7674028 0.7671647 0.5971604 0.5971604
## sp12 0.8322170 0.8264198 0.7689102 0.8859268 0.8370252 0.7019605 0.7019605
## sp13 0.6454972 0.7637626 0.8196798 0.7845499 0.6516846 0.6192712 0.6192712
## sp14 0.9128709 0.7637626 0.7395100 0.8605338 0.9172565 0.7959251 0.8466582
## sp15 0.8684391 0.7100609 0.7423352 0.8468228 0.8274814 0.7637626 0.8164966
## sp16 0.8660254 0.8660254 0.6808389 0.6290397 0.7892837 0.7664984 0.8190563
## sp17 0.2886751 0.5773503 0.7671647 0.7771900 0.6107116 0.5809646 0.5809646
## sp18 0.6652707 0.5160455 0.7689102 0.8859268 0.7307151 0.7019605 0.6398556
## sp19 0.7892837 0.6755013 0.8184398 0.7432689 0.8164966 0.7754948 0.7197631
## sp20 0.6149610 0.6479535 0.5381778 0.7674028 0.7671647 0.5971604 0.7233721
## sp21 0.6755013 0.6755013 0.7658396 0.7432689 0.8164966 0.7197631 0.6593372
## sp22 0.6107116 0.6107116 0.7658396 0.8480067 0.7637626 0.5927834 0.5927834
## sp23 0.6573806 0.5906064 0.6921158 0.7395100 0.8199027 0.7096247 0.7096247
## sp24 0.5292494 0.5948301 0.7155949 0.8234825 0.7375481 0.5252793 0.5252793
## sp25 0.8404178 0.7345988 0.7093497 0.7432689 0.6454972 0.7754948 0.7197631
## sp26 0.8886143 0.7892837 0.8184398 0.7973594 0.7071068 0.8274814 0.7754948
## sp27 0.8347284 0.8289488 0.7716277 0.7712263 0.7920294 0.6652707 0.6652707
## sp28 0.6454972 0.7637626 0.7671647 0.8290703 0.7345988 0.5809646 0.5809646
## sp29 0.7137190 0.6839065 0.7100609 0.7660943 0.5971604 0.7107801 0.5818433
## sp30 0.8347284 0.8289488 0.7155949 0.6543114 0.7920294 0.6652707 0.6652707
## sp31 0.8259233 0.6573806 0.8035489 0.8447140 0.8199027 0.7096247 0.7660943
## sp32 0.9213121 0.8259233 0.9013088 0.7938566 0.9159187 0.8680825 0.8186781
## sp33 0.8414675 0.7842952 0.8794158 0.8705685 0.8355590 0.7276576 0.7276576
## sp34 0.7520331 0.7456127 0.6813161 0.6558429 0.7252023 0.6787075 0.5421967
## sp35 0.8322170 0.8264198 0.7689102 0.8859268 0.8370252 0.7019605 0.7019605
## sp36 0.6944209 0.7995447 0.7942658 0.6558429 0.6652707 0.6142561 0.7375481
##          sp8          sp9          sp10          sp11          sp12          sp13          sp14
## sp2
## sp3
## sp4
## sp5
## sp6
## sp7
```

```

## sp8
## sp9 0.5160455
## sp10 0.5448624 0.4177992
## sp11 0.6867080 0.7182727 0.7127595
## sp12 0.5160455 0.7071068 0.7126637 0.4272966
## sp13 0.7071068 0.6652707 0.6495191 0.5887001 0.6652707
## sp14 0.7637626 0.8264198 0.7938566 0.7449169 0.5913006 0.8660254
## sp15 0.6192712 0.6950778 0.6560273 0.8051504 0.6950778 0.8945186 0.5478718
## sp16 0.7637626 0.8753873 0.8447140 0.8184398 0.7185423 1.0000000 0.6454972
## sp17 0.9574271 0.8322170 0.8196798 0.7925762 0.8808623 0.6454972 0.9574271
## sp18 0.7743619 0.5773503 0.6515798 0.6577099 0.7071068 0.6652707 0.7185423
## sp19 0.8213969 0.7241058 0.7988958 0.7395100 0.7795271 0.6516846 0.7689990
## sp20 0.8971628 0.8261854 0.7127595 0.6454972 0.7182727 0.7164039 0.6230847
## sp21 0.8213969 0.7795271 0.8494515 0.6808389 0.7241058 0.5842598 0.7689990
## sp22 0.7127595 0.7241058 0.7449169 0.5448624 0.5980489 0.4179627 0.7689990
## sp23 0.8814629 0.8425832 0.8605338 0.7432689 0.7915890 0.6660657 0.7259317
## sp24 0.7526397 0.7417301 0.7470295 0.5253079 0.6192712 0.3992684 0.8061016
## sp25 0.7127595 0.6640752 0.7449169 0.7395100 0.7241058 0.8706470 0.7127595
## sp26 0.5842598 0.6640752 0.7449169 0.8447140 0.7241058 0.9172565 0.7127595
## sp27 0.6950778 0.8466582 0.8513047 0.5253079 0.5478718 0.7590006 0.6950778
## sp28 0.8164966 0.7805458 0.7671647 0.6793455 0.7805458 0.5773503 0.8660254
## sp29 0.8248465 0.6256088 0.7417301 0.7197631 0.8008660 0.7452327 0.8248465
## sp30 0.7526397 0.7959251 0.8008660 0.5994011 0.7417301 0.7590006 0.7526397
## sp31 0.7259317 0.6781935 0.6381106 0.7432689 0.7370752 0.7812235 0.6002584
## sp32 0.7259317 0.7370752 0.8106696 0.8480067 0.7370752 0.8814629 0.6002584
## sp33 0.6136882 0.6660657 0.6719621 0.7165635 0.6002584 0.7435692 0.6136882
## sp34 0.8312616 0.7689990 0.8751661 0.5078283 0.6516846 0.7856703 0.6640752
## sp35 0.5160455 0.7071068 0.7126637 0.4272966 0.0000000 0.6652707 0.5913006
## sp36 0.7241058 0.8213969 0.7182727 0.7126637 0.7127595 0.7307151 0.7241058
##          sp15      sp16      sp17      sp18      sp19      sp20      sp21
## sp2
## sp3
## sp4
## sp5
## sp6
## sp7
## sp8
## sp9
## sp10
## sp11
## sp12
## sp13
## sp14
## sp15
## sp16 0.6487320
## sp17 0.9151611 0.9128709

```

```

## sp18 0.6950778 0.9217572 0.7252023
## sp19 0.8763897 0.8404178 0.7345988 0.6640752
## sp20 0.8051504 0.7658396 0.6793455 0.6577099 0.6808389
## sp21 0.8763897 0.7892837 0.7345988 0.6640752 0.4082483 0.6808389
## sp22 0.7754948 0.8886143 0.6755013 0.5980489 0.5773503 0.6808389 0.4082483
## sp23 0.8186781 0.7738320 0.7179711 0.6136882 0.5055430 0.6849200 0.2981729
## sp24 0.7743619 0.8777753 0.6028584 0.6192712 0.6063788 0.6652932 0.4483621
## sp25 0.6593372 0.6755013 0.7892837 0.7241058 0.8164966 0.8447140 0.8660254
## sp26 0.5177440 0.6755013 0.8404178 0.7241058 0.8660254 0.9381942 0.9128709
## sp27 0.7185423 0.7214496 0.7832230 0.8466582 0.7859359 0.7805650 0.8372745
## sp28 0.7664984 0.9128709 0.5773503 0.7252023 0.7345988 0.7381353 0.7345988
## sp29 0.7395100 0.7964891 0.6527339 0.6889991 0.6939744 0.7197631 0.8051504
## sp30 0.7743619 0.7770604 0.7832230 0.7959251 0.6715866 0.7252230 0.7310006
## sp31 0.5035545 0.8259233 0.7738320 0.6136882 0.7110371 0.7432689 0.8199027
## sp32 0.6482545 0.6573806 0.8749186 0.7915890 0.5821572 0.7973594 0.7110371
## sp33 0.5194903 0.7842952 0.7904013 0.7259317 0.7224104 0.7725261 0.8297852
## sp34 0.8372745 0.6874627 0.8055353 0.7127595 0.7185423 0.6515798 0.6580043
## sp35 0.6950778 0.7185423 0.8808623 0.7071068 0.7795271 0.7182727 0.7241058
## sp36 0.6063788 0.6874627 0.7520331 0.8213969 0.8753873 0.7689102 0.7743619
##          sp22          sp23          sp24          sp25          sp26          sp27          sp28
## sp2
## sp3
## sp4
## sp5
## sp6
## sp7
## sp8
## sp9
## sp10
## sp11
## sp12
## sp13
## sp14
## sp15
## sp16
## sp17
## sp18
## sp19
## sp20
## sp21
## sp22
## sp23 0.5055430
## sp24 0.1853695 0.5194903
## sp25 0.8164966 0.8199027 0.8372745
## sp26 0.8660254 0.8692374 0.8856421 0.4082483
## sp27 0.7310006 0.8774225 0.7071068 0.6715866 0.6715866

```

```

## sp28 0.6107116 0.7738320 0.5292494 0.7892837 0.7892837 0.6028584
## sp29 0.7516208 0.8468228 0.7126637 0.5611897 0.6939744 0.6515798 0.5854299
## sp30 0.7310006 0.7766618 0.7071068 0.7310006 0.7310006 0.5000000 0.5292494
## sp31 0.7110371 0.7637626 0.7210202 0.5821572 0.5821572 0.6607094 0.5906064
## sp32 0.8199027 0.7637626 0.8285752 0.7674028 0.6498003 0.6607094 0.7179711
## sp33 0.7224104 0.8753873 0.7096247 0.7224104 0.5959950 0.5035545 0.5398157
## sp34 0.7185423 0.7224104 0.7197631 0.6580043 0.7743619 0.5927834 0.7520331
## sp35 0.5980489 0.7915890 0.6192712 0.7241058 0.7241058 0.5478718 0.7805458
## sp36 0.6580043 0.7779525 0.6593372 0.7743619 0.7185423 0.6593372 0.7520331
##          sp29          sp30          sp31          sp32          sp33          sp34          sp35
## sp2
## sp3
## sp4
## sp5
## sp6
## sp7
## sp8
## sp9
## sp10
## sp11
## sp12
## sp13
## sp14
## sp15
## sp16
## sp17
## sp18
## sp19
## sp20
## sp21
## sp22
## sp23
## sp24
## sp25
## sp26
## sp27
## sp28
## sp29
## sp30 0.7126637
## sp31 0.6834536 0.5943093
## sp32 0.7419179 0.5943093 0.5773503
## sp33 0.6543114 0.5804313 0.4277495 0.4277495
## sp34 0.6652932 0.5177440 0.7779525 0.7224104 0.7674028
## sp35 0.8008660 0.7417301 0.7370752 0.7370752 0.6002584 0.6516846
## sp36 0.7805650 0.7754948 0.7779525 0.8297852 0.7110371 0.7637626 0.7127595

```

*# O índice "Functional evenness" só funciona para comunidades com 3 ou mais espécies*  
*# Caso você tenha comunidades com 1 ou 2 espécies, a análise irá retornar o valor NA*

```
feve <- dbFD(dist_mist, comun_dat)$FEve
```

```
## FRic: Dimensionality reduction was required. The last 17 PCoA axes (out of 19 in total) were r
## FRic: Quality of the reduced-space representation = 0.3243851
## CWM: When 'x' is a distance matrix, CWM cannot be calculated.
```

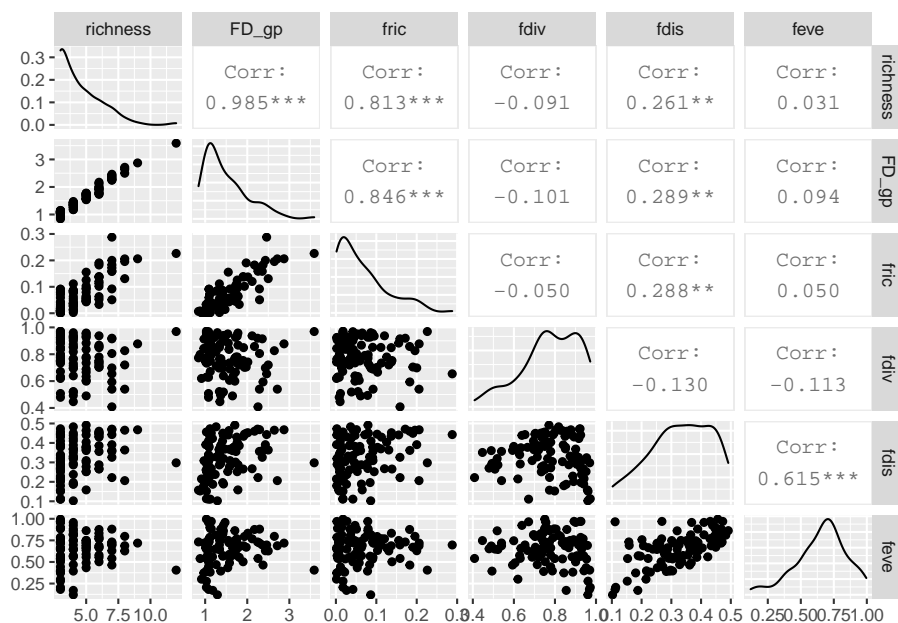
*### Você pode criar uma tabela com os resultados de todas as métricas*

```
locais <- rownames(comun_dat)
```

```
metricas <- data.frame(richness=richness,
                      FD_gp = FD,
                      fric = fric,
                      fdiv = fdiv,
                      fdis = fdis,
                      feve = feve)
```

*### Gráfico para comparar o comportamento das métricas*

```
ggpairs(metricas)
```



## 8.4 2.3 Métricas de diversidade funcional (beta)

```

library(FD)
library(tidyverse)
library(ecodados)
library(vegan)
library(GGally)
library(betapart)
library(vegan)

### Partição da Diversidade beta (Método Baselga)

comun_fren_pa <- as.matrix(decostand(comun_fren_dat, "pa"))
trait <- as.matrix(trait_fren_dat)

rowSums(comun_fren_pa) > ncol(trait)

##      1      2      3      4      5      6      7      8      9     10     11     12     13     14     15     17
## TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##     18     19     20     21     22     23     24     25     26     27     28     29     31     32     33     34
## TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##     41     42     43     44     45     46     47     48     49     50     51     53     54     55
## TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
colnames(comun_fren_pa) == rownames(trait)

## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [16] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [31] TRUE TRUE

fun_beta <- functional.beta.pair(x = comun_fren_pa, trait=trait, index="jaccard") # Es

# Os comandos abaixo permitem extrair a matriz de distância (par-a-par) com a partição

fun_turnover <- fun_beta$funct.beta.jne
fun_nestedness <- fun_beta$funct.beta.jne

### Composição Funcional: Community Weighted Means (CWM)

cwm_fren <- functcomp(trait_pad, as.matrix(comun_fren_dat))
cwm_fren

##           LA           SLA           LDMC           LN15           LCC
## 1 -0.24117001 -0.34855154  0.197452005  0.187400303 -0.53673676
## 2 -0.39773712  0.23266219 -0.090932698 -0.285977720  0.16431901
## 3 -0.18571342  0.20107562 -0.398772647 -0.125064251 -0.43046175

```

```

## 4 -0.22840640 0.16041005 0.804963068 -0.370425340 0.71938526
## 5 -0.16647900 0.34869559 0.022322127 -0.204193092 0.20513913
## 6 -0.32588213 0.36645830 0.049968287 -0.335257153 0.47130894
## 7 -0.43197058 -0.25619948 0.882004760 0.096681649 0.42065181
## 8 -0.33857863 1.12929965 0.278478893 -0.339493573 0.35571124
## 9 -0.40637654 0.66868021 0.377931295 -0.759687717 0.67746099
## 10 -0.33465801 0.26219601 0.683907823 -0.531993528 0.48456065
## 11 -0.35362891 0.44031044 -0.098078566 -0.447057493 0.25490829
## 12 -0.17542712 0.46363786 -0.252119368 -0.375615215 0.03807885
## 13 -0.23670889 0.64165706 -0.243427434 -0.510792769 0.16037083
## 14 -0.17552077 0.38780017 -0.221048163 -0.169623873 -0.11858724
## 15 -0.13274039 0.36304974 -0.049173336 -0.254725433 -0.03772986
## 17 0.01771351 0.87327677 -0.145441057 -0.489096566 0.08335821
## 18 1.03239586 0.50046791 -0.426467881 -0.416310755 0.02468425
## 19 0.37208764 1.45649761 -0.804115966 0.388010050 -0.20704973
## 20 1.13432791 0.75003603 -0.294438976 -0.499286730 -0.06243781
## 21 1.35442421 -0.70515495 2.082205988 -1.038022002 0.81956154
## 22 -0.36252857 -0.11915470 0.470138892 -0.349437843 0.55666135
## 23 -0.29349787 0.19703824 0.679820129 -0.474449656 0.71142400
## 24 -0.33836216 0.08304760 0.099230530 -0.121483271 0.35394645
## 25 0.31915111 -0.25335758 1.057707296 -0.542238715 0.63662030
## 26 -0.24534578 -0.21751810 0.695937334 -0.007139066 0.47627258
## 27 -0.20960264 -0.58180437 0.394392498 -0.518763303 0.37889078
## 28 -0.26982346 -0.42908491 0.450840875 -0.202161091 0.02367337
## 29 -0.32058272 -0.29318472 0.453475403 -0.212532176 0.10144194
## 31 -0.38641680 -0.38582263 0.554819836 0.173638980 0.67867808
## 32 -0.31712244 -0.11381172 0.791643497 0.096268499 0.32899087
## 33 -0.32213376 -0.48103699 1.038751044 0.205328512 0.30745238
## 34 -0.33013501 -0.65581527 1.293037498 0.231480875 0.26741920
## 41 -0.21326809 -0.50951184 0.308211438 0.621624073 -0.02416227
## 42 -0.26935232 0.21916126 0.572210470 -0.361020644 0.59665773
## 43 -0.22062318 0.11645607 0.340636401 0.151210561 0.27108509
## 44 -0.26028393 -0.26173176 0.172017848 0.244822665 0.23292430
## 45 -0.23379351 -0.25943724 0.278658873 0.250480329 0.21130657
## 46 -0.33192013 0.74464867 0.130344698 -0.929207679 0.45776144
## 47 0.34630198 0.63422133 -0.042007884 -0.730381834 0.17446721
## 48 0.39131426 1.01748971 -0.358288180 0.112660259 -0.12592045
## 49 0.42699578 0.72427942 0.001705993 -0.893869698 0.29221358
## 50 -0.13939830 0.90998495 -0.158000274 -0.706916062 0.41680064
## 51 -0.27746336 -0.02797998 0.838975452 -0.400596491 0.68387800
## 53 0.42752854 -0.40185894 0.806015665 -0.411297259 0.40868170
## 54 0.03358955 -0.54014447 0.465576291 -0.403175303 0.33532396
## 55 0.40337693 -0.42511472 1.072441309 -0.737290886 0.79404728

```

### Exemplo 1: diversidade funcional (alfa) de 34 espécies de plantas em um gradiente de aridez

```
# Dados
```

```
comun_fren_dat # matriz de espécies por localidade
```

##	sp1	sp2	sp3	sp5	sp6	sp7	sp8	sp9	sp10	sp11	sp12	sp13	sp14	sp15	sp17	sp18	sp19
## 1	0	74	0	5	5	0	0	0	0	47	3	2	0	0	14	0	0
## 2	0	10	0	6	0	0	0	0	2	4	2	24	0	0	47	0	0
## 3	0	1	0	0	0	0	0	0	0	0	0	0	1	0	24	0	0
## 4	0	3	0	0	0	0	0	0	0	0	0	2	0	0	3	0	0
## 5	0	0	0	15	0	0	0	0	0	27	0	2	0	0	0	0	0
## 6	23	0	22	2	0	0	7	0	4	0	0	0	0	0	0	0	0
## 7	2	0	0	3	0	0	0	0	0	0	0	2	0	2	0	0	0
## 8	7	0	0	7	0	0	10	0	43	0	0	0	0	2	0	0	0
## 9	0	0	5	8	0	0	8	0	1	0	0	0	0	0	0	0	0
## 10	1	0	0	4	0	0	16	0	0	0	0	0	0	1	0	0	5
## 11	0	0	0	2	0	0	0	0	0	0	1	0	0	0	21	0	0
## 12	0	5	0	1	0	0	0	0	0	0	0	0	0	0	35	0	0
## 13	0	7	0	3	0	0	0	0	0	0	0	0	0	0	29	0	0
## 14	0	1	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0
## 15	0	8	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
## 17	0	0	0	34	0	0	29	0	8	0	0	0	6	0	0	3	0
## 18	0	0	0	3	0	0	0	59	0	0	0	9	9	0	0	0	30
## 19	0	0	0	21	3	0	128	0	1	0	0	3	1	0	0	0	0
## 20	0	0	0	8	0	0	0	38	6	0	0	7	1	0	0	1	0
## 21	0	0	0	1	1	0	0	0	0	0	0	17	0	0	0	138	0
## 22	0	0	0	0	1	0	0	0	0	0	4	17	0	0	6	0	14
## 23	0	0	0	0	4	0	0	0	0	0	0	12	0	0	1	0	0
## 24	0	0	0	1	1	0	0	0	3	0	2	1	0	0	50	0	5
## 25	0	0	0	0	4	0	0	0	0	0	0	8	0	0	22	36	0
## 26	0	0	0	0	0	0	0	0	0	47	25	0	0	0	6	0	0
## 27	0	11	0	1	0	29	0	0	0	0	221	0	0	0	9	0	0
## 28	0	47	0	1	24	0	0	0	0	47	78	0	0	0	28	0	0
## 29	0	18	0	1	0	0	0	0	0	0	66	0	0	0	23	0	0
## 31	15	0	134	11	0	0	0	0	0	0	0	0	0	0	0	0	0
## 32	0	0	1	82	0	0	5	0	0	0	0	0	0	46	0	0	0
## 33	0	0	0	53	0	0	1	0	0	0	0	0	0	57	0	0	7
## 34	0	0	0	15	0	0	2	0	3	0	0	0	0	26	0	0	0
## 41	0	0	0	8	3	0	0	0	0	171	15	1	0	0	0	0	0
## 42	0	0	0	19	0	0	0	0	2	32	0	0	0	0	0	0	0
## 43	0	0	2	3	1	0	0	0	12	62	2	1	4	0	0	0	0
## 44	0	0	11	4	12	0	0	0	2	66	12	31	0	0	0	0	19
## 45	0	0	0	10	2	0	0	0	0	84	6	0	3	0	0	0	0
## 46	0	0	0	4	20	0	16	0	10	0	0	1	3	0	0	0	0
## 47	0	0	0	12	13	0	0	45	14	0	0	0	11	0	0	0	0
## 48	0	0	0	7	0	0	133	0	1	0	0	13	3	0	0	18	0



## 49	0	0	0	1	5	0	0	50	0	0	0	1	0	0	0	0	0
## 50	0	0	0	10	22	0	75	0	4	0	0	0	45	0	0	0	0
## 51	0	0	0	1	2	0	0	0	0	3	0	44	0	0	0	0	0
## 53	0	0	0	0	13	0	0	1	0	0	0	109	0	0	0	66	0
## 54	0	0	1	0	12	0	0	0	0	0	113	78	0	0	41	45	0
## 55	0	0	16	0	55	0	0	0	0	0	0	11	0	0	23	94	36
##	sp20	sp21	sp22	sp23	sp24	sp25	sp27	sp28	sp29	sp30	sp31	sp32	sp33	sp34	sp35		
## 1	1	0	0	0	21	5	0	0	1	4	0	113	0	0	0	4	
## 2	0	0	0	0	5	21	0	0	0	63	0	29	0	0	0	0	
## 3	0	9	1	0	64	12	0	1	30	6	0	21	0	0	0	0	
## 4	1	5	0	0	8	1	0	0	0	0	0	298	0	0	0	0	
## 5	6	14	1	0	42	6	0	5	0	15	0	21	0	0	0	1	
## 6	0	0	14	13	0	0	0	0	0	42	0	0	2	4	0	0	
## 7	0	0	1	2	0	0	41	0	0	71	0	0	80	19	0	0	
## 8	0	0	3	0	0	0	0	0	0	59	0	0	13	12	0	0	
## 9	0	0	1	2	0	0	14	0	0	139	0	0	9	3	0	0	
## 10	0	0	0	17	0	0	0	0	0	135	0	0	75	16	0	0	
## 11	12	7	3	0	16	10	0	9	0	48	1	5	0	0	0	0	
## 12	6	5	11	0	103	0	0	0	0	27	0	17	0	0	0	0	
## 13	1	22	14	0	70	0	0	0	0	63	0	6	0	0	0	0	
## 14	48	5	2	0	63	1	0	12	0	3	0	7	0	0	0	0	
## 15	9	11	1	0	57	0	0	12	0	0	1	31	0	0	0	0	
## 17	0	0	0	68	0	0	0	0	0	28	0	0	0	0	0	0	
## 18	0	0	0	52	0	0	0	0	0	31	5	0	0	0	0	0	
## 19	0	0	0	6	0	1	0	0	0	41	0	0	0	0	0	0	
## 20	0	0	0	35	0	1	0	0	0	24	1	0	0	0	0	0	
## 21	0	1	6	0	0	0	0	0	0	1	0	80	0	0	0	2	
## 22	0	0	7	1	0	1	0	0	0	7	0	46	0	2	19	0	
## 23	0	0	2	0	0	0	0	0	0	20	2	112	0	0	0	0	
## 24	0	0	2	2	0	1	0	0	0	8	0	73	0	0	0	0	
## 25	0	0	2	1	0	0	0	0	0	2	0	97	0	1	1	0	
## 26	0	0	0	0	0	2	0	3	0	2	0	142	0	0	0	0	
## 27	0	0	0	0	0	0	0	0	0	0	0	13	0	0	1	0	
## 28	0	1	2	0	34	1	0	1	0	0	0	124	0	9	38	0	
## 29	2	1	2	0	3	0	0	0	12	0	0	110	0	20	0	0	
## 31	0	0	1	0	0	0	0	0	0	8	0	0	67	65	0	0	
## 32	0	0	0	0	0	0	0	0	0	2	0	0	123	10	0	0	
## 33	0	0	0	0	0	0	8	0	0	5	0	0	68	213	0	0	
## 34	0	0	0	0	0	0	0	0	0	1	0	0	106	140	0	0	
## 41	2	7	1	0	4	24	0	8	0	28	0	16	0	3	0	0	
## 42	12	3	0	0	3	25	0	0	0	46	0	194	0	0	0	0	
## 43	26	16	0	0	0	7	0	0	0	1	3	107	0	0	0	0	
## 44	3	12	0	11	0	3	0	1	0	2	1	34	0	12	0	0	
## 45	1	2	1	0	0	64	0	0	0	2	0	57	0	1	0	0	
## 46	0	0	0	17	0	7	0	0	0	162	24	0	0	1	0	0	
## 47	0	0	2	113	0	18	0	0	0	100	3	0	0	0	0	0	

```
## 48    0    0    2  59    0    1    0    0    0    61    0    0    0    0    0
## 49    0    0    0  33    0    2    0    0    0   162   20    0    0    0    0
## 50    0    0    0  21    0    3    0    0    0   182   11    0    0    0    0
## 51    0    0    0    3    0    2    0    0    0    1    0  315    0    0   49
## 53    0    1    0    0    0    0    0   10    0    6    3   36    0    0    0
## 54    0    0    0    0    0    1    0    1    0    0    0    1    0    0    0
## 55    0    0    2    0    0    0    0    0    0    0    0  165    0    0    0
```

```
ambie_fren_dat # matriz de variáveis ambientais por localidade
```

```
##      Grazing Aridity Exclosure Bare.ground.2009 Altitude
## 1      Grazed  5.003          0              77.6      928
## 2      Grazed  5.003          0              84.9      976
## 3      Grazed  5.109          0              85.7      936
## 4      Grazed  5.109          0              59.2      937
## 5      Grazed  5.003          0              85.0      937
## 6      Grazed  2.625          0              78.8     1628
## 7      Grazed  2.625          0              75.2     1599
## 8      Grazed  2.985          0              73.7     1611
## 9      Grazed  2.625          0              71.8     1598
## 10     Grazed  2.985          0              72.5     1593
## 11     Grazed  5.109          0              81.5      917
## 12     Grazed  5.109          0              81.3      929
## 13     Grazed  5.109          0              75.2      941
## 14     Grazed  5.109          0              84.9      945
## 15     Grazed  5.109          0              92.1      927
## 17     Grazed  4.254          0              75.3     1286
## 18     Grazed  4.088          0              76.7     1314
## 19     Grazed  4.425          0              62.8     1302
## 20     Grazed  4.425          0              74.8     1302
## 21     Grazed  4.536          0              82.6      978
## 22     Grazed  4.533          0              90.1      919
## 23     Grazed  4.533          0              85.7      921
## 24     Grazed  4.397          0              90.3      911
## 25     Grazed  3.903          0              84.4      938
## 26 Ungrazed  5.003          13              75.7      950
## 27 Ungrazed  5.003          13              75.3      946
## 28 Ungrazed  5.003          13              75.3      947
## 29 Ungrazed  5.003          13              59.8      955
## 31 Ungrazed  2.625           5              76.5     1628
## 32 Ungrazed  2.625           5              43.5     1617
## 33 Ungrazed  2.625           5              57.0     1611
## 34 Ungrazed  2.625           5              68.5     1611
## 41 Ungrazed  4.965          50              84.8     1050
## 42 Ungrazed  4.965          50              45.8     1048
## 43 Ungrazed  4.965          50              68.3     1063
```

```
## 44 Ungrazed 4.965      50      83.9    1064
## 45 Ungrazed 4.965      50      74.0    1068
## 46 Ungrazed 4.254       1      70.0    1291
## 47 Ungrazed 4.254       1      65.5    1297
## 48 Ungrazed 4.254       1      55.5    1293
## 49 Ungrazed 4.254       1      60.3    1302
## 50 Ungrazed 4.425       1      52.3    1302
## 51 Ungrazed 4.536      11      52.5     988
## 53 Ungrazed 3.906      11      80.5    1034
## 54 Ungrazed 3.906      11      79.3    1034
## 55 Ungrazed 3.906      11      59.3    1034
```

```
trait_fren_dat # matriz de atributos contínuos por espécie
```

```
##      LA      SLA      LDMC      LN15      LCC
## sp1  22.201  5.393 257.969  3.210 432.680
## sp2  27.803  3.642 194.628  6.257 281.276
## sp3  14.960  9.524 236.038  5.051 468.355
## sp5  30.828 15.949 244.372 -0.421 450.266
## sp6  50.863  4.505 315.679 -1.904 463.876
## sp7  286.552  8.619 201.724  3.193 450.805
## sp8  182.924 20.532 119.861  7.623 389.473
## sp9  629.033 15.377 143.187  3.340 393.802
## sp10 17.820 26.600 247.820  4.549 419.162
## sp11 55.044  3.751 301.926  8.052 404.560
## sp12 14.048  5.933 325.822  1.669 443.315
## sp13  3.643  9.843 209.086  4.929 420.098
## sp14 35.037 13.796 178.776 -0.400 462.883
## sp15 37.975  8.975 306.213 10.228 419.355
## sp17 10.163  8.379 154.654  5.182 403.914
## sp18 430.583  1.906 598.311 -1.167 462.227
## sp19  9.744  8.334 210.496  3.841 458.277
## sp20 25.373 11.740 231.553  4.023 405.811
## sp21 76.946 20.456 139.723  3.147 433.548
## sp22  4.296 12.278 227.964  3.833 426.892
## sp23 71.477 11.130 270.938  2.336 408.617
## sp24 78.794 13.624 235.067  1.997 420.955
## sp25 25.448 10.468 234.652  3.026 433.149
## sp27  2.255  5.392 279.853 10.324 425.468
## sp28  3.851  3.513 301.856  6.186 401.162
## sp29 23.231  4.510 158.147  5.936 294.205
## sp30  6.978 14.850 311.852 -0.157 457.671
## sp31  9.564 10.946 159.904  3.021 369.946
## sp32 40.091 10.578 358.145  2.499 455.474
## sp33 18.449  4.445 432.189  4.480 432.073
## sp34 29.501  4.860 410.760  3.689 433.906
```

```
## sp35 21.275 3.577 441.302 0.170 448.661
```

```
# Hipótese e predições:
```

```
# Hipótese: O gradiente de aridez influencia a divergência e regularidade funcional de  
# Predição 1: locais mais áridos possuem menor divergência funcional de plantas (métri  
# Predição 2: locais mais úmidos possuem menor regularidade funcional de plantas (métri
```

```
## Passo 1: calcular a distância funcional
```

```
trait_pad <- decostand(trait_fren_dat, "standardize")  
euclid_dis <- vegdist(trait_pad, "euclidean")
```

```
## Passo 2: calcular a Divergência funcional (FDis) e Regularidade Funcional (FEve)
```

```
fdis <- dbFD(euclid_dis, comun_fren_dat)$FDis # Fdis=0 em locais com somente uma espécie
```

```
## FRic: No dimensionality reduction was required. All 5 PCoA axes were kept as 'traits'.  
## CWM: When 'x' is a distance matrix, CWM cannot be calculated.
```

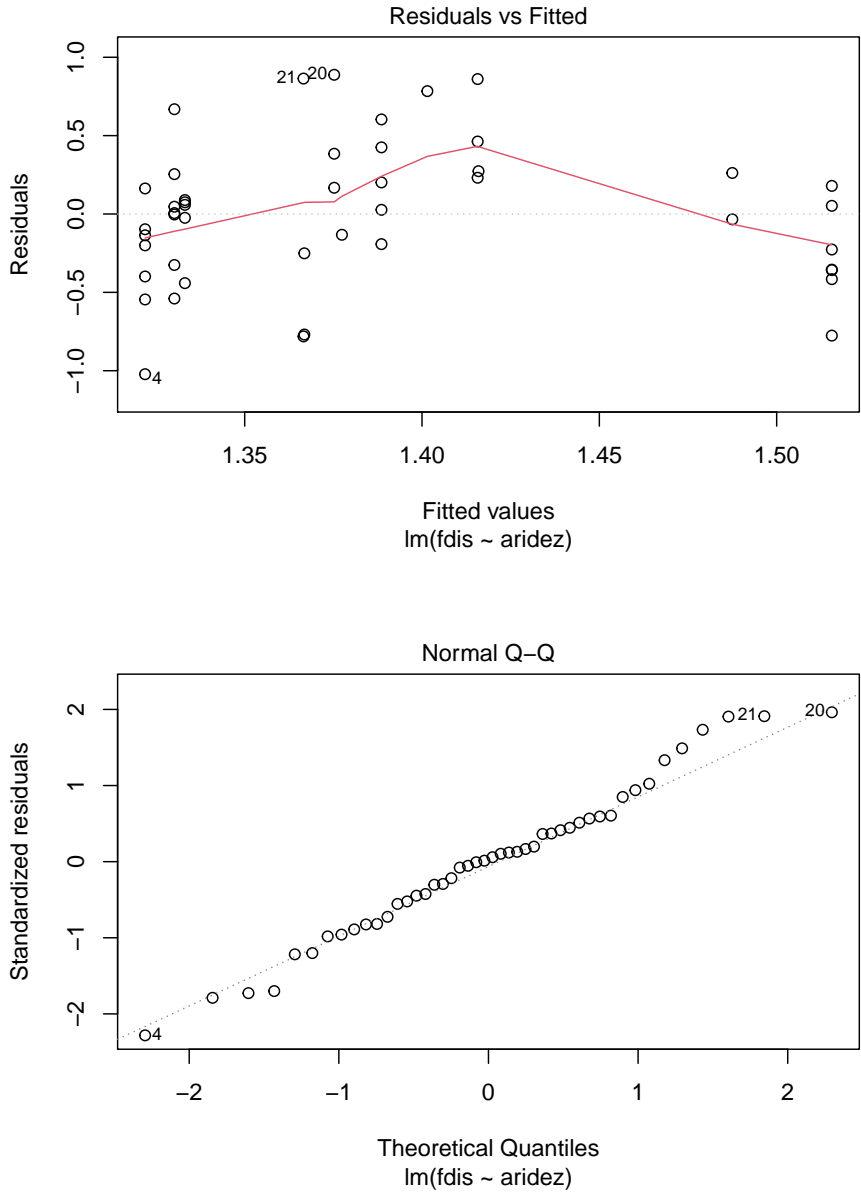
```
feve <- dbFD(euclid_dis, comun_fren_dat)$FEve
```

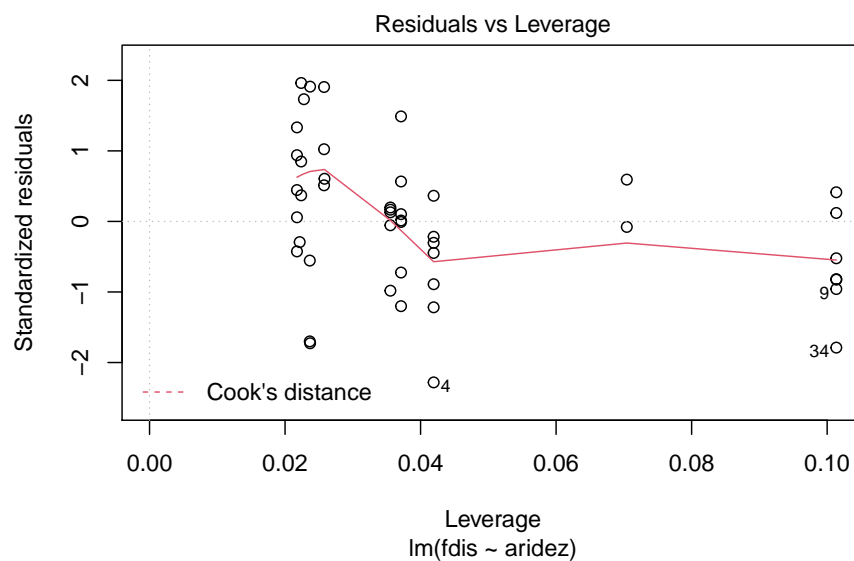
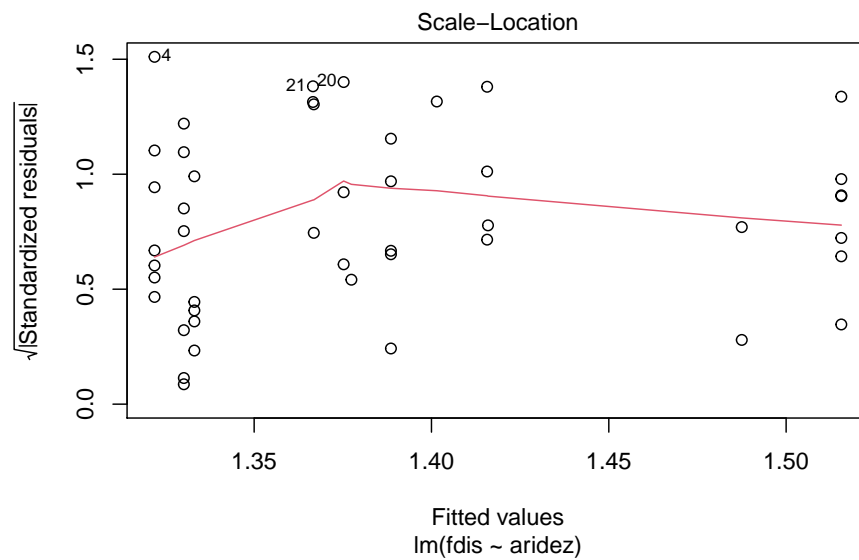
```
## FRic: No dimensionality reduction was required. All 5 PCoA axes were kept as 'traits'.  
## CWM: When 'x' is a distance matrix, CWM cannot be calculated.
```

```
## Passo 3: Utilizar um modelo linear para comparar o efeito da aridez sobre FDis (pre
```

```
lm_dat <- data.frame(aridez = ambie_fren_dat$Aridity, fdis = fdis, feve = feve)
```

```
mod1 <- lm(fdis ~ aridez, data = lm_dat)  
plot(mod1) # conferir os pressupostos da análise
```





```
anova(mod1) # Conclusão: a aridez não tem efeito sobre a divergência funcional
```

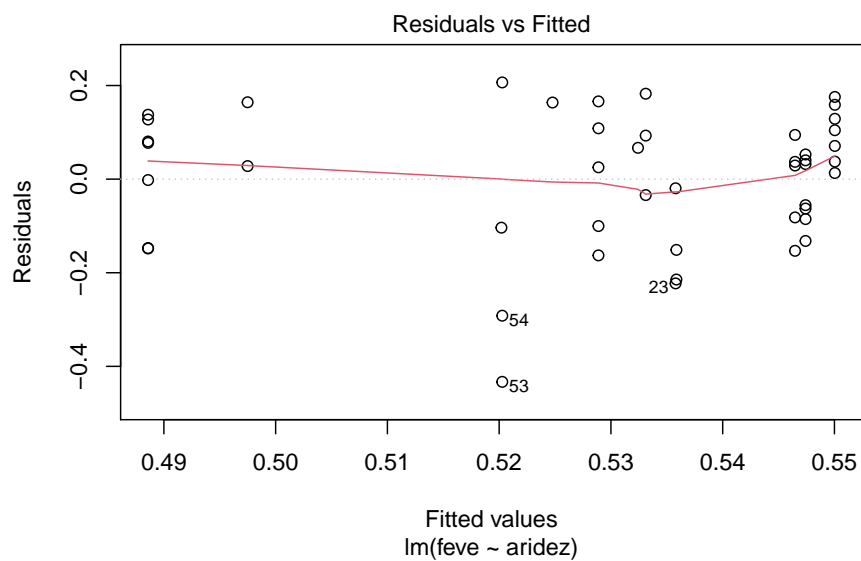
```
## Analysis of Variance Table
```

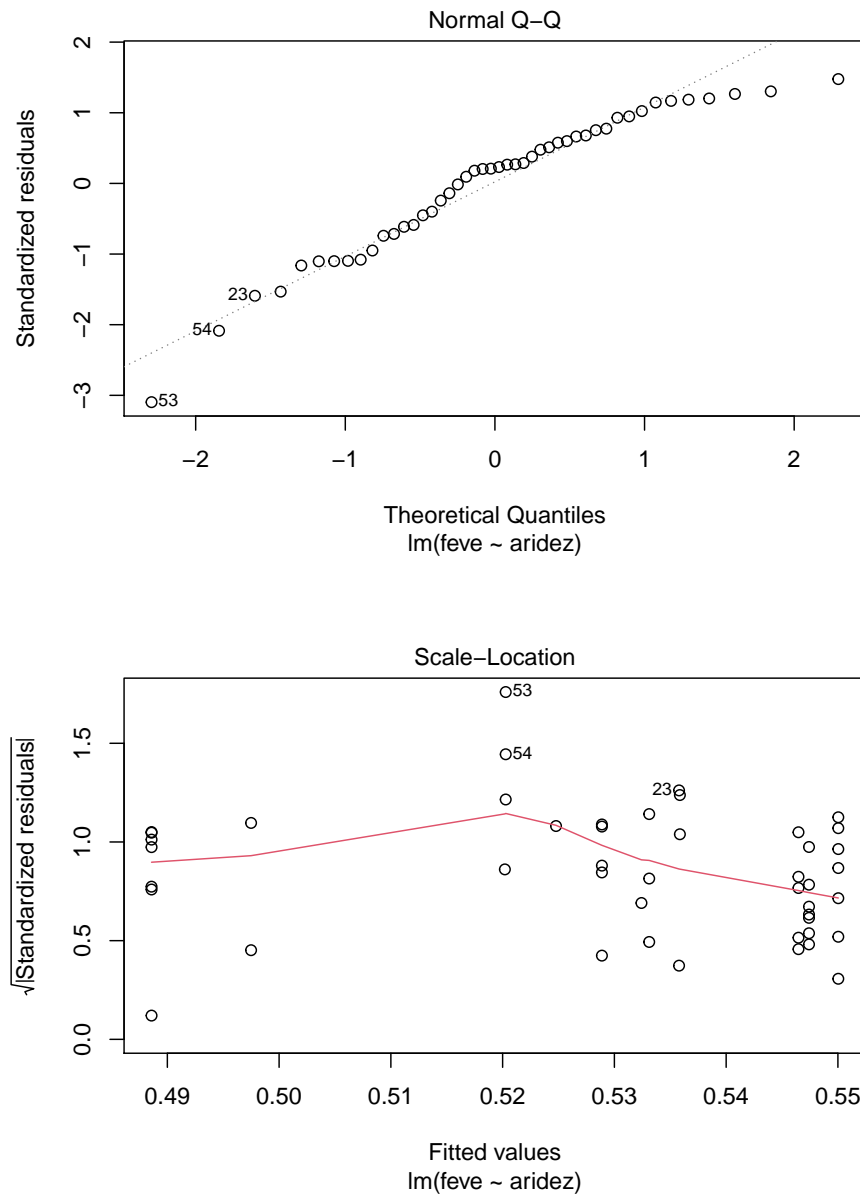
```
##
```

```
## Response: fdis
```

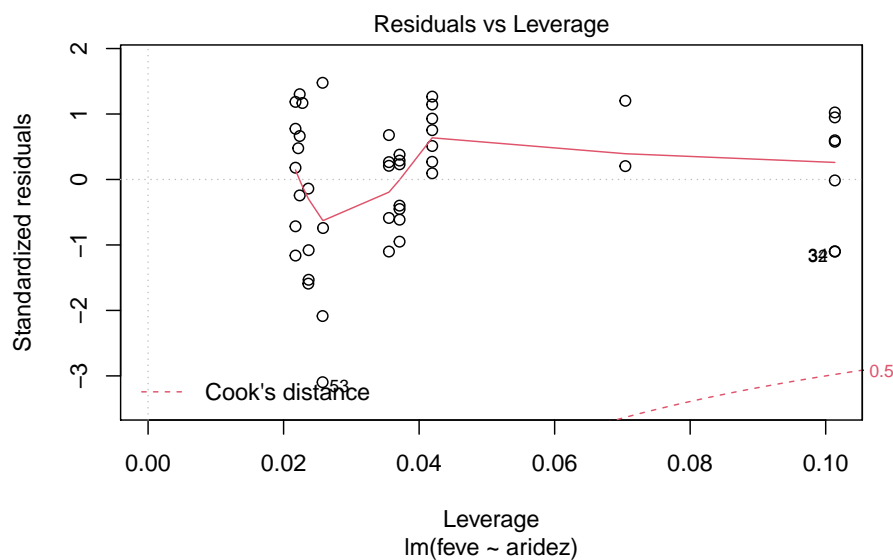
```
##           Df Sum Sq Mean Sq F value Pr(>F)
## aridez     1 0.2083 0.20834   0.9945 0.3241
## Residuals 44 9.2179 0.20950

mod2 <- lm(feve ~ aridez, data = lm_dat)
plot(mod2) # conferir os pressupostos da análise
```







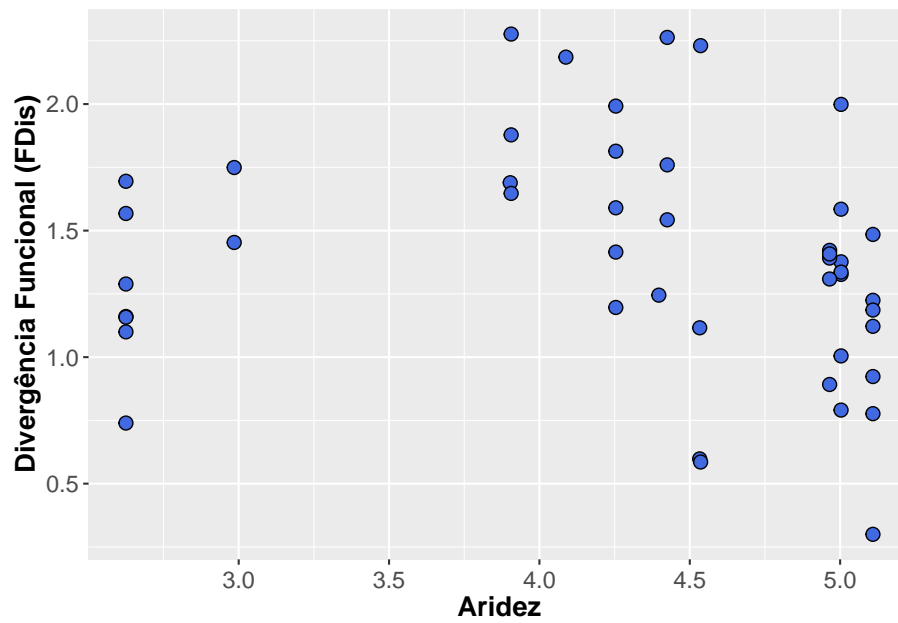


```
anova(mod2) # Conclusão: a aridez não tem efeito sobre a regularidade funcional
```

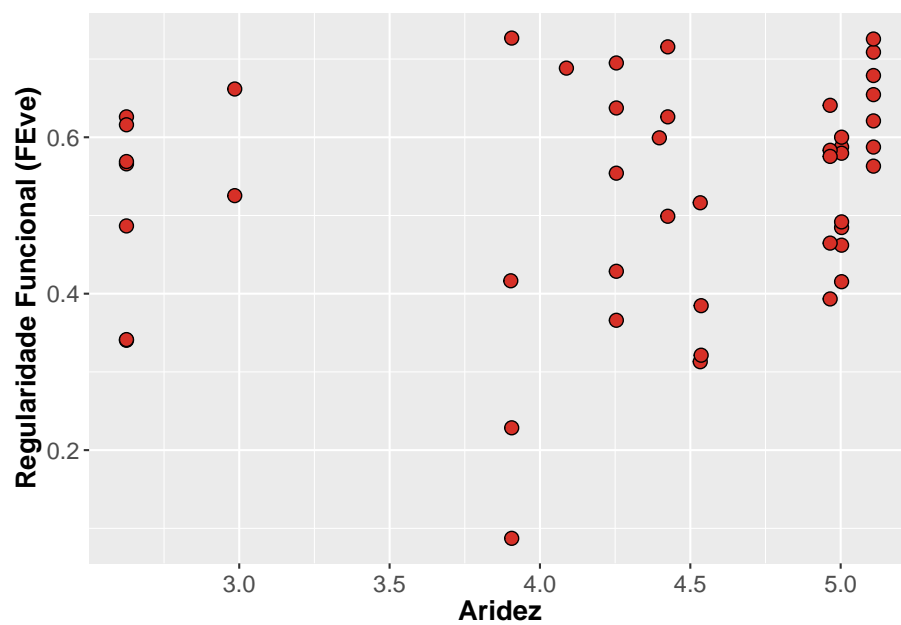
```
## Analysis of Variance Table
##
## Response: feve
##           Df Sum Sq Mean Sq F value Pr(>F)
## aridez      1  0.02098  0.020979   1.0447  0.3123
## Residuals 44  0.88353  0.020080

## Passo 4: gráfico para visualizar os dois resultados

lm_dat %>%
  ggplot(aes(x=aridez, y=fdis)) +
  geom_point(pch=21, size=3, color = "black", fill="royalblue") +
  xlab("Aridez") + ylab("Divergência Funcional (FDis)") +
  theme(axis.title.x = element_text(face="bold", size=14),
        axis.text.x = element_text(vjust=0.5, size=12)) +
  theme(axis.title.y = element_text(face="bold", size=14),
        axis.text.y = element_text(vjust=0.5, size=12)) +
  theme(legend.position = "top", legend.title=element_blank()) -> plot_pred1
plot_pred1
```

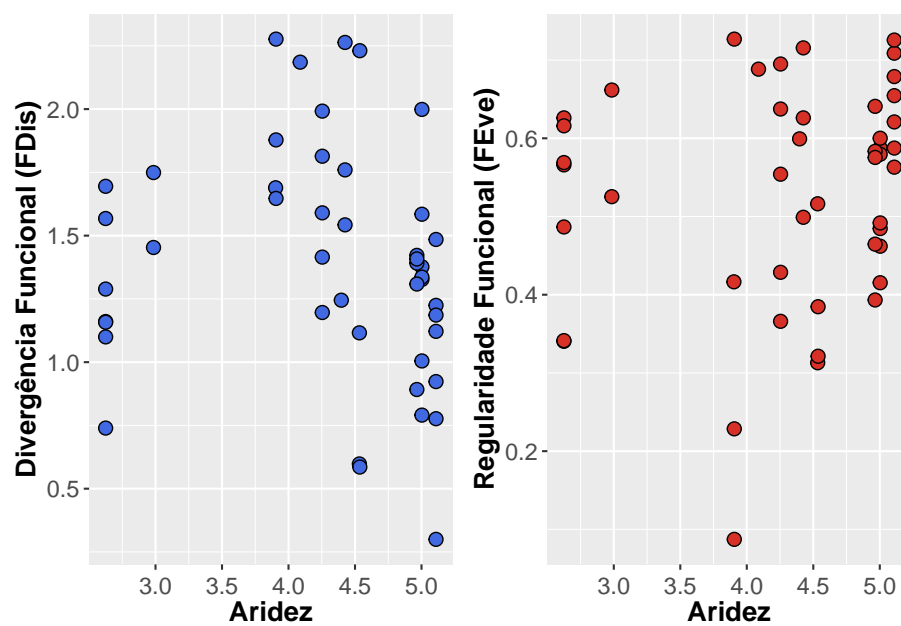


```
lm_dat %>%
  ggplot(aes(x=aridez, y=feve)) +
  geom_point(pch=21, size=3, color = "black", fill="#d73027") +
  xlab("Aridez") + ylab("Regularidade Funcional (FEve)") +
  theme(axis.title.x = element_text(face="bold", size=14),
        axis.text.x = element_text(vjust=0.5, size=12)) +
  theme(axis.title.y = element_text(face="bold", size=14),
        axis.text.y = element_text(vjust=0.5, size=12)) +
  theme(legend.position = "top", legend.title=element_blank()) -> plot_pred2
plot_pred2
```



```
## Visualização dos dois gráficos em um única janela
```

```
grid.arrange(plot_pred1, plot_pred2, ncol=2)
```



```
## Exportar resultad em pdf. Você pode mudar .pdf por .png ou .svg.
ggsave("plot_pred1.pdf", plot_pred1, height = 14, width = 14, dpi = 600, units = "cm")
ggsave("plot_pred2.pdf", plot_pred2, height = 14, width = 14, dpi = 600, units = "cm")
```

```
### Exemplo 2: o pastejo influencia a diversidade beta funcional de 34 espécies de plantas
```

```
# Dados
```

```
comun_fren_dat # matriz de espécies por localidade
```

##	sp1	sp2	sp3	sp5	sp6	sp7	sp8	sp9	sp10	sp11	sp12	sp13	sp14	sp15	sp17	sp18	sp19
## 1	0	74	0	5	5	0	0	0	0	47	3	2	0	0	14	0	0
## 2	0	10	0	6	0	0	0	0	2	4	2	24	0	0	47	0	0
## 3	0	1	0	0	0	0	0	0	0	0	0	0	1	0	24	0	0
## 4	0	3	0	0	0	0	0	0	0	0	0	2	0	0	3	0	0
## 5	0	0	0	15	0	0	0	0	0	27	0	2	0	0	0	0	0
## 6	23	0	22	2	0	0	7	0	4	0	0	0	0	0	0	0	0
## 7	2	0	0	3	0	0	0	0	0	0	0	2	0	2	0	0	0
## 8	7	0	0	7	0	0	10	0	43	0	0	0	0	2	0	0	0
## 9	0	0	5	8	0	0	8	0	1	0	0	0	0	0	0	0	0
## 10	1	0	0	4	0	0	16	0	0	0	0	0	0	1	0	0	5
## 11	0	0	0	2	0	0	0	0	0	0	1	0	0	0	21	0	0
## 12	0	5	0	1	0	0	0	0	0	0	0	0	0	0	35	0	0
## 13	0	7	0	3	0	0	0	0	0	0	0	0	0	0	29	0	0
## 14	0	1	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0
## 15	0	8	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
## 17	0	0	0	34	0	0	29	0	8	0	0	0	6	0	0	3	0
## 18	0	0	0	3	0	0	0	59	0	0	0	9	9	0	0	0	30
## 19	0	0	0	21	3	0	128	0	1	0	0	3	1	0	0	0	0
## 20	0	0	0	8	0	0	0	38	6	0	0	7	1	0	0	1	0
## 21	0	0	0	1	1	0	0	0	0	0	0	17	0	0	0	138	0
## 22	0	0	0	0	1	0	0	0	0	0	4	17	0	0	6	0	14
## 23	0	0	0	0	4	0	0	0	0	0	0	12	0	0	1	0	0
## 24	0	0	0	1	1	0	0	0	3	0	2	1	0	0	50	0	5
## 25	0	0	0	0	4	0	0	0	0	0	0	8	0	0	22	36	0
## 26	0	0	0	0	0	0	0	0	0	47	25	0	0	0	6	0	0
## 27	0	11	0	1	0	29	0	0	0	0	221	0	0	0	9	0	0
## 28	0	47	0	1	24	0	0	0	0	47	78	0	0	0	28	0	0
## 29	0	18	0	1	0	0	0	0	0	0	66	0	0	0	23	0	0
## 31	15	0	134	11	0	0	0	0	0	0	0	0	0	0	0	0	0
## 32	0	0	1	82	0	0	5	0	0	0	0	0	0	46	0	0	0
## 33	0	0	0	53	0	0	1	0	0	0	0	0	0	57	0	0	7
## 34	0	0	0	15	0	0	2	0	3	0	0	0	0	26	0	0	0
## 41	0	0	0	8	3	0	0	0	0	171	15	1	0	0	0	0	0
## 42	0	0	0	19	0	0	0	0	2	32	0	0	0	0	0	0	0

## 43	0	0	2	3	1	0	0	0	12	62	2	1	4	0	0	0	0
## 44	0	0	11	4	12	0	0	0	2	66	12	31	0	0	0	0	19
## 45	0	0	0	10	2	0	0	0	0	84	6	0	3	0	0	0	0
## 46	0	0	0	4	20	0	16	0	10	0	0	1	3	0	0	0	0
## 47	0	0	0	12	13	0	0	45	14	0	0	0	11	0	0	0	0
## 48	0	0	0	7	0	0	133	0	1	0	0	13	3	0	0	18	0
## 49	0	0	0	1	5	0	0	50	0	0	0	1	0	0	0	0	0
## 50	0	0	0	10	22	0	75	0	4	0	0	0	45	0	0	0	0
## 51	0	0	0	1	2	0	0	0	0	3	0	44	0	0	0	0	0
## 53	0	0	0	0	13	0	0	1	0	0	0	109	0	0	0	66	0
## 54	0	0	1	0	12	0	0	0	0	0	113	78	0	0	41	45	0
## 55	0	0	16	0	55	0	0	0	0	0	0	11	0	0	23	94	36
##	sp20	sp21	sp22	sp23	sp24	sp25	sp27	sp28	sp29	sp30	sp31	sp32	sp33	sp34	sp35		
## 1	1	0	0	0	21	5	0	0	1	4	0	113	0	0	0	4	
## 2	0	0	0	0	5	21	0	0	0	63	0	29	0	0	0	0	
## 3	0	9	1	0	64	12	0	1	30	6	0	21	0	0	0	0	
## 4	1	5	0	0	8	1	0	0	0	0	0	298	0	0	0	0	
## 5	6	14	1	0	42	6	0	5	0	15	0	21	0	0	0	1	
## 6	0	0	14	13	0	0	0	0	0	42	0	0	2	4	0	0	
## 7	0	0	1	2	0	0	41	0	0	71	0	0	80	19	0	0	
## 8	0	0	3	0	0	0	0	0	0	59	0	0	13	12	0	0	
## 9	0	0	1	2	0	0	14	0	0	139	0	0	9	3	0	0	
## 10	0	0	0	17	0	0	0	0	0	135	0	0	75	16	0	0	
## 11	12	7	3	0	16	10	0	9	0	48	1	5	0	0	0	0	
## 12	6	5	11	0	103	0	0	0	0	27	0	17	0	0	0	0	
## 13	1	22	14	0	70	0	0	0	0	63	0	6	0	0	0	0	
## 14	48	5	2	0	63	1	0	12	0	3	0	7	0	0	0	0	
## 15	9	11	1	0	57	0	0	12	0	0	1	31	0	0	0	0	
## 17	0	0	0	68	0	0	0	0	0	28	0	0	0	0	0	0	
## 18	0	0	0	52	0	0	0	0	0	31	5	0	0	0	0	0	
## 19	0	0	0	6	0	1	0	0	0	41	0	0	0	0	0	0	
## 20	0	0	0	35	0	1	0	0	0	24	1	0	0	0	0	0	
## 21	0	1	6	0	0	0	0	0	0	1	0	80	0	0	0	2	
## 22	0	0	7	1	0	1	0	0	0	7	0	46	0	2	19	0	
## 23	0	0	2	0	0	0	0	0	0	20	2	112	0	0	0	0	
## 24	0	0	2	2	0	1	0	0	0	8	0	73	0	0	0	0	
## 25	0	0	2	1	0	0	0	0	0	2	0	97	0	1	1	0	
## 26	0	0	0	0	0	2	0	3	0	2	0	142	0	0	0	0	
## 27	0	0	0	0	0	0	0	0	0	0	0	13	0	0	1	0	
## 28	0	1	2	0	34	1	0	1	0	0	0	124	0	9	38	0	
## 29	2	1	2	0	3	0	0	0	12	0	0	110	0	20	0	0	
## 31	0	0	1	0	0	0	0	0	0	8	0	0	67	65	0	0	
## 32	0	0	0	0	0	0	0	0	0	2	0	0	123	10	0	0	
## 33	0	0	0	0	0	0	8	0	0	5	0	0	68	213	0	0	
## 34	0	0	0	0	0	0	0	0	0	1	0	0	106	140	0	0	
## 41	2	7	1	0	4	24	0	8	0	28	0	16	0	3	0	0	

```

## 42 12 3 0 0 3 25 0 0 0 46 0 194 0 0 0
## 43 26 16 0 0 0 7 0 0 0 1 3 107 0 0 0
## 44 3 12 0 11 0 3 0 1 0 2 1 34 0 12 0
## 45 1 2 1 0 0 64 0 0 0 2 0 57 0 1 0
## 46 0 0 0 17 0 7 0 0 0 162 24 0 0 1 0
## 47 0 0 2 113 0 18 0 0 0 100 3 0 0 0 0
## 48 0 0 2 59 0 1 0 0 0 61 0 0 0 0 0
## 49 0 0 0 33 0 2 0 0 0 162 20 0 0 0 0
## 50 0 0 0 21 0 3 0 0 0 182 11 0 0 0 0
## 51 0 0 0 3 0 2 0 0 0 1 0 315 0 0 49
## 53 0 1 0 0 0 0 0 10 0 6 3 36 0 0 0
## 54 0 0 0 0 0 1 0 1 0 0 0 1 0 0 0
## 55 0 0 2 0 0 0 0 0 0 0 0 165 0 0 0

```

```

ambie_fren_dat # matriz de variáveis ambientais por localidade

```

```

##      Grazing Aridity Exclosure Bare.ground.2009 Altitude
## 1      Grazed 5.003          0              77.6      928
## 2      Grazed 5.003          0              84.9      976
## 3      Grazed 5.109          0              85.7      936
## 4      Grazed 5.109          0              59.2      937
## 5      Grazed 5.003          0              85.0      937
## 6      Grazed 2.625          0              78.8     1628
## 7      Grazed 2.625          0              75.2     1599
## 8      Grazed 2.985          0              73.7     1611
## 9      Grazed 2.625          0              71.8     1598
## 10     Grazed 2.985          0              72.5     1593
## 11     Grazed 5.109          0              81.5      917
## 12     Grazed 5.109          0              81.3      929
## 13     Grazed 5.109          0              75.2      941
## 14     Grazed 5.109          0              84.9      945
## 15     Grazed 5.109          0              92.1      927
## 17     Grazed 4.254          0              75.3     1286
## 18     Grazed 4.088          0              76.7     1314
## 19     Grazed 4.425          0              62.8     1302
## 20     Grazed 4.425          0              74.8     1302
## 21     Grazed 4.536          0              82.6      978
## 22     Grazed 4.533          0              90.1      919
## 23     Grazed 4.533          0              85.7      921
## 24     Grazed 4.397          0              90.3      911
## 25     Grazed 3.903          0              84.4      938
## 26 Ungrazed 5.003          13              75.7      950
## 27 Ungrazed 5.003          13              75.3      946
## 28 Ungrazed 5.003          13              75.3      947
## 29 Ungrazed 5.003          13              59.8      955
## 31 Ungrazed 2.625           5              76.5     1628

```

```
## 32 Ungrazed 2.625      5      43.5    1617
## 33 Ungrazed 2.625      5      57.0    1611
## 34 Ungrazed 2.625      5      68.5    1611
## 41 Ungrazed 4.965     50      84.8    1050
## 42 Ungrazed 4.965     50      45.8    1048
## 43 Ungrazed 4.965     50      68.3    1063
## 44 Ungrazed 4.965     50      83.9    1064
## 45 Ungrazed 4.965     50      74.0    1068
## 46 Ungrazed 4.254      1      70.0    1291
## 47 Ungrazed 4.254      1      65.5    1297
## 48 Ungrazed 4.254      1      55.5    1293
## 49 Ungrazed 4.254      1      60.3    1302
## 50 Ungrazed 4.425      1      52.3    1302
## 51 Ungrazed 4.536     11      52.5     988
## 53 Ungrazed 3.906     11      80.5    1034
## 54 Ungrazed 3.906     11      79.3    1034
## 55 Ungrazed 3.906     11      59.3    1034
```

```
trait_fren_dat # matriz de atributos contínuos por espécie
```

```
##      LA      SLA      LDMC      LN15      LCC
## sp1  22.201  5.393 257.969  3.210 432.680
## sp2  27.803  3.642 194.628  6.257 281.276
## sp3  14.960  9.524 236.038  5.051 468.355
## sp5  30.828 15.949 244.372 -0.421 450.266
## sp6  50.863  4.505 315.679 -1.904 463.876
## sp7  286.552  8.619 201.724  3.193 450.805
## sp8  182.924 20.532 119.861  7.623 389.473
## sp9  629.033 15.377 143.187  3.340 393.802
## sp10 17.820 26.600 247.820  4.549 419.162
## sp11 55.044  3.751 301.926  8.052 404.560
## sp12 14.048  5.933 325.822  1.669 443.315
## sp13  3.643  9.843 209.086  4.929 420.098
## sp14 35.037 13.796 178.776 -0.400 462.883
## sp15 37.975  8.975 306.213 10.228 419.355
## sp17 10.163  8.379 154.654  5.182 403.914
## sp18 430.583  1.906 598.311 -1.167 462.227
## sp19  9.744  8.334 210.496  3.841 458.277
## sp20 25.373 11.740 231.553  4.023 405.811
## sp21 76.946 20.456 139.723  3.147 433.548
## sp22  4.296 12.278 227.964  3.833 426.892
## sp23 71.477 11.130 270.938  2.336 408.617
## sp24 78.794 13.624 235.067  1.997 420.955
## sp25 25.448 10.468 234.652  3.026 433.149
## sp27  2.255  5.392 279.853 10.324 425.468
## sp28  3.851  3.513 301.856  6.186 401.162
```

```
## sp29 23.231 4.510 158.147 5.936 294.205
## sp30 6.978 14.850 311.852 -0.157 457.671
## sp31 9.564 10.946 159.904 3.021 369.946
## sp32 40.091 10.578 358.145 2.499 455.474
## sp33 18.449 4.445 432.189 4.480 432.073
## sp34 29.501 4.860 410.760 3.689 433.906
## sp35 21.275 3.577 441.302 0.170 448.661
```

```
# Hipótese e predições:
```

```
# Hipótese: o pastejo determina a ocorrência de espécies de plantas com diferentes atr
# Predição 1: a composição funcional de plantas é diferente entre áreas com e sem past
```

```
## Passo 1: calcular a distância funcional
```

```
cwm_dis <- vegdist(cwm_fren, "euclidean")
```

```
## Passo 2: calcular os valores de composição funcional (CWM)
```

```
cwm_fren <- funtcomp(trait_pad, as.matrix(comun_fren_dat))
```

```
## Passo 3: testar se a composição funcional varia entre as áreas com uma PERMANOVA
```

```
perman_fren <- adonis(cwm_fren~Grazing, data = ambie_fren_dat)
```

```
## Passo 4: comparar a variação dentro de cada grupo com Betadisper
```

```
betad_fren <- betadisper(cwm_dis, ambie_fren_dat$Grazing)
permutest(betad_fren)
```

```
##
## Permutation test for homogeneity of multivariate dispersions
## Permutation: free
## Number of permutations: 999
##
## Response: Distances
##           Df Sum Sq Mean Sq      F N.Perm Pr(>F)
## Groups    1  0.0539 0.053858 0.1946   999  0.682
## Residuals 44 12.1763 0.276735
```

```
## Passo 5: visualização com PCoA
```

```
plot(betad_fren)
```



