

Introdução ao R com aplicações em biodiversidade e conservação

2020-06-22

Contents

1	Pré-requisitos	5
2	Rarefação	7
2.1	Backgorund da análise	7
2.2	Exemplo prático 1 - Morcegos	8
2.3	Exemplo prático 2 - Rarefação	10
2.4	Para se aprofundar	12
3	Estimadores de Riqueza	13
3.1	Backgorund da análise	13
3.2	Estimadores baseados na abundância das espécies	14
3.3	Estimadores baseados na incidência das espécies	20

Chapter 1

Pré-requisitos

Chapter 2

Rarefação

2.1 Backgorund 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.

2.2 Exemplo prático 1 - Morcegos

2.2.1 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

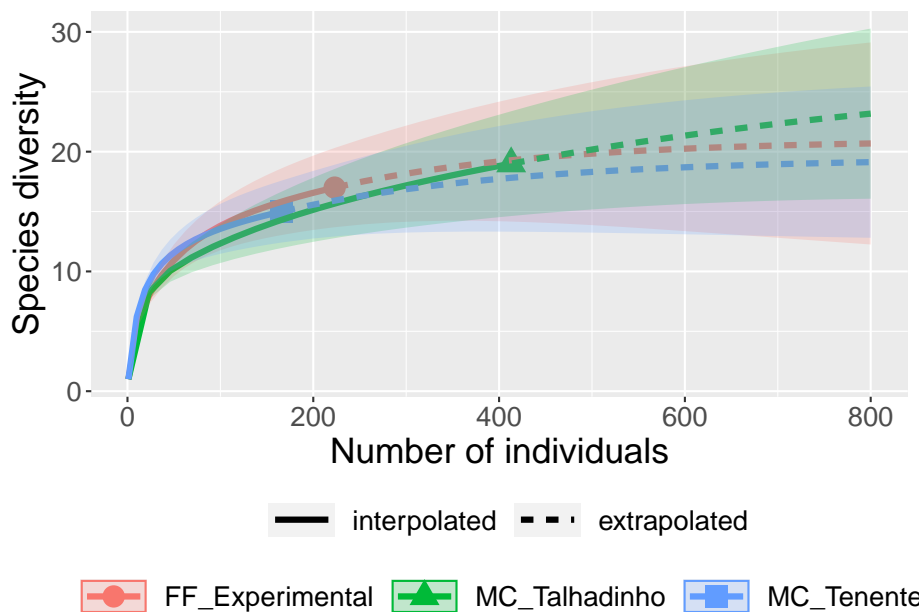
2.2.2 Análise

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
# 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)
```



2.2.3 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.

2.3 Exemplo prático 2 - Rarefação

2.3.1 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.

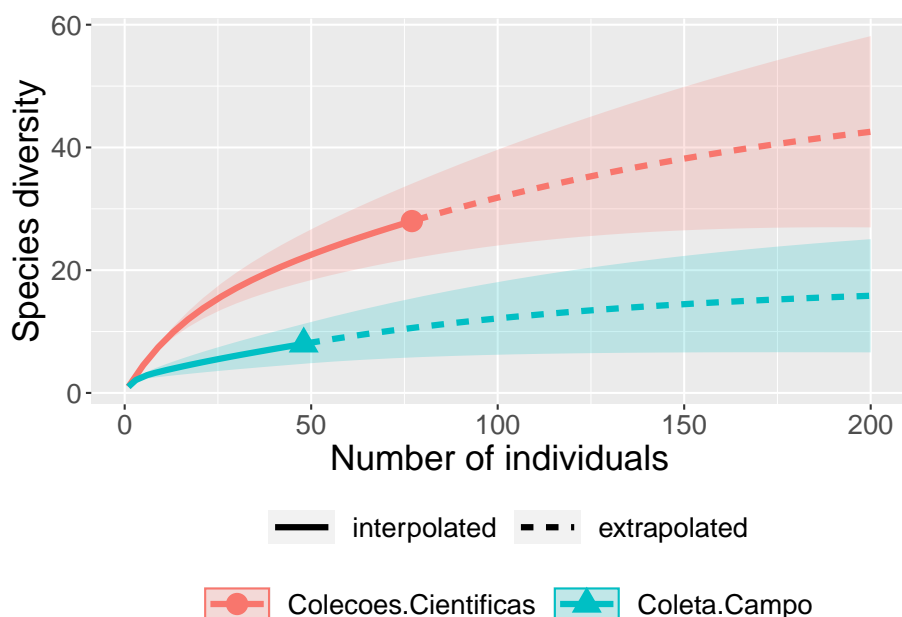
2.3.2 Análise

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)
```



2.3.3 Interpretação dos resultados

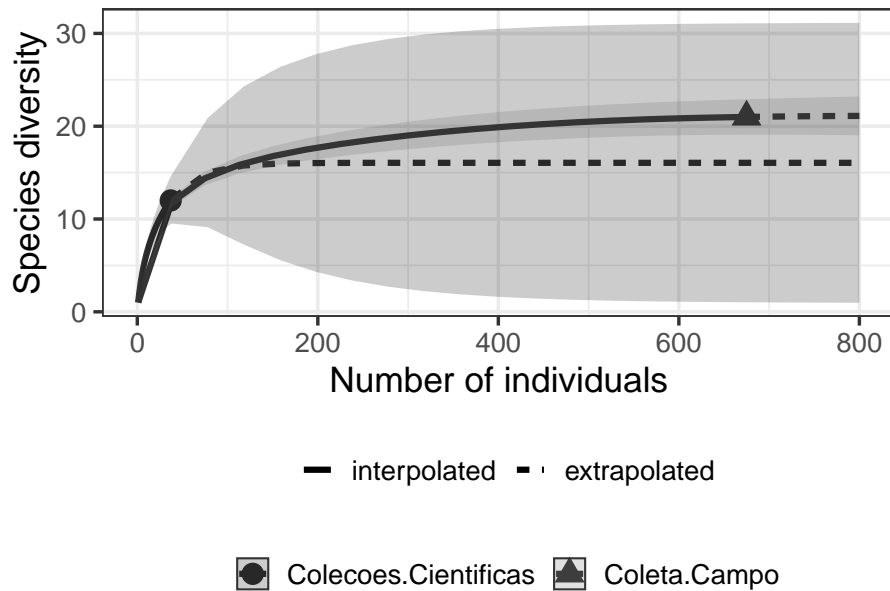
Neste exemplo, foram registradas oito espécies de répteis nas coletas de campo (40 indivíduos) e 28 espécies nas coleções científicas (77 indivíduos). Com base na rarefação, concluímos que a riqueza de espécies de répteis obtida nas coleções científicas é 2,5 vezes maior do que a obtida em coletas de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

2.4 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 3

Estimadores de Riqueza

3.1 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).

3.1.0.1 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.

3.2 Estimadores baseados na abundância das espécies

3.2.1 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,
- F1 = número de espécies observadas com abundância de um indivíduo (espécies *singleton*),
- F2 = 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.

3.2.1.1 Exemplo prático - Chao 1

3.2.1.1.1 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 aumentasse 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 uma 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

3.2.2 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_6  1  6.687500  3.0000 12.33333 2.708178
## Dia_4  2  9.362095  6.0000 15.69375 2.647881
## Dia_9  3 11.042500  6.0000 17.52500 2.754868
## Dia_12 4 11.933333  9.0000 16.00000 2.212130
## Dia_14 5 12.846667  9.0000 17.52500 2.289350
## Dia_13 6 14.155000 10.0000 21.05000 2.908852
## Dia_2  7 14.816667 10.0000 22.00000 3.047096
## Dia_8  8 15.760000 11.7375 22.00000 2.929715
## Dia_1  9 16.435000 13.0000 22.00000 2.638413
## Dia_7 10 16.821667 13.0000 22.00000 2.614629
## Dia_11 11 17.518333 13.7375 22.00000 2.325733
## Dia_3 12 18.490000 14.0000 22.00000 2.309379
## Dia_5 13 19.460000 15.5000 22.00000 1.664665
## Dia_10 14 20.000000 20.0000 20.00000 0.000000
##
## attr(,"class")
## [1] "summary.poolaccum"
```

Visualizar os resultados com intervalo de confiança de 95%.

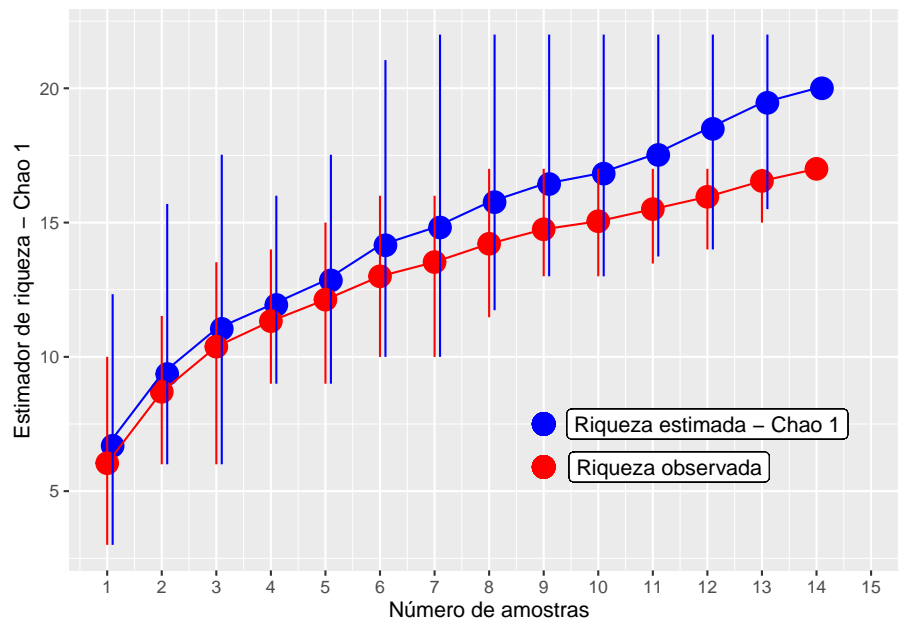
```
library(ggplot2)
# preparando os dados para fazer o gráfico
resultados <- summary(est_chao1, display = c("S", "chao"))
res_chao <- cbind(resultados$chao[,1:4], resultados$S[,2:4])
res_chao <- as.data.frame(res_chao)
colnames(res_chao) <- c("Amostras", "Chao", "C_inferior", "C_superior", "Riqueza",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

```



3.2.2.1 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.

3.2.3 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.

3.2.3.1 Exemplo prático - ACE

3.2.3.1.1 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 aumentasse 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 uma 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

3.2.4 Análise

Calculo do estimador de riqueza - ACE

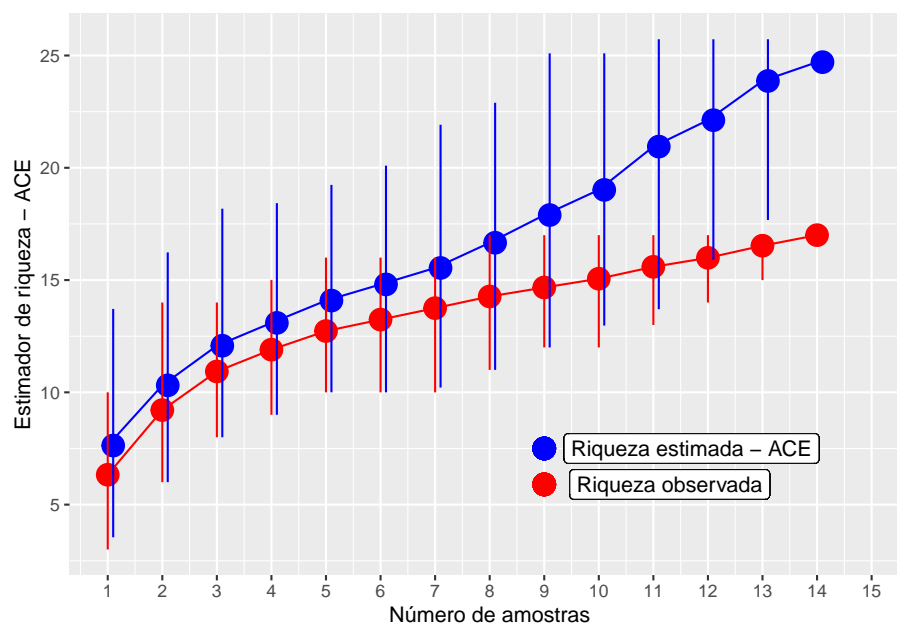
```
library(vegan)
dados_coleta <- poca_anuros
est_ace <- estaccumR(dados_coleta, permutations = 100)
summary(est_ace, display = "ace")
```

```
## $ace
##      N      ACE      2.5%      97.5% Std.Dev
## Dia_13  1  7.630291  3.54519 13.71429 2.905391
## Dia_5   2 10.311844  6.00000 16.23283 2.962658
## Dia_3   3 12.075617  8.00000 18.17742 2.781726
## Dia_1   4 13.092806  9.00000 18.42378 2.752028
## Dia_12  5 14.086487 10.00000 19.23721 2.661520
## Dia_10  6 14.806023 10.00000 20.09544 2.855354
## Dia_4   7 15.539865 10.21409 21.91599 2.882226
## Dia_8   8 16.661493 11.00000 22.89409 3.287395
## Dia_11  9 17.895562 12.00000 25.09679 3.751350
## Dia_2  10 19.015305 12.97421 25.09679 3.804308
## Dia_14 11 20.943241 13.70221 25.72368 3.855821
## Dia_9   12 22.119592 15.91136 25.72368 3.418184
## Dia_6   13 23.865516 17.67647 25.72368 2.347762
## Dia_7   14 24.703704 24.70370 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")
```



3.2.4.1 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.

3.3 Estimadores baseados na incidência das espécies

3.3.1 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 somente 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:

- Sobs = 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.

3.3.1.1 Exemplo prático - Chao 2

3.3.1.1.1 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 aumentasse 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 uma habitat reprodutivo

Checklist

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

3.3.2 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 12.84406  7.333333 22.00000 4.234087
## [2,]  4 13.84358  9.062500 23.27812 4.007125
## [3,]  5 15.64160  9.590000 34.12500 5.533777
## [4,]  6 17.56660 10.153646 35.41667 6.598625
## [5,]  7 19.92464 11.482143 36.52500 6.862326
## [6,]  8 21.18469 12.003646 36.50781 6.730503
## [7,]  9 22.62185 12.830556 36.82500 6.746445
## [8,] 10 25.18525 13.800000 42.20000 7.581706
## [9,] 11 27.95045 13.861364 42.45455 7.076700
## [10,] 12 29.12146 18.666667 39.45833 6.270209
## [11,] 13 31.36077 22.384615 39.61538 5.087386
## [12,] 14 33.71429 33.714286 33.71429 0.000000
##
## attr(,"class")
## [1] "summary.poolaccum"
```

Visualizar os resultados com intervalo de confiança de 95%

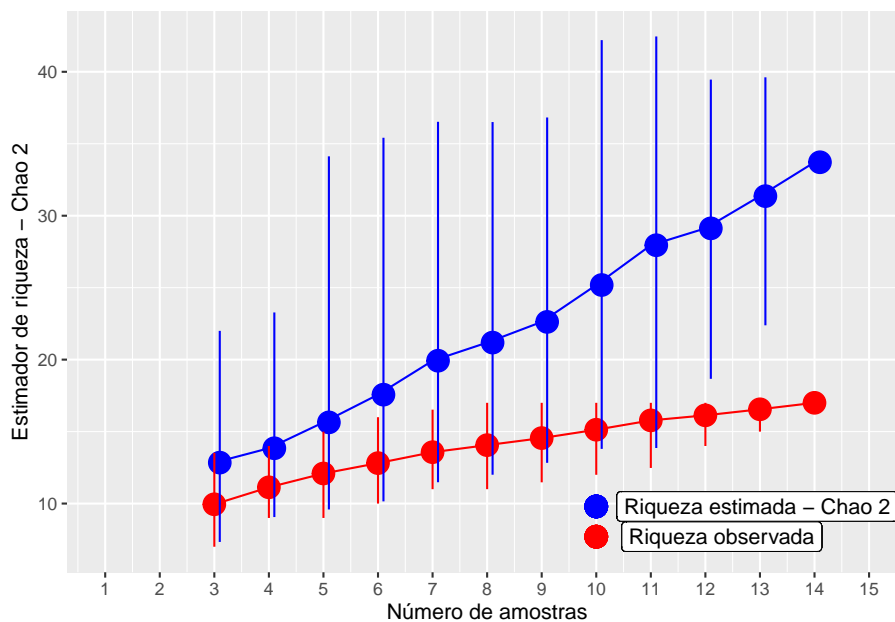
```
library(ggplot2)
# preparando os dados para fazer o gráfico
resultados_chao2 <- summary(est_chao2, display = c("S", "chao"))
res_chao2 <- cbind(resultados_chao2$chao[,1:4], resultados_chao2$S[,2:4])
res_chao2 <- as.data.frame(res_chao2)
colnames(res_chao2) <- c("Amostras", "Chao2", "C_inferior",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

```



3.3.2.1 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.

3.3.3 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.

3.3.3.1 Exemplo prático - Jackknife 1

3.3.3.1.1 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 aumentasse 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 uma habitat reprodutivo

Checklist

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

3.3.4 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      13.82000  8.333333 18.84167 2.881662
## [2,] 4      14.72250  9.750000 20.25000 2.832419
## [3,] 5      15.35200 11.180000 20.60000 2.579973
## [4,] 6      16.34167 12.062500 21.35833 2.512636
## [5,] 7      17.37143 12.714286 22.07500 2.633999
## [6,] 8      18.23375 13.225000 23.12500 2.696428
## [7,] 9      18.95222 14.666667 22.80000 2.325880
## [8,] 10     19.79000 15.227500 23.30000 2.305834
## [9,] 11     20.52364 16.202273 23.36364 2.135337
## [10,] 12     21.04833 17.666667 23.41667 1.877348
## [11,] 13     21.90692 18.692308 23.46154 1.373925
## [12,] 14     22.57143 22.571429 22.57143 0.000000
##
## attr(,"class")
## [1] "summary.poolaccum"
```

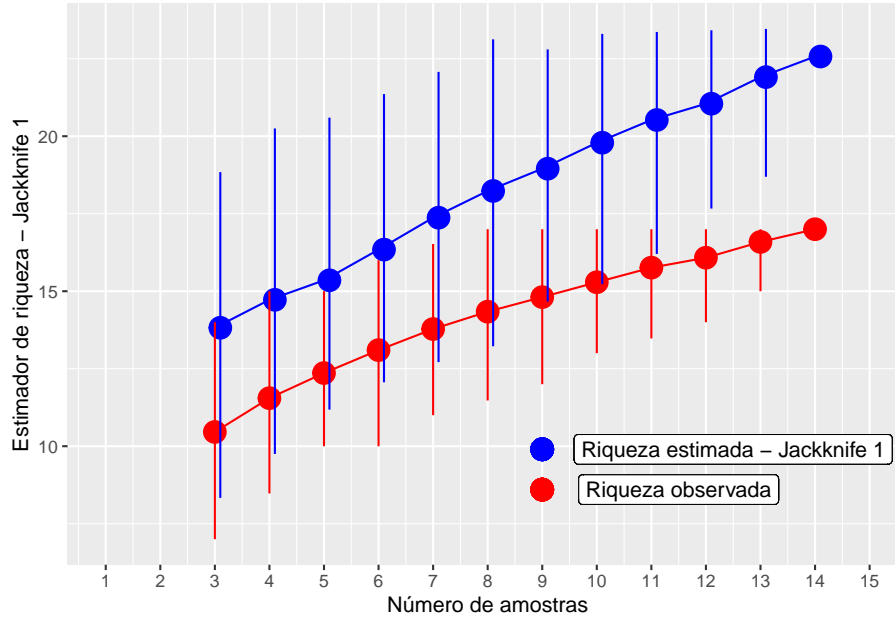
Visualizar os resultados com 95% intervalo de confiança

```
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")
```



3.3.4.1 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.

3.3.5 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:

- Sobs = 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*).

3.3.5.1 Exemplo prático - Jackknife 2

3.3.5.1.1 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 aumentasse 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 uma habitat reprodutivo

Checklist

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

3.3.6 Análise

Calculo do estimador de riqueza - Jackknife 2

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

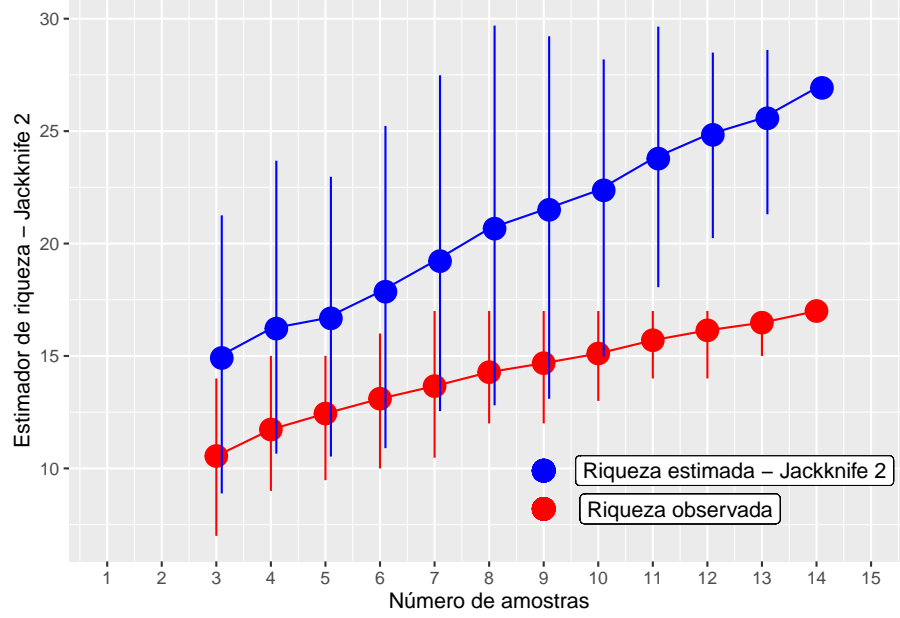
```
## $jack2
##      N Jackknife 2      2.5%    97.5% Std.Dev
## [1,]  3    14.91167  8.88750 21.25417 3.375944
## [2,]  4    16.21750 10.65833 23.68333 3.631004
```

```
## [3,] 5      16.67400 10.52750 22.96875 3.306609
## [4,] 6      17.85700 10.90083 25.23083 3.864266
## [5,] 7      19.21762 12.54762 27.48333 3.868007
## [6,] 8      20.66232 12.80580 29.69464 4.371045
## [7,] 9      21.50042 13.09722 29.21840 4.217831
## [8,] 10     22.36722 14.97778 28.18889 4.051951
## [9,] 11     23.77555 18.05886 29.64795 3.438322
## [10,] 12    24.83424 20.24242 28.49242 2.646664
## [11,] 13    25.56987 21.30128 28.60897 2.132235
## [12,] 14    26.92308 26.92308 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")
```



3.3.6.1 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.

3.3.7 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.

3.3.7.1 Exemplo prático - Bootstrap

3.3.7.1.1 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 aumentasse 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 uma habitat reprodutivo

Checklist

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

3.3.8 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")
```

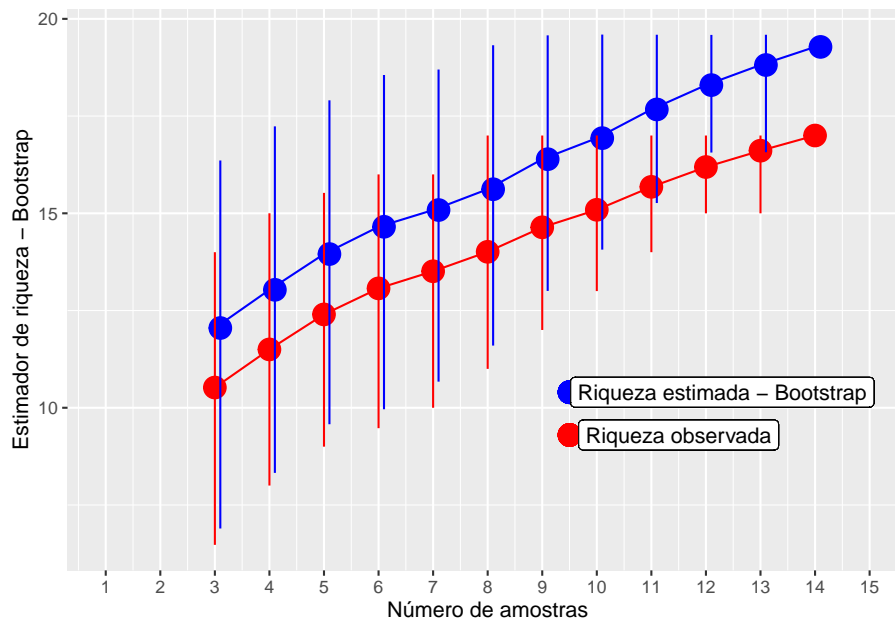
```
## $boot
##      N Bootstrap      2.5%    97.5%   Std.Dev
## [1,]  3  12.05074  6.895370 16.35833 2.4215370
## [2,]  4  13.03246  8.325684 17.23535 2.3246919
## [3,]  5  13.95204  9.576704 17.90485 2.1804251
## [4,]  6  14.65318  9.961938 18.55596 2.2006485
## [5,]  7  15.08415 10.671610 18.69690 2.1384204
## [6,]  8  15.61869 11.598992 19.32114 1.9593186
## [7,]  9  16.39149 13.004322 19.57541 1.7373933
## [8,] 10  16.93291 14.066321 19.59567 1.6126194
## [9,] 11  17.67048 15.264332 19.59145 1.3629752
## [10,] 12  18.29642 16.561824 19.58719 1.1907495
## [11,] 13  18.81307 16.570376 19.59107 0.9189265
```

```
## [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")
```



3.3.8.1 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.

3.3.9 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%.

3.3.9.1 Exemplo prático

3.3.9.1.1 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 aumentasse 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 uma habitat reprodutivo

Checklist

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

3.3.10 Análise

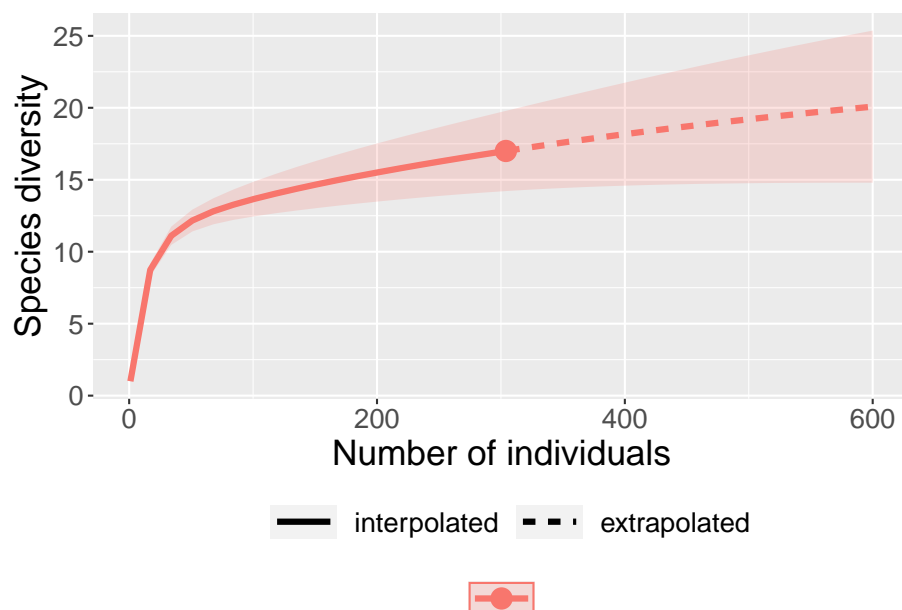
Calculo da extrapolação da riqueza com base no número de indivíduos

```
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)
```

3.3.10.1 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.

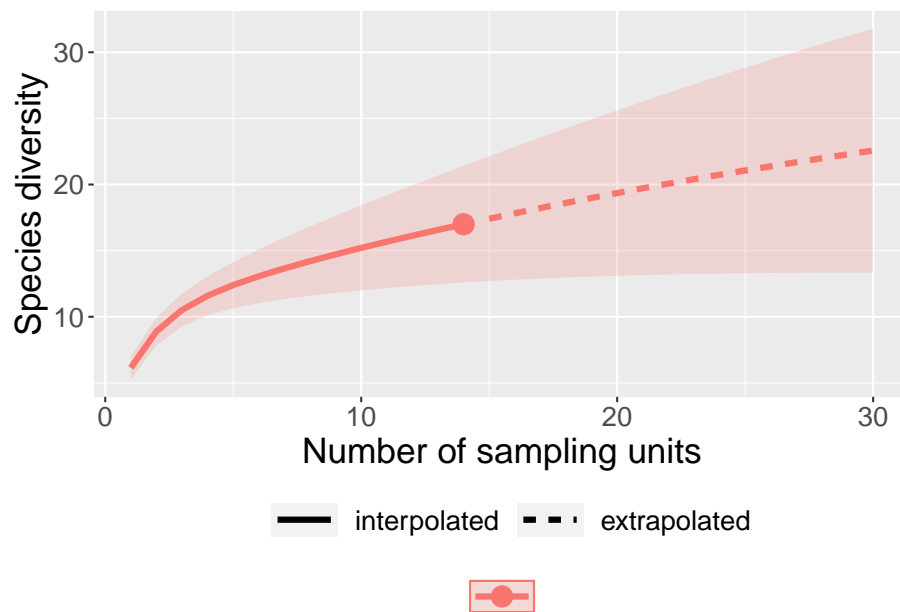
Calculo da extrapolação da riqueza com base no número de amostras

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

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



3.3.10.2 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.

3.3.11 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.