

Introdução ao R com aplicações em biodiversidade e conservação

2020-06-25

Contents

1	Pré-requisitos	5
2	Rarefação	7
2.1	Background da análise	7
2.2	Exemplo prático 1 - Morcegos	8
2.3	Exemplo prático 2 - Rarefação	11
2.4	Para se aprofundar	13
3	Estimadores de Riqueza	15
3.1	Backgorund da análise	15
3.2	Estimadores baseados na abundância das espécies	16
3.3	Estimadores baseados na incidência das espécies	23

Chapter 1

Pré-requisitos

Chapter 2

Rarefação

2.1 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.

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

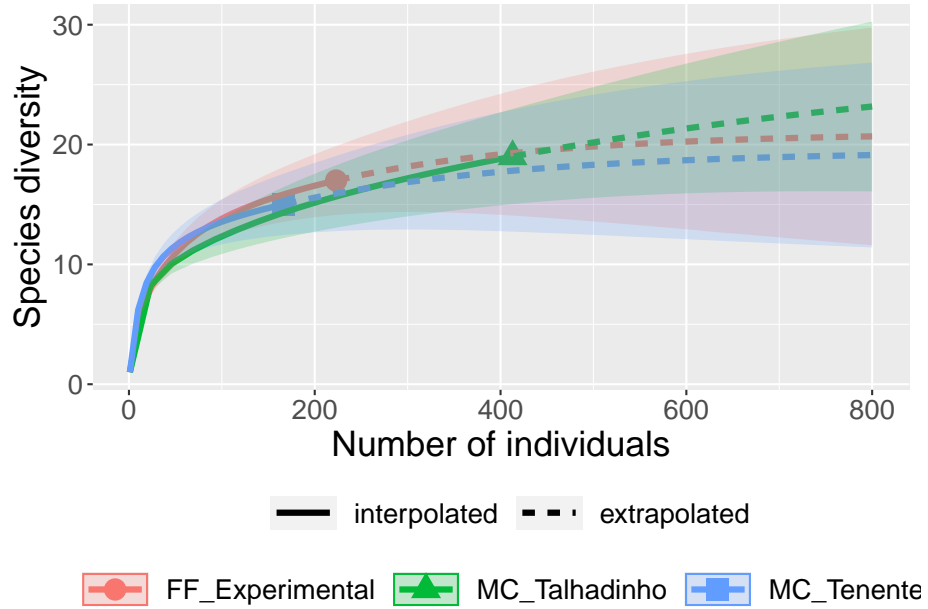
2.2.2 Análise

Cálculo 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 no 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.

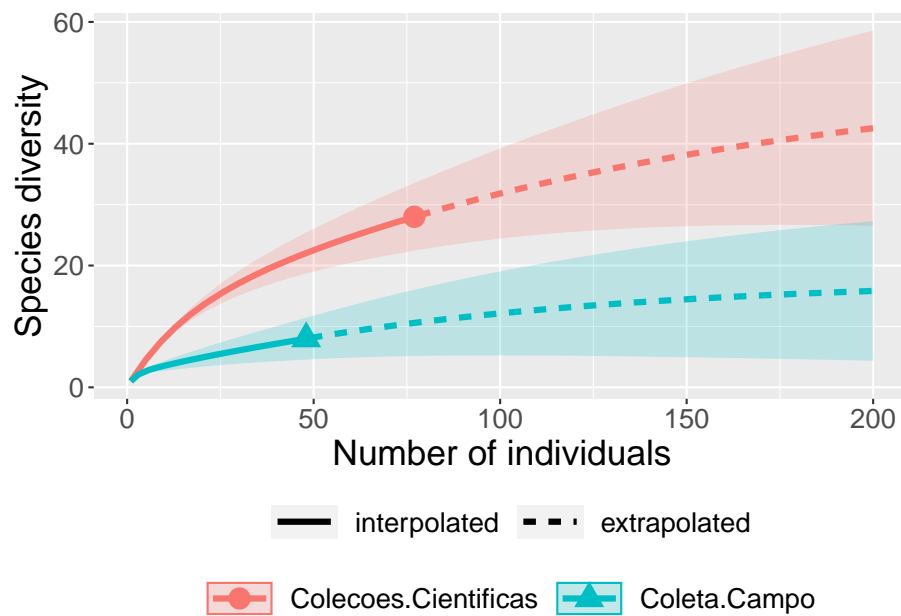
2.3.2 Análise

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)
```



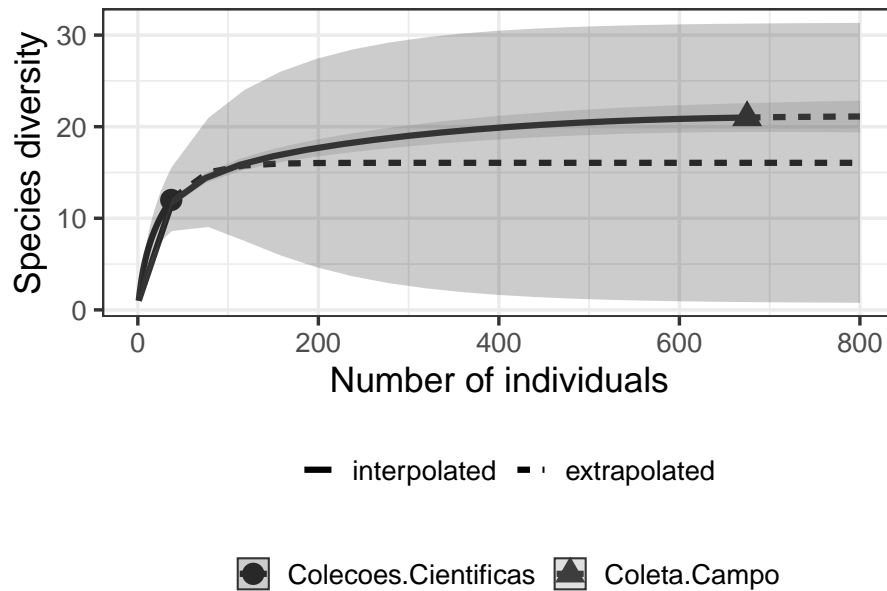
2.3.3 Interpretação dos resultados

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 =
# 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:

- S_{obs} = 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.

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

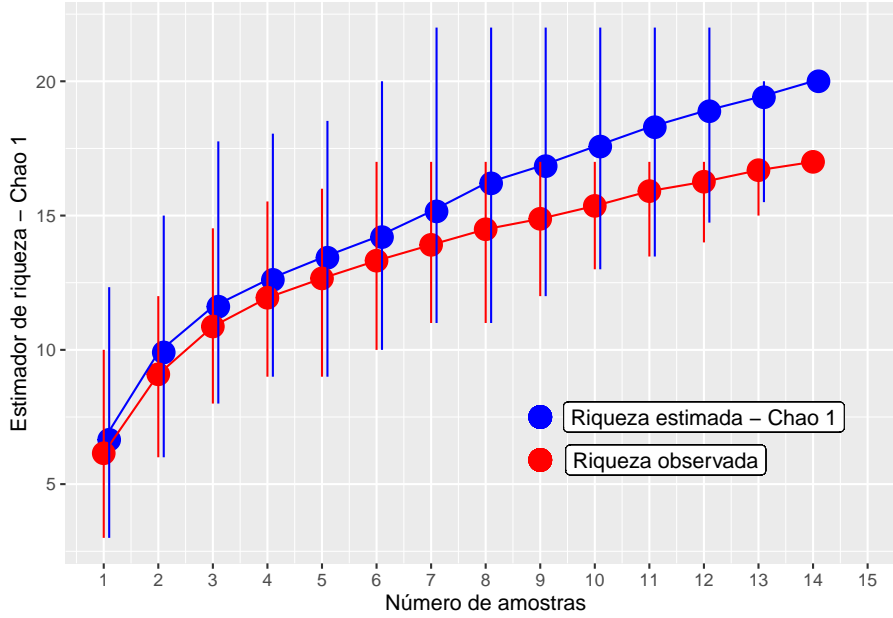
```
library(ecodados)
library(vegan)
dados_coleta <- poca_anuros
est_chao1 <- estaccumR(dados_coleta, permutations = 100)
summary(est_chao1, display = "chao")
```

```
## $chao
##           N      Chao    2.5%   97.5% Std.Dev
## Dia_12    1  6.651667  3.0000 12.33333 2.679640
## Dia_10    2  9.904167  6.0000 15.00000 2.655262
## Dia_3     3 11.607667  8.0000 17.76250 2.674637
## Dia_1     4 12.605833  9.0000 18.05000 2.526233
## Dia_4     5 13.430000  9.0000 18.52500 2.579399
## Dia_6     6 14.198333 10.0000 20.00000 2.652286
## Dia_9     7 15.155000 11.0000 22.00000 2.813533
## Dia_8     8 16.200000 11.0000 22.00000 2.840010
## Dia_2     9 16.840000 12.0000 22.00000 2.898868
## Dia_5    10 17.570000 13.0000 22.00000 2.651491
## Dia_14   11 18.290000 13.4750 22.00000 2.408088
## Dia_11   12 18.885000 14.7375 22.00000 2.106813
## Dia_7    13 19.405000 15.5000 20.00000 1.419018
## Dia_13   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 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

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  7.612776  3.761224 13.71429 2.853244
## Dia_2  2  9.733455  6.000000 15.74635 2.561675
## Dia_14 3 11.259911  8.000000 15.87885 2.078870
## Dia_7  4 12.425328  8.121225 18.54988 2.663095
## Dia_12 5 13.553731  8.475000 19.87776 2.973525
## Dia_10 6 14.446920  9.167937 20.02742 2.894626
## Dia_9  7 15.050844 10.000000 20.44469 2.805688
## Dia_6  8 15.997108 10.000000 22.76955 3.258119
## Dia_1  9 17.809288 12.000000 24.93470 3.776383
## Dia_8 10 19.377831 13.005455 25.09679 3.997948
## Dia_3 11 21.046964 13.686710 25.72368 3.862266
## Dia_4 12 22.454068 16.144275 25.72368 3.404189
## Dia_5 13 23.841765 17.676471 25.72368 2.416507
## Dia_11 14 24.703704 24.703704 24.70370 0.000000
##
## attr(,"class")
## [1] "summary.poolaccum"
```

Visualizar os resultados com intervalo de confiança de 95%

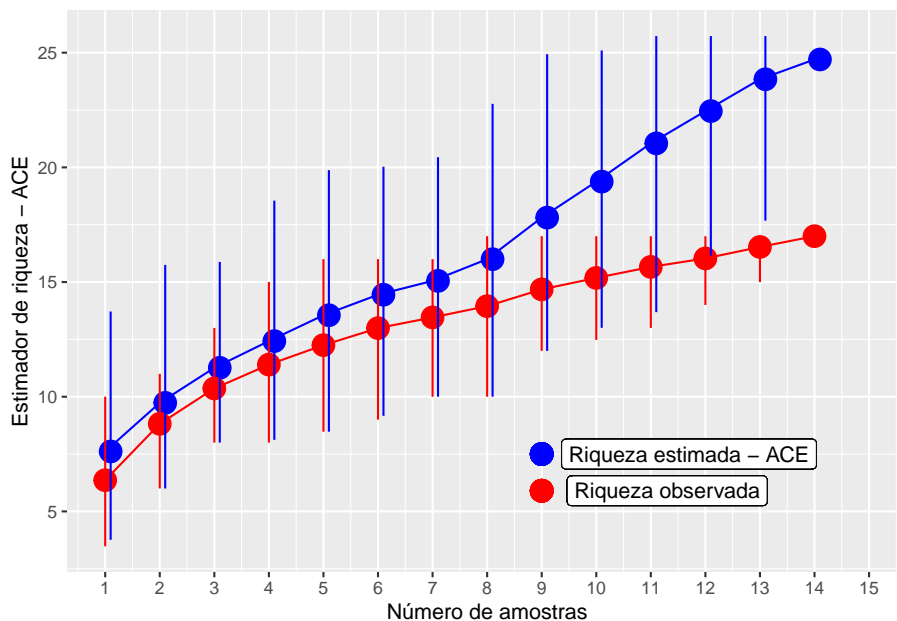
```
library(ggplot2)
# preparando os dados para fazer o gráfico
resultados_ace <- summary(est_ace, display = c("S", "ace"))
res_ace <- cbind(resultados_ace$ace[,1:4], resultados_ace$S[,2:4])
res_ace <- as.data.frame(res_ace)
colnames(res_ace) <- c("Amostras", "ACE", "ACE_inferior", "ACE_superior", "Riqueza",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

```



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

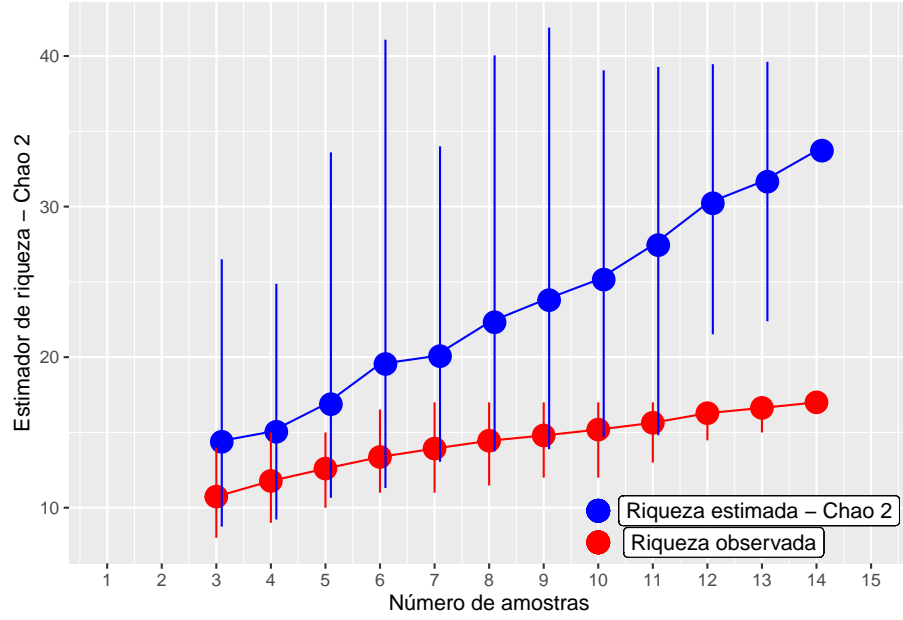
```
library(vegan)
dados_coleta <- poca_anuros
est_chao2 <- poolaccum(dados_coleta, permutations = 100)
summary(est_chao2, display = "chao")
```

```
## $chao
##      N      Chao      2.5%      97.5% Std.Dev
## [1,]  3 14.38152  8.75000 26.50000 4.559114
## [2,]  4 15.04590  9.21750 24.87656 4.407826
## [3,]  5 16.87697 10.66000 33.60000 5.472025
## [4,]  6 19.54569 11.30729 41.08333 7.397860
## [5,]  7 20.06679 13.04643 34.00000 5.732353
## [6,]  8 22.31917 13.74375 40.04531 6.941101
## [7,]  9 23.78593 13.88889 41.88889 7.477713
## [8,] 10 25.15075 14.70000 39.05000 6.931083
## [9,] 11 27.44227 14.81818 39.27273 6.403143
## [10,] 12 30.22021 21.52135 39.45833 5.415158
## [11,] 13 31.65308 22.38462 39.61538 4.480768
## [12,] 14 33.71429 33.71429 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:

- S_{obs} = 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

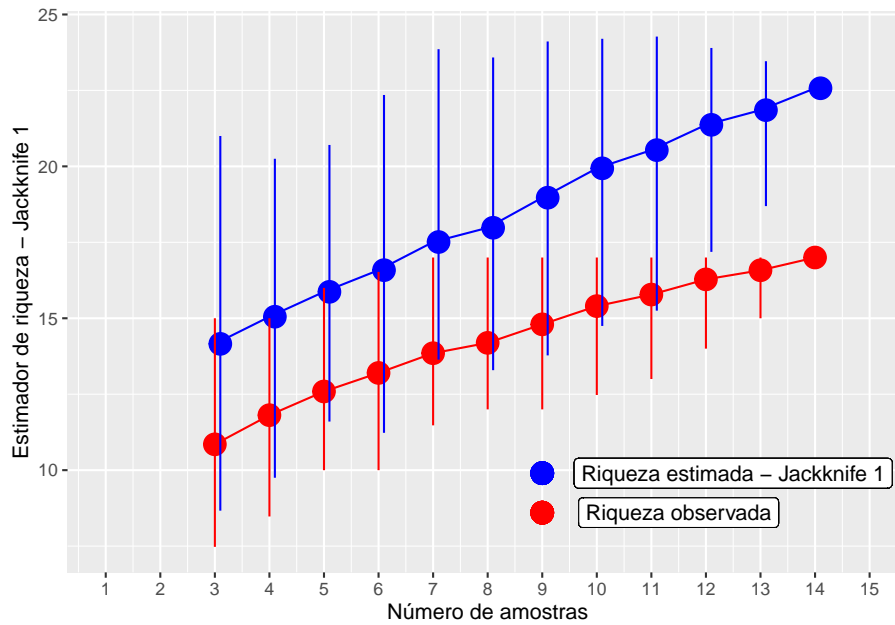
```
library(vegan)
dados_coleta <- poca_anuros
est_jack1 <- poolaccum(dados_coleta, permutations = 100)
summary(est_jack1, display = "jack1")
```

```
## $jack1
##      N Jackknife 1      2.5%      97.5% Std.Dev
## [1,] 3      14.15667  8.666667 21.00000 3.187537
## [2,] 4      15.05000  9.750000 20.25000 2.908321
## [3,] 5      15.87000 11.600000 20.70500 2.826766
## [4,] 6      16.58333 11.229167 22.35000 2.901923
## [5,] 7      17.51000 13.639286 23.85714 2.684791
## [6,] 8      17.97875 13.290625 23.58437 2.631219
## [7,] 9      18.96889 13.777778 24.11111 2.886258
## [8,] 10     19.93600 14.747500 24.20000 2.630840
## [9,] 11     20.53455 15.250000 24.27273 2.429309
## [10,] 12    21.36750 17.185417 23.89792 1.905751
## [11,] 13    21.85077 18.692308 23.46154 1.464411
## [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:

- S_{obs} = o número de espécies na comunidade,
- m = número de amostragens,
- Q_1 = 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

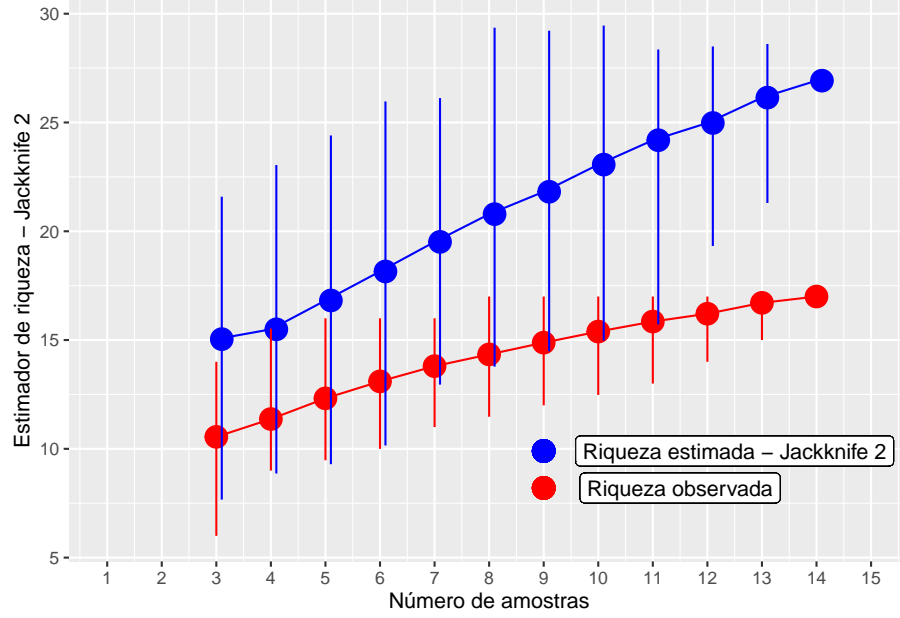
```
library(vegan)
dados_coleta <- poca_anuros
est_jack2 <- poolaccum(dados_coleta, permutations = 100)
summary(est_jack2, display = "jack2")
```

```
## $jack2
##      N Jackknife 2      2.5%    97.5% Std.Dev
## [1,] 3      15.04167  7.666667 21.58750 3.445534
## [2,] 4      15.49250  8.866667 23.04375 3.688470
## [3,] 5      16.81950  9.293750 24.40250 3.661352
## [4,] 6      18.15233 10.153333 25.96667 4.144035
## [5,] 7      19.51214 12.952381 26.12202 4.030475
## [6,] 8      20.78768 13.776786 29.35714 4.230203
## [7,] 9      21.81639 14.485069 29.21840 4.114662
## [8,] 10     23.06389 14.977778 29.45472 3.759535
## [9,] 11     24.18464 15.718182 28.35455 3.418031
## [10,] 12    24.97659 19.323674 28.49242 2.715636
## [11,] 13    26.14160 21.301282 28.60897 1.738847
## [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")
```



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

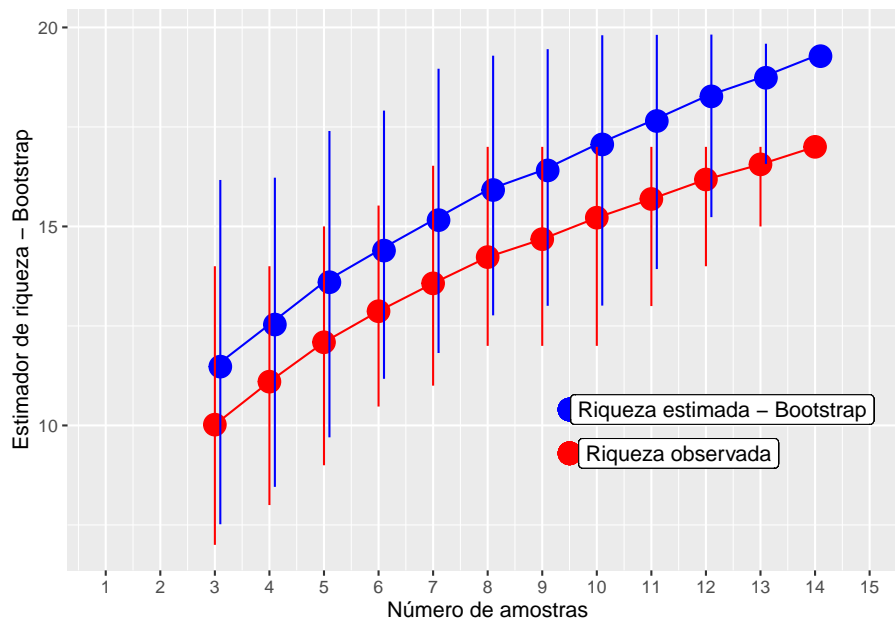
```
library(vegan)
dados_coleta <- poca_anuros
est_boot <- poolaccum(dados_coleta, permutations = 100)
summary(est_boot, display = "boot")
```

```
## $boot
##      N Bootstrap      2.5%      97.5%      Std.Dev
## [1,]  3  11.47778   7.517593 16.16759  2.2218138
## [2,]  4  12.53539   8.457031 16.22490  2.2228959
## [3,]  5  13.59872   9.701288 17.39817  1.9848709
## [4,]  6  14.39329  11.172274 17.91056  1.7777948
## [5,]  7  15.15844  11.820110 18.96258  1.8767058
## [6,]  8  15.91556  12.766099 19.29035  1.8632981
## [7,]  9  16.40730  13.006320 19.45435  1.7932676
## [8,] 10  17.06085  13.011781 19.80358  1.8958075
## [9,] 11  17.64928  13.929216 19.81378  1.7133537
## [10,] 12  18.26775  15.233580 19.82092  1.3823778
## [11,] 13  18.73539  16.570376 19.59107  0.9919236
## [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

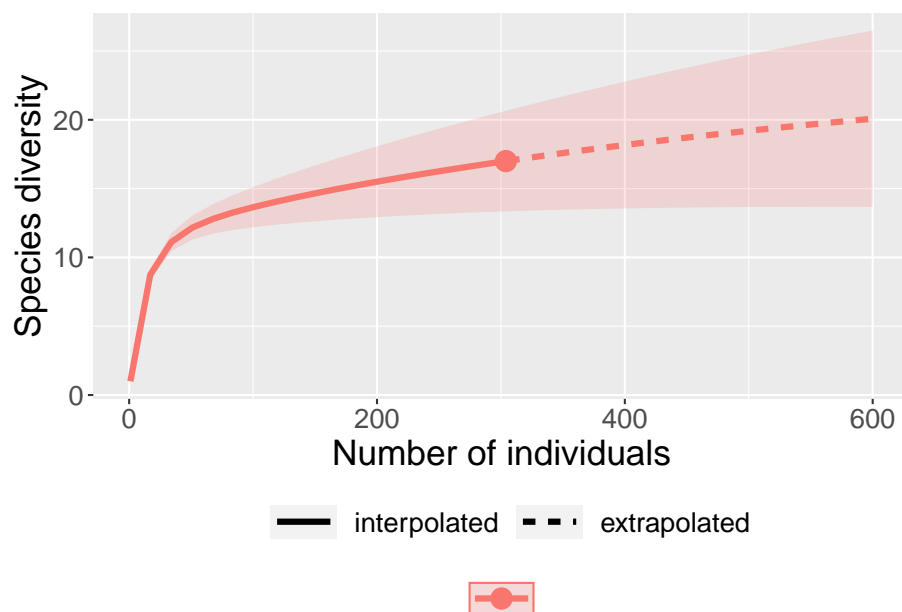
Calculo da extrapolação da riqueza com base no número de indivíduos

```
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.

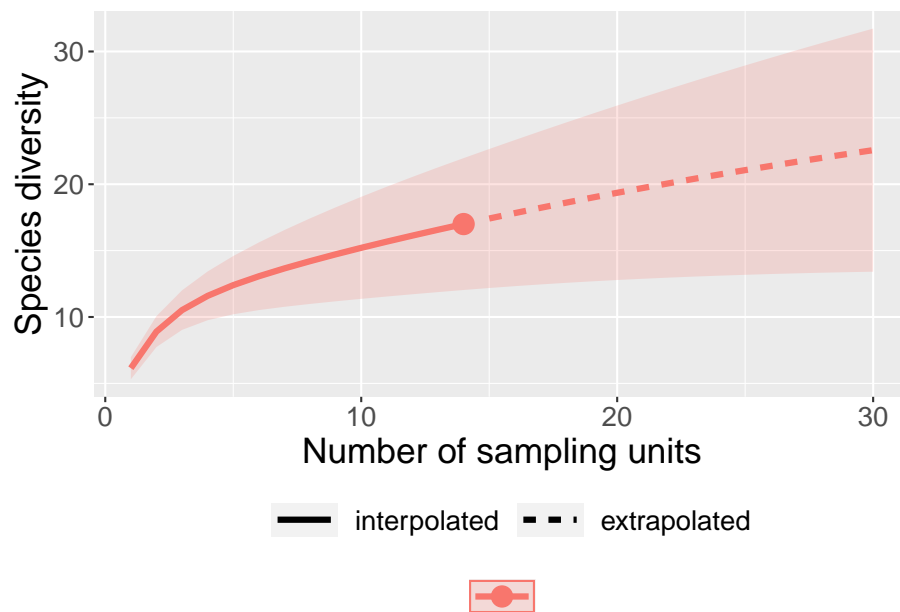
Calculo da extrapolação da riqueza com base no número de amostras

```
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.