Juan Pablo Hurtado-Gómez

Sou Biólogo da Universidade de Antioquia em Medellín, Colômbia. A minha área de interesse e a sistemática taxonomia e biogeografia de escamados (cobras e lagartos). Atualmente estou fazendo o mestrado no Laboratório de Herpetologia do Museu de Zoologia da USP com sistemática dum gênero de cobras, utilizando evidencia morfológica e molecular.

exe

Propostas para função

Proposta A. Resumo variação morfológica.

Na hora de preparar os manuscritos para publicar trabalhos taxonômicos, baseados em morfologia, geralmente incluem-se tabelas que resumem a variação dos caracteres, onde se fazem comparações entre os diferentes grupos (p.ex. sexo, espécie, região geográfica). Além disso, incluem-se gráficos de boxplot ou histogramas para mostrar as diferenças entre a distribuição dos caracteres que foram informativos.

A proposta é fazer uma função que:

- A) Crie uma tabela resumo, a partir de uma matriz de dados, comparando os diferentes grupos, em até dois níveis, selecionados pelo usuário (p.ex. espécie e sexo). Para cada caráter de cada nível serão informados os seguintes dados, amplitude de variação, média e o desvio padrão, min-max (media ± sd).

A tabela será salva em um arquivo no formato csv para que o usuário possa abrir e formatar em outros programas como excel ou word.

- B) Gerar gráficos comparativos para os caracteres dos grupos (p. ex. comprimento total por espécie ou por sexo), onde o usuário escolha os níveis, o tipo de gráfico e os caracteres.

Proposta B. Escolha de caracteres a utilizar em análises multivariadas.

Um dos problemas mais comuns na hora de fazer uma análises multivariadas são os tamanhos amostrais pequenos e muitos caracteres. Uma das soluções para esse problema é reduzir a quantidade de caracteres a serem incluídos na análise. A ideia é incluir apenas os caracteres que sejam mais informativos. O modo mais recomendável para fazer isso é realizar análises de variância para cada variável entre os grupos e aquelas que apresentem diferenças mais significativas serão incluídas na matriz final, para posteriormente serem utilizadas na análise multivariada (ver Sanders et al, 2004).

A proposta é fazer uma função que realize para cada variável uma analises de variância, considerando se apresentam ou não uma distribuição normal. No caso de haver diferença significativa, a variável será incluída na nova matriz a ser utilizada na análise multivariada. No final, a função salvará a nova matriz em um arquivo no formato csv (comma-separated values) contendo apenas as variáveis que foram informativas para a separação dos grupos nas análises de variância.

Sanders KL, Malhotra A, Thorpe RS. 2006. Combining molecular, morphological and ecological data to infer species boundaries in a cryptic tropical pitviper. Biological Journal of the Linnean Society 87: 343–364.

> Olá Juan, sua proposta A parece mais interessante. Com ela você poderá fazer uma análise exploratória de seus dados, criar gráficos, observar intervalos de confiança e jogar com os argumentos.
>
> A proposta B pode trazer dificuldades na execução, pois você tenderia a recriar análises básicas de variância e os resultados não seriam tão interessantes, apesar de úteis em análises multivariadas. A questão é que você pode realizar essa escolha de variáveis em poucas linhas de código, o que poderia trazer questionamentos sobre a utilidade da função.
>
> Eu optaria pela proposta A, e tentaria incrementá-la brincando com os gráficos comparativos e tentando incluir informações geográficas nestes mapas, já que muitas vezes abrimos mão das classificações existentes para caracterizar as unidades taxonômicas e entender padrões clinais, por exemplo. Qualquer dúvida, entre em contato. —-*Glaucia Del-Rio*

Obrigado Glaucia. Vou fazer a função A então. Se não tiver muito problema, então vou la incrementar com o que vc me falo.

TRABALHO FINAL

Obsevaçoes:Não consegui fazer as análises em dois nives pois nao encontrei o jeito de concatenar os resultados de tipo data.frame num loop dentro da função, mas continuare na procura duma solução. **1**Ver embaixo uma versão melhorada que trabalha em dois niveis, postada 10 horas depois do prazo.

Função Variação Morfologica (var.morf)

Script

```
var.morf=function(matrix, gru1, vars, name="summary")
{
  if(!is.data.frame(matrix)) # Pergunta se a matriz nao e um data.frame
    stop ("matrix must be a data.frame object")# No caso de nao ser data
frame, mostrar erro
  if(length(vars)>1)# Pergunta se a variabel tem so um valor
  {
    for (x in vars)
      if (x==vars[1])
      {
        mar=matrix[vars]
      }
    else
    {
      mat=cbind(mar,matrix[x])
    }
    if(!is.numeric(gru1)|!is.numeric(vars))# Pergunta se o valor inserido na
variavel e numerico
      stop ("gru1, gru2 or vars are not numeric vectors")# No caso de nao
ser numerico, mostrar erro
    if(length(gru1)>1)# Pergunta se a variabel tem so um valor
      stop("gru1 must be of length '1'")# No caso de ser maior que 1,
mostrar erro
    res=aggregate(mat, list(matrix[[gru1]]),  function(x) c(n=length(x),
Min=min(x, na.rm=T), Max=max(x, na.rm=T), sd=sd(x, na.rm=T),  Mean=mean(x,
na.rm=T)))#Tabela de sumario das variaveis selecionadas pelo usuario
    res[,sapply(res, is.numeric)] <-round(res[,sapply(res, is.numeric)],2)#
arredonda os valores das variaveis numericas para dois decimais.
    write.csv(res,paste(name,".csv", sep=""))#Salva a tabela em formato
'.csv' com o nomen escolhido pelo usuario
    return("A matriz foi salva na sua pasta de trabalho") #Messagem que
confirma que a funcao foi executada como sucesso
  } else
  {
    if(!is.numeric(gru1)|!is.numeric(vars))# Pergunta se o valor inserido na
variavel e numerico
      stop ("gru1 or vars are not numeric vectors")# No caso de nao ser
numerico, mostrar erro
    if(length(gru1)>1)# Pergunta se a variabel tem so um valor
      stop("gru1 must be of length '1'")# No caso de ser maior que 1,
mostrar erro
    res=aggregate(matrix[[vars]], list(matrix[[gru1]]),  function(x)
c(n=length(x), Min=min(x, na.rm=T), Max=max(x, na.rm=T), sd=sd(x, na.rm=T),
Mean=mean(x, na.rm=T)))#Tabela de sumario das variaveis selecionadas pelo
usuario
    res[,sapply(res, is.numeric)] <-round(res[,sapply(res, is.numeric)],2)#
arredonda os valores das variaveis numericas para dois decimais.
    write.csv(res,paste(name,".csv", sep=""))#Salva a tabela em formato
'.csv' com o nomen escolhido pelo usuario
    return("A matriz foi salva na sua pasta de trabalho")
  }
```

```
}
```

HELP

```
    Var.morf        package:none        R Documentation
    Resumo de Variaão morfologica
    Description:

Realiza uma tabela resumo da variação de um ou vários carateres para os
grupos de uma variável (p.ex. espécie, sexo, região geográfica). A função
salva uma tabela em formato .csv com o tamnho da amostra (n), o valor mínimo
(min), o valor máximo (max), desvio padrão (sd), e a média aritmética (mean)
para cada variável dentro de cada grupo.
    Usage:
      var.morf(matrix, gru1, vars, name)
    var.morf(matrix, gru1, vars, name="summary")
Arguments:
matrix:     objeto de classe “data frame” previamente criado, que contém as
variáveis nas colunas e os indivíduos nas linhas.
gru1:   Número da coluna que contém a lista dos grupos nos que se pretende
resumir a variação morfológica. Essa variável só pode ser um (longitude =1)
número.
vars:   Número(s) da(s) coluna(s) nas que se encontram as variáveis a serem
resumidas.
name:   Vector de caráteres (“entre aspas”) no qual se coloca o nomen
desejado para o arquivo .csv. O nomen predeterminado para o arquivo é
summary e ficará salvo na pasta de trabalho como summary.csv
Details:
matrix deve ser um arquivo de tipo data frame que contém as variáveis a
serem resumidas.
A coluna indicada no parâmetro gru1 deve incluir os nomes dos grupos que
desejam ser resumidos e podem ser numéricos o carateres. A função mostrara
erro quando não seja de tipo numérico ou tenha um comprimento maior a um.
Os números indicados no argumento vars deveram corresponder as colunas dos
caráters a ser resumidos, os quais devem ser de tipo numérico. A função
mostrara erro quando não sejam de tipo numérico.
No argumento name o usuário não precisa especificar a extenção do archivo.
Ela é .csv por padrão.
A função opera com funções do pacote base do R.

Author(s):
Juan Pablo Hurtado Gómez
Juanphg@usp.br

Examples:
## Abra o aquivo “ex_var.morf.csv”
evmf=read.csv(“ex_var.morf.csv”)
source(“var.morf.r”)
```

```
# Para resumir uma variavel das especies
var.morf(matrix=evmf, gru1=1, vars=5, name="summarysp")
#Para resumir varias variaveis por localidade
var.morf(matrix=evmf, gru1=2, vars=c(5:15), name="summaryloc")
# ou varias variaveis por sexo
var.morf(matrix=evmf, gru1=3, vars=c(5, 4, 6,15), name="summarysex")
```

Arquivos

Os arquivos estam comprimidos no seguente arquivo:var.morf.rar ou no seguente link: https://copy.com/fF10fv0ZZyQcCDcV

Função Variação Morfologica (var.morf) V11.1

Script

```
var.morf=function(matrix, gru1, gru2=NULL, vars, name="summary")
{
  if(!is.data.frame(matrix)) # Pergunta se a matriz nao e um data.frame
    stop ("matrix must be a data.frame object")# No caso de nao ser data
frame, mostrar erro
  if(is.null(gru2)) # Se o usuario so quer resumir as variaveis num nivel
  {
    if(length(vars)>1)# Pergunta se a variabel tem so um valor
    {
      for (x in vars) #cria um loop para juntar as variaveis escolhidas a
serem resumidas
        if (x==vars[1]) # No caso que o loop estivese na primeira variavel
        {
          mar=matrix[vars]# criar um vetor inicial para a posterior
concatenacao
        }
      else # Se e outra variavel
      {
        mat=cbind(mar,matrix[x])# concatenar a primeira variavel com a
variavel seguente
      }
      mat=mat[,-(length(vars)+1)] # tirar a ultima coluna que se repete.
      if(!is.numeric(gru1)|!is.numeric(vars))# Pergunta se o valor inserido
na variavel e numerico
        stop ("gru1 or vars are not numeric vectors")# No caso de nao ser
numerico, mostrar erro
      if(length(gru1)>1)# Pergunta se a variabel tem so um valor
        stop("gru1 must be of length '1'")# No caso de ser maior que 1,
mostrar erro
      res=aggregate(mat, list(matrix[[gru1]]),  function(x) c(n=length(x),
Min=min(x, na.rm=T), Max=max(x, na.rm=T), sd=sd(x, na.rm=T),  Mean=mean(x,
na.rm=T)))#Tabela de sumario das variaveis selecionadas pelo usuario
```

```
      res[,sapply(res, is.numeric)] <-round(res[,sapply(res,
is.numeric)],2)# arredonda os valores das variaveis numericas para dois
decimais.
      write.csv(res,paste(name,".csv", sep=""))#Salva a tabela em formato
'.csv' com o nomen escolhido pelo usuario
      return("A matriz foi salva na sua pasta de trabalho") #Messagem que
confirma que a funcao foi executada como sucesso
    } else
    {
      if(!is.numeric(gru1)|!is.numeric(vars))# Pergunta se o valor inserido
na variavel e numerico
        stop ("gru1 or vars are not numeric vectors")# No caso de nao ser
numerico, mostrar erro
      if(length(gru1)>1)# Pergunta se a variabel tem so um valor
        stop("gru1 must be of length '1'")# No caso de ser maior que 1,
mostrar erro
      res=aggregate(matrix[[vars]], list(matrix[[gru1]]),  function(x)
c(n=length(x), Min=min(x, na.rm=T), Max=max(x, na.rm=T), sd=sd(x, na.rm=T),
Mean=mean(x, na.rm=T)))#Tabela de sumario das variaveis selecionadas pelo
usuario
      res[,sapply(res, is.numeric)] <-round(res[,sapply(res,
is.numeric)],2)# arredonda os valores das variaveis numericas para dois
decimais.
      write.csv(res,paste(name,".csv", sep=""))#Salva a tabela em formato
'.csv' com o nomen escolhido pelo usuario
      return("A matriz foi salva na sua pasta de trabalho")# Messagem final
que diz que a funcao termino o processo e salvo os resultados.
    }
  }
  else
  {
    if(!is.numeric(gru1)|!is.numeric(vars))# Pergunta se o valor inserido na
variavel e numerico
      stop ("gru1 or vars are not numeric vectors")# No caso de nao ser
numerico, mostrar erro
    if (length(vars)>1) # para dos grupos (niveis) so e posivel fazer para
ua variavel.
      stop("two level analysis is only possible for one variable: vars must
be of length=1")# Mostrar Erro quando tenta fazer analises de duas variaveis
em dois niveis
    if(length(gru1)>1)# Pergunta se a variabel tem so um valor
      stop("gru1 must be of length '1'")# No caso de ser maior que 1,
mostrar erro
    res=aggregate(matrix[[vars]], list(matrix[[gru1]]),  function(x)
c(n=length(x), Min=min(x, na.rm=T), Max=max(x, na.rm=T), sd=sd(x, na.rm=T),
Mean=mean(x, na.rm=T)))#Tabela de sumario das variaveis selecionadas pelo
usuario
    res[,sapply(res, is.numeric)] <-round(res[,sapply(res, is.numeric)],2)#
arredonda os valores das variaveis numericas para dois decimais.
    write.csv(res,paste(name,".csv", sep=""))#Salva a tabela em formato
```

```
'.csv' com o nomen escolhido pelo usuario
    return("A matriz foi salva na sua pasta de trabalho") # Messagem final
que diz que a funcao termino o processo e salvo os resultados.
  }
}
```

HELP

```
    Var.morf        package:none        R Documentation
    Resumo de Variaão morfologica
    Description:

Realiza uma tabela resumo da variação de um ou vários carateres para os
grupos de uma variável (p.ex. espécie, sexo, região geográfica) ou em dois
niveis (p. ex. espécie e sexo) só para uma variavel. A função salva uma
tabela em formato .csv com o tamnho da amostra (n), o valor mínimo (min), o
valor máximo (max), desvio padrão (sd), e a média aritmética (mean) para
cada variável dentro de cada grupo.
    Usage:
      var.morf(matrix, gru1, gru2, vars, name)
    var.morf(matrix, gru1, gru2=NULL, vars, name="summary")
Arguments:
matrix:     objeto de classe “data frame” previamente criado, que contém as
variáveis nas colunas e os indivíduos nas linhas.
gru1:   Número da coluna que contém a lista dos grupos nos que se pretende
resumir a variação morfológica. Essa variável só pode ser um número
(longitude =1).
gru2:   Número da coluna que contém a lista dos grupos nos que se pretende
subagrupar a variação morfológica. Essa variável só pode ser um número
(longitude =1).
vars:   Número(s) da(s) coluna(s) nas que se encontram as variáveis a serem
resumidas.
name:   Vector de caráteres (“entre aspas”) no qual se coloca o nomen
desejado para o arquivo .csv. O nomen predeterminado para o arquivo é
summary e ficará salvo na pasta de trabalho como summary.csv
Details:
matrix deve ser um arquivo de tipo data frame que contém as variáveis a
serem resumidas.
A coluna indicada no parâmetro gru1 deve incluir os nomes dos grupos que
desejam ser resumidos e podem ser numéricos o carateres. A função mostrara
erro quando não seja de tipo numérico ou tenha um comprimento maior a um.
Os números indicados no argumento vars deveram corresponder as colunas dos
caráters a ser resumidos, os quais devem ser de tipo numérico. A função
mostrara erro quando não sejam de tipo numérico.
Para trabalhar em dois niveis, a função so consegue analizar uma vararivavel.
No argumento name o usuário não precisa especificar a extenção do archivo.
Ela é .csv por padrão.
A função opera com funções do pacote base do R.

Author(s):
```

```
Juan Pablo Hurtado Gómez
Juanphg@usp.br

Examples:
## Abra o aquivo “ex_var.morf.csv”
evmf=read.csv(“ex_var.morf.csv”)
source(“var.morf.r”)
# Para resumir uma variavel das especies
var.morf(matrix=evmf, gru1=1, vars=5, name="summarysp")
#Para resumir varias variaveis por localidade
var.morf(matrix=evmf, gru1=2, vars=c(5:15), name="summaryloc")
# ou varias variaveis por sexo
var.morf(matrix=evmf, gru1=3, vars=c(5, 4, 6,15), name="summarysex")
# Em dois niveis p. ex. espécie e sexo
var.morf(matrix=evmf, gru1=1, gru2=3, vars=5, name="summaryspsex")
```

Arquivos V11.1

Os arquivos estam comprimidos no seguente arquivo:var.morf_v11.1.rar ou no link :https://copy.com/JCItNsCQT24m3X2S

From:
http://ecor.ib.usp.br/ - **ecoR**

Permanent link:
http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2014:alunos:trabalho_final:juanphg:start

Last update: **2020/08/12 06:04**