

MARVIN ANGANROY



Sou estudante de doutorado em Zoologia, Instituto de Biociências, USP. Eu trabalho em sistemática filogenética de um grupo de sapos neotropicais, do gênero *Hyloxalus*, da família Dendrobatidae.

[exec](#)

Trabalho final

Proposta A - Análise intraespecífico de diferentes variáveis quantitativas

Dentro das espécies há variação intraespecífica. Quantificar a variação é importante para conhecer se há mudanças entre populações, determinar se há dimorfismo sexual, e estimar a variação das entidades biológicas, o que vai ser de utilidade em diferentes áreas da biologia. A função analisa diferentes variáveis quantitativas de duas ou mais amostras (populações) e determina se há diferenças entre as variáveis, além de isso, determinar qual das variáveis aporta mais ao a variação por meio de um pca e vai fazer um análise discriminante das populações. Os objetos de entrada serão um dataframe que contem as variáveis por população. O objeto de saída vai ser gráficos de análise exploratório (boxplot e histograma), e os gráficos do pca e de análise discriminante, e também uma lista com os produtos de anova, pca e análise discriminante.

Proposta A - Análise estadístico das variáveis morfométricas de sapos. Info:

Dados das medidas de diferentes variáveis de diferentes populações

Passos:

Análise exploratório.

Anovas (comparando entre sexo, populações)

PCA e o Análise discriminante

Produtos:

Gráficos de análises exploratório, pca e de análise discriminante.

Uma lista com os productos de anova, pca e análise discriminante.

Oi Marvin, consigo entender melhor agora.

Ainda tenho dúvidas em relação ao que é uma análise discriminante.

Como você vai contar quais variáveis entrarão no pca?

Imaginemos duas populações A e B. A função calculará a probabilidade de serem duas populações distintas com base em várias características morfológicas certo? O que acontece se para "tamanho de olho" as populações forem consideradas diferentes, mas

para “cor” as populações forem consideradas iguais? Você vai contar isso pro usuário de que forma? Qq dúvida, pode chamar. bjs

Rena

Oi Marvin,

não deu pra entender muito bem o que seu plano A vai fazer exatamente. Me parece uma análise exploratória um pouco mais profunda. É interessante, mas tem que ser melhor explicada, como a Rena disse, não está claro exatamente o que o usuário vai receber.

O plano B parece igual ao A mas com girinos, e o C não parece uma boa ideia se você não sabe como realizar o procedimento. Minha sugestão é explicar melhor o plano A.

—*Danilo G Muniz*

Como a proposta A foi a sugestão para continuar desenvolvendo, então estuo enviando os passos mais detalhados da proposta. Objetos de entrada

Dataframe com as variáveis quantitativas por populações (dois ou mais)

Passos:

a-Analise exploratório para todas as variáveis (meia, desvio padrão, mínimo e maximo)

b-Determinar se há diferenças significativas entre as variáveis por população (Anova)

c-Escolher as variáveis que tem diferenças

d-Com as variáveis de “c” fazer o pca e o analise discriminante

Produtos

Os gráficos de analises exploratório (Histograma, boxplot), pca, analise discriminante.

Uma lista com os produtos de anova, pca e analise discriminante.

Respondendo a pergunta de Renata: eu acho que se uma variável não é significativa e outra sim, não tem importância para a função. Se a função vai fazer o pca e o analise discriminante com variáveis que biologicamente não tem significado, então é um critério do investigador determinar se vai tomar como certo os resultados da função. Abraços Marvin

Proposta B. - Analise do crescimento das larvas de anuros

Os anuros tem uma fase larvária, conhecido como girino. Nela ocorrem mudanças das características no tempo de desenvolvimento dos organismos. Aquelas variações são estudadas pela heterocronia. Essas variações tem impacto em a diversidade fenotípica dos organismos (Gould 1977). A função analisa a relação de diferentes variáveis quantitativas através do desenvolvimento do girino. Os dados de entrada são as medidas quantitativas em diferentes estágios de desenvolvimento (tabela, dataframe). A função compara diferentes variáveis quantitativas respeito a comprimento total do girino e comparações por médio de regressões lineais. Os produtos de saída são uma lista dos valores e gráficos das regressões lineais, nos quais reflectara a variação em a ontogenia dos organismos.

Proposta B - Crescimento dos girinos de sapos

Info:

Procurando mais informação, sobre as taxas de crescimento, eu preciso de dados temporais, mas nesta proposta eu quero comparar diferentes variáveis através do desenvolvimento do girino. Os dados são uma tabela com 8 a 10 medidas do corpo de girino através do desenvolvimento da larva.

Passos:

O analise exploratório das variáveis.

A relação entre as variáveis, com regressão lineal.

Comparação entre regressões lineais

Produtos:

Uma lista com os produtos da função (coeficientes, os valores da anova)

Os gráficos das regressões lineais.

Oi Marvin,

Se entendi bem, sua função fará uma série de $lm()$ entre variáveis contínuas. Além disso, plotará os gráficos de dispersão e as retas das regressões.

Acho que seria legal se vc colocasse um argumento para que o usuário pudesse escolher quais as jogar comparar (ex. use apenas as colunas 3, 4 e 8 do meu conjunto de dados).

Você tem co-variáveis no seu conjunto de dados? Seria interessante se o usuário pudesse indicar a existência de co-variáveis e sua função calculasse efeitos a partir de resíduos (um pouco parecido com o exercício da iris, lembra?).

De qualquer forma, vamos esperar os comentários do Danilo, pois ele é mais expert nisso.

bjs

Rena

Proposta C - Mapeamento de variáveis discretas em topologias filogenéticas

Neste caso como eu não tenho muito conhecimento de como fazer, então não envio o procedimento. Eu acho que a informação para fornecer é uma filogenia já gerada, e uma matriz de dados que quero mapear na filogenia. E os produtos podem ser um árvore com o mapeamento dos caracteres e as variáveis discretas.

Não parece interessante investir seu tempo em aprender isso tudo pra função e acho que com as propostas A e B já dá pra gente trabalhar. Ainda assim, prefiro esperar a opinião do Danilo pra depois decidir qual proposta seguir, ok?

bjs

Rena

Função Morfometria

Página de ajuda ou help

Morfometria package:unknown R Documentation

Análise da variação morfométrica ao nível intraespecífico

Description:

A função faz um análise da variação morfométrica de variáveis quantitativas para determinar se existem diferenças entre populações de uma espécie, por meio de um análise de variância, um análise de componentes principais e um análise de função discriminante.

Usage:

```
morfometria (x, nlocs)
```

Arguments:

x Uma tabela ou dataframe com as variáveis morfométricas discriminadas por localidade. As variáveis estão nas colunas e as observações nas linhas. A primeira coluna é para as localidades, a segunda coluna é o identificador das observações e desde a terceira coluna são as variáveis morfométricas. As linhas devem estar organizadas por localidade

nlocs O número de localidades.

Details:

A função morfometria analisa variáveis de tipo quantitativas. A função precisa dos pacotes vegan, stats e MASS, mas se ainda não tem os pacotes instalados, a função faz a instalação.

Value:

Os produtos são:

Uma lista de:

comp1 : Os valores da análise exploratório (valor mínimo, máximo, meia e desvio padrão) das variáveis e discriminadas por localidade

comp2 : Os valores p para cada variável do análise de variâncias (95% de probabilidade) e as variáveis com variação significativa

comp3 : A variação explicada de cada PCA ou valores Eigen

comp4 : Os produtos do análise discriminante (matriz, porcentagem por localidade e porcentagem total dos acertos do modelo)

Gráficos de:

graf 1 : Boxplots das variáveis discriminado por localidade

graf 2 : Histogramas das variáveis

graf 3 : O gráfico do primer e segundo PCA

graf 4 : O aporte dos PCA ao a variância explicada

graf 5 : O gráfico de dispersão do FD1 e FD2

Warning:

As células das variáveis quantitativas podem só conter valores numéricos, do contrario, morfometria vai gerar erro

Author(s):

Marvin Anganoy-Criollo

marvinanganoy@gmail.com

References:

Jolliffe, I.T. 2002. Principal Components Analysis. Second Edition, Springer Series in Statistics, 457p.

Hill, T., and P. Lewicki. 2006. Statistics: Methods and Applications, a Comprehensive Reference for Science, Industry, and Data Mining. Statsoft Inc. 832 p.

Código da função

```
##### Função Morfometria #####
morfometria<-function(x, nlocs)
{
  #####1A- O análise exploratório dos dados####
  #Resumo da meia, desvio padrão, mínimo e o máximo para cada variável
  resumo<-summary(x[,c(3:length(x))])
  #Gerar objeto localidade e variáveis quantitativas
  localities<-x[,1] #vetor localidade
  vari.quant<-x[,3:length(x)] #objeto das variáveis quantitativas
  #A meia por localidade
  result.mean<-matrix(NA, nrow=nlocs,
ncol=length(vari.quant[,1:length(vari.quant)]))
  for(i in 1:length(vari.quant))
```

```
{m<-tapply(vari.quant[,i], localities, FUN = mean)
result.mean[,i]<-m
}
#0 valor mínimo por localidade
result.min<-matrix(NA, nrow=nlocs,
ncol=length(vari.quant[,1:length(vari.quant)]))
for(i in 1:length(vari.quant))
{mi<-tapply(vari.quant[,i], localities, FUN = min)
result.min[,i]<-mi
}
#0 valor máximo por localidade
result.max<-matrix(NA, nrow=nlocs,
ncol=length(vari.quant[,1:length(vari.quant)]))
for(i in 1:length(vari.quant))
{ma<-tapply(vari.quant[,i], localities, FUN = max)
result.max[,i]<-ma
}
#0 desvio padrão por localidade
result.sd<-matrix(NA, nrow=nlocs,
ncol=length(vari.quant[,1:length(vari.quant)]))
for(i in 1:length(vari.quant))
{rsd<-tapply(vari.quant[,i], localities, FUN = sd)
result.sd[,i]<-rsd
}
####1B- 0 analise de variância, Anova ####
result.anv<-rep(NA, length(vari.quant))
for(i in 1:length(result.anv))
{
  anov.a<-lm(vari.quant[,i]~localities)
  sumar.a<-summary(anov.a)
  result.anv[i]<-sumar.a$coefficients[1,4]
}
#A variáveis quantitativas com diferenças significativas
vari.signif<-which((vari.signif=result.anv<=0.05))
####1C- 0 analise de componentes principais, PCA####
#1C.1- Conferir se os pacotes "vegan", "stat" e "MASS" estão instalados
#0s pacotes são necessários para o PCA e AD
pacote<- function(pacote)
{is.element(pacote, installed.packages()[,1])}
if(pacote("vegan")==FALSE)
{install.packages("vegan")
  cat("\t", "0 pacote vegan foi instalado!\n")} #Se o pacote vegan não é
instalado, a função vai gerar uma mensagem
if(pacote("vegan")==TRUE)
{require ("vegan", quietly=TRUE)}
if(pacote("stats")==FALSE)
{install.packages("stats")
  cat("\t", "0 pacote stats foi instalado!\n")}
if(pacote("stats")==TRUE)
{require ("stats", quietly=TRUE)}
```

```

if(paquete("MASS")==FALSE)
{install.packages("MASS")
  cat("\t", "O pacote MASS foi instalado!\n")}
if(paquete("MASS")==TRUE)
{require ("MASS", quietly=TRUE)}
#1C.2- Análise de componentes principais - PCA
quant.pca<-rda(vari.quant)
eig.pca<-(quant.pca$CA$eig) #Os valores Eigen
vari.pca<-eig.pca/(sum(eig.pca)) #O aporte ao varianca explicada por cada
PCA
####1D- O Analise de função discriminante, AD ####
resul.afd<-lda(localities~.,vari.quant)
resul.coun<-resul.afd$counts #Número de observações por localidade
resul.pred<-predict(resul.afd) #Predicto do modelo
resul.tab=table(localities, resul.pred$class) #A tabela de precisão do
modelo
resul.diag<-diag(prop.table(resul.tab, 1)) #Porcentagem dos acertos por
localidade
resul.sum<-sum(diag(prop.table(resul.tab))) #Porcentagem total de acertos
do modelo
ADList<-list(resul.coun, resul.tab, resul.diag, resul.sum) #Gerando uma
lista de dados dos produtos principais de AD
names(ADList)<-c("Número de observações por localidade", "Matriz dos
acertos do modelo", "Porcentagem dos acertos por localidade", "Porcentagem
total do modelo")
####2- Os gráficos da função####
##2.1- O gráficos do analise exploratório
#2.1A- Os boxplots das variáveis por localidade
for(i in 1:length(vari.quant)){
  boxplot(vari.quant[,i]~localities, col="lightgray",
xlab=names(vari.quant[i]))}
#2.1B- O histograma por variável
for(i in 1:length(vari.quant)){hist(vari.quant[,i], main = "", xlab =
names(vari.quant[i]), ylab = "Frequência", col = "lightgray")}
##2.2- O biplot dos PCA
biplot(quant.pca, scaling = "symmetric", col = c("black", "red"))
##2.3- O screeplot ou aporte dos PCA
screeplot(quant.pca, type=c("lines"), col= "red", main = "", xlab = "Eixos
PCA", ylab = "Variância")
##2.4- O plot dos dois primeiros vetores discriminantes
LD1<-resul.pred$x[,1]
LD2<-resul.pred$x[,2]
plot(LD1, LD2, type="n")
text(cbind(LD1,LD2), labels=unclass(localities),cex =0.8)
#### Os Produtos da Função ####
Resultado<-list(resumo, result.mean, result.min, result.max, result.sd,
result.anv, vari.signif, vari.pca, ADList)
names(Resultado)<-c("Resumo das variáveis", "A media por localidade das
variáveis ", "O mínimo por localidade das variáveis", "O máximo por
localidade das variáveis", "O desvio padrão por localidade das variáveis",
"O valor-p da anova das variáveis", "As variáveis significativas da anova",

```

```
"Valores Eigen o variação explicado dos PCA", "Lista do Analise Discriminante")  
  return(Resultado)  
}
```

Examples: [exemplo1.csv](#)

Arquivo da função [funcao_morfometria.r](#)

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

Permanent link:
http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2017:alunos:trabalho_final:marvinanganoy:start 

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