

Paula Alves Condé



Mestranda em Ecologia, Instituto de Biologia, USP.

Março de 2009

Proposta

Principal

Uma função para análise exploratória de dados que sejam dois vetores através de apresentações gráficas, resumos e opções de simulação de uma distribuição nula para teste unicaudal, bicaudal e teste t. Apresenta gráficos para exploração das variáveis, como histograma, boxplot, plot, qqnorm, e também retorna um gráfico de simulação de uma distribuição nula para teste unicaudal, bicaudal e teste t. É uma função integrando a análise exploratória de dados através de apresentações gráficas e opções de simulação usadas na função “simula”.

(Reformulada para responder aos comentários abaixo)

Comentários

Paulo

Idéia promissora, mas algumas coisas não ficaram claras para mim, veja se estão para você:

1. boxplot e qqnorm são gráficos para uma variável e sua função receberá duas. Um gráfico para cada uma?
2. Que simulação será feita e que gráfico será retornado? Há vários pontos importantes aqui, a começar pela pergunta que se pretende responder com a simulação, o que vai determinar o tipo de simulação feita, o melhor modo de mostrar os resultados, etc.

Plano B

Uma função para planilha de dados (data frame) organizados com cada indivíduo da amostra nas linhas e com uma coluna com o nome das espécies correspondente a cada indivíduo.

A função retorna uma matriz e gráfico do rank de abundância das espécies, e terá argumento de opção para aplicar a função para apenas algum ou vários subconjuntos da coleta dos dados (como locais, armadilhas, meses ou parcelas) os quais você queira saber essas informações (abundância de cada espécie).

E quem sabe também apresente cálculos de medidas ecológicas (como índices de diversidade e

equitabilidade).

Comentários

Paulo

Bem geral e factível. Se entendi sua planilha tem uma linha para cada indivíduo, obrigatoriamente um fator que identifica a espécie, e outro fatores opcionais que podem ser usados para separar os dados em subconjuntos, é isto? Uma dica: resista à tentação de usar a função `sort`, e opte pela `order`, que parece mais difícil, mas é bem mais poderosa. Só ela pode resolver problemas de ordenação mais complicados como este. Outra sugestão: você pode oferecer a opção da planilha ter uma linha por espécie, e um campo de abundância de cada espécie.

Página de Ajuda

analix

package:nenhum

R Documentation

Análise exploratória de duas variáveis e opção para simulação nula com teste unicaudal, bicaudal e teste t.

Description:

Apresenta sumário das variáveis e os principais gráficos para uma análise exploratória de dados, índice de correlação entre duas variáveis, e retorna uma simulação de uma distribuição nula com testes estatísticos.

Usage:

```
analix <- function(dados1,dados2,teste="N")
```

Arguments:

`dados1`: Vetor numérico. Valores de uma amostra ou variável.

`dados2`: numérico. Valores de uma amostra ou variável.

`teste`: lógico. Qual o tipo de teste na simulação nula?

Details:

Os valores das amostras (`dados1`, `dados2`) são plotados em gráficos de histograma, boxplot, qqnorm e `"dados2"` em função de `"dados1"`, com linha da regressão linear simples.

É apresentado um sumário de cada variável (ou amostra) com o índice de correlação entre elas.

Também são simulados uma hipótese nula para os dados com opção de teste unicaudal (`teste="uni"`), teste bicaudal (`teste="bi"`), teste t de Student (`teste="t"`) ou sem simulação e teste estatístico (`teste="N"`).

As simulações são feitas com repetição de 1000 de uma distribuição uniforme dos dados.

Value:

São gerados um conjunto de seis gráficos para análise exploratória dos dados, e um gráfico com o resultado da simulação de hipótese nula com o teste estatístico escolhido.

Author(s):

Paula Alves Condé

paula.conde@usp.br

References:

Exploratory data analysis.

http://en.wikipedia.org/wiki/Exploratory_data_analysis

Null hypothesis. http://en.wikipedia.org/wiki/Null_hypothesis

Statistical hypothesis testing.

http://en.wikipedia.org/wiki/Hypothesis_testing

See Also:

'qqnorm' para analisar a normalidade dos dados.

'simula' simulação da hipótese nula com testes estatísticos.

Examples:

```
dados1=round(rnorm(10,mean=6,sd=3),2)
dados2=round(rnorm(10,mean=7.5,sd=3.2),2)
```

```
## Sem simulações e teste, apenas exploração gráfica das variáveis:
  analix(dados1,dados2,teste="N")

## Exploração gráfica + Simulação de uma distribuição nula com teste
bicaudal:
  analix(dados1,dados2,teste="bi")

## Exploração gráfica + Simulação de uma distribuição nula com teste
unicaudal:
  analix(dados1,dados2,teste="uni")

## Exploração gráfica + Simulação de uma distribuição nula com teste t:
  analix(dados1,dados2,teste="t")
```

Código da Função

```
analix <- function(dados1,dados2,teste="N")
{
  ### Função utilizada para analise exploratória de duas variáveis.
  ### A entrada de dados deve ser feita por dois objetos vetores de mesmo
  tamanho.
  ### Também faz simulação de uma distribuição nula para teste unicaudal,
  bicaudal
  ## e teste t.

  cat("Para fazer simulação de uma distribuição nula, no caso de teste
  unicaudal o
  + primeiro vetor deve ser o dos dados que seriam maiores.\n As opções de
  teste são
  + (entre aspas):\n\t uni (normal unicaudal)\n\t bi (normal bicaudal) \n\t
  + t (distribuição t)\n")

  ### Autora: Paula Alves Condé
  ### Data: 01/04/2009
  ### Trabalho final
  ### Disciplina "Uso da Linguagem R para Analise de Dados Ecologicos" -
  BIE5782.

  x11()
  par(mfrow=c(3,2),bty="l")
  ## Produz Histograma dos "dados1"
  hist(dados1,col="red",main=paste(c("Histograma",substitute(dados1))))
  ## Produz Histograma dos "dados2"
  hist(dados2,col="blue",main=paste(c("Histograma",substitute(dados2))))
  ## Produz Boxplot das duas variáveis
  boxplot(dados1,dados2,main="Boxplot",names=c(substitute(dados1),
  + substitute(dados2)))
  ## Plota "dados2" em função dos "dados1" com a reta de regressão
```

```

## linear simples para ver a relação entre as variáveis
plot(dados2~dados1,main=paste(c(substitute(dados2),"~",substitute(dados1)))
+ ,pch=16,xlab=deparse(substitute(dados1)),ylab=deparse(substitute(dados2)))
  abline(lm(dados2~dados1),col="purple",lwd=2)
  ## Gráfico QQNorm para "Dados1"
  qqnorm(dados1)
  qqline(dados1,col="red")
  mtext(deparse(substitute(dados1)),side=3)
  ## Gráfico QQNorm para "Dados2"
  qqnorm(dados2)
  qqline(dados2,col="blue")
  mtext(deparse(substitute(dados2)),side=3)
  ## Sumário das variáveis e índice de correlação
  name.xy <- paste(deparse(substitute(dados1)), "e",
deparse(substitute
+ (dados2)))
  cat("Dados:", name.xy, "\n")
  n <- length(dados1)
  sumx <- sum(dados1^2) - sum(dados1)^2/n
  sumy <- sum(dados2^2) - sum(dados2)^2/n
  sumxy <- sum(dados1 * dados2) - sum(dados1) * sum(dados2)/n
  correl <- sumxy/sqrt(sumx * sumy)
  concluindo<- list(summary(dados1),summary(dados2),correl)
  names(concluindo)<- c(substitute(dados1), substitute(dados2),
+ "correlação")

### Fazendo simulação de uma distribuição nula para teste unicaudal,
bicaudal
## ou teste t.
  nsim=1000
  resultado<-rep(NA,nsim)
  dif = mean(dados1) - mean(dados2)
  dif.abs = round(abs(dif), 1)
  cat("\n Diferença absoluta observada entre as médias das variáveis =
",dif.abs,
+ "\n")
  v1 = var(dados1)
  v2 = var(dados2)
  n1 = length(dados1)
  n2 = length(dados2)
  s12 = sqrt((v1/n1) + (v2/n2))
  tvalor = abs(dif/s12)
  med = mean(c(dados1, dados2))
  des = sd(c(dados1, dados2))
  res = rep(NA,nsim)
  arco = rainbow(nsim)
x11()

  if (teste == "bi")
  {

```

```
meu.cex = 900/nsim
plot(runif(50, 0, (dif.abs * 1.5)), 0:49, type = "n",
     xlim = c(0, dif.abs * 1.5), ylim = c(0, (0.08 * nsim)),
     xlab = "Diferença absoluta", ylab = "Frequência",
     main = "Simulação")
for (i in 1:nsim) {
  res[i] = abs(round(mean(rnorm(n1, mean = med, sd = des)) -
                    mean(rnorm(n2, mean = med, sd = des)), 1))
  stripchart(res, method = "stack", add = T, cex = meu.cex,
             pch = 15, col = arco[i])
}
res.menor <- res[res < dif.abs]
stripchart(res.menor, method = "stack", add = T, cex = meu.cex,
           pch = 15, col = "dark green")
legend(5, (6 * nsim), legend = paste(sum(res >= dif.abs),
   "valores >= ", round(dif.abs, 1)), bty = "n", text.col = "red")
abline(v = dif.abs, col = "red")
}
if (teste == "uni")
{
  meu.cex = 1700/nsim
  plot(runif(50, (dif.abs * -1.5), (dif.abs * 1.5)), 0:49,
       type = "n", xlim = c((dif * -1.5), (dif * 1.5)),
       ylim = c(0, (0.04 * nsim)), xlab = "Diferença Absoluta",
       ylab = "Frequência", main = "Simulação")
  for (i in 1:nsim) {
    res[i] = round(mean(rnorm(n1, mean = med, sd = des)) -
                  mean(rnorm(n2, mean = med, sd = des)), 1)
    stripchart(res, method = "stack", add = T, cex = meu.cex,
               pch = 15, col = arco[i])
  }
  res.menor <- res[res < dif.abs & res > (-1 * dif.abs)]
  stripchart(res.menor, method = "stack", add = T, cex = meu.cex,
             pch = 15, col = "dark green")
  legend((dif.abs * 0.5), (0.038 * nsim), legend = paste(sum(res >=
    dif.abs), "valores >= ", round(dif, 1)), bty = "n",
    text.col = "red")
  legend((dif.abs * -1.4), (0.038 * nsim), legend = paste(sum(res <=
    (dif.abs * -1)), "valores <= -", round(dif, 1)),
    bty = "n", text.col = "red")
  abline(v = dif.abs, col = "red")
  abline(v = (-1 * dif.abs), col = "red")
}
if (teste == "t")
{
  meu.cex = 1300/nsim
  cat("\t\nValor t observado = ", tvalor, "\n")
  plot(runif(50, (-0.7 * dif.abs), (0.7 * dif.abs)), runif(50,
    0, (nsim/20)), type = "n", xlim = c((-0.8 * dif.abs),
    (0.8 * dif.abs)), ylim = c(0, (nsim/20)), xlab = "valor t ",
```

```
ylab = "Frequência", main = "Simulação")
for (i in 1:nsim) {
  simula1 = rnorm(n1, mean = med, sd = des)
  simula2 = rnorm(n2, mean = med, sd = des)
  difs = mean(simula1) - mean(simula2)
  vs1 = var(simula1)
  vs2 = var(simula2)
  ssl2 = sqrt((vs1/n1) + (vs2/n2))
  tsimula = difs/ssl2
  res[i] = round(tsimula, 1)
  stripchart(res, method = "stack", add = T, cex = meu.cex,
    pch = 15, col = arco[i])
}
tvalor.vf = res < round(tvalor, 1) & res > round((-1 *
  tvalor), 1)
ntvalor = sum(tvalor.vf == F)
res.menor = res[tvalor.vf]
stripchart(res.menor, method = "stack", add = T, cex = meu.cex,
  pch = 15, col = "dark green")
legend((tvalor * 0.2), (nsim/22), legend = paste(sum(res >=
  round(tvalor, 1)), "valores >= ", round(tvalor, 1)),
  bty = "n", text.col = "red")
legend((tvalor * -0.9), (nsim/22), legend = paste(sum(res <=
  round((tvalor * -1), 1)), "valores <= -", round(tvalor,
  1)), bty = "n", text.col = "red")
abline(v = tvalor, col = "red")
abline(v = (-1 * tvalor), col = "red")
cat("\n\tProbabilidade erro I = ", ntvalor/nsim, "\n")
}
return(concluindo)
}
```

Arquivo da Função

[analix.r](#)

From:

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

Permanent link:

http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:alunos:trabalho_final:paulac



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