



Olá, meu nome é Diana!

Terminei a minha graduação no ano passado, e estou estudando para o mestrado. Bom o R é muito bom! Fiz parte das análises da minha iniciação utilizando o R, mas com a ajuda de amigos que sabiam manipulá-lo. Agora pretendo fazer a disciplina para tentar terminar as análises!

## Meus exercícios

[lista1.r](#) Lista 1 [lista2.r](#) Lista 2 [lista3.r](#) Lista 3 [lista4.r](#) Lista 4 [lista5.r](#) Lista 5 [lista6.r](#) Lista 6 [lista7.r](#) Lista 7 [lista9.r](#) Lista 9

## Trabalho final

### Proposta principal

Considerando o seguinte conjunto de dados: medidas de número de folhas verdadeiras, número de folhas cotiledonares e altura das plântulas de *Myrcia bicarinata*, tomadas ao longo de um período de tempo ( $t_0$  a  $t_{11}$ ). Sendo que as plântulas estão submetidas a dois tratamentos: condição de luz (muita e pouca luz) e origem das sementes (Floresta de Restinga Alta e Floresta de Restinga Baixa). As perguntas em relação a esses dados são: “As plântulas de *Myrcia bicarinata* que estão na condição de muita luz crescem mais do que na condição de pouca luz?” “Elas crescem mais rápido?” “A sobrevivência das plântulas sob muita luz é maior do que a das plântulas que cresceram sob pouca luz?” “As plântulas sob pouca luz tem maior número de folhas do que as plântulas sob muita luz?”

Com base nessas considerações gostaria de construir uma função que me ajudasse a responder essas perguntas. Que calcule a média para cada variável em cada tempo e relacione essa média com cada fator (luz e origem). Que fizesse o final menos o inicial para cada variável e também relacionando com as combinações dos fatores, como para cada um separado. E para essas medidas, selecionasse as bandejas referentes a cada amostra. Além, disso que ciclasse a minha estatística de interesse e construísse cenários nulos para calcular o p e fazer o teste de significância.

### Plano B

Uma função que me ajudasse a fazer a análise exploratória dos dados.

**PS:** preciso articular melhor as idéias da minha proposta para os meus objetivos

## Comentários

Ale: Di, acho que precisa mesmo organizar melhor as idéias. Posso sugerir algo, mas acho que vc. deve dar o start. São idéias muito abertas e sem objetividade. Pense em apenas uma das fases do seu trabalho e pense em uma tarefa para automatizar... Algo que vc. teve que fazer muitas vezes e poderia fazer uma função para evitar repetir a tarefa (com outras espécies p.ex). Talvez uma AED

seja uma alternativa interessante, mas precisa me dizer o que a função irá fazer especificamente!

Daniel: Estava com problemas de adicionar os comentários na sua página, mas agora consegui! Concordo que tem muitas questões soltas na sua proposta de função. Na última parte você explica um pouco o que quer fazer com seus dados. Seria interessante expandir esta parte, sem se preocupar, em um primeiro momento, nas questões específicas do seu projeto, mas sim em quais manipulações sua função será capaz de fazer. Calcular médias e outras estatísticas para um dado vetor (valores quantitativos) considerando tratamentos (valores qualitativos) e implementar um modelo nulo me parece dois bons elementos para sua função.

## Resposta aos comentários

Estou trabalhando os dados de numero de folhas verdadeiras. Bom consegui fazer a média, maximo de folhas para cada fator (luz e origem). Gostaria de fazer um plot com o aumento do número de folhas ao longo do tempo, mas não estou conseguindo! Outra coisa, Musgo, não entendi a questão de implementar um modelo nulo! Seria construir esse modelo com base na variação do número de folhas ao longo do tempo? Valeu!

## Página de Ajuda

AED  
Documentation

package:bie5782\_curso2011

R

Análise exploratória de dados

Description:

Produz um conjunto de gráficos (boxplot,histogramas,qqplot)tanto para dados contínuos como para contagens.

Usage:

AED(dados,na.rm=FALSE)

Arguments:

dados: data.frame em que cada linha representa a observação para um indivíduo;

- a primeira coluna contém o nome da espécie;
- a segunda e terceira coluna representam dois tratamentos;
- a quarta coluna contém a identificação das amostras;
- a quinta coluna contém a identificação dos indivíduos;
- a partir da sexta coluna estão as variáveis respostas.

na.rm=FALSE: argumento lógico, indicando que se há NA no data.frame, é

preciso que seja retirado dos dados antes destes serem computados na função.

#### Details:

Os dados devem ser apenas para uma espécie de cada vez. Para contagens, é preciso que os valores sejam da classe dos números inteiros.

#### Value:

Um data.frame com a média dos valores das variáveis respostas e uma série de gráficos são gerados. Se os dados são contagem, são produzidos quatro dispositivos gráficos:

Dispositivo 1: Um boxplot com a distribuição da variável resposta dentro de cada amostra para t<sub>final</sub>.

Dispositivo 2 : São gerados três boxplots, com a variável resposta sendo o valor máximo das contagens para cada amostra (para t<sub>final</sub>), em relação aos fatores e às combinações destes.

Dispositivo 3: São gerados três boxplots, com a variável resposta sendo a média das contagens para cada amostra (para t<sub>final</sub>), em relação aos fatores e às combinações destes.

Dispositivo 4: Um histograma de probabilidade, com a linha de densidade e a curva normal. Um qqplot para verificar a normalidade das médias de contagem para cada amostra.

Se os dados são contínuos, são gerados três dispositivos gráficos:

Dispositivo 1 : Um boxplot com a distribuição da variável resposta para de cada amostra.

Dispositivo 2 : Quatro boxplots com a distribuição das médias da variável resposta em cada tratamento.

Dispositivo 3: Dois histogramas de probabilidade, com a linha de densidade e a curva normal. Dois qqplots, um para cada variável resposta.

#### Warning:

Se nas contagens conter números que não são da classe dos inteiros, é preciso arredá-los antes de computar os dados na função.

#### Author(s):

Diana Cristina Cesar da Graça

#### References:

Crawley, M. J. The R Book. Wiley, New York, 2007

#### See Also:

'qqnorm' e 'qqline', para o gráfico de quantil padrão do R.  
'hist' para obter mais detalhes sobre os histogramas e a linha de densidade.

Examples:

```
f.verd=read.table("plantula_Mb_fv.csv",header=TRUE,sep=";",as.is=TRUE)#  
dados de contagem  
f.verd$fv8=round(f.verd$fv8)# arredondando os números  
f.verd$fv9=round(f.verd$fv9)  
f.verd$fv10=round(f.verd$fv10)  
f.verd$fv11=round(f.verd$fv11)  
head(f.verd)  
altura=read.table("plantula_Mb_h.csv",header=TRUE,as.is=TRUE,sep=";")#dados  
contínuos  
head(altura)  
AED(f.verd)  
AED(altura)
```

## Código da Função

```
AED=function(dados,na.rm=FALSE)  
{  
  colnames(dados)[c(2,3,4)]=c("fator1","fator2","amostra") # altera nome  
de certas colunas  
  if(class(dados[,6])=="integer") # se os dados são contagens  
  {  
    x=dados  
    x$dif=x[,length(x)]-x[,6] # cria coluna, diferença de contagens  
tfinal-tinicial  
    x11()  
    boxplot(x[,length(x)]~amostra,data=x,xlab="Amostras",ylab="Nº de  
folhas")  
    max.1=aggregate(x[,6:length(x)],by=list(fator1=x$fator1,amostra=x$amostra,fator2=x$fator2),FUN=max)  
    x11()  
    par(mfrow=c(1,3))  
    boxplot(max.1[,length(max.1)]~fator2+fator1,data=max.1,ylab="Nº max  
de ____ para tfinal")  
    boxplot(max.1[,length(max.1)]~fator2,data=max.1,col=c("grey50","grey80"))  
    boxplot(max.1[,length(max.1)]~fator1,data=max.1,col=c("grey50","grey80"))  
    par(mfrow=c(1,1))  
    media=aggregate(x[,6:length(x)],data=x,by=list(fator1=x$fator1,amostra=x$amostra,fator2=x$fator2),FUN=mean) # média contagens por amostra  
    f.1=aggregate(x[,6:length(x)],data=x,by=list(fator1=x$fator1,amostra=x$amostra),FUN=mean) # média de contagens pelo fator 1  
    x11()  
  }
```

```

        par(mfrow=c(1,3))
        boxplot(media[,length(media)]~fator1+fator2,data=media,ylab="Nº médio
de ____ em tfinal")
        boxplot(f.1[,length(f.1)]~fator1,data=f.1,ylab="Nº médio de ____ em
tfinal")
        boxplot(media[,length(media)]~fator2,data=media,ylab="Nº médio de ____
em tfinal")
        par(mfrow=c(1,1))
        x11()
        par(mfrow=c(1,2))
        hist(media[,length(media)],prob=TRUE,main="Probabilidade de médias do
nº de ____ no tfinal")
        lines(density(media[,length(media)]),col="blue")
        curve(expr =
dnorm(x,mean=mean(media[,length(media)]),sd=sd(media[,length(media)])),add=T
, col="red")
        qqnorm(media[,length(media)])
        qqline(media[,length(media)])
        par(mfrow=c(1,1))
        return(media) # retorna um data.frame com as médias das contagens
    }
    else
    {
        y=dados #dados contínuos
        y$cresc=(y[,length(y)])-(y[,6])#cria coluna "crescimento"
tamanhofinal-tamanhoinicial
        x11()
        boxplot(y[,length(y)-1]~amostra,data=y,ylab="Altura em
tfinal",xlab="amostras")
        media.dados=
aggregate(y[,6:length(y)],data=y,by=list(fator1=y$fator1,fator2=y$fator2,amo
stra=y$amostra),FUN=mean)# médias v.resposta por amostra
        x11()
        par(mfrow=c(2,2))
        boxplot(media.dados[,length(media.dados)-1]~fator1,data=media.dados,ylab="Al
tura média em tfinal")
        boxplot(media.dados[,length(media.dados)]~fator1,data=media.dados,ylab="Cres
cimento médio em tfinal")
        boxplot(media.dados[,length(media.dados)-1]~fator2,data=media.dados,ylab="Al
tura média em tfinal")
        boxplot(media.dados[,length(media.dados)]~fator2,data=media.dados,ylab="Cres
cimento médio em tfinal")
        par(mfrow=c(1,1))
        X11()
        par(mfrow=c(2,2))
        hist(y[,length(y)-1],prob=TRUE,xlab="altura em tfinal")
        lines(density(y[,length(y)-1]),col="blue")
        curve(expr =
dnorm(x,mean=mean(y[,length(y)-1]),sd=sd(y[,length(y)-1])),
        add=T, col="red")
        qqnorm(y[,length(y)-1])

```

```
    qqline(y[,length(y)-1])
    hist(media.dados[,length(media.dados)-1],prob=TRUE,xlab="Crescimento
em t final-t inicial")
    lines(density(media.dados[,length(media.dados)-1]),col="blue")
    curve(expr =
dnorm(x,mean=mean(media.dados[,length(media.dados)-1]),sd=sd(media.dados[,le
ngth(media.dados)-1])),
        add=T, col="red")
    qqnorm(media.dados[,length(media.dados)-1])
    qqline(media.dados[,length(media.dados)-1])
    return(media.dados) #retorna um data.frame das médias da variáveis
resp. por amostra
  }
}
```

## Arquivos da Função

### Arquivo da função AED

[funcao\\_aed\\_funcionando.r](#)

### Arquivos de dados

[plantula\\_mb\\_fv.csv](#) contagen

[plantula\\_mb\\_h.csv](#) contínuo

From:

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

Permanent link:

[http://ecor.ib.usp.br/doku.php?id=05\\_curso\\_antigo:r2011:alunos:trabalho\\_final:diana:start](http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2011:alunos:trabalho_final:diana:start)



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