



# ANÁLISES EXPLORATÓRIAS DE DADOS

## Preparação dos dados e programa

As análises abaixo serão realizadas em ambiente R e para isso teremos que instalar alguns pacotes, abaixo estão todos os comandos necessários para a realização da atividade.

1) Crie um diretório (pasta), copie os arquivos de dados abaixo para esse diretório e faça a descompactação no mesmo diretório:

- [autocorr.zip](#)
- [univar.zip](#)

2) Abra o R no seu computador e mude o diretório de trabalho para o diretório (*i.e.* a pasta) que você criou, usando o menu **Arquivo > mudar dir...**

3) Instale os pacotes *car* e *lattice*.

Para isso, basta copiar e colar os comandos que estão nas caixas de cor cinza:

```
install.packages("car")
```

Espera finalizar todo o processo de instalação desse pacote para iniciar o próximo:

```
install.packages("lattice")
```

4) Agora carregue os pacotes:

```
library(car)  
library(lattice)  
library(graphics)
```

## ANALISANDO DADOS UNIVARIADOS

1) importe o conjunto de dados para o R

```
univar1<-read.csv("univar1.csv")
```

2) Use a função *head* para visualizar as 5 primeiras linhas do conjunto de dados

```
head(univar1)
```

3) Inspecione o resumo dos dados

```
summary(univar1)
```

## Conhecendo os dados:

4) Se quiser, visualize o conjunto de dados como uma planilha convencional

```
edit(univar1)
```

## Análises gráficas

```
par(mfrow = c(2,2)) ##Aqui estamos criando um layout para colocar os quatro  
gráficos juntos  
hist(univar1$COMPRIMENTO_BICO)  
hist(univar1$BIOMASSA_AVE)  
hist(univar1$BIOMASSA_INSETOS)  
hist(univar1$TAMANHO_SEMENTES)  
par(mfrow=c(1,1)) ## voltando ao padrão de apresentar apenas 1 gráfico por  
página
```

```
#use o argumento breaks para determinar o número de classes  
par(mfrow = c(2,2))  
hist(univar1$COMPRIMENTO_BICO, breaks = 20)  
hist(univar1$BIOMASSA_AVE, breaks = 20)  
hist(univar1$BIOMASSA_INSETOS, breaks = 20)  
hist(univar1$TAMANHO_SEMENTES, breaks = 20)  
par(mfrow=c(1,1))
```

```
par(mfrow = c(2,2))  
hist(univar1$COMPRIMENTO_BICO, breaks = 10)  
hist(univar1$BIOMASSA_AVE, breaks = 10)  
hist(univar1$BIOMASSA_INSETOS, breaks = 10)  
hist(univar1$TAMANHO_SEMENTES, breaks = 10)  
par(mfrow=c(1,1))
```

```
par(mfrow = c(2,2))  
plot(density(univar1$COMPRIMENTO_BICO))  
plot(density(univar1$BIOMASSA_AVE))  
plot(density(univar1$BIOMASSA_INSETOS))  
plot(density(univar1$TAMANHO_SEMENTES))  
par(mfrow=c(1,1))
```

Podemos juntar esses dois gráficos em um só. Para isso, use o código abaixo:

```
par(mfrow = c(2,2))  
hist(univar1$COMPRIMENTO_BICO, prob=T )  
lines(density(univar1$COMPRIMENTO_BICO))
```

```
hist(univar1$BIOMASSA_AVE, prob=T)
lines(density(univar1$BIOMASSA_AVE))

hist(univar1$BIOMASSA_INSETOS, prob=T)
lines(density(univar1$BIOMASSA_INSETOS))

hist(univar1$TAMANHO_SEMENTES, prob=T)
lines(density(univar1$TAMANHO_SEMENTES))
par(mfrow=c(1,1))
```

Podemos também mostrar, na parte inferior do gráfico de densidade, o número de observações em cada faixa do gráfico. Para isso vamos usar a função *rug()*

```
par(mfrow = c(2,2))
plot(density(univar1$COMPRIMENTO_BICO))
rug(univar1$COMPRIMENTO_BICO, side=1)

plot(density(univar1$BIOMASSA_AVE))
rug(univar1$BIOMASSA_AVE, side=1)

plot(density(univar1$BIOMASSA_INSETOS))
rug(univar1$BIOMASSA_INSETOS, side=1)

plot(density(univar1$TAMANHO_SEMENTES))
rug(univar1$TAMANHO_SEMENTES, side=1)

par(mfrow=c(1,1))
```

Todas essas informações nos auxiliam para identificarmos a quais distribuições teóricas nossos dados se ajustam.

```
(sort(univar1$COMPRIMENTO_BICO))
```

Mas temos uma função que faz isso por nós:

```
boxplot(univar1$COMPRIMENTO_BICO, range=0)
```

Vamos fazer um boxplot modificado com os nossos dados de `COMPRIMENTO_BICO`

```
boxplot(univar1$COMPRIMENTO_BICO)
```

```
boxplot(univar1$BIOMASSA_INSETOS ~ univar1$NIVEL_DISTURBIO)
```

```
boxplot(univar1$BIOMASSA_INSETOS ~ univar1$NIVEL_DISTURBIO, notch=TRUE)
```

**E agora, você está mais seguro(a) para afirmar se a biomassa de insetos**

## difere ou não entre os dois níveis de distúrbio?

### CHECANDO O AJUSTE DOS DADOS A UMA DISTRIBUIÇÃO

Vamos então aplicar as funções abaixo aos nossos dados:

```
par(mfrow = c(2,2))

qqnorm(univar1$COMPRIMENTO_BICO)
qqline(univar1$COMPRIMENTO_BICO)

qqnorm(univar1$BIOMASSA_AVE)
qqline(univar1$BIOMASSA_AVE)

qqnorm(univar1$BIOMASSA_INSETOS)
qqline(univar1$BIOMASSA_INSETOS)

qqnorm(univar1$TAMANHO_SEMENTES)
qqline(univar1$TAMANHO_SEMENTES)

par(mfrow=c(1,1))
```

### AVALIANDO AUTOCORRELAÇÃO

Para essa parte do tutorial, importe o conjunto de dados "autocorr.csv" para o R e inspecione os dados:

```
autocorr<-read.csv("autocorr.csv")
head(autocorr)
summary(autocorr)
```

```
lag.plot(autocorr$x1, do.lines = FALSE, diag=FALSE)
```

```
lag.plot(autocorr$x2, do.lines = FALSE, diag=FALSE)
```

### ANALISANDO DADOS BIVARIADOS

```
bivar<-read.csv("bivar.csv")
head(bivar)
summary (bivar)
```

```
plot(bivar$y.l ~ bivar$x.l)
```

```
plot(bivar$y.l ~ bivar$x.l)
lines(lowess(bivar$y.l ~ bivar$x.l))
```

```
plot(bivar$y.n ~ bivar$x.n)
lines(lowess(bivar$y.n ~ bivar$x.n))
```

```
#grafico do pacote car
scatterplot (bivar$y.l ~ bivar$x.l)
```

```
scatterplot (bivar$y.n ~ bivar$x.n)
```

## Transformando os dados

```
scatterplot(univar1$COMPRIMENTO_BICO ~ univar1$BIOMASSA_AVE)
```

Como podemos observar pelos boxplots laterais, nesse caso, aparentemente são os dados da variável Y que parecem estar afetando a linearidade da relação. Então, vamos transformar os dados de Y pelo logaritmo natural e ver se o ajuste melhora.

```
scatterplot (log(univar1$COMPRIMENTO_BICO) ~ univar1$BIOMASSA_AVE)
```

From:

<http://labtrop.ib.usp.br/> - **Laboratório de Ecologia de Florestas Tropicais**

Permanent link:

<http://labtrop.ib.usp.br/doku.php?id=cursos:planeco2020:roteiro:05-descr>



Last update: **2021/03/01 15:59**