

# Bioestatística Florestal - Aula de campo

## Análise descritiva dos dados obtidos em aula

Eduardo E. R. Junior & Clarice G. B. Demétrio

30 de março de 2018

Este arquivo contém as análises realizadas a partir do conjunto de dados mensurados na aula de campo. Na aula de campo, formaram-se quatro grupos em que cada um mensurou, em linha, o comprimento à altura do peito (CAP) e se há ou não bifurcação (bifur) em árvores Sapucaia.

Os dados foram anotados em uma folha e organizados em uma planilha eletrônica (Excel, LibreOffice, Gnumeric, etc.) e posteriormente exportados em formato texto, nesse documento o formato do arquivo exportado .tsv, cujo separador é a tabulação. Abaixo são apresentados os códigos em R para análise dos dados

1. O primeiro passo é a leitura (ou carregamento) dos dados na sessão R, isso é feito com a função `read.table(...)`. Note os argumentos da função, o que cada um representa é documentado na ajuda sobre ela (`help(read.table)`).

```
# Lê os dados
dados <- read.table("dados.tsv",
                    header = TRUE,
                    dec = ".",
                    sep = "\t",
                    encoding = "UTF-8")

# Verifica a estrutura do objeto (note o número de observações, número
# de variáveis e o tipo das variáveis)
str(dados)
```

```
#> 'data.frame': 33 obs. of 3 variables:
#> $ CAP : num 86.2 111 70.5 76.6 52.2 ...
#> $ bifur: Factor w/ 2 levels "Não","Sim": 2 2 2 1 1 1 2 1 1 1 ...
#> $ grupo: Factor w/ 4 levels "A","B","C","E": 1 1 1 1 1 1 1 2 2 2 ...
```

2. Agora desejamos obter o DAP. Embora tenhamos mensurado o CAP, comumente reporta-se os resultados na escala do DAP.

```
# Adiciona uma coluna com o diâmetro à altura do peito
dados$DAP <- dados$CAP / pi
str(dados)
```

```
#> 'data.frame': 33 obs. of 4 variables:
#> $ CAP : num 86.2 111 70.5 76.6 52.2 ...
#> $ bifur: Factor w/ 2 levels "Não","Sim": 2 2 2 1 1 1 2 1 1 1 ...
#> $ grupo: Factor w/ 4 levels "A","B","C","E": 1 1 1 1 1 1 1 2 2 2 ...
#> $ DAP : num 27.4 35.3 22.4 24.4 16.6 ...
```

3. Com a variável de interesse (DAP) já calculada, agora podemos verificar algumas estatísticas descritivas.

```
# Calcula estatísticas de posição para DAP (geral)
cbind(c("Média" = mean(dados$DAP),
        "Mediana" = median(dados$DAP),
        "1º quartil" = quantile(dados$DAP, prob = 1/4),
```

```

"3º quartil" = quantile(dados$DAP, prob = 3/4),
"Mínimo" = min(dados$DAP),
"Máximo" = max(dados$DAP)))

#>           [,1]
#> Média       27.00425
#> Mediana     26.41972
#> 1º quartil.25% 19.73521
#> 3º quartil.75% 33.61352
#> Mínimo      11.61831
#> Máximo      58.98282

# Calcula estatísticas de dispersão para DAP (geral)
cbind(c("Variância" = var(dados$DAP),
       "Desvio Padrão" = sd(dados$DAP),
       "Amplitude" = max(dados$DAP) - min(dados$DAP),
       "Amplitude inter-quartilica" = IQR(dados$DAP),
       "Coef. de variação" = sd(dados$DAP)/mean(dados$DAP)))

#>           [,1]
#> Variância      104.3204691
#> Desvio Padrão   10.2137392
#> Amplitude       47.3645111
#> Amplitude inter-quartilica 13.8783110
#> Coef. de variação 0.3782271

4. Uma hipótese levantada sobre o estudo realizado é que árvores com bifurcação apresentam um menor volume de madeira do que as árvores não bifurcadas. Portanto, podemos verificar as estatísticas descritivas para as árvores bifurcadas e não bifurcadas.

# Calcula estatísticas descritivas para DAP (estratificado por bifurcação)
tapply(dados$DAP, dados$bifur, length) # número de árvores

#> Não Sim
#> 18 15

tapply(dados$DAP, dados$bifur, mean) # média

#> Não Sim
#> 25.48955 28.82190

tapply(dados$DAP, dados$bifur, sd) # variância

#> Não Sim
#> 6.664957 13.342283

# Calcula várias estatísticas de uma vez (posição)
tapply(dados$DAP, dados$bifur, function(x) {
  cbind(c("Variância" = var(x),
        "Desvio Padrão" = sd(x),
        "Amplitude" = max(x) - min(x),
        "Amplitude inter-quartilica" = IQR(x),
        "Coef. de variação" = sd(x)/mean(x)))
})

#> $Não
#>           [,1]
#> Variância      44.4216572

```

```

#> Desvio Padrão          6.6649574
#> Amplitude              21.5177483
#> Amplitude inter-quartílica  9.4697191
#> Coef. de variação        0.2614781
#>
#> $Sim
#>                                [,1]
#> Variância                  178.0165238
#> Desvio Padrão              13.3422833
#> Amplitude                  47.3645111
#> Amplitude inter-quartílica  20.6901426
#> Coef. de variação          0.4629217

# Calcula várias estatísticas de uma vez (dispersão)
tapply(dados$DAP, dados$bifur, function(x) {
  cbind(c("Média" = mean(x),
          "Mediana" = median(x),
          "1º quartil" = quantile(x, prob = 1/4),
          "3º quartil" = quantile(x, prob = 3/4),
          "Mínimo" = min(x),
          "Máximo" = max(x)))
})

```

```

#> $Não
#>                                [,1]
#> Média                25.48955
#> Mediana              25.89451
#> 1º quartil.25%      20.25247
#> 3º quartil.75%      29.72219
#> Mínimo               13.81465
#> Máximo               35.33240
#>
#> $Sim
#>                                [,1]
#> Média                28.82190
#> Mediana              26.41972
#> 1º quartil.25%      17.98451
#> 3º quartil.75%      38.67465
#> Mínimo               11.61831
#> Máximo               58.98282

```

5. Uma forma mais interessante de ver os resultados é a partir de gráficos. Nesse documento os gráficos serão realizados a partir do pacote `ggplot2` que possui facilidades para gráficos estratificados além de fornecer um tema padrão bastante agradável e o pacote `cowplot` para exibir gráficos lado a lado

```

# Instala o pacote ggplot2
install.packages(c("ggplot2", "cowplot"))

# Carrega o pacote ggplot na sessão
library(ggplot2)
library(cowplot)

```

Com os pacotes devidamente instalados e carregados, faz-se:

```

# Histograma (geral)
g1 <- ggplot(dados, aes(x = DAP)) +

```

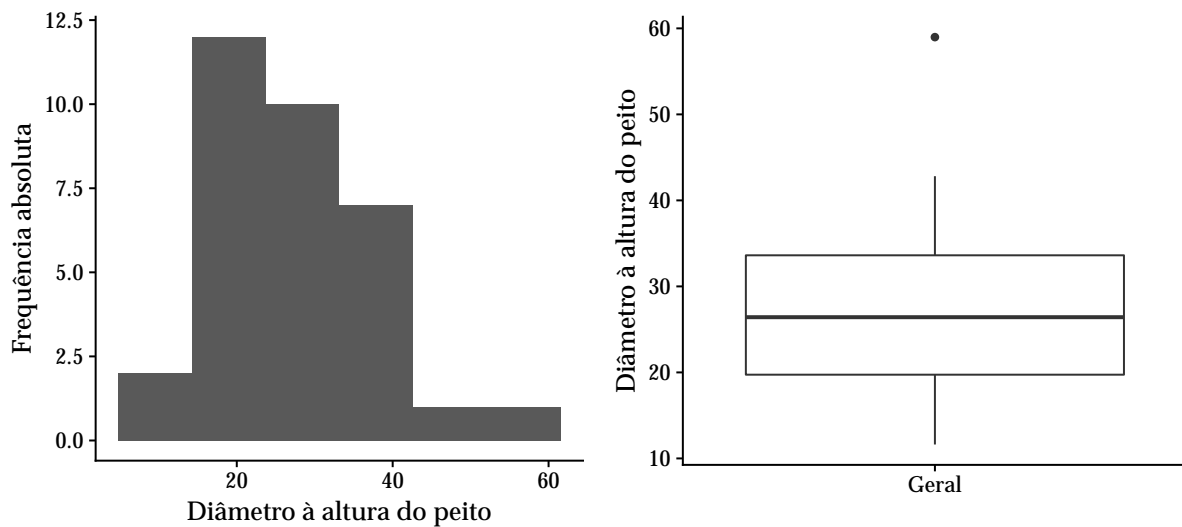
```

geom_histogram(bins = 6) +
labs(x = "Diâmetro à altura do peito",
     y = "Frequência absoluta")

# Boxplot (geral)
g2 <- ggplot(dados, aes(y = DAP, x = "Geral")) +
  geom_boxplot() +
  labs(x = "",
       y = "Diâmetro à altura do peito")

# Exibe os gráficos na mesma janela
plot_grid(g1, g2)

```



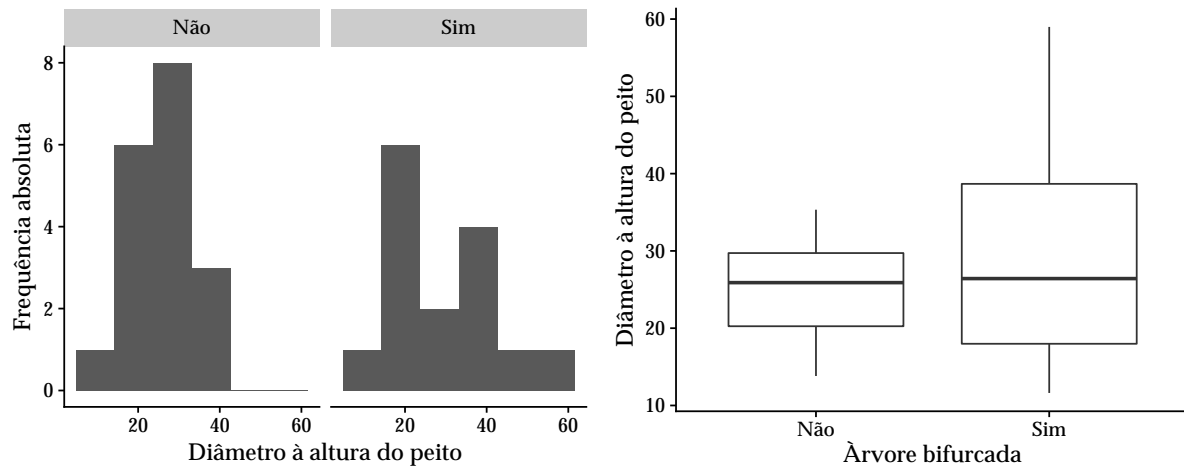
```

# Histograma (estratificado por bifurcação)
g1 <- ggplot(dados, aes(x = DAP)) +
  geom_histogram(bins = 6) +
  labs(x = "Diâmetro à altura do peito",
       y = "Frequência absoluta") +
  facet_wrap(~bifur, ncol = 2)

# Boxplot (estratificado por bifurcação)
g2 <- ggplot(dados, aes(y = DAP, x = bifur)) +
  geom_boxplot() +
  labs(x = "Árvore bifurcada",
       y = "Diâmetro à altura do peito")

# Exibe os gráficos na mesma janela
plot_grid(g1, g2)

```



6. Outra curiosidade é se os resultados são muito distintos em cada linha, como cada grupo ficou responsável por mensurar os resultados em cada linha precisamos apenas estratificar os resultados por grupo. Aqui só faremos os gráficos, embora as estatísticas pontuais também possam ser calculadas.

```
# Histograma (estratificado por bifurcação)
g1 <- ggplot(dados, aes(x = DAP)) +
  geom_histogram(bins = 6) +
  labs(x = "Diâmetro à altura do peito",
       y = "Frequência absoluta") +
  facet_wrap(~grupo, ncol = 4)

# Boxplot (estratificado por bifurcação)
g2 <- ggplot(dados, aes(y = DAP, x = grupo)) +
  geom_boxplot() +
  labs(x = "Grupo",
       y = "Diâmetro à altura do peito")

# Exibe os gráficos na mesma janela
plot_grid(g1, g2, ncol = 1)
```

