

## LABORATÓRIO 2: Regressão Linear Simples

Fernando Bispo, Jeff Caponero

# Sumário

Introdução . . . . .	2
Primeira parte . . . . .	2
Apresentação . . . . .	2
Objetivo . . . . .	2
Análise dos dados . . . . .	2
Segunda parte . . . . .	9
Apresentação . . . . .	9
Objetivos . . . . .	9
Análise dos dados . . . . .	9
Conclusão . . . . .	13
Apêndice . . . . .	14
Script em linguagem R . . . . .	14

## Introdução

O presente relatório está subdividido em duas partes, tendo em vista terem sido disponibilizados dois arquivos para análise, este relatório vem trazendo as análises conforme os pré-requisitos solicitados para cada conjunto de dados. Tendo como principal objetivo a introdução das técnicas de Regressão Linear Simples e a prática da elaboração de relatórios analíticos fundamentadas na Análise Exploratória de Dados.

## Primeira parte

### Apresentação

Nesta primeira parte a análise de dará sobre os dados contendo medidas morfológicas de 104 gambás das montanhas, que foram capturados através de armadilhas em sete localizações na Inglaterra. As variáveis contidas nesse arquivo são:

- sexo (**sex**);
- largura do crânio (**skullw**);
- comprimento total (**totlngth**).

### Objetivo

O objetivo dessa análise visa responder as seguintes questões:

- (a) Descrição do comportamento de cada uma das variáveis, a partir das medidas morfológicas segundo o sexo.
- (b) Representação gráfica da distribuição do sexo.
- (c) Apresentação de um histograma para as variáveis morfológicas.
- (d) Discussão da relação entre as variáveis morfológicas.
- (e) Tratamento dos dados.
- (f) Avaliação do ajuste de um modelo linear de regressão.
- (g) Caso o ajuste seja adequado, apresentar a reta ajustada pelo modelo.

### Análise dos dados

Composto por três características (variáveis) morfológicas dos gambás, em que duas destas são classificadas como aritméticas contínuas, sendo estas a largura do crânio (**skullw**) e o comprimento total (**totlngth**) e uma variável classificada como categórica ordinal, sendo esta sexo(**sex**).

Table 1: Medidas Resumo para o sexo feminino.

	Min	Q1	Med	Média	Q3	Max	D. Padrão	CV
Comprimento Total	75,0	85	88,5	87,91	90,5	96,5	4,18	0,05
Largura Crânio	51,5	55	56,4	56,59	57,7	67,7	2,57	0,05

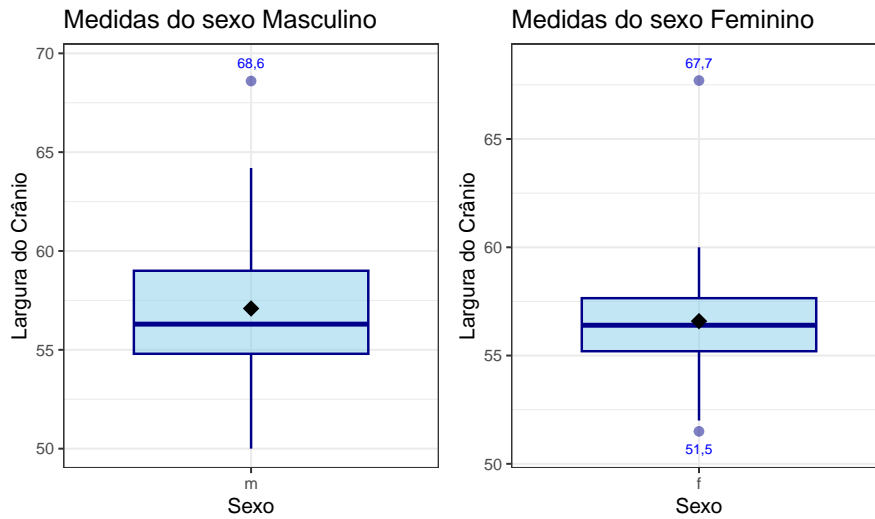
Table 2: Medidas Resumo para o sexo masculino

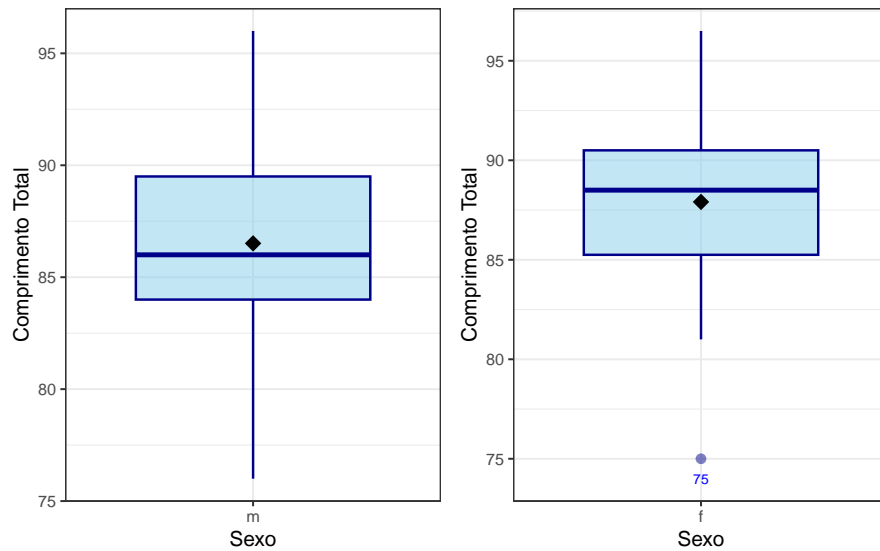
	Min	Q1	Med	Média	Q3	Max	D. Padrão	CV
Comprimento Total	76	84,0	86,0	86,51	89,5	96,0	4,34	0,05
Largura Crânio	50	54,8	56,3	57,09	59,0	68,6	3,45	0,06

A seguir são apresentadas as tabelas com as principais medidas resumo por sexo dos gambás.

Nota-se que as medidas de resumo não apresentam diferenças significativas entre os dois sexos, avaliando o Coeficiente de Variação de Pearson, sendo esta a medida que avalia o grau de variabilidade dos dados em relação a média, constata-se que sua classificação é baixa, possuindo o valor de 5%, menor que o limite considerado baixo (15%), para todas as características independente do sexo. A fim de se possibilitar uma análise visual, as Figuras 1 e 2 trazem os gráficos de caixa (*BoxPlot*) com as características morfológicas dos gambas separadas por sexo.

Figura 1: Comparação das medidas morfológicas por sexo

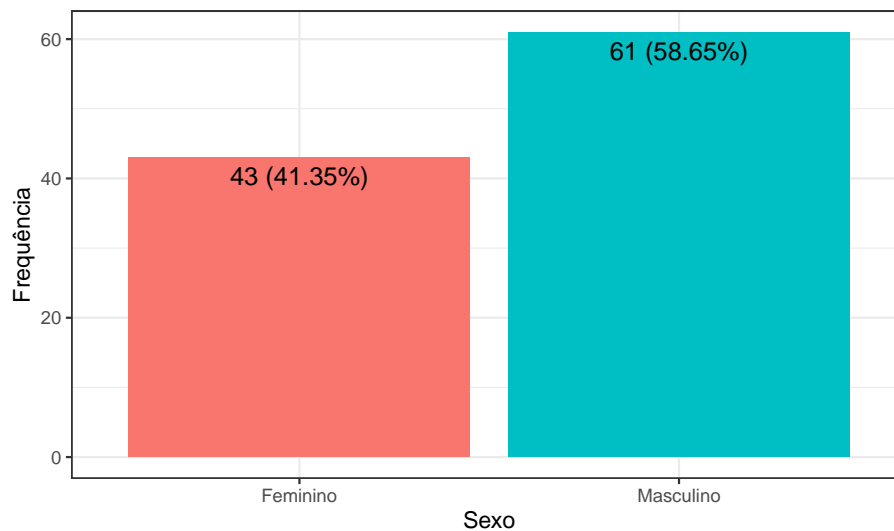




A Figura 1 confirma a percepção advinda das medidas resumo, que não há, em princípio, diferenças significativas entre os sexos, apesar de haver alguns pontos atípicos. É interessante notar a baixa variabilidade entre a largura do crânio para o sexo feminino e seu comportamento aparentemente simétrico, tendo em vista a média estar representada pelo ponto na região central do gráfico, possibilitando uma percepção visual acerca da assimetria tendo em vista os posicionamentos da média e da mediana. Cabe então uma avaliação mais apurada dos dados.

A Figura 2 retrata a distribuição do sexo, possibilitando identificar o quantitativo de informações por sexo. Note-se que a quantidade de dados de indivíduos do sexo masculino prevalece. É possível identificar pontos atípicos (*outliers*) presentes nos gráficos em ambas as medidas morfológicas e em ambos os sexos dos gambás. Apenas o comprimento dos indivíduos do sexo masculino não apresentou esses pontos atípicos.

Figura 2: Distribuição da variável sexo



A fim de se identificar com maior clareza a distribuição dos dados das variáveis constantes nos histogramas seguem gráficos de densidade juntamente com as medias e medianas sobrepostas.

Figura 3: Histograma das medidas morfológicas

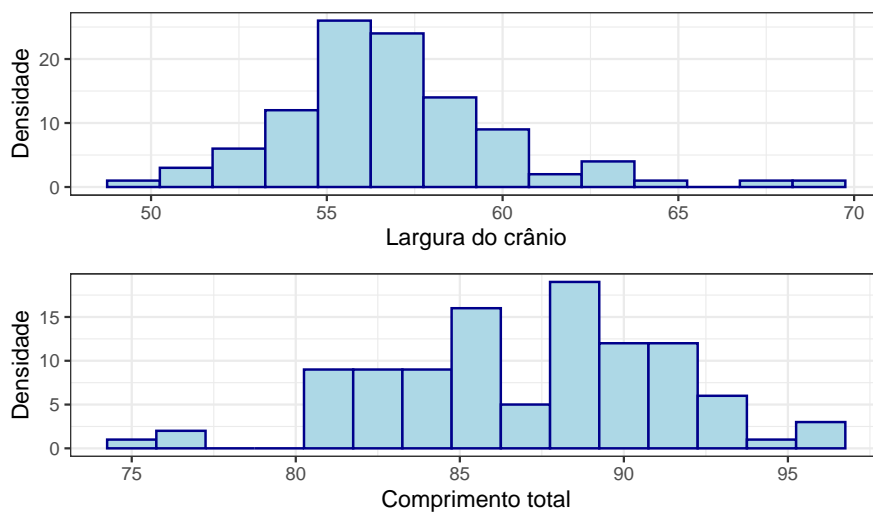


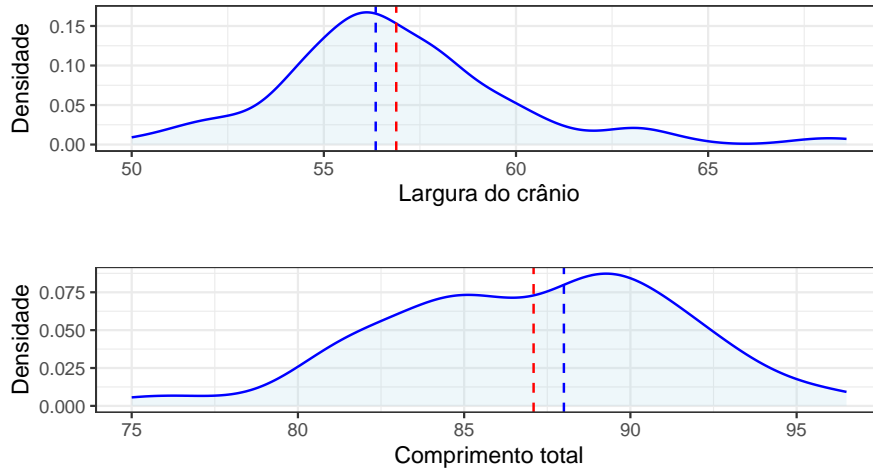
Table 3: Coeficientes de Assimetria

Variável	Coefficiente
Largura do Crânio	0,99
Comprimento Total	-0,28

Table 4: Coeficientes de Curtose

Variável	Coefficiente
Largura do Crânio	5,30
Comprimento Total	2,84

Figura 4: Média, mediana e densidade das medidas morfológicas



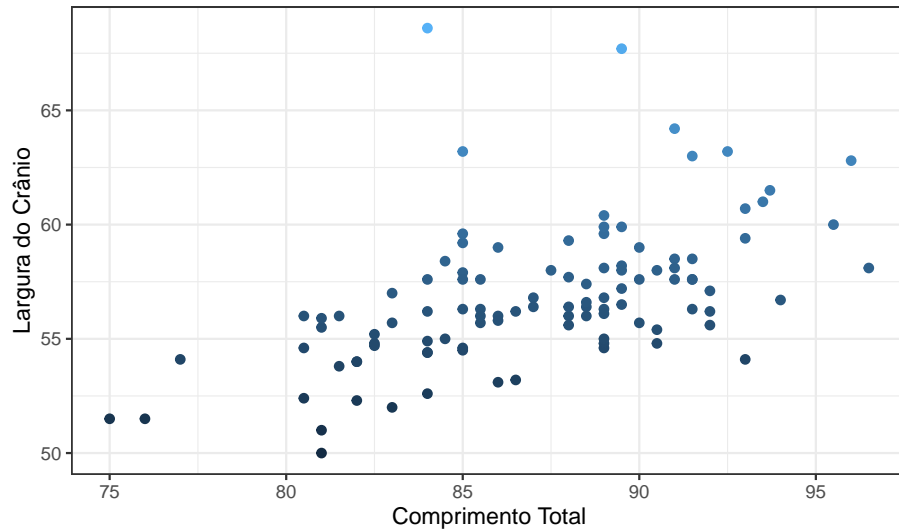
Analisando as Figuras 3 e 4 é possível identificar uma leve assimetria positiva para os dados referentes a largura do crânio dos gambás, para os dados referentes ao comprimento total se identifica um comportamento bimodal e uma leve assimetria negativa. A fim de dirimir quaisquer dúvidas sobre a assimetria e curtose dos dados em análise, seguem as tabelas com os valores calculados.

Baseado nos valores obtidos nas Tabelas 3 e 4, no que diz respeito ao coeficiente de assimetria é possível concluir que a distribuição dos dados referentes ao **comprimento total** apresenta um coeficiente de assimetria positivo, tendo em vista seu valor maior que 0, indicando que a maioria dos valores são menores que a média, caracterizada pela linha tracejada vertical vermelha presente nos gráficos de densidade, já os dados referentes a **largura do crânio** são considerados de assimetria negativa, ou seja, a maioria dos valores dos dados são maiores que a média, em virtude do seu valor menor que zero. Referente ao coeficiente

de curtose é possível concluir que a distribuição dos dados referentes ao **comprimento total** possui valor menor que 3, caracterizando um comportamento Platocúrtico, ou seja, mais achatado, já a distribuição dos dados referentes a **largura do crânio** apresenta um coeficiente maior que 3, caracterizando um comportamento leptocúrtico, demonstrando possuir um pico mais acentuado no gráfico, característica identificada nos gráficos de densidade.

Com o intuito de comparar a relação entre as variáveis largura do crânio (**skullw**) e comprimento total (**totlngth**), foi construída a Figura 5.

Figura 5: Relação entre Comprimento Total e Largura do Crânio



É possível identificar uma relação linear positiva entre as características em análise, ao se calcular o coeficiente de correlação de Pearson  $\hat{\rho}$  estimado obteve-se o valor de 0.5264 caracterizando uma relação **regular**.

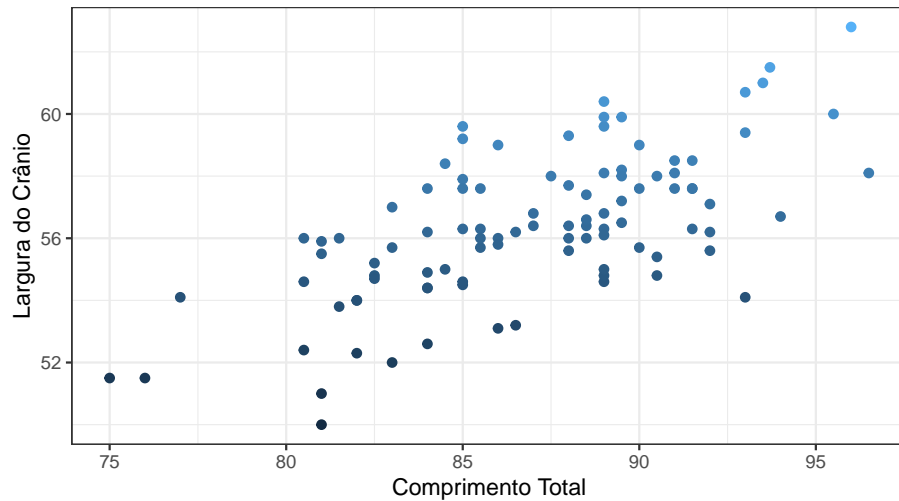
Para testar a significância da correlação linear ao nível de significância de 5% foi realizado o teste de hipótese para correlação linear, em que a hipótese nula foi rejeitada, podendo assumir com 95% de confiança que há correlação linear entre as variáveis estudadas.

Através da Figura 5 é possível identificar com mais clareza pontos atípicos na relação entre as variáveis morfológicas, a fim de identificar o quão influente são



estes pontos no comportamento geral dos dados, a Figura 6 foi construída sem estes pontos.

Figura 6: Relação entre Comprimento Total e Largura do Crânio com remoção dos pontos atípicos



Analisando o comportamento da relação sem os pontos atípicos não se nota diferença significativa acerca do comportamento da relação, permanecendo uma possível relação positiva. Ao se calcular o novo coeficiente de correlação de Pearson  $\hat{\rho}$  estimado obteve-se o valor de 0.6542, ligeiramente maior que o anteriormente calculado com todos os dados inseridos no cálculo, caracterizando ainda assim uma relação **regular**.

Em seguida foi feito o teste hipótese que avalia a significância da correlação linear ao nível de 5%, em que a hipótese nula foi rejeitada, podendo assumir com 95% de confiança que permanece a correlação linear entre as variáveis estudadas.

## Segunda parte

### Apresentação

Com base nos dados sobre a eleição presidencial de 2000 nos Estados Unidos, referentes ao número de votos de cada um dos candidatos por condado no estado da Flórida. Deseja-se investigar a relação entre o número de votos que Bush recebeu em relação ao número de votos recebidos por Buchanan, bem como, trazer um pouco de luz sobre o debate referente aos votos recebidos por Buchanan que poderiam ter sido de Al Gore, se o primeiro não estivesse no pleito. Como Bush e Gore foram os candidatos principais daquela eleição, é de interesse avaliar a relação entre os votos recebidos por Bush e Buchanan na Flórida, que é um Estado importante na corrida presidencial dos EUA. Para isto, ajuste um modelo de regressão linear no qual o número de votos de Bush é usado para prever o número de votos de Buchanan. Os dados estão disponíveis no arquivo “florida.csv”.

### Objetivos

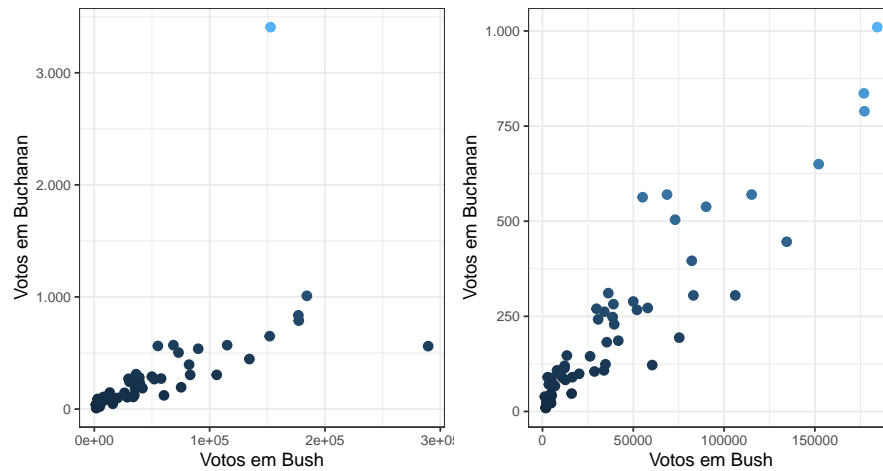
O objetivo dessa análise responderá as seguintes questões:

- (a) Discutir a relação entre os votos recebidos por Bush e por Buchanan através do uso de diagramas de dispersão.
- (b) Avaliar a relevância do argumento que os votos de Buchanan poderiam ser direcionados a Al Gore, caso Buchanan não tivesse participado do pleito.
- (c) Tratar dados atípicos.
- (d) Obter uma reta ajustada aos dados tratados e prever a votação de Buchanan caso Bush receba 152.846 votos em determinado condado.
- (e) Criar um programa baseado em estimativas de mínimos quadrados, prever a votação de Buchanan, sob as mesmas condições e compará-las.

### Análise dos dados

O gráfico de dispersão sugere certa correlação positiva entre os votos de Bush e Buchanan, muito embora se observe que no condado de Palm Beach a votação de Buchanan (3.407 votos) represente um dado inesperado, bem como, em menor medida, a votação de Bush em Dade (289.456 votos). Retirados estes valores discrepantes, esta provável correlação parece ainda mais certa.

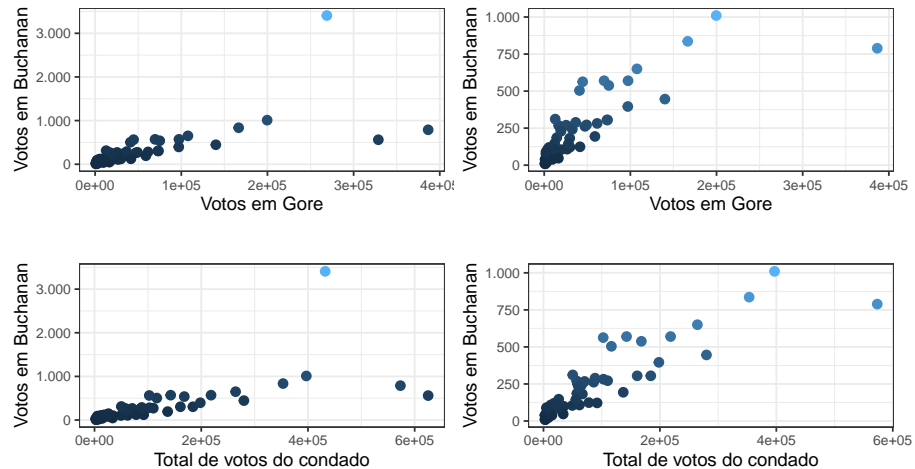
Figura 6: Relação entre votos recebidos por Bush e por Buchanan nos condados da Fló



Fonte: 2000 OFFICIAL PRESIDENTIAL GENERAL ELECTION RESULTS – USA

Embora a Figura 1 sugira certa correlação positiva entre os votos de Bush e Buchanan, a análise destes mesmos votos em relação a votação recebida por Al Gore e os votos totais dos condados parece retratar outra realidade (ver Figura 2), qual seja, que a correlação positiva é mais propriamente deviada a um fator externo, e não avaliado inicialmente, que corresponde ao aumento de eleitores nos condados. Desta forma, não parece ser plausível afirmar que a votação dada ao candidato Buchanan seria direcionada a qualquer dos candidatos caso este não participasse do pleito.

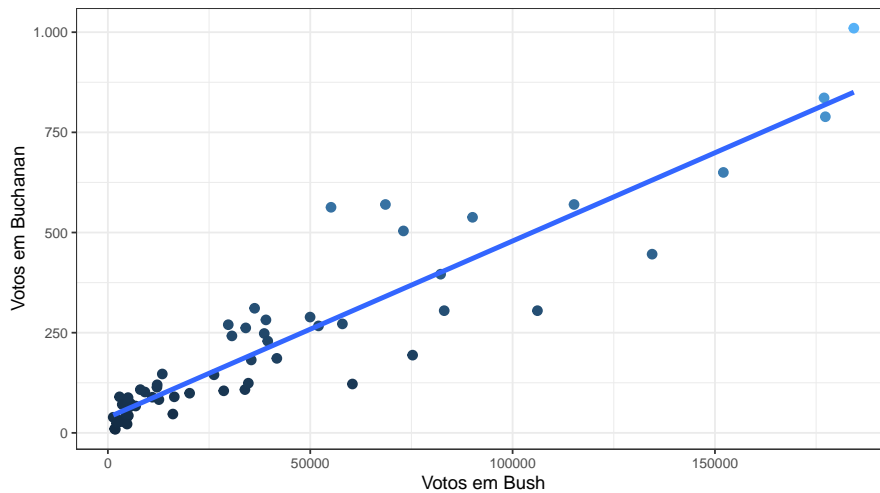
Figura 7: Relação entre votos recebidos por Gore e por Buchanan e com o total de voto



Fonte: 2000 OFFICIAL PRESIDENTIAL GENERAL ELECTION RESULTS – USA

Assumindo que haja uma correlação entre os votos dos dois candidatos é possível realizar uma regressão linear a partir dos dados dos votos dos condados. A Figura 3 mostra a reta de regressão linear obtida. Com base nesta regressão é possível inferir que caso o candidato Bush tivesse recebido 152.846 votos em um determinado condado, a votação do candidato Buchanan seria de aproximadamente 791 votos.

Figura 8: Relação entre votos recebidos por Bush e por Buchanan nos condados da Fló



É possível elaborar um código de programação em linguagem R que realize esta mesma previsão utilizando-se do método de Estimativas de Mínimos Quadrados, para se obter  $\hat{\beta}_0$  e  $\hat{\beta}_1$  e assim, prever um valor de votação do candidato Buchanan sob as mesmas condições, como se segue:

```
x = dados$BUSH
y = dados$BUCHANAN
n = length(x)
beta1 = (sum(x*y) - as.numeric(sum(x)) * as.numeric(sum(y))/n) /
  (sum(x^2) - ((sum(x))^2)/n)
beta0 = mean(y) - beta1 * mean(x)
votos = round(beta0 + 152846 * beta1, 0)

cat("Caso o candidato Bush tivesse recebido 152.846 votos em um
determinado condado, a votação do candidato Buchanan seria
de aproximadamente", votos," votos.")
```

Executando o código acima, obtém-se a estimação por mínimos quadrados de 797 votos. Observa-se que por regressão linear a resposta foi de 791 votos, desta

forma, percebe-se que ambos os métodos foram similares na predição da votação do candidato Buchanan.

## Conclusão

Parte 1.

A atividade permite avaliar um modelo de predição que tem boa aderência aos dados disponibilizados, ao mesmo tempo que mostra a fragilidade da técnica quando assume que a correlação entre as votações dos candidatos pode ser explicada apenas pelos votos observados.

## Apêndice

### Script em linguagem R

```
# PACOTES ----
set.seed(7)
setwd("~/Dropbox/Estatística/2023.1/Regressão/Lab/Lab2")

library(pacman)
library(tidyverse)
library(janitor)
library(kableExtra)
library(moments)
library(ggthemes)
library(patchwork)
library(glue)
library(formattable)
library(gridExtra)

#####
#
#                                     PRIMEIRA PARTE
#
#####

# DADOS ----
dados <- read.csv2("possum_red.csv")

dados <- dados |>
  mutate(
    skullw = as.numeric(skullw),
    totlngth = as.numeric(totlngth),
    sex = forcats::as_factor(sex)
  )

# MEDIDAS DE RESUMO ----
dados|>
  filter(sex == "f")|>
  rename("Largura Crânio" = skullw, "Comprimento Total" = totlngth)|>
  summarytools::descr(
    stats = c("min", "q1", "med", "mean", "q3", "max", "sd", "cv"),
    justify = "c", style = "grid", transpose = T
  )|>
```

```

kbl(
  caption = "Medidas Resumo para o sexo feminino.",
  format.args=list(big.mark=".", decimal.mark=","),
  digits = 2, align = "c", row.names = T,
  col.names =
    c("Min", "Q1", "Med", "Média", "Q3", "Max", "D. Padrão", "CV")
)|>
kable_material(c("striped", "hover", "condensed"))|>
kable_material()

dados|>
filter(sex == "m")|>
rename("Largura Crânio" = skullw, "Comprimento Total" = totlngth)|>
summarytools::descr(
  stats = c("min", "q1", "med", "mean", "q3", "max", "sd", "cv"),
  justify = "c", style = "grid", transpose = T
) |>
kbl(
  caption = "Medidas Resumo para o sexo masculino",
  format.args=list(big.mark=".", decimal.mark=","),
  digits = 2, align = "c",
  row.names = T,
  col.names =
    c("Min", "Q1", "Med", "Média", "Q3", "Max", "D. Padrão", "CV")
)|>
kable_material(c("striped", "hover", "condensed"))|>
kable_material()

# BOXPLOTS ----
b1 <- dados|>
  filter(sex == "m")|>
  ggplot(aes(x = sex, y = skullw)) +
  geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
  labs(
    title = 'Medidas do sexo Masculino',
    x = "Sexo",
    y = "Largura do Crânio"
  ) +
  scale_y_continuous(
    labels = scales::number_format(
      dig.mark = ".",
      decimal.mark = ",")+
  theme_bw()

```



```

b2 <- dados|>
  filter(sex == "f")|>
  ggplot(aes(x = sex, y = skullw)) +
  geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
  labs(
    title = 'Medidas do sexo Feminino',
    x = "Sexo",
    y = "Largura do Crânio"
  ) +
  scale_y_continuous(
    labels = scales::number_format(
      dig.mark = ".",
      decimal.mark = ",")+
  theme_bw()

b3 <- dados|>
  filter(sex == "m")|>
  ggplot(aes(x = sex, y = totlngth)) +
  geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
  labs(
    x = "Sexo",
    y = "Comprimento Total"
  ) +
  scale_y_continuous(
    labels = scales::number_format(
      dig.mark = ".",
      decimal.mark = ",")+
  theme_bw()

b4 <- dados|>
  filter(sex == "f")|>
  ggplot(aes(x = sex, y = totlngth)) +
  geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
  labs(
    x = "Sexo",
    y = "Comprimento Total"
  ) +
  scale_y_continuous(
    labels = scales::number_format(
      dig.mark = ".",
      decimal.mark = ",")+
  theme_bw()

(b1+b2)/(b3+b4)+

```

```

plot_annotation(
  title = "Figura 1: Comparação das medidas morfológicas por sexo",
) & theme_bw(base_size = 9) &
theme(
  plot.tag.position = c(0, 1),
  plot.tag = element_text(size = 12, hjust = 0, vjust = 0)
)

# DADOS POR SEXO ----
dados|>
  count(sex)|>
  mutate(
    sex = forcats::fct_reorder(sex, n),
    sex = lvls_revalue(sex, c("Feminino", "Masculino")),
    pct = prop.table(n) * 100|>
  ggplot(aes(x = sex, y = n, fill = sex))+
  geom_col() +
  geom_text(
    aes(
      label = glue::glue("{n} ({round(pct, 2)}%)"), vjust = 1.5)+
  labs(
    title = "Figura 2: Distribuição da variável sexo",
    x = "Sexo", y = "Frequência"
  ) + theme_bw(base_size = 10) +
  theme(legend.position = "none")

# HISTOGRAMAS POR SEXO ----
h1 <- dados|>
  ggplot() +
  aes(x = skullw) +
  geom_histogram(
    binwidth = 1.5,
    fill = "lightblue",
    colour = "darkblue") +
  labs(
    x = "Largura do crânio",
    y = "Densidade"
  )

h2 <- dados|>
  ggplot() +
  aes(x = totlngth) +
  geom_histogram(
    binwidth = 1.5,

```

```

    fill = "lightblue",
    colour = "darkblue") +
  labs(
    x = "Comprimento total",
    y = "Densidade"
  )

(h1/h2)+
  plot_annotation(
    title = "Figura 3: Histograma das medidas morfológicas por sexo",
  ) & theme_bw(base_size = 10) &
  theme(
    plot.tag.position = c(0, 1),
    plot.tag = element_text(size = 12, hjust = 0, vjust = 0)
  )

# DISTRIBUIÇÕES POR MM ----
d1 <- dados |>
  ggplot(aes(x = skullw)) +
  geom_density(
    fill = "lightblue",
    colour = "blue",
    alpha = 0.2) +
  geom_vline(
    xintercept = mean(dados$skullw),
    color = "red",
    linetype = "dashed"
  ) +
  geom_vline(
    xintercept = quantile(dados$skullw, 0.5),
    color = "blue",
    linetype = "dashed"
  ) +
  labs(
    title = "",
    x = "Largura do crânio",
    y = "Densidade"
  ) +
  scale_x_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    )
  )

```

```

d2 <- dados |>
  ggplot(aes(x = totlngth)) +
  geom_density(
    fill = "lightblue",
    colour = "blue",
    alpha = 0.2) +
  geom_vline(
    xintercept = mean(dados$totlngth),
    color = "red",
    linetype = "dashed") +
  geom_vline(
    xintercept = quantile(dados$totlngth, 0.5),
    color = "blue",
    linetype = "dashed"
  ) +
  labs(
    title = "",
    x = "Comprimento total",
    y = "Densidade"
  ) +
  scale_x_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    )
  )

(d1/d2)+
  plot_annotation(
    title = "Figura 4: Média, mediana e distribuição das medidas morfológicas",
  ) & theme_bw(base_size = 10) &
  theme(
    plot.tag.position = c(0, 1),
    plot.tag = element_text(size = 12, hjust = 0, vjust = 0),
    legend.position = "none"
  )

# RELAÇÃO COMPRIMENTO X LARGURA ----
dados|>
  ggplot(aes(
    y = skullw,
    x = totlngth, color = skullw)) +
  geom_point()+
  labs(
    title = 'Figura 5: Relação entre Comprimento Total e Largura do Crânio',

```

```

    x = 'Comprimento Total',
    y = 'Largura do Crânio',
  )+
  theme_bw(base_size = 10)+
  theme(legend.position = "none")

#####
#                                                                 #
#                               SEGUNDA PARTE                     #
#                                                                 #
#####

# DADOS ----
dados <- read.csv2("florida.csv")
m.bh = max(dados$BUSH)
m.bn = max(dados$BUCHANAN)
aux = dados %>% dplyr::filter(dados$BUSH < m.bh)
dados.sem = aux %>% dplyr::filter(aux$BUCHANAN < m.bn)

# GRÁFICOS DE DISPERSÃO BUBH X BUCHANAN ----

d1 <- dados |>
  ggplot(aes(
    x = BUSH,
    y = BUCHANAN, color = BUCHANAN)) +
  geom_point()+
  labs(
    title = '',
    x = 'Votos em Bush',
    y = 'Votos em Buchanan'
  )+
  scale_y_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    )
  )

d1.1 <- dados.sem |>
  ggplot(aes(
    x = BUSH,
    y = BUCHANAN, color = BUCHANAN)) +
  geom_point()+

```

```

labs(
  title = '',
  x = 'Votos em Bush',
  y = 'Votos em Buchanan'
)+
scale_y_continuous(
  labels = scales::number_format(
    big.mark = ".",
    decimal.mark = ",",
  ))
d1+d1.1+plot_annotation(
  title = "Figura 6: Relação entre votos recebidos por Bush e
por Buchanan nos condados da Flórida. (a) Dados completos,
(b) Sem dados discrepantes",
  caption = "Fonte: 2000 OFFICIAL PRESIDENTIAL GENERAL ELECTION
RESULTS - USA") &
theme_bw(base_size = 8) &
theme(
  legend.position = "none",
  plot.tag.position = c(0, 1),
  plot.tag = element_text(size = 8, hjust = 0, vjust = 0)
)

# GRÁFICOS DE DISPERSÃO GORE X BUCHANAN ----

d1 <- dados |>
ggplot(aes(
  x = GORE,
  y = BUCHANAN, color = BUCHANAN)) +
geom_point()+
labs(
  title = '',
  x = 'Votos em Gore',
  y = 'Votos em Buchanan'
)+
scale_y_continuous(
  labels = scales::number_format(
    big.mark = ".",
    decimal.mark = ",",
  ))
))

d1.1 <- dados.sem |>
ggplot(aes(

```

```

      x = GORE,
      y = BUCHANAN, color = BUCHANAN)) +
geom_point()+
labs(
  title = '',
  x = 'Votos em Gore',
  y = 'Votos em Buchanan'
)+
scale_y_continuous(
  labels = scales::number_format(
    big.mark = ".",
    decimal.mark = ",",
  ))

d2 <- dados |>
ggplot(aes(
  x = TOTAL,
  y = BUCHANAN, color = BUCHANAN)) +
geom_point()+
labs(
  title = '',
  x = 'Total de votos do condado',
  y = 'Votos em Buchanan'
)+
scale_y_continuous(
  labels = scales::number_format(
    big.mark = ".",
    decimal.mark = ",",
  ))

d2.2 <- dados.sem |>
ggplot(aes(
  x = TOTAL,
  y = BUCHANAN, color = BUCHANAN)) +
geom_point()+
labs(
  title = '',
  x = 'Total de votos do condado',
  y = 'Votos em Buchanan'
)+
scale_y_continuous(
  labels = scales::number_format(
    big.mark = ".",
    decimal.mark = ",",
  ))

```

```

    ))
(d1+d1.1)/(d2+d2.2) + plot_annotation(
  title = "Figura 7: Relação entre votos recebidos por Gore e por
  Bachanan e com o total de votos nos condados da Flórida.
  (a) Dados completos, (b) Sem dados discrepantes",
  caption = "Fonte: 2000 OFFICIAL PRESIDENTIAL GENERAL ELECTION
  RESULTS - USA") &
  theme_bw(base_size = 8) &
  theme(
    legend.position = "none",
    plot.tag.position = c(0, 1),
    plot.tag = element_text(size = 8, hjust = 0, vjust = 0)
  )

# REGRESSÃO LINEAR ----

reg = lm(dados.sem$BUSH ~ dados.sem$BUCHANAN)
intercepto = reg$coefficients[1][[1]]
c.angular = reg$coefficients[2][[1]]
resposta = round((152846 - intercepto)/c.angular, 0)

d1<- dados.sem |>
  ggplot(aes(
    x = BUSH,
    y = BUCHANAN, color = BUCHANAN)) +
  geom_point()+
  labs(
    title = '',
    x = 'Votos em Bush',
    y = 'Votos em Buchanan'
  )+
  scale_y_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    )
  )+
  geom_smooth(method=lm, se=FALSE)

d1+plot_annotation(
  title = "Figura 8: Relação entre votos recebidos por Bush e por
  Bachanan nos condados da Flórida e sua regressão linear.") &
  theme_bw(base_size = 8) &
  theme(

```



```

        legend.position = "none",
        plot.tag.position = c(0, 1),
        plot.tag = element_text(size = 8, hjust = 0, vjust = 0)
    )

# MÍNIMOS QUADRADOS ----

x = dados$BUSH
y = dados$BUCHANAN
n = length(x)
beta1 = (sum(x*y) - as.numeric(sum(x)) * as.numeric(sum(y))/n) /
        (sum(x^2) - ((sum (x))^2)/n)
beta0 = mean(y) - beta1 * mean(x)
votos = round(beta0 + 152846 * beta1, 0)

cat("Caso o candidato Bush tivesse recebido 152.846 votos em um
    determinado condado, a votação do candidato Buchanan seria
    de aproximadamente", votos," votos.")

```