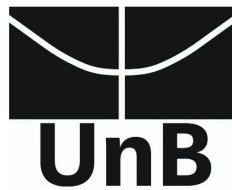


Allan Victor Almeida Faria (190127180), Ananda Almeida de Sá (150117345),  
Bruno Kevyn Andrade de Souza

## **Trabalho de Regressão Linear**

Brasília, DF

21/02/2021



Allan Victor Almeida Faria (190127180), Ananda Almeida de Sá (150117345),  
Bruno Kevyn Andrade de Souza

## **Trabalho de Regressão Linear**

Trabalho de Regressão Linear de Análise  
de dados hospitalares.

Universidade de Brasília (UnB)  
Instituto de Ciências Exatas (IE)  
Departamento de Estatística (DE)

Brasília, DF

21/02/2021

# Resumo

resumo aqui

**Palavras-chaves:** 1. Análise de dados.

# Lista de ilustrações

Figura 1 – Gráfico de box-plot das variáveis dos dados. . . . .	11
Figura 2 – Gráfico de calor da correlação entre as variáveis dos dados. . . . .	12

# Lista de tabelas

Tabela 1 – Descrição dos códigos da tabela com a seguinte indentificação da variável.	8
Tabela 2 – Medidas descritivas para boxplots . . . . .	10

# Lista de abreviaturas e siglas

INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
SAEB	Sistema de Avaliação da Educação Básica

# Lista de símbolos

$\sigma$  Letra grega minúscula sigma

$\mu$  Letra grega minúscula mu

# Sumário

<b>1</b>	<b>RESULT</b>	<b>8</b>
<b>1.1</b>	<b>Introdução</b>	<b>8</b>
1.1.1	Leitura de dados	8
1.1.1.1	Descrição das variáveis	8
1.1.1.2	Correlação entre as variáveis	11
<b>1.2</b>	<b>Objetivo</b>	<b>13</b>
1.2.1	Número de enfermeira(o)s	13
1.2.2	Número de enfermeira(o)s	16
	<b>REFERÊNCIAS</b>	<b>19</b>
	<b>ANEXOS</b>	<b>20</b>
	<b>ANEXO A – AMOSTRA</b>	<b>21</b>





## 1 RESULT

### 1.1 Introdução

#### 1.1.1 Leitura de dados

O programa utilizado para analisar os dados disponibilizados em Excel será o R Studio, versão 4.2.0, importados como um data frame (planilha), onde as colunas representam as variáveis de estudo e cada linha representa um hospital dos Estados Unidos no período de 1975-1976.

```
# Tabela de nomes X1: Nome variavel
knitr::kable(cbind(names(data),names(datax)),
              caption = 'Descrição dos códigos da tabela com a seguinte indentificação
```

Tabela 1 – Descrição dos códigos da tabela com a seguinte indentificação da variável.

Número de Identificação	ID
Duração da Internação	X1
Idade	X2
Risco de Infecção	X3
Proporção de Culturas de Rotina	X4
Proporção de Raio-X de Tórax de Rotina	X5
Número de leitos	X6
Filiação a Escola de Medicina	X7
Região	X8
Média diária de pacientes	X9
Número de enfermeiro(s)	X10
Facilidades e serviços disponíveis	X11

##### 1.1.1.1 Descrição das variáveis

- **Duração da Internação**



A duração de internação é uma variável quantitativa contínua que representa a duração média da internação de todos os pacientes no hospital (em dias).

- **Idade**

A idade é uma variável quantitativa contínua que representa a idade média dos pacientes de cada hospital.

- **Risco de Infecção**

O risco de infecção é uma variável quantitativa contínua que representa a probabilidade média estimada de adquirir infecção no hospital (em %).

- **Proporção de Culturas de Rotina**

A proporção de culturas de rotina é uma variável quantitativa contínua que representa a razão do número de culturas realizadas com relação ao número de pacientes sem sinais ou sintomas de infecção adquirida no hospital, vezes 100.

- **Proporção de Raio-X de Tórax de Rotina**

A proporção de raio-X de tórax de rotina é uma variável quantitativa contínua que representa a razão do número de raio-X de tórax realizados com relação ao número de pacientes sem sinais ou sintomas de pneumonia, vezes 100.

- **Número de leitos**

O número de leitos é uma variável quantitativa discreta que representa o número médio de leitos no hospital durante o período de estudo.

- **Filiação a Escola de Medicina**

A filiação a escola de medicina é uma variável qualitativa ordinal onde o 1 significa que a escola tem filiação, e 2 que não tem.

- **Região**



A região é uma variáveis qualitativas nominal onde “NE” se refere ao Nordeste, “W” ao Oeste, “S” ao Sul e “NC” ao ?????.

- **Média diária de pacientes**

O média diária de pacientes é uma variável quantitativa discreta que representa o número médio de pacientes no hospital por dia durante o período do estudo.

- **Número de enfermeiro(s)**

O número de enfermeiro(s) é uma variável quantitativa discreta que representa o Número médio de enfermeiros(as) de tempo-integral ou equivalente registrados e licenciados durante o período de estudo (número de tempos integrais+metade do número de tempo parcial).

- **Facilidades e serviços disponíveis**

A facilidades e serviços disponíveis é uma variável quantitativa contínua que representa a porcentagem de 35 potenciais facilidades e serviços que são fornecidos pelo hospital.

Tabela 2 – Medidas descritivas para boxplots

Variaveis	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Duração da Internação	6.700	8.340	9.420	9.648	10.470	19.560
Idade	38.80	50.90	53.20	53.23	56.20	65.90
Risco de Infecção	1.300	3.700	4.400	4.355	5.200	7.800
Proporção de Culturas de Rotina	1.60	8.40	14.10	15.79	20.30	60.50
Proporção de Raio-X de Tórax de Rotina	39.60	69.50	82.30	81.63	94.10	133.50
Número de leitos	29.0	106.0	186.0	252.2	312.0	835.0
Média diária de pacientes	20.0	68.0	143.0	191.4	252.0	791.0
Número de enfermeiro(s)	14.0	66.0	132.0	173.2	218.0	656.0
Facilidades e serviços disponíveis	5.70	31.40	42.90	43.16	54.30	80.00

```
datax2 <-datax %>%  
  select(X5,X2,X4,X11)  
datax3 <-datax %>%  
  select(X1,X3)
```



```
datax1 <-datax %>%  
  select(X6,X9,X10)
```

```
par(mfrow = c(1,3))  
boxplot(datax1)  
boxplot(datax2)  
boxplot(datax3)
```

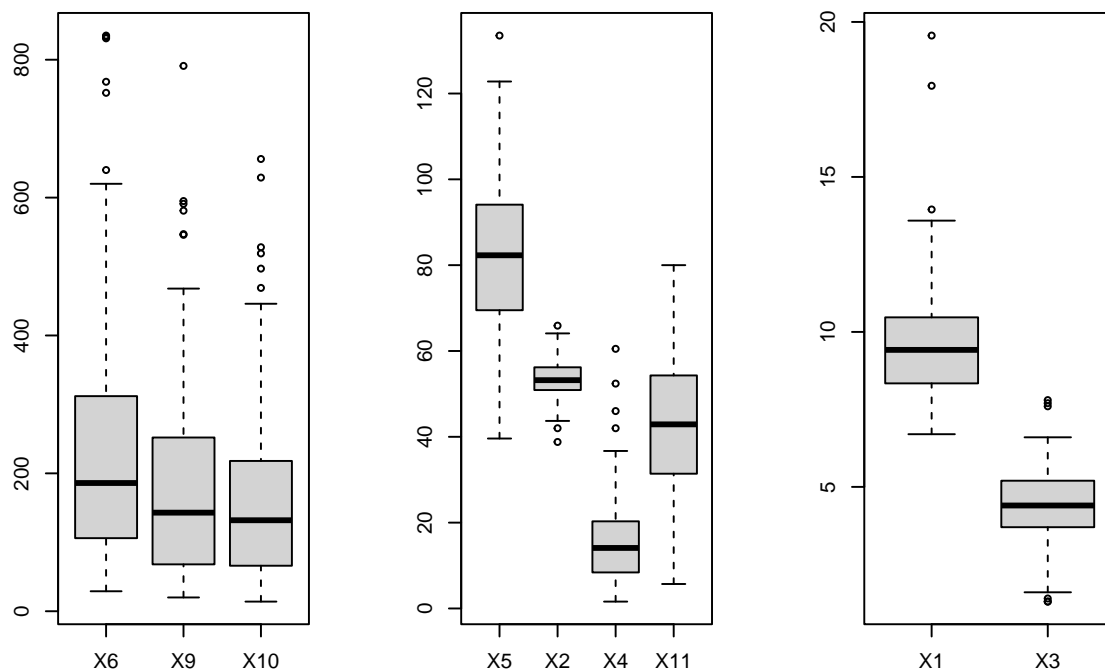


Figura 1 – Gráfico de box-plot das variáveis dos dados.

### 1.1.1.2 Correlação entre as variáveis

Para verificar a natureza e a força da relação entre as variáveis e identificar lacunas e pontos discrepantes no conjunto de dados, utiliza-se a matriz de correlação aplicado no script a seguir.



```
library(ggcorrplot)
```

```
## Carregando pacotes exigidos: ggplot2
```

```
library(dplyr)
```

```
pmat = datax %>% select_if(is.numeric) %>% cor_pmat()
```

```
datax %>% select_if(is.numeric) %>% cor(.) %>%  
  ggcorrplot( type = "lower", p.mat = pmat, hc.order = TRUE)
```

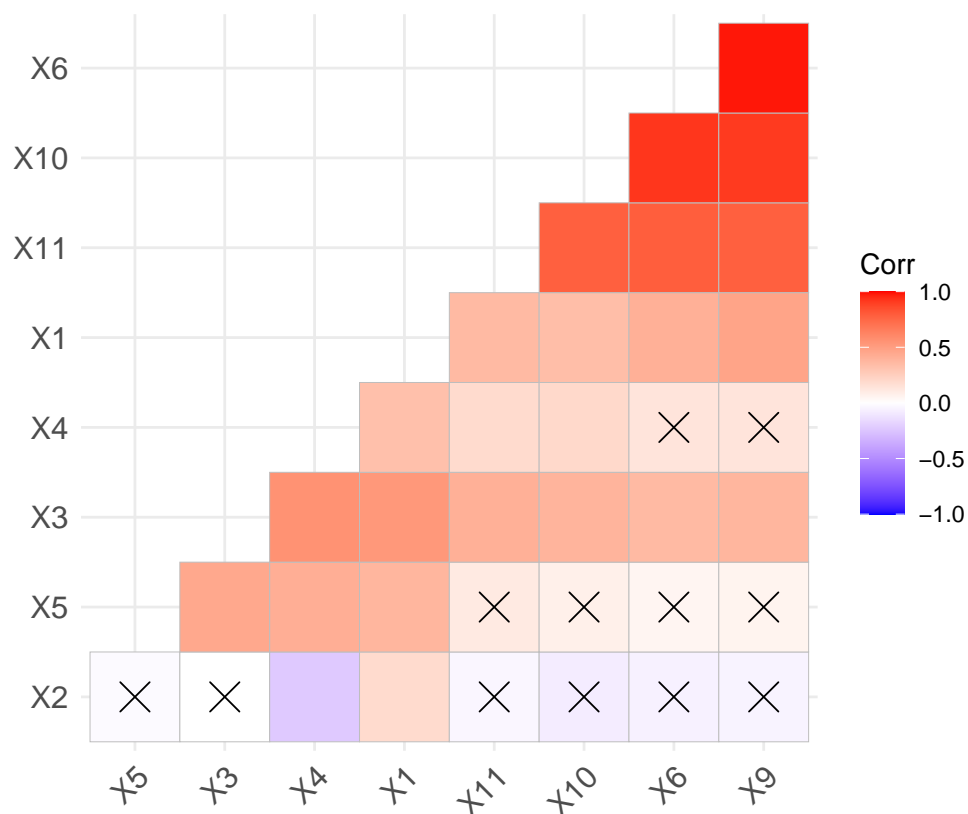


Figura 2 – Gráfico de calor da correlação entre as variáveis dos dados.



## 1.2 Objetivo

### 1.2.1 Número de enfermeira(o)s

Deseja-se estudar se o número de enfermeira(o)s está relacionado às instalações, ou seja, os números de leitos do hospital, e se há diferenças entre os serviços disponíveis pelos hospitais. Neste caso, a variável resposta é o número de enfermeira(o)s e as duas outras variáveis são explicativas. Para isso, faz-se necessário a aplicação da regressão linear múltipla realizada no script a seguir:

```
# teste de ausencia de regresao (significativo)
summary(aov(`Número de enfermeiro(s)` ~ `Número de leitos` + `Facilidades e serviços d
```

```
##                                Df  Sum Sq Mean Sq F value Pr(>F)
## `Número de leitos`           1 1820644 1820644 601.362 <2e-16 ***
## `Facilidades e serviços disponíveis` 1   18550   18550   6.127 0.0148 *
## Residuals                    110   333029    3028
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
modcomp<-lm(`Número de enfermeiro(s)` ~ `Número de leitos` +
            I(`Número de leitos`^2)+
            `Facilidades e serviços disponíveis`+
            I(`Facilidades e serviços disponíveis`^2)+
            `Facilidades e serviços disponíveis`*`Número de leitos`
            , data = data)
```

```
summary(modcomp)
```

```
##
## Call:
## lm(formula = `Número de enfermeiro(s)` ~ `Número de leitos` +
##     I(`Número de leitos`^2) + `Facilidades e serviços disponíveis` +
##     I(`Facilidades e serviços disponíveis`^2) + `Facilidades e serviços disponíveis`
##     `Número de leitos`, data = data)
##
```



```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -162.973  -24.758   -2.339   21.880   210.964
##
## Coefficients:
##                                Estimate Std. Error
## (Intercept)                  -3.125e+01  3.698e+01
## `Número de leitos`           4.217e-01  2.023e-01
## I(`Número de leitos`^2)      -5.006e-04  2.686e-04
## `Facilidades e serviços disponíveis`  2.212e+00  2.350e+00
## I(`Facilidades e serviços disponíveis`^2) -3.646e-02  3.770e-02
## `Número de leitos`:`Facilidades e serviços disponíveis`  9.797e-03  5.920e-03
##                                t value Pr(>|t|)
## (Intercept)                  -0.845    0.3999
## `Número de leitos`           2.084    0.0395 *
## I(`Número de leitos`^2)      -1.864    0.0651 .
## `Facilidades e serviços disponíveis`  0.941    0.3486
## I(`Facilidades e serviços disponíveis`^2) -0.967    0.3356
## `Número de leitos`:`Facilidades e serviços disponíveis`  1.655    0.1009
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54.89 on 107 degrees of freedom
## Multiple R-squared:  0.8516, Adjusted R-squared:  0.8447
## F-statistic: 122.8 on 5 and 107 DF,  p-value: < 2.2e-16
```

```
shapiro.test(residuals(modcomp))
```

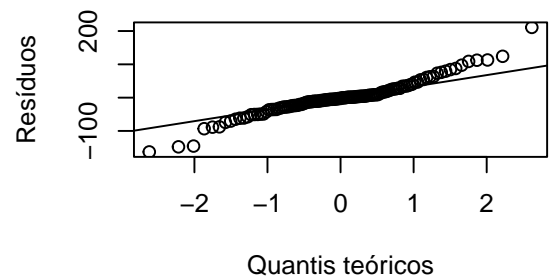
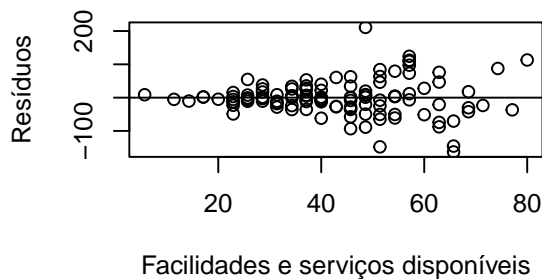
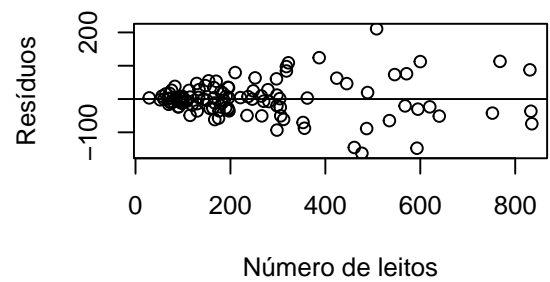
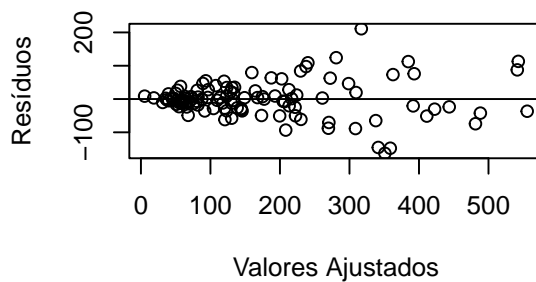
```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(modcomp)
## W = 0.94595, p-value = 0.0001791
```



```
# windows()
par(mfrow = c(2, 2))
plot(fitted(modcomp), residuals(modcomp), xlab="Valores Ajustados", ylab="Resíduos")
abline(h=0)
plot(data$`Número de leitos`, residuals(modcomp), xlab="Número de leitos", ylab="Resíduos")
abline(h=0)

plot(data$`Facilidades e serviços disponíveis`, residuals(modcomp), xlab="Facilidades e serviços disponíveis", ylab="Resíduos")
abline(h=0)

qqnorm(residuals(modcomp), ylab="Resíduos", xlab="Quantis teóricos", main="")
qqline(residuals(modcomp))
```







## 1.2.2 Número de enfermeira(o)s

```
# Teste de ausencia de regressao
summary(aov(`Duração da Internação` ~
            `Proporção de Raio-X de Tórax de Rotina`+
            `Proporção de Culturas de Rotina`,
            data = data))
```

```
##                                Df Sum Sq Mean Sq F value    Pr(>F)
## `Proporção de Raio-X de Tórax de Rotina`      1    59.9    59.86   19.605 2.26e-05
## `Proporção de Culturas de Rotina`              1    13.5    13.46    4.406  0.0381
## Residuals                                110   335.9     3.05
##
## `Proporção de Raio-X de Tórax de Rotina` ***
## `Proporção de Culturas de Rotina`             *
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Regressao
modcomp<-lm(`Duração da Internação` ~
            `Proporção de Raio-X de Tórax de Rotina`+
            `Proporção de Culturas de Rotina`,
            data = data)
summary(modcomp)
```

```
##
## Call:
## lm(formula = `Duração da Internação` ~ `Proporção de Raio-X de Tórax de Rotina` +
##     `Proporção de Culturas de Rotina`, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8987 -1.0846 -0.2387  0.6384  8.9177
##
```



```
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        6.66147    0.71664   9.295 1.61e-15
## `Proporção de Raio-X de Tórax de Rotina` 0.02935    0.00942   3.116 0.00234
## `Proporção de Culturas de Rotina`        0.03741    0.01782   2.099 0.03809
##
## (Intercept)                        ***
## `Proporção de Raio-X de Tórax de Rotina` **
## `Proporção de Culturas de Rotina`        *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.747 on 110 degrees of freedom
## Multiple R-squared:  0.1792, Adjusted R-squared:  0.1642
## F-statistic: 12.01 on 2 and 110 DF, p-value: 1.922e-05
```

```
# Teste de normalidade
```

```
shapiro.test(residuals(modcomp))
```

```
##
## Shapiro-Wilk normality test
##
## data: residuals(modcomp)
## W = 0.86649, p-value = 1.146e-08
```

```
# bartlett.test(residuals(modcomp))
```

```
# windows()
```

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

```
plot(fitted(modcomp), residuals(modcomp), xlab="Valores Ajustados", ylab="Resíduos")
```

```
abline(h=0)
```

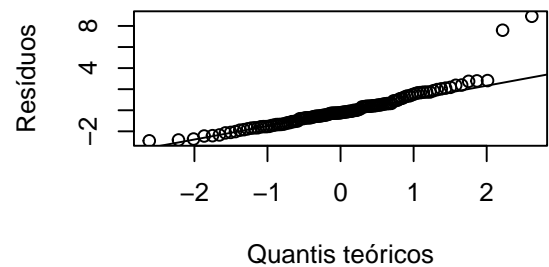
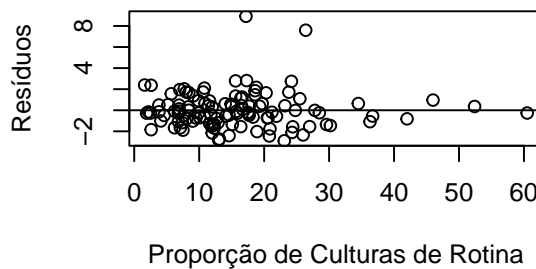
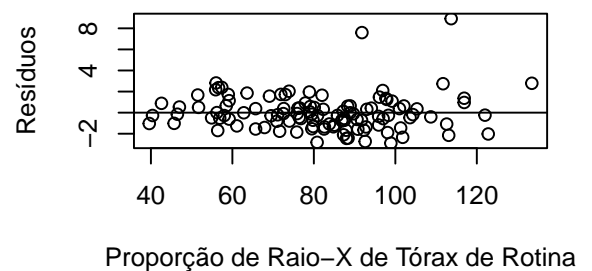
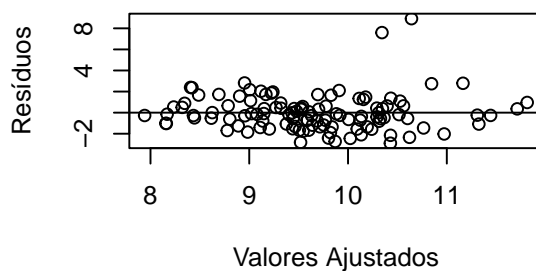
```
plot(data$`Proporção de Raio-X de Tórax de Rotina`, residuals(modcomp), xlab="Proporção de Raio-X de Tórax de Rotina")
```

```
abline(h=0)
```



```
plot(data$`Proporção de Culturas de Rotina`, residuals(modcomp), xlab="Proporção de Cu",  
abline(h=0))
```

```
qqnorm(residuals(modcomp), ylab="Resíduos", xlab="Quantis teóricos", main="")  
qqline(residuals(modcomp))
```



-> -> ->



## Referências

## Anexos



## ANEXO A – Amostra