

# Tópicos de estatística aplicada utilizando R\*

Uma abordagem gerando relatórios reprodutíveis usando RStudio, *LaTeX*, RMarkdown e Quarto

Emerson Scheidegger

2023-06-05

Neste documento tento abordar de forma simples como trabalhar com a Regressão Linear Simples utilizando a linguagem R e a suíte do RStudio. Para gerar este documento utilizei o Rstudio + RMarkdown + LaTeX + Quarto.

## Sumário

<b>1</b>	<b>Características do R</b>	<b>3</b>
<b>2</b>	<b>Importação de dados</b>	<b>4</b>
2.1	Importando arquivos .csv . . . . .	4
2.2	Importando arquivos .xls ou .xlsx . . . . .	4
2.3	<b>Distribuições de Frequência</b> . . . . .	4
<b>3</b>	<b>Funções estatísticas</b>	<b>5</b>
3.1	Tabelas . . . . .	5
3.2	Tabelas de proporções . . . . .	5
3.3	Medidas de resumo . . . . .	6
3.4	Summary . . . . .	6
3.4.1	Média . . . . .	6
3.4.2	Mediana . . . . .	6
3.4.3	Variância . . . . .	7
3.4.4	Desvio-padrão . . . . .	7
<b>4</b>	<b>Testes de Hipótese</b>	<b>8</b>
4.1	Testes para a média populacional e para a comparação de duas médias . . . . .	8
4.2	Testes para uma proporção populacional e para comparação de duas proporções . . . . .	8
4.3	Testes para Normalidade . . . . .	8
4.4	Testes para comparação de variâncias . . . . .	8
<b>5</b>	<b>Funções Matemáticas</b>	<b>8</b>

---

\*Agradeço ao R CORE TEAM (2023) e a todos que dedicaram seu tempo no desenvolvimento de funções e pacotes para o ambiente RStudio principalmente a Yihui Xie <https://yihui.org> pela milhares de horas dedicadas a melhorar nossas vidas.

<b>6</b>	<b>Gráficos</b>	<b>8</b>
<b>7</b>	<b>Probabilidade</b>	<b>8</b>
<b>8</b>	<b>Exemplos</b>	<b>8</b>
	<b>Referências</b>	<b>8</b>

## **Lista de Figuras**

# 1 Características do R

- Não foi feito para manipulação de dados em larga escala.
- Forma mais fácil e direta de acessar os dados é convertê-los para texto e importar.
- Salva a sessão em um arquivo .RDATA, que armazena todos os objetos R, possibilitando que um projeto seja retomado posteriormente ou intercambiado com colaboradores.
- Acessa bancos de dados e planilhas Microsoft Excel via ODBC e outros bancos de dados por servidor SQL, ampliando a capacidade de trabalhar com dados em larga escala.
- A partir da versão 2.1.1 possui um editor de script, que facilita a execução de comandos diretamente de dentro do R.
- Possui pacotes com funções específicas que podem ser instalados pela Internet, através do próprio programa.
- Conta com inúmeros colaboradores no mundo inteiro que criam, testam e corrigem as funções que podem ser usadas por qualquer pessoa.
- Gera gráficos em diferentes formatos para as mais diversas utilizações.
- O Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>.

## 2 Importação de dados

### 2.1 Importando arquivos .csv

```
#download.file("https://www.ime.usp.br/~pam/dados.RData", "dados.RData")
#load("dados.RData")

tab2_1<-read.table("tabela2_1.csv", dec=",", sep=";", h=T)

names(tab2_1)
```

```
[1] "N"                "estado_civil"    "grau_instrucao"  "n_filhos"
[5] "salario"          "idade_anos"      "idade_meses"     "reg_procedencia"
```

```
summary(tab2_1$salario)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
4.000	7.553	10.165	11.122	14.060	23.300

### 2.2 Importando arquivos .xls ou .xlsx

### 2.3 Distribuições de Frequência

```
# Calcula a tabela de frequências absolutas e armazena o resultado em 'ni'
ni<-table(tab2_1$grau_instrucao)
fi<-prop.table(ni) # Tabela de frequências relativas (f_i)
p-fi<-100*prop.table(ni) # Porcentagem (100 f_i)

# Adiciona linhas de total
ni<-c(ni,sum(ni))
fi<-c(fi,sum(fi))
p-fi<-c(p-fi,sum(p-fi))
names(ni)[4]<-"Total"
tab2_2<-cbind(ni,fi=round(fi,digits=2),p-fi=round(p-fi,digits=2))
tab2_2
```

	ni	fi	p-fi
ensino fundamental	12	0.33	33.33
ensino médio	18	0.50	50.00
superior	6	0.17	16.67
Total	36	1.00	100.00

## 3 Funções estatísticas

### 3.1 Tabelas

```
#Grau de instrução
table(tab2_1$grau_instrucao)
```

ensino fundamental	ensino médio	superior
12	18	6

```
#Grau de instrução x Estado Civil
table(tab2_1$grau_instrucao, tab2_1$estado_civil)
```

	casado	solteiro
ensino fundamental	5	7
ensino médio	12	6
superior	3	3

```
table(tab2_1$n_filhos, tab2_1$grau_instrucao)
```

	ensino fundamental	ensino médio	superior
0	1	2	1
1	1	4	0
2	2	5	0
3	1	0	2
5	0	1	0

### 3.2 Tabelas de proporções

```
# Grau de instrução x Estado Civil
prop.table(table(tab2_1$grau_instrucao, tab2_1$estado_civil))
```

	casado	solteiro
ensino fundamental	0.13888889	0.19444444
ensino médio	0.33333333	0.16666667
superior	0.08333333	0.08333333

```
# Com duas casas decimais
round(prop.table(table(tab2_1$grau_instrucao, tab2_1$estado_civil)), 2)
```

	casado	solteiro
ensino fundamental	0.14	0.19
ensino médio	0.33	0.17
superior	0.08	0.08

### 3.3 Medidas de resumo

#### 3.4 Summary

Resume a variável quantitativa em: mínimo, máximo, média, mediana, 1º quartil, 3º quartil e dados não preenchidos. Caso a variável seja qualitativa, é informado o número de observações para cada nível.

Medidas de resumo dos salários:

```
summary(tab2_1$salario)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
4.000	7.553	10.165	11.122	14.060	23.300

Resumo da variável salário apenas para casados

```
summary(tab2_1$salario[tab2_1$estado_civil=="casado"])
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
4.560	8.742	11.925	12.123	15.030	23.300

#### ! Importante

Observação: Caso a variável desejada seja qualitativa numérica, é possível que o R interprete-a como sendo uma variável quantitativa. Para evitar que isso aconteça, utilize a função `as.factor()`.  
Ex: `summary(as.factor(dados$sexo))`

##### 3.4.1 Média

```
# Média dos salários  
cat('A média de salário é:', mean(tab2_1$salario), '\n')
```

A média de salário é: 11.12222

```
# Média de idade  
cat('A média de idade é:', mean(tab2_1$idade_anos), '\n')
```

A média de idade é: 34.58333

##### 3.4.2 Mediana

```
# Mediana dos salários
cat('A mediana de salário é:', median(tab2_1$salario), '\n')
```

A mediana de salário é: 10.165

```
# Mediana de idade
cat('A mediana de idade é:', median(tab2_1$idade_anos), '\n')
```

A mediana de idade é: 34.5

### 3.4.3 Variância

```
# Variância dos salários
cat('A variância de salário é:', var(tab2_1$salario), '\n')
```

A variância de salário é: 21.04477

### 3.4.4 Desvio-padrão

**sintaxe:** `sd(variável)`

**opções:**

`na.rm`: TRUE, calcula o desvio padrão considerando apenas os dados existentes, ignora os dados faltantes.

FALSE, calcula o desvio padrão apenas se todos os valores estiverem preenchidos, caso contrário retorna NA.

**Exemplo:**

```
cat('O desvio-padrão dos salários é:', sd(tab2_1$salario), '\n')
```

O desvio-padrão dos salários é: 4.587458

Erro em `var(tab2_1$n_filhos)` : observações faltantes em cov/cor

```
sd(tab2_1$n_filhos)
```

[1] NA

Tratando as observações faltantes

```
sd(tab2_1$n_filhos, na.rm=TRUE)
```

```
[1] 1.268028
```

## 4 Testes de Hipótese

### 4.1 Testes para a média populacional e para a comparação de duas médias

### 4.2 Testes para uma proporção populacional e para comparação de duas proporções

### 4.3 Testes para Normalidade

### 4.4 Testes para comparação de variâncias

## 5 Funções Matemáticas

## 6 Gráficos

## 7 Probabilidade

## 8 Exemplos

## Referências

R CORE TEAM. [R: A Language and Environment for Statistical Computing](#). Vienna, Austria: R Foundation for Statistical Computing, 2023.