

Atividade Teste de hipótese

2023-12-03

Alunos: Arthur Vinícius e Gabriel Henrique

Coletânea de dados

```
#Carregar o arquivo csv dos dados
library(readr)
resultados_1 <- read_csv("resultados_1.csv")

## Rows: 81 Columns: 7
## -- Column specification -----
## Delimiter: ","
## chr (2): rendimento_escolar, serie
## dbl (5): id_aluno, nota, cod_escola, num_sala, ano_letivo
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
#Dados fictícios de resultados de alunos de escolas das 27 capitais brasileiras
#Fonte: Banco de dados elaborado pelo nosso grupo da cadeira de BD
knitr::kable(resultados_1)
```

id_aluno	nota	rendimento_escolar	cod_escola	num_sala	ano_letivo	serie
1	8.5	Aprovado	11000201	101	2023	1A
2	7.2	Aprovado	11000201	102	2023	2B
3	9.0	Aprovado	11000201	103	2023	3C
4	6.8	Reprovado	12009172	201	2023	1A
5	8.2	Aprovado	12009172	202	2023	2B
6	7.5	Aprovado	12009172	203	2023	3C
7	9.5	Aprovado	27035018	301	2023	MA
8	5.5	Reprovado	27035018	302	2023	7B
9	8.7	Aprovado	27035018	303	2023	3C
10	7.0	Aprovado	16000641	401	2023	1A
11	8.9	Aprovado	16000641	402	2023	2B
12	6.4	Reprovado	16000641	403	2023	3C
13	9.2	Aprovado	13027107	501	2023	1A
14	8.0	Aprovado	13027107	502	2023	2B
15	7.8	Aprovado	13027107	503	2023	3C
16	7.7	Aprovado	29190290	601	2023	1A
17	8.6	Aprovado	29190290	602	2023	2B
18	6.3	Reprovado	29190290	603	2023	3C
19	9.7	Aprovado	23065591	701	2023	1A
20	7.4	Aprovado	23065591	702	2023	2B
21	8.1	Aprovado	23065591	703	2023	3C
22	6.9	Reprovado	53002458	801	2023	1A

id_aluno	nota	rendimento_escolar	cod_escola	num_sala	ano_letivo	serie
23	8.3	Aprovado	53002458	802	2023	2B
24	7.1	Aprovado	53002458	803	2023	3C
25	8.8	Aprovado	32040121	901	2023	1A
26	6.2	Reprovado	32040121	902	2023	2B
79	9.3	Aprovado	17046807	2701	2023	1A
80	6.8	Reprovado	17046807	2702	2023	2B
81	8.2	Aprovado	17046807	2703	2023	3C
27	9.4	Aprovado	32040121	903	2023	3C
28	7.3	Aprovado	52034682	1001	2023	1A
29	8.4	Aprovado	52034682	1002	2023	2B
30	6.5	Reprovado	52034682	1003	2023	3C
31	9.1	Aprovado	21015473	1101	2023	1A
32	7.6	Aprovado	21015473	1102	2023	2B
33	8.6	Aprovado	21015473	1103	2023	3C
34	6.7	Reprovado	51056500	1201	2023	MA
35	8.5	Aprovado	51056500	1202	2023	6B
36	7.0	Aprovado	51056500	1203	2023	9C
37	9.0	Aprovado	50031198	1301	2023	MA
38	7.8	Aprovado	50031198	1302	2023	8B
39	8.3	Aprovado	50031198	1303	2023	1C
40	7.2	Aprovado	31000019	1401	2023	1A
41	8.9	Aprovado	31000019	1402	2023	2B
42	6.6	Reprovado	31000019	1403	2023	3C
43	9.5	Aprovado	15588947	1501	2023	1A
44	7.1	Aprovado	15588947	1502	2023	2B
45	8.0	Aprovado	15588947	1503	2023	3C
46	6.8	Reprovado	25124528	1601	2023	1A
47	8.7	Aprovado	25124528	1602	2023	2B
48	7.4	Aprovado	25124528	1603	2023	3C
49	9.2	Aprovado	41132173	1701	2023	MA
50	6.3	Reprovado	41132173	1702	2023	9B
51	8.1	Aprovado	41132173	1703	2023	3C
52	7.7	Aprovado	26139618	1801	2023	9A
53	8.4	Aprovado	26139618	1802	2023	2B
54	6.9	Reprovado	26139618	1803	2023	3C
55	8.6	Aprovado	22022236	1901	2023	1A
56	6.5	Reprovado	22022236	1902	2023	1B
57	9.3	Aprovado	22022236	1903	2023	2A
58	7.8	Aprovado	33230021	2001	2023	MA
59	8.2	Aprovado	33230021	2002	2023	MB
60	6.7	Reprovado	33230021	2003	2023	MC
61	9.0	Aprovado	24058769	2101	2023	4A
62	7.4	Aprovado	24058769	2102	2023	5B
63	8.5	Aprovado	24058769	2103	2023	7C
64	6.9	Reprovado	43209262	2201	2023	MA
65	8.7	Aprovado	43209262	2202	2023	MB
66	7.2	Aprovado	43209262	2203	2023	MC
67	9.1	Aprovado	14000814	2301	2023	5A
68	7.6	Aprovado	14000814	2302	2023	6B
69	8.4	Aprovado	14000814	2303	2023	7C
70	6.7	Reprovado	42141184	2401	2023	MA
71	8.8	Aprovado	42141184	2402	2023	2B

id_aluno	nota	rendimento_escolar	cod_escola	num_sala	ano_letivo	serie
72	7.1	Aprovado	42141184	2403	2023	3C
73	9.4	Aprovado	35107273	2501	2023	1A
74	6.6	Reprovado	35107273	2502	2023	2B
75	8.3	Aprovado	35107273	2503	2023	3C
76	7.9	Aprovado	28034570	2601	2023	1A
77	8.1	Aprovado	28034570	2602	2023	2B
78	6.5	Reprovado	28034570	2603	2023	3C

Medidas resumo

```
library(psych)
notas <- resultados_1$nota
medidas <- describe(notas)
print(medidas)
```

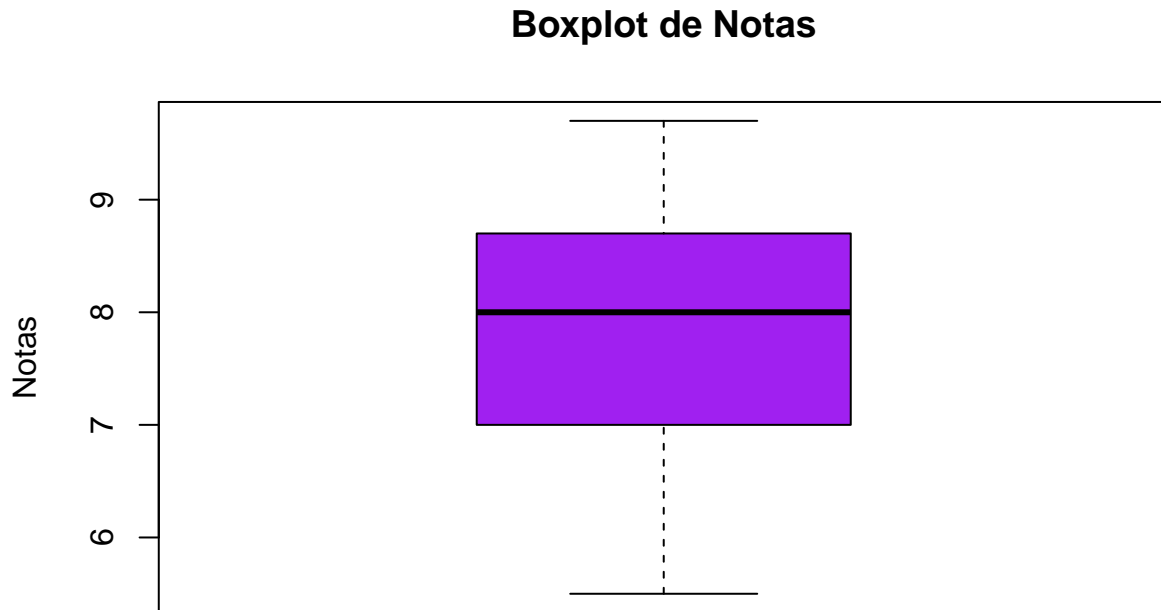
```
##      vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1      1 81 7.88 0.99      8    7.88 1.19 5.5 9.7   4.2 -0.11   -1.05 0.11
```

Gráficos

```
hist(notas, col='lightgreen', main='Histograma de notas')
```



```
boxplot(notas, main="Boxplot de Notas", ylab="Notas", col="purple", border="black")
```



```
categorias <- cut(notas, breaks = c(0, 5, 7, 9, 10), labels = c("0-5", "6-7", "8-9", "10"))
```

```
contagem <- table(categorias)
```

```
pie(contagem, main="Gráfico de Pizza das Notas", col=rainbow(length(contagem)), labels=contagem, cex=0.8)
```

Gráfico de Pizza das Notas

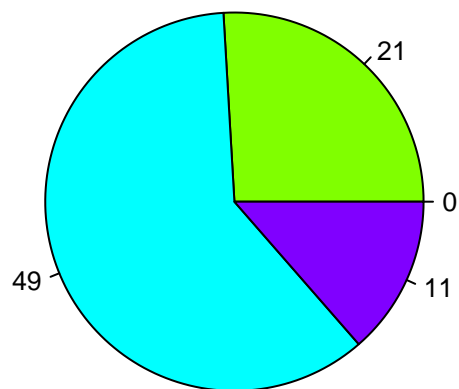


Gráfico de densidade

```
media = mean(notas)
mediana = median(notas)
f = density(notas)
moda = f$x[which(f$y == max(f$y))]

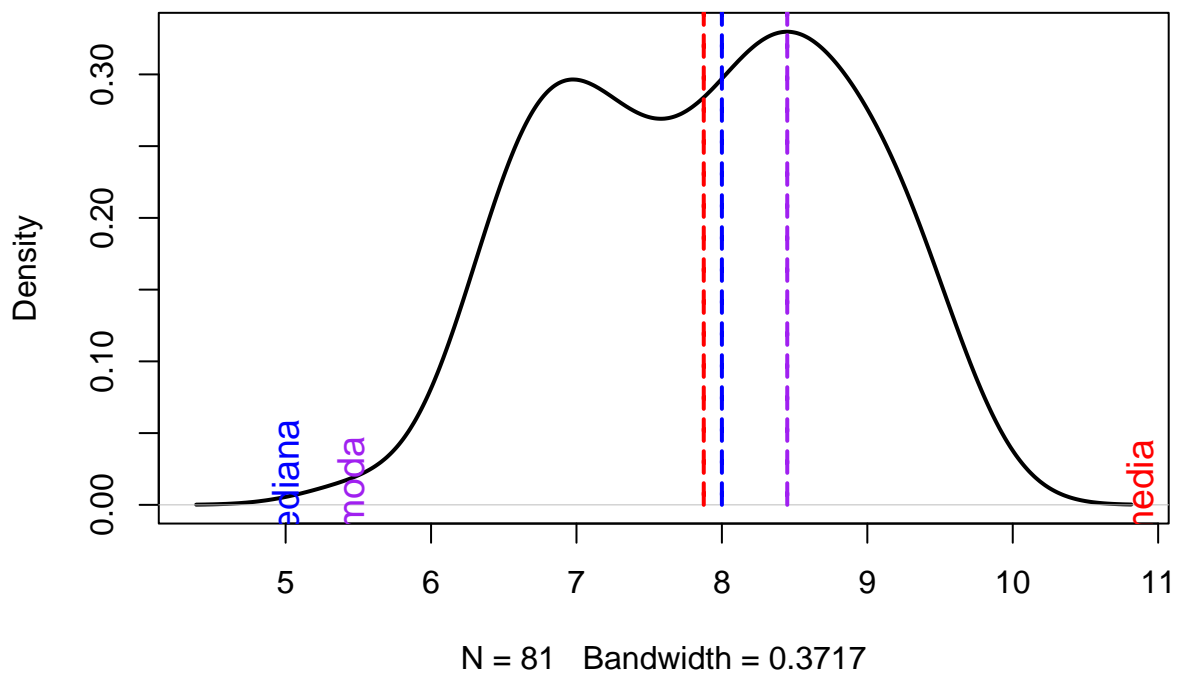
plot(density(notas), main='Densidade alisada de x', lwd=2)
```

```
xx <- c(media, media)
yy <- c(0, 1)
polygon(xx, yy, col = "white", border = "red", lty=2, lwd=2)
text(media+3, 0.01, 'media', col='red', srt=90, cex = 1.2)

xx <- c(mediana, mediana)
yy <- c(0, 1)
polygon(xx, yy, col = "white", border = "blue", lty=2, lwd=2)
text(mediana-3, 0.01, 'mediana', col='blue', srt=90, cex = 1.2)

xx <- c(moda, moda)
yy <- c(0, 1)
polygon(xx, yy, col = "white", border = "purple", lty=2, lwd=2)
text(moda-3, 0.015, 'moda', col='purple', srt=90, cex = 1.2)
```

Densidade alisada de x



Inferência em questão

Tendo em vista os dados referenciados anteriormente, é plausível dizer que a maioria dos alunos obtiveram nota maior que 7.3?

Hipótese Nula (H_0)

H_0 : A média de notas dos alunos é igual ou menor que 7.3.

Hipótese Alternativa (H_1)

H_1 : A média de notas dos alunos é maior que 7.3.

Teste de Hipótese

```
teste1 <- t.test(notas, mu = 7.3, alternative = "less")
print(teste1)
```

```
##
## One Sample t-test
##
## data:  notas
## t = 5.2055, df = 80, p-value = 1
## alternative hypothesis: true mean is less than 7.3
## 95 percent confidence interval:
##      -Inf 8.059227
## sample estimates:
## mean of x
##  7.875309
```

Traduzindo os resultados:

t-value (valor-t): O valor t é 5.2055. Este valor mede o quão longe a média amostral está da média assumida na hipótese nula. Quanto maior o valor t, maior a evidência contra a hipótese nula.

Graus de liberdade (df): Os graus de liberdade são 80. Este é um valor importante para calcular o valor-p.

p-value: O valor-p é 1. Isso significa que o valor-p é maior que o nível de significância comum de 0.05. Como resultado, não há evidências suficientes para rejeitar a hipótese nula.

Hipótese Alternativa: A hipótese alternativa é que a verdadeira média é menor que 7.3.

Intervalo de Confiança: O intervalo de confiança de 95% para a média verdadeira vai de -Infinito até 8.059227. O lado esquerdo do intervalo é -Infinito, o que indica que a parte inferior do intervalo não é limitada.

Estimativas da Amostra: A média das amostras é 7.875309.

Conclusão

O valor-p é alto (1), indicando que não há evidências suficientes para rejeitar a hipótese nula. Portanto, com base nos dados fornecidos, não podemos afirmar com confiança que a média das notas é menor que 7.3. O intervalo de confiança inclui valores acima de 7.3, o que reforça a falta de evidência para rejeitar a hipótese nula. A média amostral é 7.875309, o que está acima de 7.3, mas sem evidências suficientes para considerar uma diferença significativa.

Observação: Como temos 81 notas ou seja, $n > 30$, não será necessário realizar o shapiro.test