

UNIVERSIDADE DA REGIÃO DE JOINVILLE - UNIVILLE

Bacharelado em Engenharia de Software (BES)

Estatística para computação

Professora Priscila Ferraz Franczak

Alunos: David Bryan, Felipe Cristian, Robertha Rezende, Victor Araújo

Plano de apresentação

1. Introdução

- Apresentação do problema: Fatores que influenciam a performance estudantil.
- Contexto do projeto.

2. data.frame

- Construção de um data.frame com 1000 observações e 10 variáveis.
- Análise estatística da relação entre o nível de escolaridade dos pais e o desempenho dos filhos.

3. Metodologia

- Descrição das variáveis analisadas.
- Visualização de dados: Boxplots.
- Aplicação de testes estatísticos (ANOVA e Teste de Tukey).

4. Resultados e Discussões

- Interpretação dos gráficos.
- Impactos observados:
 - Matemática.
 - Leitura.
 - Escrita.

5. Conclusões

- Resumo das principais descobertas.
- Relação entre níveis educacionais dos pais e desempenho dos filhos.

Problema

Fatores que influenciam a performance estudantil

Nível de Escolaridade dos Pais

- Os alunos cujos pais têm níveis mais altos de escolaridade (como **bacharelado** ou **mestrado**) tendem a apresentar **médias de desempenho superiores** . Isso sugere que o nível educacional dos pais pode impactar positivamente o suporte acadêmico e a valorização da educação no ambiente familiar.

data.frame do excel

data.frame com todas as informações contidas na tabela excel

O data.frame contém:

- **1000 observações** (linhas).
- **10 variáveis** (colunas).

```
> Excel_estatistica$average_score <- as.numeric(gsub("\\.", "", Excel_estatistica$average_score))
> str(Excel_estatistica)
'data.frame': 1000 obs. of 10 variables:
 $ gender          : int  0 0 0 1 1 0 0 1 1 0 ...
 $ race_ethnicity  : chr   "group B" "group C" "group B" "group A" ...
 $ parental_level_of_education: chr   "bachelor's degree" "some college" "master's degree" "associate's degree" ...
 $ lunch           : int  1 1 1 0 1 1 1 0 0 0 ...
 $ test_preparation_course    : int  0 1 0 0 0 0 1 0 1 0 ...
 $ math_score       : int  72 69 90 47 76 71 88 40 64 38 ...
 $ reading_score    : int  72 90 95 57 78 83 95 43 64 60 ...
 $ writing_score     : int  74 88 93 44 75 78 92 39 67 50 ...
 $ total_score      : int  218 247 278 148 229 232 275 122 195 148 ...
 $ average_score    : num  7.27e+15 8.23e+15 9.27e+15 4.93e+16 7.63e+15 ...
> |
```

Summary

```
> print('Influência do nível de escolaridade dos pais nas notas')
[1] "Influência do nível de escolaridade dos pais nas notas"
> summary(Excel_estatistica$parental_level_of_education)
  Length      Class      Mode 
  1000 character character
> print('Verificar as colunas de notas')
[1] "Verificar as colunas de notas"
> summary(Excel_estatistica[, c("math_score", "reading_score", "writing_score")])
  math_score  reading_score  writing_score
Min.   : 0.00   Min.   : 17.00   Min.   : 10.00
1st Qu.: 57.00   1st Qu.: 59.00   1st Qu.: 57.75
Median : 66.00   Median : 70.00   Median : 69.00
Mean   : 66.09   Mean   : 69.17   Mean   : 68.05
3rd Qu.: 77.00   3rd Qu.: 79.00   3rd Qu.: 79.00
Max.   :100.00   Max.   :100.00   Max.   :100.00
> |
```

- **Resultado** : Mostra que a variável `parental_level_of_education` é do tipo **caractere** e tem 1000 valores.
- **Bachelor's degree** : Bacharelado
- **Some college** : Faculdade Incompleta
- **Master's degree** : Mestrado
- **Associate's degree** : Tecnólogo
- **High school** : Ensino Médio
- **Some high school** : Ensino Médio Incompleto
- Isso indica que ela contém informações categóricas, como o nível de educação dos pais (ex.: “bacharelado”, “alguma faculdade”).

Esse resultado mostra as médias das notas em matemática, leitura e escrita para diferentes níveis de escolaridade dos pais.

```
> print('Calcular a média de cada nota por nível de escolaridade')
[1] "Calcular a média de cada nota por nível de escolaridade"
> media_math <- aggregate(math_score ~ parental_level_of_education, data = Excel_estatistica, mean)
> media_reading <- aggregate(reading_score ~ parental_level_of_education, data = Excel_estatistica, mean)
> media_writing <- aggregate(writing_score ~ parental_level_of_education, data = Excel_estatistica, mean)
> print('Exibir as médias')
[1] "Exibir as médias"
```

```
> print(media_math)
parental_level_of_education math_score
1      associate's degree    67.88288
2      bachelor's degree    69.38983
3      high school          62.13776
4      master's degree      69.74576
5      some college         67.12832
6      some high school     63.49721
> print(media_reading)
parental_level_of_education reading_score
1      associate's degree    70.92793
2      bachelor's degree    73.00000
3      high school          64.70408
4      master's degree      75.37288
5      some college         69.46018
6      some high school     66.93855
> print(media_writing)
parental_level_of_education writing_score
1      associate's degree    69.89640
2      bachelor's degree    73.38136
3      high school          62.44898
4      master's degree      75.67797
5      some college         68.84071
6      some high school     64.88827
```

Há uma relação evidente entre o nível de escolaridade dos pais e o desempenho acadêmico dos filhos em todas as áreas avaliadas (matemática, leitura e escrita).

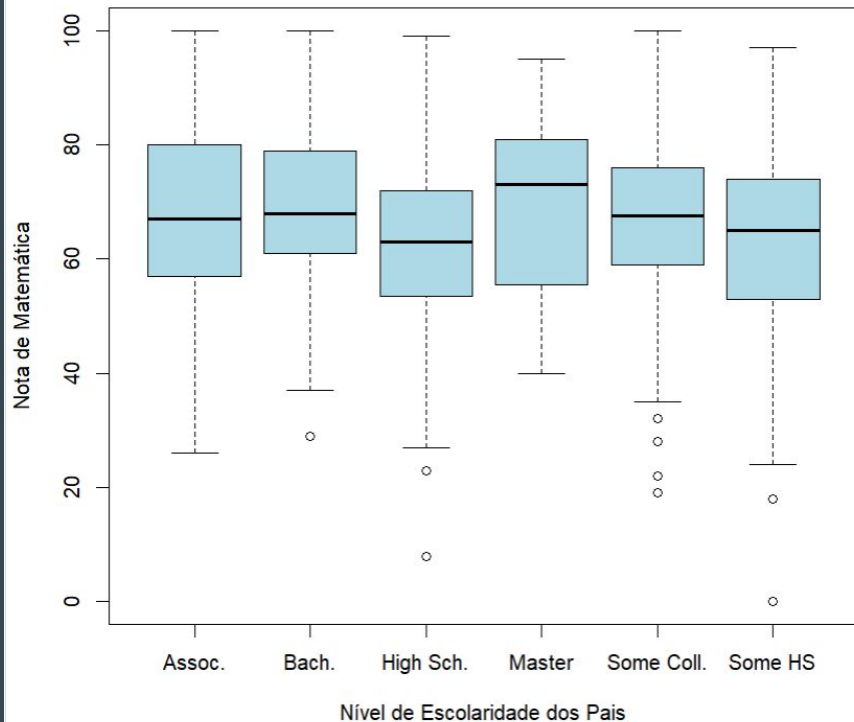
Filhos de pais com níveis mais altos de escolaridade, como bacharelado e mestrado, apresentam desempenho significativamente superior. Em contrapartida, filhos de pais com níveis mais baixos de escolaridade, como ensino médio completo ou incompleto, tendem a obter resultados inferiores.

Box plot das disciplinas

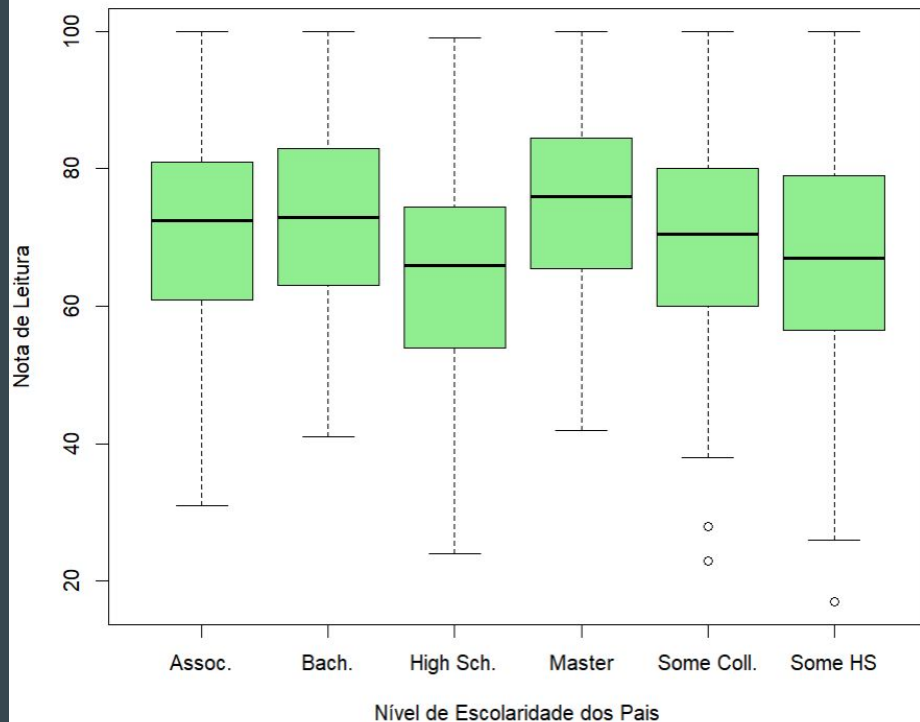
```
> print('Boxplot para Matematica')
[1] "Boxplot para Matemática"
> boxplot(math_score ~ parental_level_of_education,
+         data = Excel_estatistica,
+         main = "Nível de Escolaridade dos Pais x Nota de Matemática",
+         xlab = "Nível de Escolaridade dos Pais",
+         ylab = "Nota de Matemática",
+         col = "lightblue",
+         names = c("Assoc.", "Bach.", "High Sch.", "Master","Some Coll.", "Some HS"))
> print('Boxplot para Leitura')
[1] "Boxplot para Leitura"
> boxplot(reading_score ~ parental_level_of_education, data = Excel_estatistica,
+         main = "Nível de Escolaridade dos Pais x Nota de Leitura",
+         xlab = "Nível de Escolaridade dos Pais", ylab = "Nota de Leitura",
+         col = "lightgreen",
+         names = c("Assoc.", "Bach.", "High Sch.", "Master","Some Coll.", "Some HS"))
> print('Boxplot para Escrita')
[1] "Boxplot para Escrita"
> boxplot(writing_score ~ parental_level_of_education, data = Excel_estatistica,
+         main = "Nível de Escolaridade dos Pais x Nota de Escrita",
+         xlab = "Nível de Escolaridade dos Pais", ylab = "Nota de Escrita",
+         col = "lightpink",
+         names = c("Assoc.", "Bach.", "High Sch.", "Master","Some Coll.", "Some HS"))
> |
```

Interpretação dos gráficos

Nível de Escolaridade dos Pais x Nota de Matemática

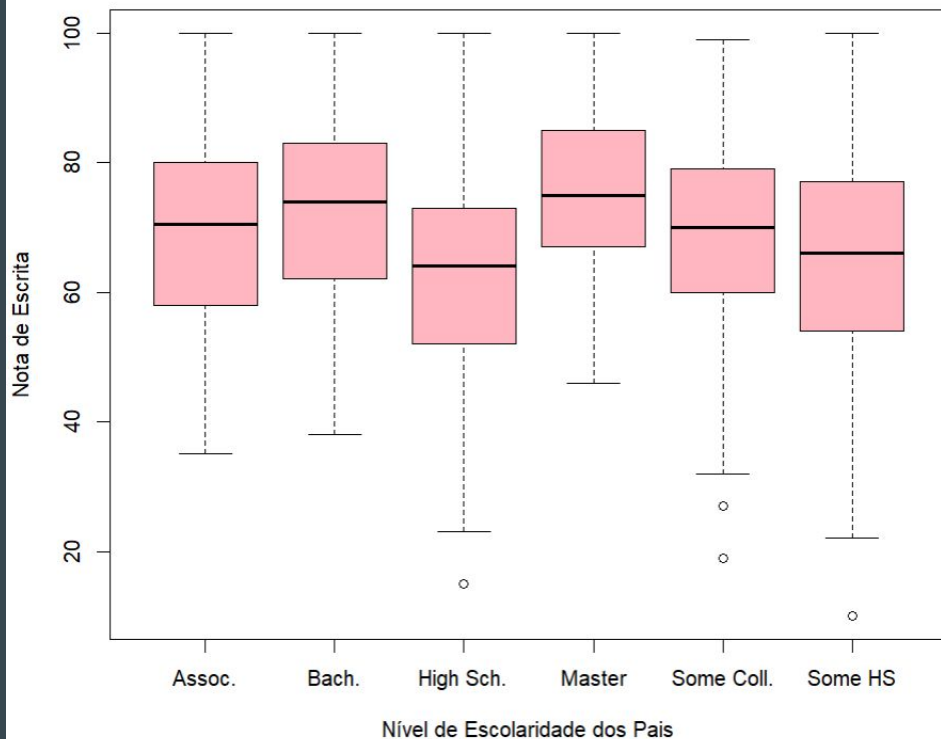


Nível de Escolaridade dos Pais x Nota de Leitura



Interpretação dos gráficos

Nível de Escolaridade dos Pais x Nota de Escrita



É provável que pais com maior escolaridade (como “Master” e “Bach.”) tenham filhos com desempenho mais elevado, enquanto os pais com menor escolaridade (“High Sch.” e “Some HS”) tendem a apresentar filhos com desempenho inferior.

Resultados do anova

```
> anova_math <- aov(math_score ~ parental_level_of_education, data = Excel_estatistica)
> anova_reading <- aov(reading_score ~ parental_level_of_education, data = Excel_estatistica)
> anova_writing <- aov(writing_score ~ parental_level_of_education, data = Excel_estatistica)
> print('Resumo dos resultados')
[1] "Resumo dos resultados"
> print('Matemática')
[1] "Matemática"
> summary(anova_math)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
parental_level_of_education	5	7296	1459.1	6.522	<u>5.59e-06</u> ***
Residuals	994	222394	223.7		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> print('Leitura')
[1] "Leitura"
> summary(anova_reading)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
parental_level_of_education	5	9506	1901.3	9.289	<u>1.17e-08</u> ***
Residuals	994	203446	204.7		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> print('Escrita')
[1] "Escrita"
> summary(anova_writing)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
parental_level_of_education	5	15623	3124.6	14.44	<u>1.12e-13</u> ***
Residuals	994	215054	216.4		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

Os resultados da ANOVA mostram que o nível de escolaridade dos pais influencia significativamente as notas dos filhos em matemática, leitura e escrita.

- **Matemática** : Diferença significativa entre os grupos, mas menor impacto em comparação com outras áreas.
- **Leitura** : Impacto mais forte do nível de escolaridade, com diferenças bem marcadas.
- **Escrita** : A habilidade mais influenciada pela escolaridade dos pais, com o maior impacto observado.

Teste Tukey matemática

```
> print('teste tukey para Matematica')  
[1] "Teste Tukey para Matemática"  
> tukey_math <- TukeyHSD(anova_math)  
> print(tukey_math)  
Tukey multiple comparisons of means  
95% family-wise confidence level
```

Fit: aov(formula = math_score ~ parental_level_of_education, data = Excel_estatistica)

\$parental_level_of_education

	diff	lwr	upr	p adj
bachelor's degree-associate's degree	1.5069476	-3.358681	6.37257671	0.9502834
high school-associate's degree	-5.7451278	-9.931142	-1.55911404	<u>0.0013308</u>
master's degree-associate's degree	1.8628798	-4.392702	8.11846150	0.9578456
some college-associate's degree	-0.7545643	-4.790324	3.28119570	0.9947937
some high school-associate's degree	-4.3856762	-8.675962	-0.09539047	<u>0.0417546</u>
high school-bachelor's degree	-7.2520754	-12.228447	-2.27570365	<u>0.0004918</u>
master's degree-bachelor's degree	0.3559322	-6.453904	7.16576824	0.9999897
some college-bachelor's degree	-2.2615119	-7.112174	2.58915025	0.7676188
some high school-bachelor's degree	-5.8926238	-10.957021	-0.82822687	0.0118857
master's degree-high school	7.6080076	1.265908	13.95010753	0.0083719
some college-high school	4.9905635	0.821956	9.15917095	0.0085748
some high school-high school	1.3594516	-3.056030	5.77493356	0.9514996
some college-master's degree	-2.6174441	-8.861392	3.62650328	0.8384321
some high school-master's degree	-6.2485560	-12.659958	0.16284570	0.0610615
some high school-some college	-3.6311119	-7.904416	0.64219230	0.1481784

Teste Tukey de leitura

```
> print('Teste Tukey para Leitura')  
[1] "Teste Tukey para Leitura"  
> tukey_reading <- TukeyHSD(anova_reading)  
> print(tukey_reading)  
  Tukey multiple comparisons of means  
    95% family-wise confidence level
```

```
Fit: aov(formula = reading_score ~ parental_level_of_education, data = Excel_estatistica)
```

```
$parental_level_of_education
```

	diff	lwr	upr	p adj
bachelor's degree-associate's degree	2.072072	-2.5816716	6.72581573	0.8006901
high school-associate's degree	-6.223846	-10.2275701	-2.22012251	<u>0.0001463</u>
master's degree-associate's degree	4.444953	-1.5382140	10.42812087	0.2770595
some college-associate's degree	-1.467751	-5.3277641	2.39226226	0.8871557
some high school-associate's degree	-3.989380	-8.0928354	0.11407454	0.0622049
high school-bachelor's degree	-8.295918	-13.0555821	-3.53625460	0.0000113
master's degree-bachelor's degree	2.372881	-4.1404041	8.88616683	0.9042540
some college-bachelor's degree	-3.539823	-8.1792515	1.09960551	0.2486973
some high school-bachelor's degree	-6.061453	-10.9053082	-1.21759683	<u>0.0049602</u>
master's degree-high school	10.668800	4.6028817	16.73471777	<u>0.0000090</u>
some college-high school	4.756095	0.7690199	8.74317086	0.0089453
some high school-high school	2.234466	-1.9887334	6.45766511	0.6574517
some college-master's degree	-5.912704	-11.8847442	0.05933545	0.0541073
some high school-master's degree	-8.434334	-14.5665358	-2.30213195	0.0012867
some high school-some college	-2.521630	-6.6088425	1.56558345	0.4912798

Teste Tukey escrita

```
> print('Teste Tukey para Escrita')  
[1] "Teste Tukey para Escrita"  
> tukey_writing <- TukeyHSD(anova_writing)  
> print(tukey_writing)  
  Tukey multiple comparisons of means  
    95% family-wise confidence level
```

Fit: aov(formula = writing_score ~ parental_level_of_education, data = Excel_estatistica)

\$parental_level_of_education

	diff	lwr	upr	p adj
bachelor's degree-associate's degree	3.484960	-1.2997057	8.2696248	0.2987656
high school-associate's degree	-7.447417	-11.5637755	-3.3310582	<u>0.0000043</u>
master's degree-associate's degree	5.781570	-0.3699194	11.9330588	<u>0.0794141</u>
some college-associate's degree	-1.055688	-5.0242936	2.9129167	0.9740854
some high school-associate's degree	-5.008128	-9.2270238	-0.7892327	0.0094688
high school-bachelor's degree	-10.932376	-15.8259415	-6.0388112	0.0000000
master's degree-bachelor's degree	2.296610	-4.3999105	8.9931308	0.9245528
some college-bachelor's degree	-4.540648	-9.3105953	0.2292994	0.0725881
some high school-bachelor's degree	-8.493088	-13.4732134	-3.5129622	<u>0.0000192</u>
master's degree-high school	13.228987	6.9924189	19.4655542	<u>0.0000000</u>
some college-high school	6.391728	2.2924864	10.4909704	<u>0.0001376</u>
some high school-high school	2.439289	-1.9027200	6.7812971	0.5960379
some college-master's degree	-6.837258	-12.9773065	-0.6972098	0.0189417
some high school-master's degree	-10.789698	-17.0944142	-4.4849817	0.0000177
some high school-some college	-3.952440	-8.1546364	0.2497568	0.0790042

Conclusão

- Pais com **maior escolaridade** (mestrado ou bacharelado) têm filhos com desempenho **significativamente superior** em todas as áreas avaliadas.
- O maior impacto foi apresentado em **escrita** , com mais expressivas.
- Comparações entre níveis educacionais próximos (ex.: **Licenciatura vs. Tecnólogo**) não observaram diferenças estatísticas, mas as maiores influências estão entre extremos (ex.: **Ensino médio vs. Mestrado**).
- **Influência Social (financeiro)**

Referências

Montgomery, D. C. (2020). *Design and Analysis of Experiments* (10^a ed.). Wiley.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd.

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Disponível em: <https://www.r-project.org/>

Wickham, H., & Grolemund, G. (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media.