

Análise Exploratória SAEB 2019

Tailine J. S. Nonato

2024-06-24

Análise Exploratória SAEB 2019

Carregamento de pacotes e dados

Filtragem dos dados para região Centro-Oeste

Leitura dos dados

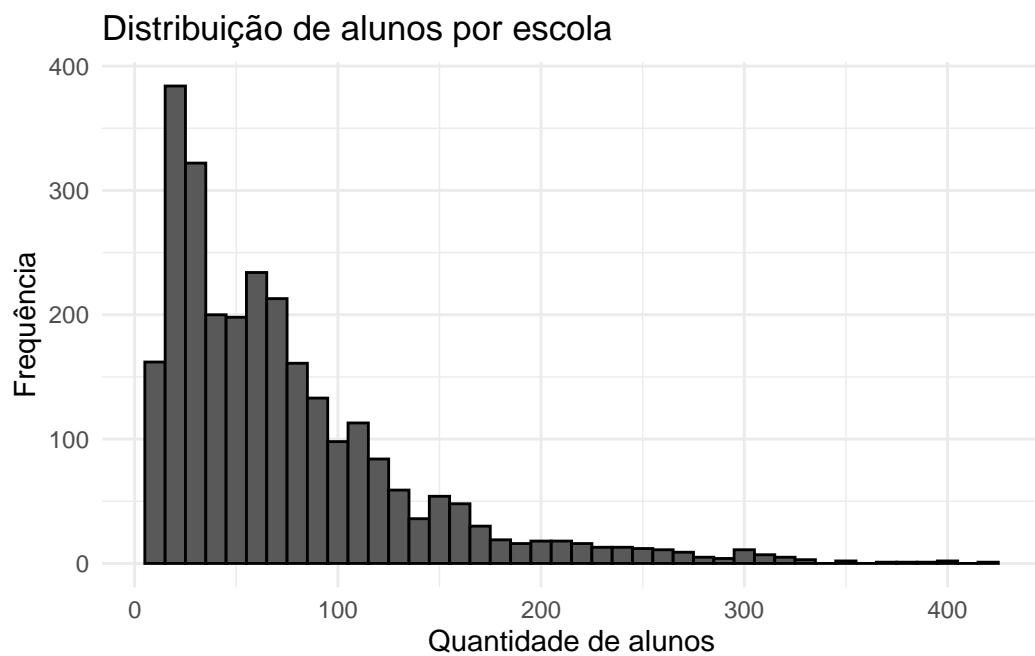
```
[1] 203694    261
```

Análise descritiva dos dados

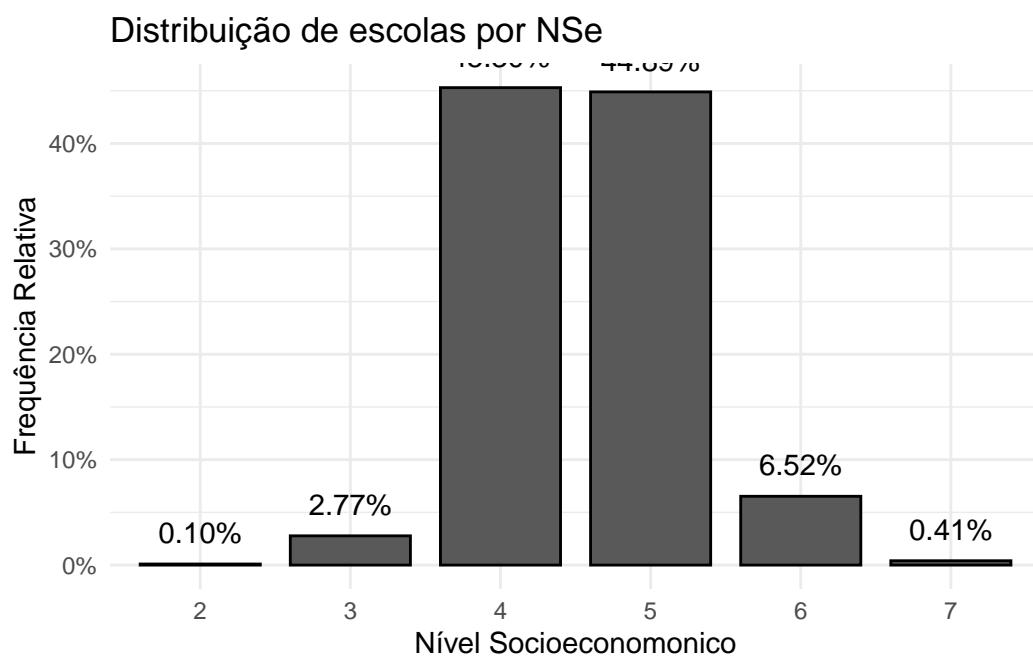
Quantidade de Escolas

```
# A tibble: 1 x 1
  n
  <int>
1 2717
```

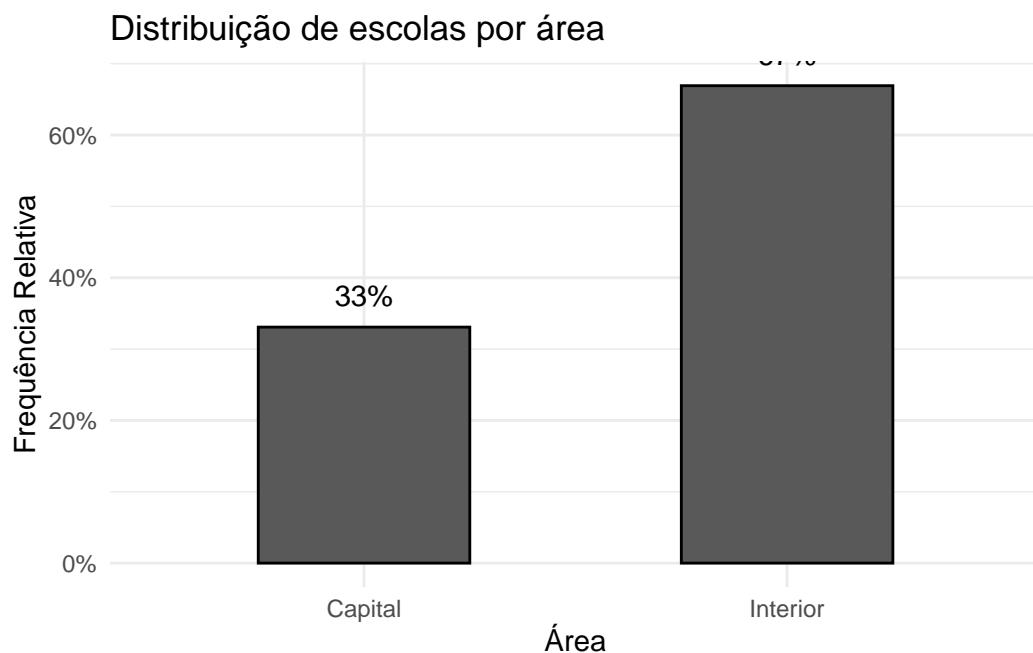
Quantidade de Alunos por Escola



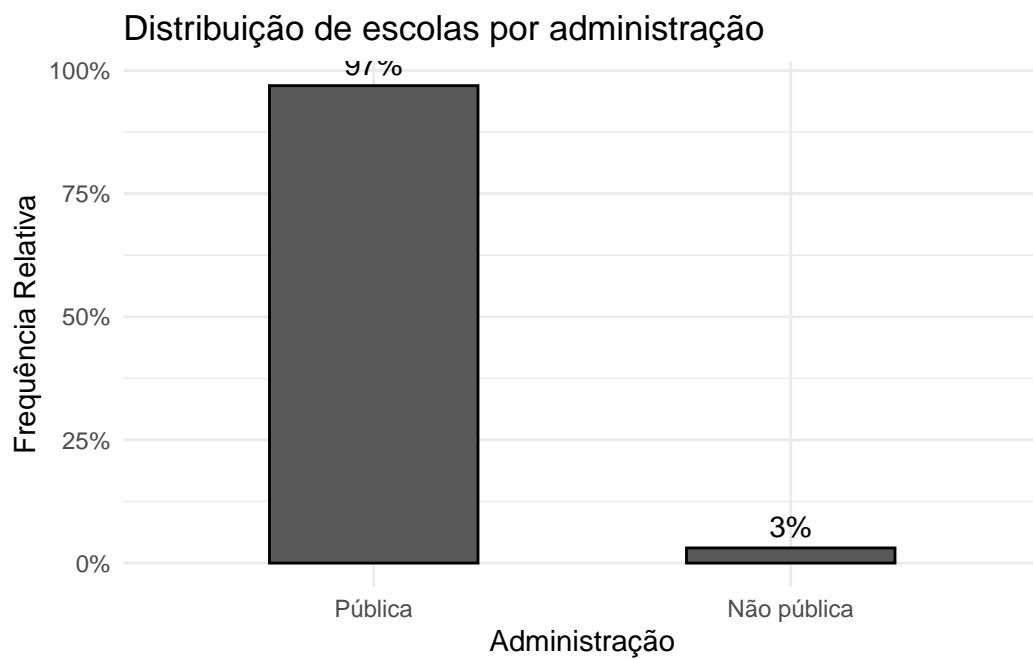
Nível Socioeconômico (NSe) das Escolas



Área das Escolas

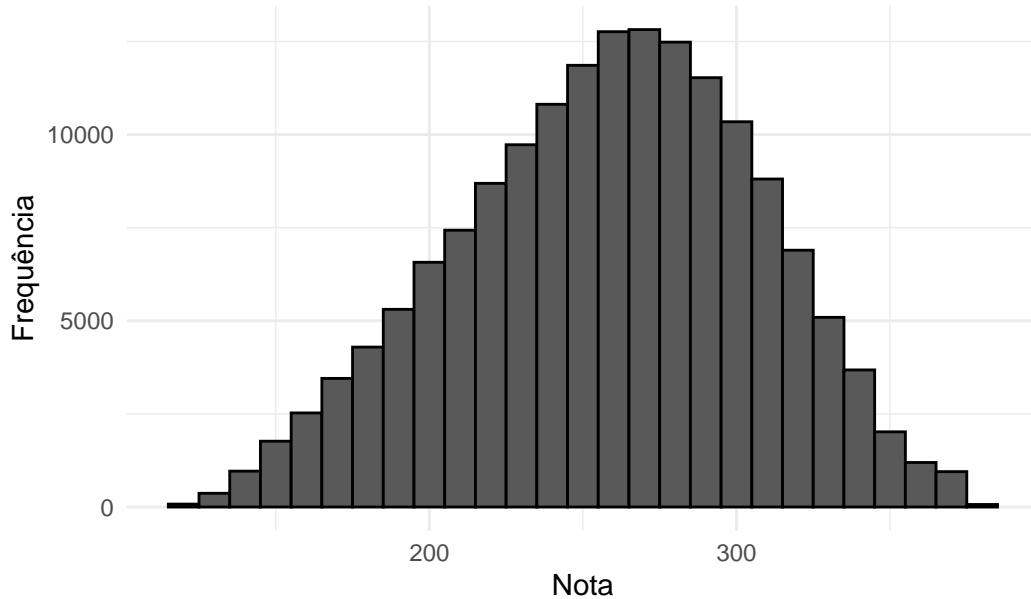


Administração das Escolas (Pública)

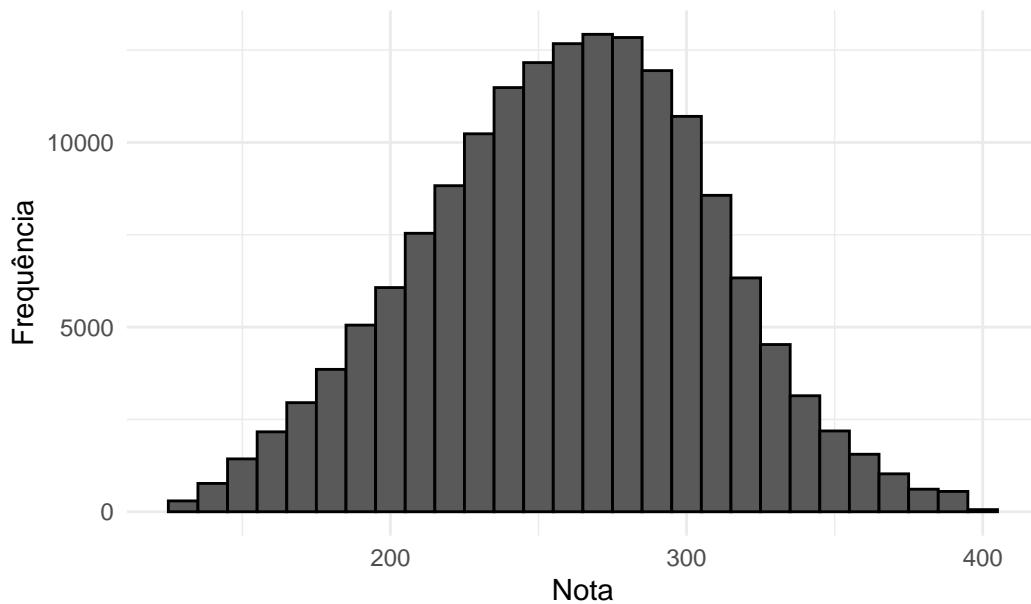


Distribuição da proficiência em Língua Portuguesa e Matemática

Distribuição de notas de LP



Distribuição de notas de MT



Quadro Resumo da proficiência em Língua Portuguesa e Matemática

Table 1: Summary Statistics for Língua Portuguesa and Matemática

Estatística	Língua Portuguesa	Matemática
Média	257,98	259,90
Desvio Padrão	48,84	48,53
Mínimo	118,39	126,33
Primeiro Quartil	224,10	226,77
Mediana	260,84	261,59
Terceiro Quartil	293,60	293,85
Máximo	376,52	397,49

```
\begin{table}
```

```
\caption{\label{tab:unnamed-chunk-10}Summary Statistics for Língua Portuguesa and Matemática}
\centering
\begin{tabular}{t|lll}
\toprule
Estatística & Língua Portuguesa & Matemática\\
\midrule
Média & 257,98 & 259,90\\
Desvio Padrão & 48,84 & 48,53\\
Mínimo & 118,39 & 126,33\\
Primeiro Quartil & 224,10 & 226,77\\
Mediana & 260,84 & 261,59\\
\addlinespace
Terceiro Quartil & 293,60 & 293,85\\
Máximo & 376,52 & 397,49\\
\bottomrule
\end{tabular}
\end{table}
```

Teste de normalidade

```
Shapiro-Wilk normality test

data: sample_LP
W = 0.99082, p-value = 2.602e-15
```

Shapiro-Wilk normality test

```
data: sample_MT  
W = 0.99768, p-value = 1.105e-05
```

Correlação entre as variáveis

Modelagem Multinível

Modelo nulo

```
Linear mixed model fit by REML ['lmerMod']  
Formula: PROFICIENCIA_LP_SAEB ~ 1 + (1 | ID_ESCOLA)  
Data: df  
  
REML criterion at convergence: 1707692  
  
Scaled residuals:  
    Min     1Q   Median     3Q     Max  
-3.5874 -0.6716  0.0499  0.7084  3.1703  
  
Random effects:  
Groups      Name        Variance Std.Dev.  
ID_ESCOLA (Intercept) 348.9    18.68  
Residual            2070.0   45.50  
Number of obs: 162498, groups: ID_ESCOLA, 2681  
  
Fixed effects:  
             Estimate Std. Error t value  
(Intercept) 256.29     0.39   657.2
```

Modelo com variáveis preditoras

```
Linear mixed model fit by REML ['lmerMod']  
Formula: PROFICIENCIA_LP_SAEB ~ NIVEL_SOCIO_ECONOMICO + ID_AREA + IN_PUBLICA +  
        (1 | ID_ESCOLA)  
Data: df  
  
REML criterion at convergence: 1656524
```

Scaled residuals:

	Min	1Q	Median	3Q	Max
	-3.5898	-0.6753	0.0488	0.7098	3.1317

Random effects:

Groups	Name	Variance	Std.Dev.
ID_ESCOLA	(Intercept)	190.7	13.81
	Residual	2076.6	45.57

Number of obs: 157711, groups: ID_ESCOLA, 2500

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	178.2591	3.0018	59.384
NIVEL_SOCIO_ECONOMICO	14.6258	0.4934	29.646
ID_AREA	6.5541	0.7989	8.204

Correlation of Fixed Effects:

(Intr)	NIVEL_
NIVEL_SOCIO	-0.890
ID_AREA	-0.722 0.343

fit warnings:

fixed-effect model matrix is rank deficient so dropping 1 column / coefficient

Análise de resíduos

