



# ANÁLISE DESCRITIVA DE DADOS - RESULTADOS DO SAEB (ANEB - PROVA BRASIL) - 2017

Bruno Gondim Toledo

15 julho, 2022

## **1. Comente sobre a coleta de dados do SAEB: população alvo, população acessível, características levantadas, método de coleta (no máximo 10 linhas).**

O SAEB é realizado de dois em dois anos por amostragem de alunos, através da Avaliação Nacional da Educação Básica (Aneb) (pesquisa amostral aplicada para os alunos do 3º ao 5º ano do Ensino Fundamental e Ensino Médio, da rede pública e particular) e a Avaliação Nacional do Rendimento Escolar (Anresc) (pesquisa censitária aplicada apenas na rede pública, em alunos do 5º e 9º ano do Ensino Fundamental.) produz informações a respeito da realidade educacional por regiões brasileiras

Fonte: Wikipedia

## **2. Liste as variáveis disponibilizadas na amostra que você recebeu (com seu número de matrícula) e classifique-as segundo o tipo.**

ANO: Variável qualitativa nominal

REGIAO: Variável qualitativa nominal

UF: Variável qualitativa nominal

MUNICIPIO: Variável qualitativa nominal

AREA: Variável qualitativa nominal

LOCALIZACAO: Variável qualitativa nominal

DEPENDENCIA\_ADM : Variável qualitativa nominal

NOTA\_LP: Variável quantitativa discreta

NOTA\_MT: Variável quantitativa discreta

SEXO: Variável qualitativa nominal

RACA\_COR: Variável qualitativa nominal

Ano de nascimento: Variável qualitativa ordinal

COMPUTADOR: Variável qualitativa ordinal

ESC\_MAE: Variável qualitativa ordinal

ESC\_PAI: Variável qualitativa ordinal

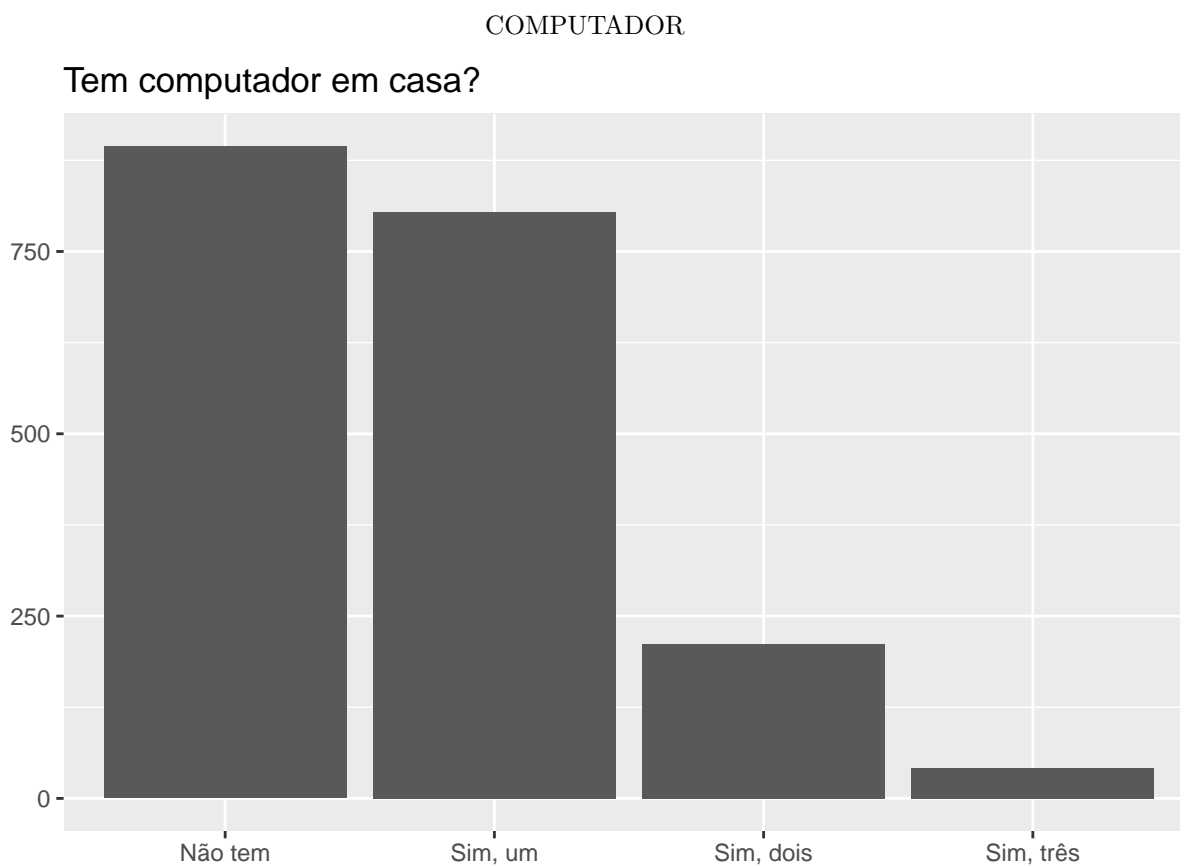
USO\_TEMPO\_TELAS: Variável qualitativa ordinal

AFAZERES\_DOM: Variável qualitativa ordinal

TRABALHO: Variável qualitativa nominal

Perspectivas: Variável qualitativa nominal

**3. Escolha CINCO variáveis categóricas e apresente-as em forma tabular e gráfica. Comente os resultados.**

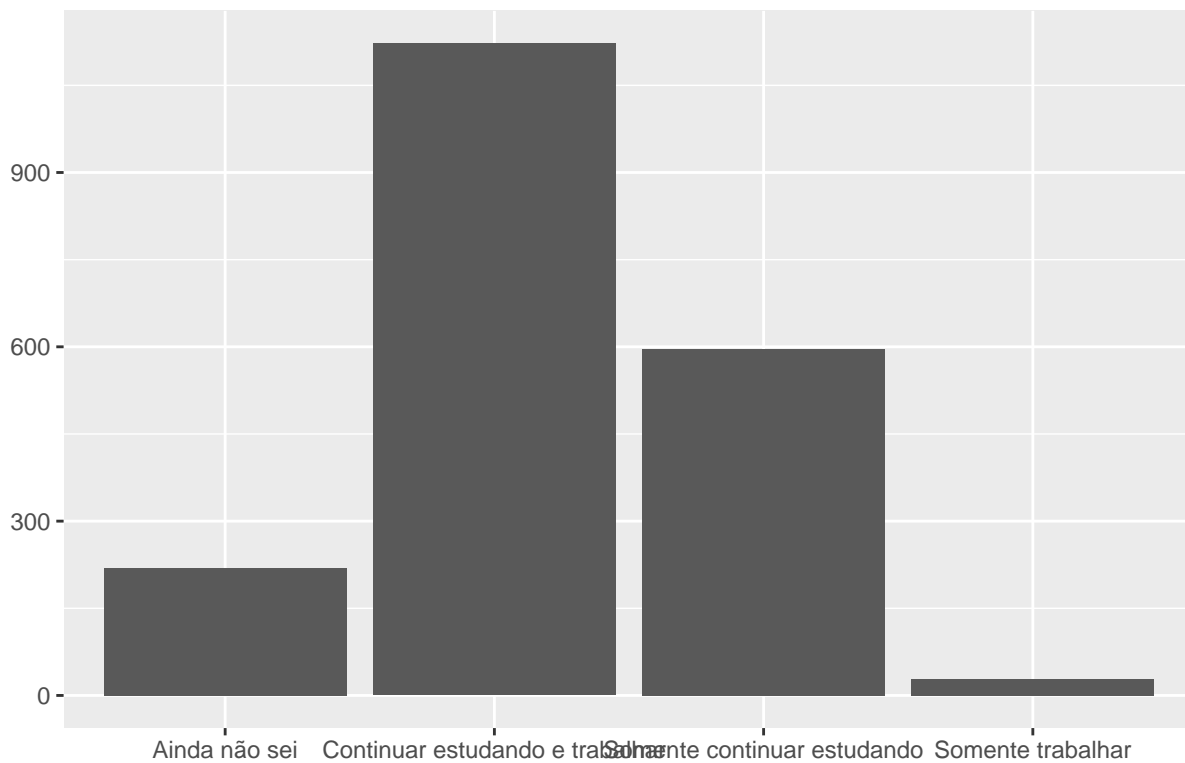


Tem computador em casa?	Frequência
Não tem	894
Sim, um	804
Sim, dois	212
Sim, três	42
quatro ou mais	0

44,7% dos estudantes amostrados não possuem computador em casa, e dentro dos que possuem, 40,2% dizem possuir apenas um.

## PERSPECTIVAS

### Perspectivas futuras

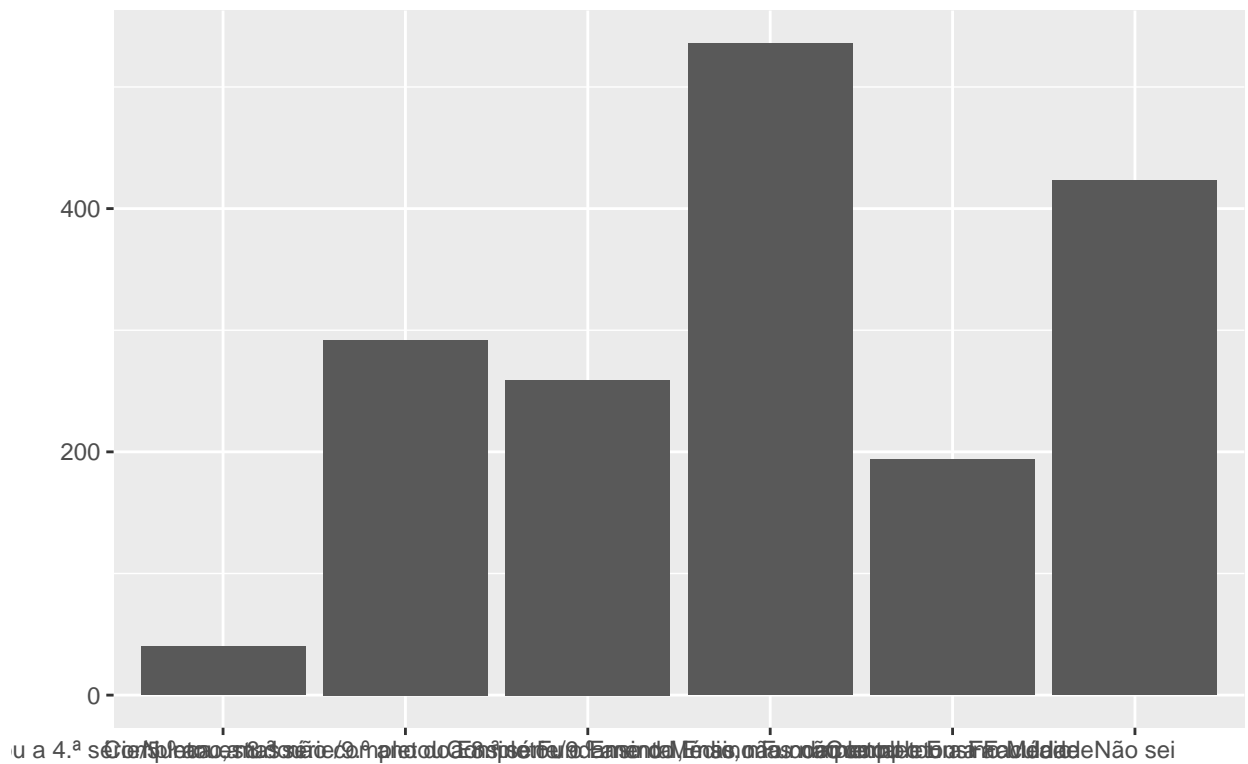


Perspectivas futuras	Frequência
Ainda não sei	219
Continuar estudando e trabalhar	1122
Somente continuar estudando	596
Somente trabalhar	28

A maior parte dos estudantes amostrados (85,9%) relatam a perspectiva de continuarem estudante, dos quais a maioria absoluta pretende também trabalhar enquanto estudam. Uma minoria considerável (10.95%) relata não saber ainda suas perspectivas futuras.

ESC\_MAE

### Escolaridade da mãe

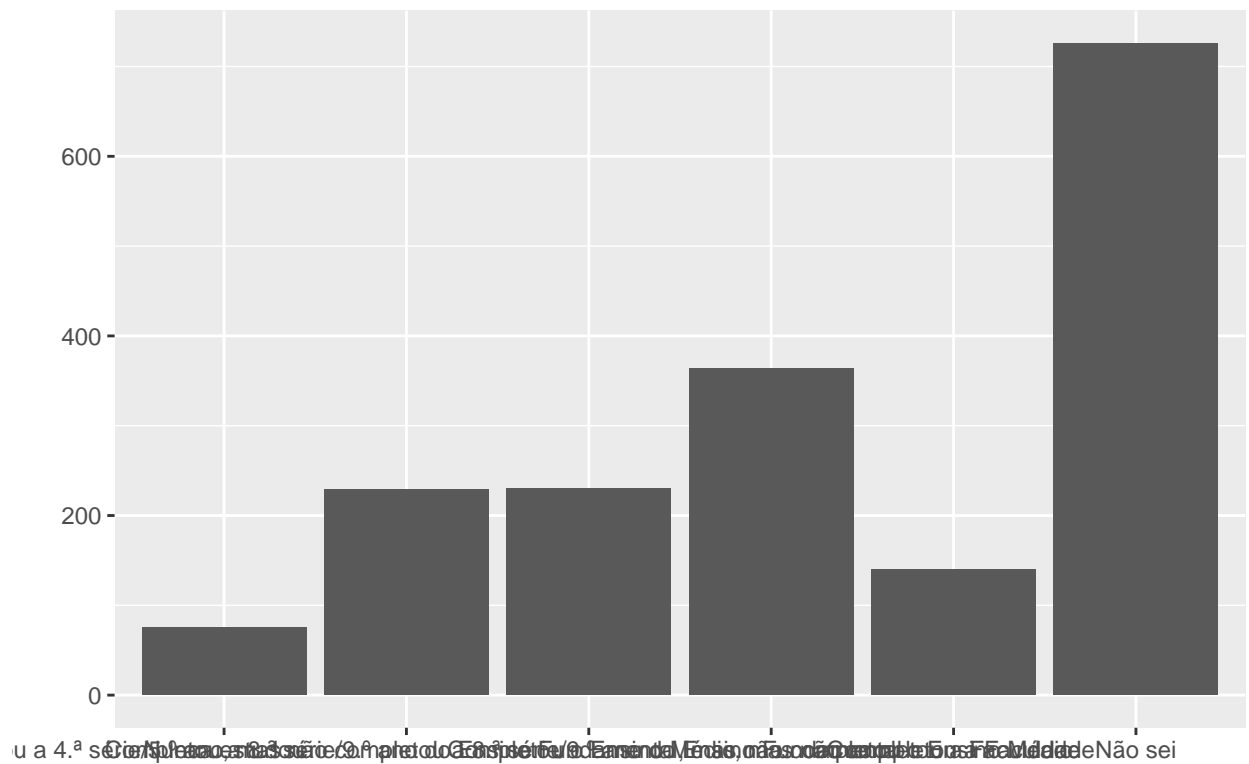


Escolaridade da mãe	Frequência
Nunca estudou	40
Não completou a 4.ª série/5.º ano do Ensino Fundamental	0
Completou a 4.ª série/5.º ano, mas não completou a 8.ª série/9.º ano do Ensino Fundamental	292
Completou a 8.ª série/9.º ano do Ensino Fundamental, mas não completou o Ensino Médio	259
Completou o Ensino Médio, mas não completou a Faculdade	536
Completou a Faculdade	194
Não sei	423

Uma parcela considerável dos estudantes amostrados (29,55%) relatam que a mãe sequer concluiu o ensino médio. Dentre as que concluíram alguma faculdade, o percentual é abaixo de 10% (9,7%)

ESC\_PA1

### Escolaridade do pai

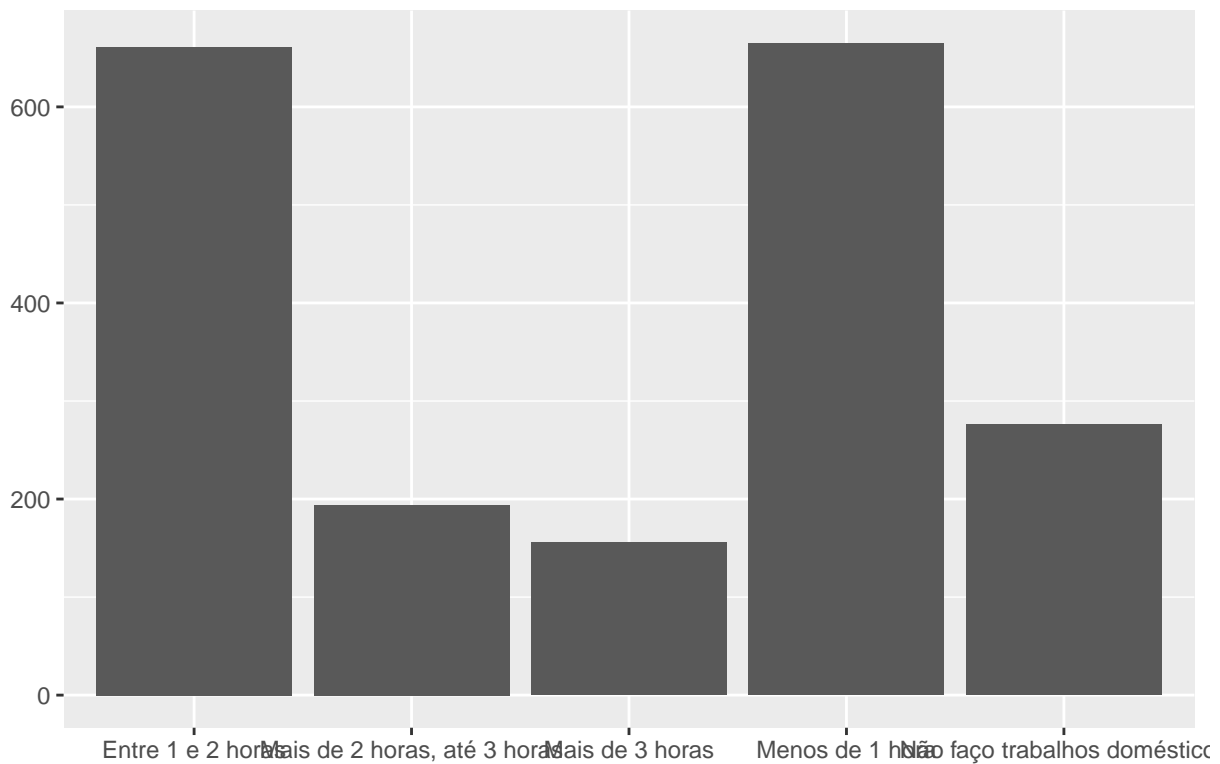


Escolaridade do pai	Frequência
Nunca estudou	75
Não completou a 4.ª série/5.º ano do Ensino Fundamental	0
Completou a 4.ª série/5.º ano, mas não completou a 8.ª série/9.º ano do Ensino Fundamental	229
Completou a 8.ª série/9.º ano do Ensino Fundamental, mas não completou o Ensino Médio	230
Completou o Ensino Médio, mas não completou a Faculdade	364
Completou a Faculdade	140
Não sei	726

No caso da escolaridade paterna dentre os estudantes amostrados, uma parcela menor é relatada não ter concluído sequer o ensino médio (26,7%), porém dentre os graduados a proporção também é menor que no caso da escolaridade materna, sendo apenas 7% dos pais formados em alguma faculdade.

# AFAZERES\_DOM

## Tempo gasto fazendo trabalhos domésticos



Tempo gasto fazendo trabalhos domésticos	Frequência
Entre 1 e 2 horas	661
Mais de 2 horas, até 3 horas	194
Mais de 3 horas	156
Menos de 1 hora	665
Não faço trabalhos domésticos	276

A grande maioria dos estudantes amostrados (86,2%) relatam terem afazeres domésticos diários.

**4. Para as variáveis quantitativas NOTA\_LP e NOTA\_MT  
apresente:**

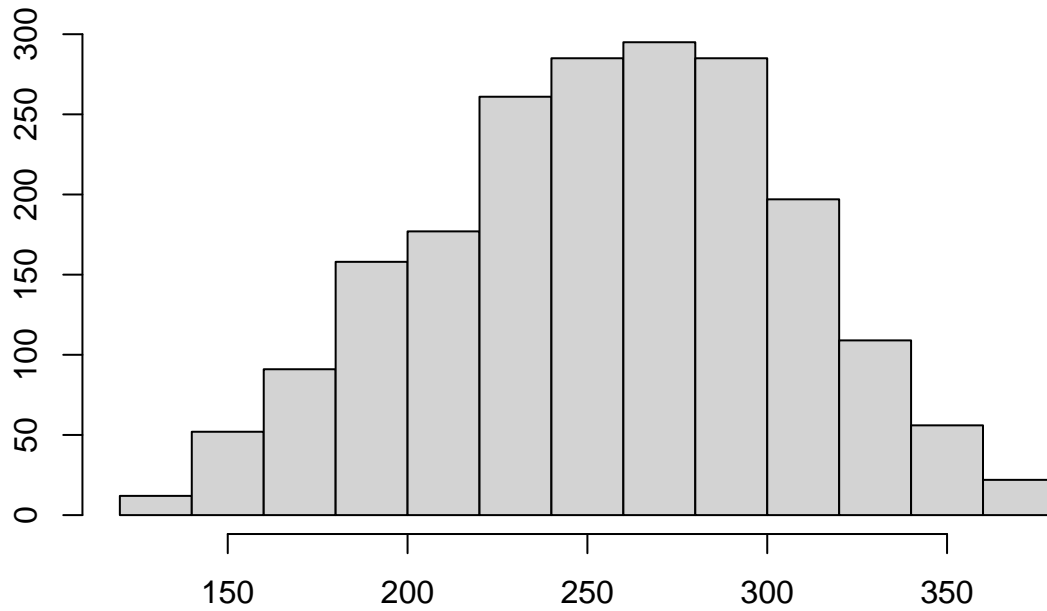
**Distribuição de frequências, com intervalos de classe**

	Até 200	200 a 250	250 a 300	300 a 350	350 ou mais
Notas em língua portuguesa	313	572	731	344	40
Notas em matemática	293	709	654	296	48

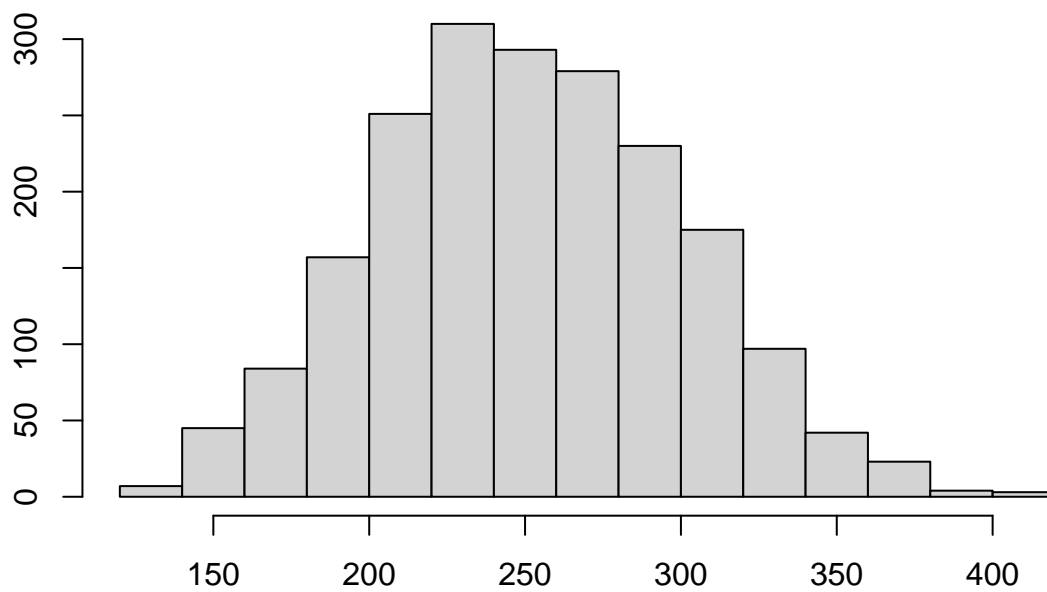


## Histograma

### Histograma – Notas em língua portuguesa



### Histograma – Notas em matemática



## Medidas de posição, variabilidade, assimetria e curtose.

### Variável NOTA\_LP

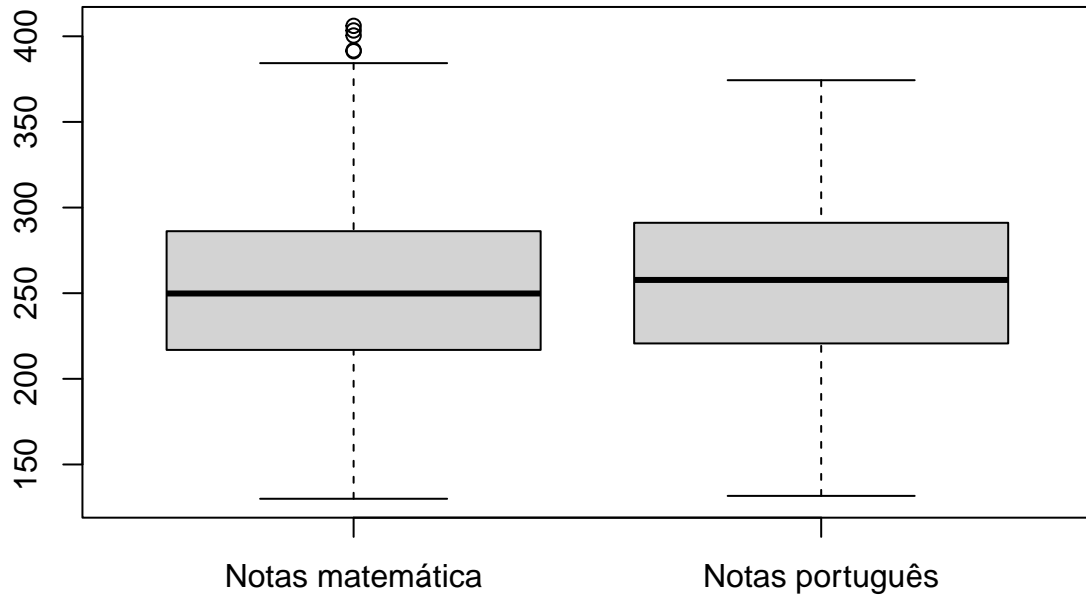
	Valores
Menor valor	131.6636190
Primeiro quartil	220.6756377
Segundo quartil	257.7346705
Terceiro quartil	291.1049268
Maior valor	374.3597440
Média	255.2336471
Mediana	257.7346705
Desvio padrão	49.6778091
Amplitude	242.6961250
Desvio absoluto	40.4905282
Amplitude interquartílica	70.4292890
Assimetria	-0.1349581
Curtose	-0.4986611

### Variável NOTA\_MT

	Valores
Menor valor	129.9821650
Primeiro quartil	216.8933378
Segundo quartil	249.8325030
Terceiro quartil	286.1915105
Maior valor	406.0774660
Média	251.6550574
Mediana	249.8325030
Desvio padrão	48.6473540
Amplitude	276.0953010
Desvio absoluto	39.5388585
Amplitude interquartílica	69.2981728
Assimetria	0.1454824
Curtose	-0.3355293

## Box-plot

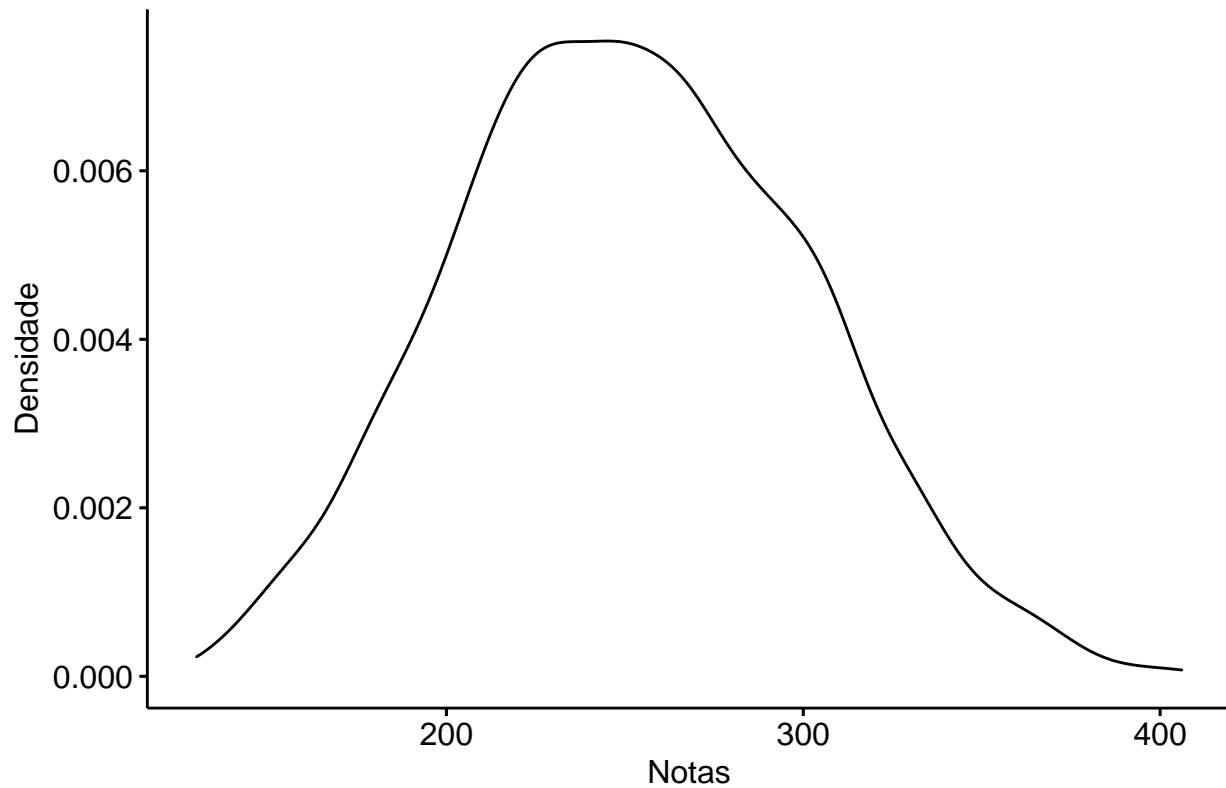
### Box-plot das notas em matemática e língua portuguesa

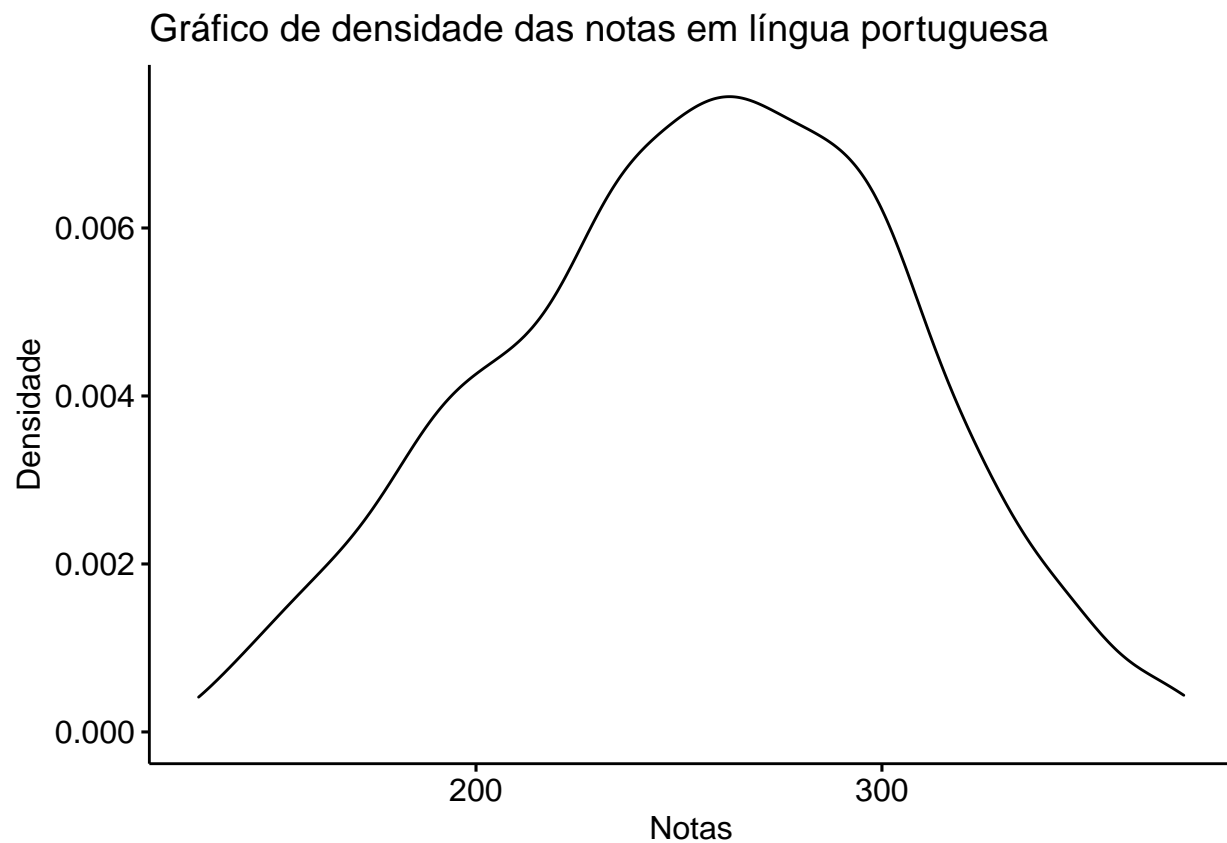


**Análise dos resultados, incluindo comentários sobre normalidade (não é para fazer testes de hipóteses).**

Primeiramente, fazendo uma análise visual para a normalidade das variáveis: Nota em Matemática, e Nota em língua portuguesa, utilizando gráficos de densidade.

**Gráfico de densidade das notas em matemática**





Estes gráficos fornecem um julgamento visual sobre se a distribuição tem forma de sino. A forma lembra o formato de sino do gráfico de densidade da distribuição normal. Mas só isso não é suficiente para tal afirmação.

Ainda em análise visual, podemos utilizar um gráfico Q-Q (quantil-quantil), que desenha a correlação entre uma determinada amostra e a distribuição normal. Uma linha de referência de 45 graus também será plotada.

Gráfico Q-Q das notas em matemática

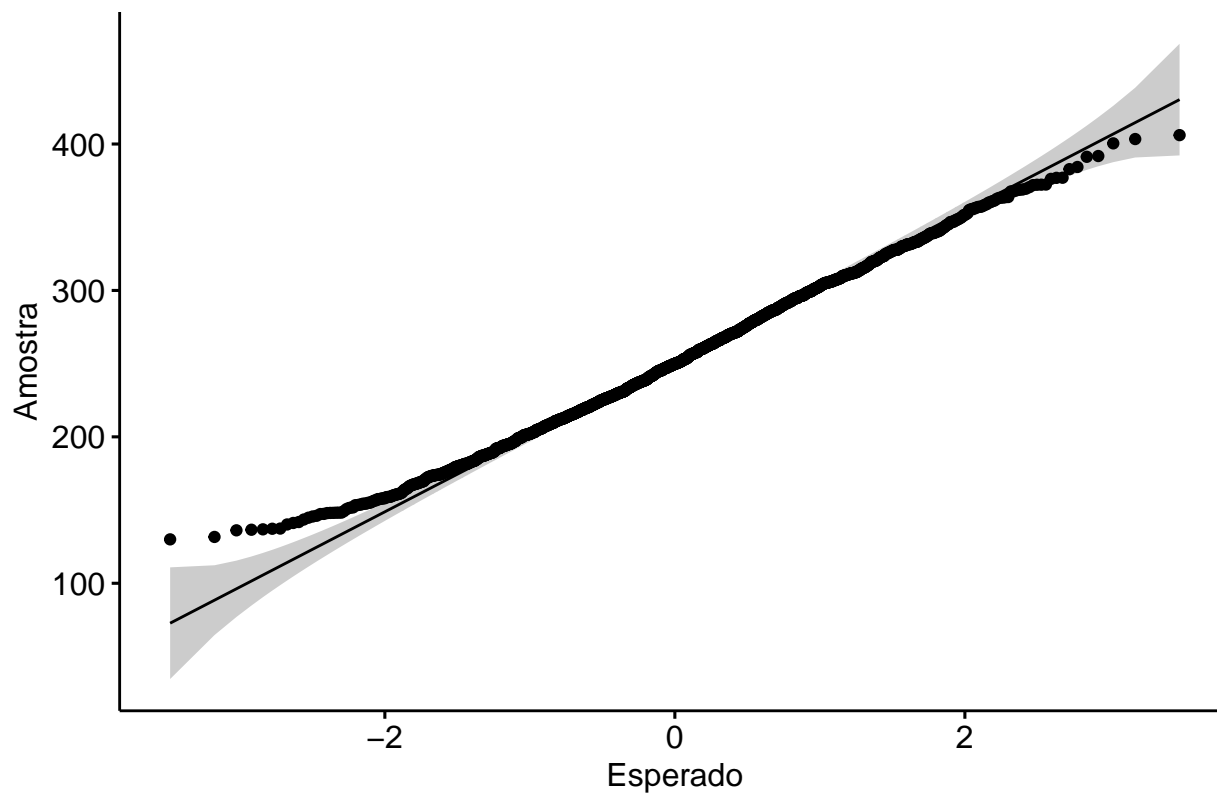
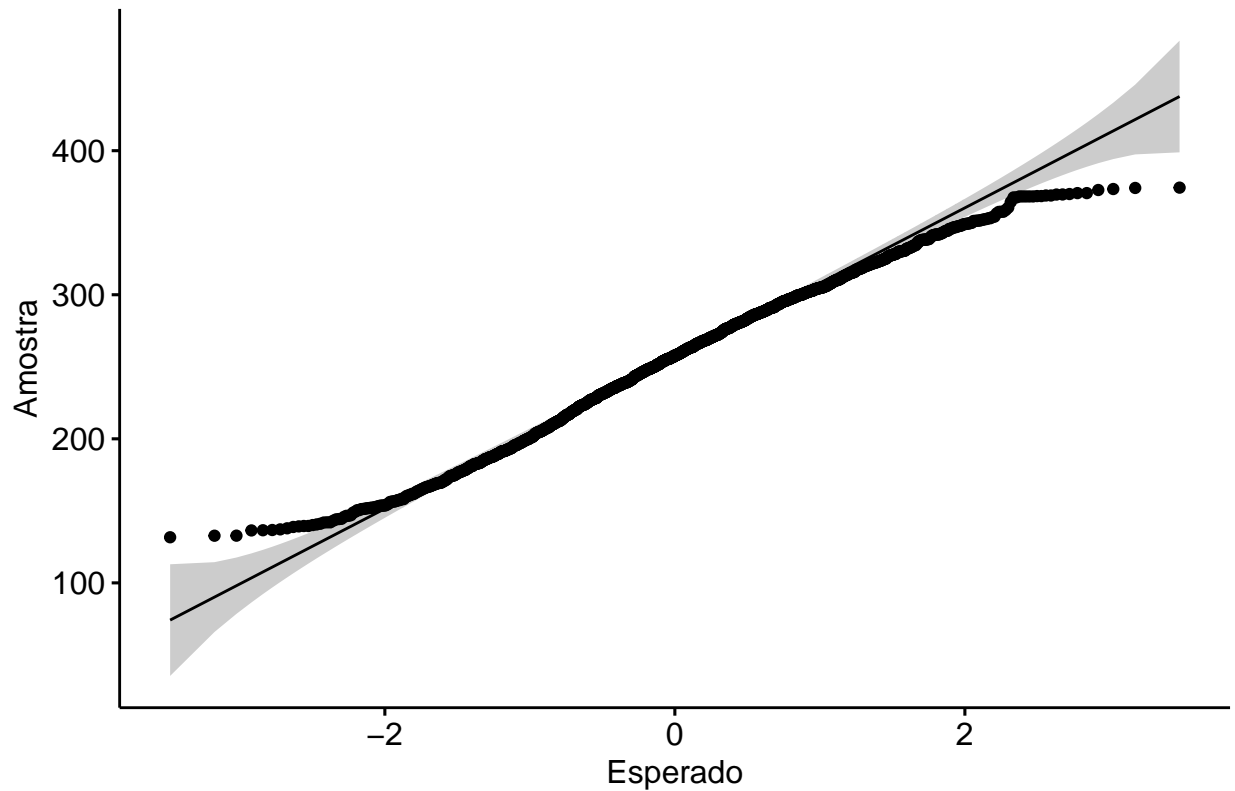


Gráfico Q–Q das notas em língua portuguesa



Para os valores mais próximos aos extremos, a amostra parece divergir um pouco do esperado para a distribuição normal para ambas as variáveis. Porém, a maioria das observações centrais aparentam convergir com a normalidade.

A análise visual da normalidade é importante para compararmos de forma didática o comportamento esperado, e o de fato observado na amostragem. Porém, geralmente não é confiável por si só. É possível usar um teste de significância comparando a distribuição da amostra com uma normal para verificar se os dados mostram ou não um desvio grave da normalidade.

Existem vários métodos para teste de normalidade, como o teste de normalidade de Kolmogorov-Smirnov e o teste de Shapiro-Wilk.

Farei, portanto, um teste de normalidade para cada variável, utilizando o teste de Shapiro-Wilk para normalidade de uma amostra.

A estatística do teste é: W, dada por:

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

```
##  
## Shapiro-Wilk normality test  
##  
## data:  amostra$NOTA_MT  
## W = 0.99618, p-value = 6.136e-05
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  amostra$NOTA_LP  
## W = 0.99275, p-value = 2.214e-08
```

Do output, o p-valor  $< 0,05$  em ambas as variáveis (Nota em matemática: 0.00006136; Nota em língua portuguesa: 0.00000002214) implica que a distribuição dos dados é significativamente diferente da distribuição normal. Em outras palavras, não podemos assumir a normalidade dessas variáveis para essa amostra.