



**Universidade de Brasília**

DEPARTAMENTO DE ESTATÍSTICA

08 agosto 2022

## **Atividade 3.2**

Prof<sup>a</sup>. Ana Maria Nogales

Métodos Estatísticos 2

Aluno: Bruno Gondim Toledo | Matrícula: 15/0167636

### **Atividade 3.2 - Análise de dados: testes de aderência**

Escolha duas subamostras geradas na atividade 2.2 (uma com  $n=20$  e a outra com  $n=200$ ):

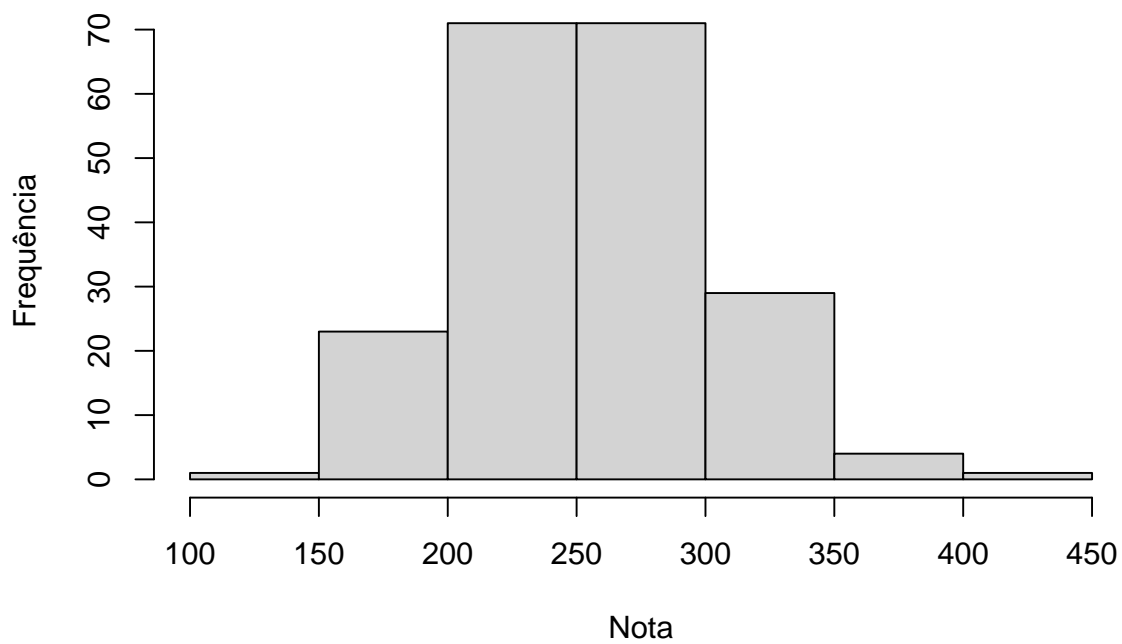
1. Para a amostra com  $n=200$ , construa uma distribuição de frequências para uma das variáveis nota em língua portuguesa ou em matemática (como na Atividade 1.3; escolher uma das variáveis) e verifique se os dados podem ser descritos pela distribuição Normal..
2. Teste normalidade das variáveis notas em língua portuguesa e matemática para a amostra ( $n=20$ ). Apresente os testes de Shapiro-Wilk, Anderson-Darling e Kolmogorov (Lilliefors). .
3. Comente os resultados obtidos, à luz da análise descritiva realizada com a amostra com 2000 observações (Atividade 1.3).

# 1 - Distribuição de frequência para a variável Nota em Matemática para a amostra n=200 e verificação de normalidade

## 1º passo: Histograma

Começarei por fazer um histograma, que além de apresentar as distribuições de frequências da variável, nos fornece um panorama visual acerca de sua semelhança com o formato de uma distribuição normal, se for o caso.

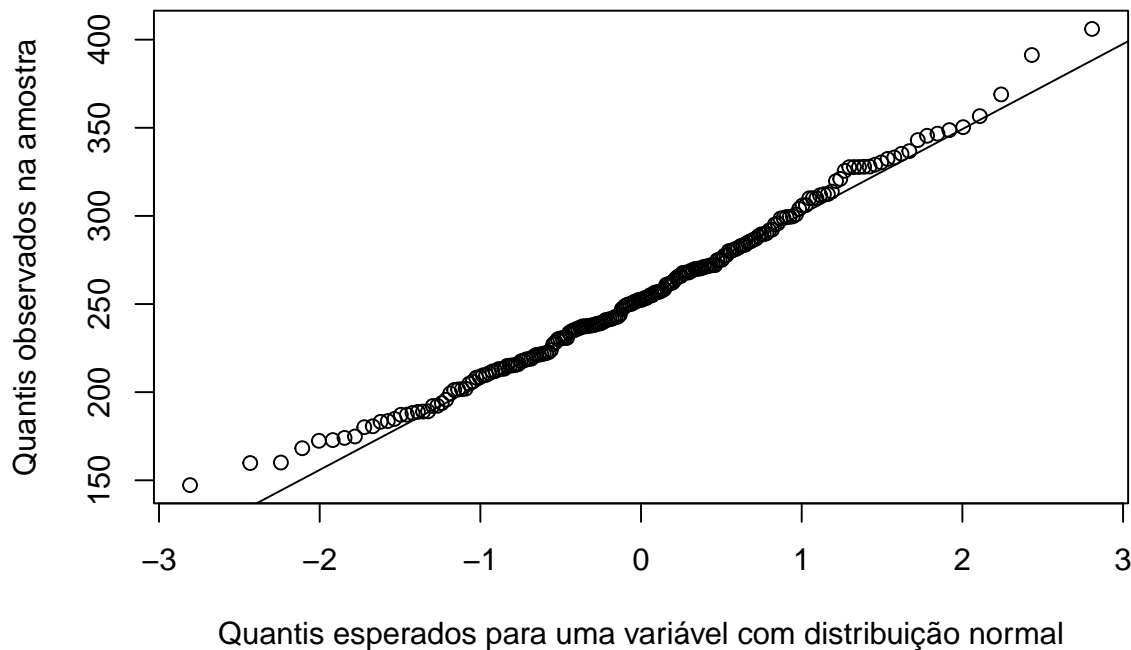
**Histograma Notas Matemática – Amostra n=200**



## 2º passo: Gráfico *Quantil-Quantil*

Seguidamente, farei um gráfico Q-Q, que traz mais uma perspectiva visual do esperado para uma distribuição normal (linha cortando o gráfico) com o observado (pontos)

### Gráfico Q-Q da variável Nota em Matemática – Amostra n=200



Essas duas abordagens nos ajudam a balizar a decisão. Pela análise visual, a distribuição aparenta ser aproximadamente normal. Porém isso não é suficiente para inferirmos uma resposta. Para tal, é necessário realizar testes de normalidade.

## 3º passo: Realizando testes de normalidade

O teste de *Shapiro-Wilk* apresentou estatística de teste **W=0.9900705**, com **p-valor=0.1829695**. O teste de *Anderson-Darling* apresentou estatística de teste **A=0.444664**, com **p-valor=0.2816183**. O teste de *Kolmogorov (Lilliefors)* apresentou estatística de teste **D=0.0472191**, com **p-valor=0.3397955**.

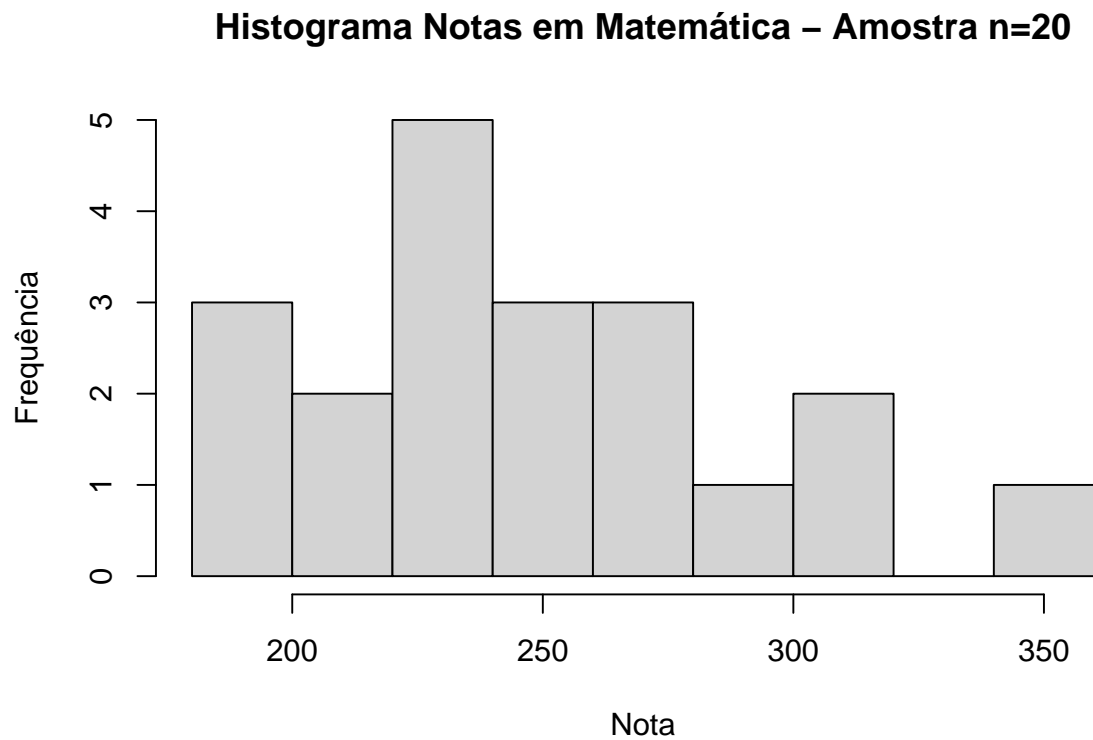
Com esses resultados, devemos **aceitar a hipótese de normalidade da amostra**.

## 2 - Testando normalidade das variáveis notas em língua portuguesa e matemática para a amostra (n=20)

De forma análoga à questão anterior, seguirei os 3 passos para testar a normalidade de cada uma das variáveis.

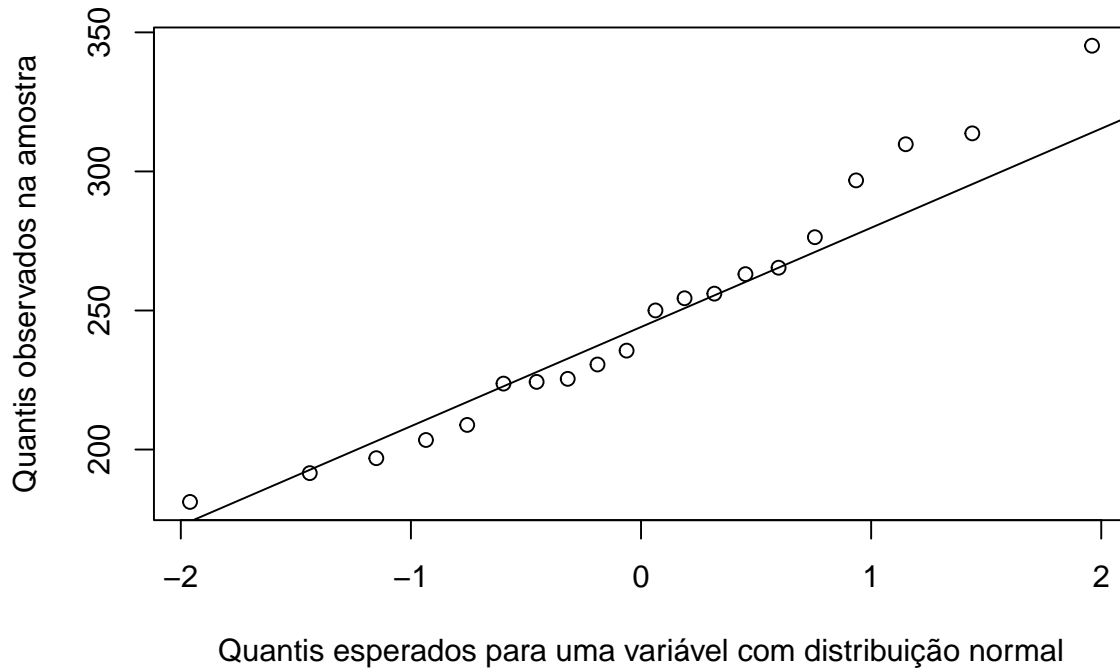
### Variável nota em Matemática

1º passo: Histograma



2º passo: Gráfico *Quantil-Quantil*

**Gráfico Q–Q da variável Nota em Matemática – Amostra n=200**



3º passo: Realizando testes de normalidade

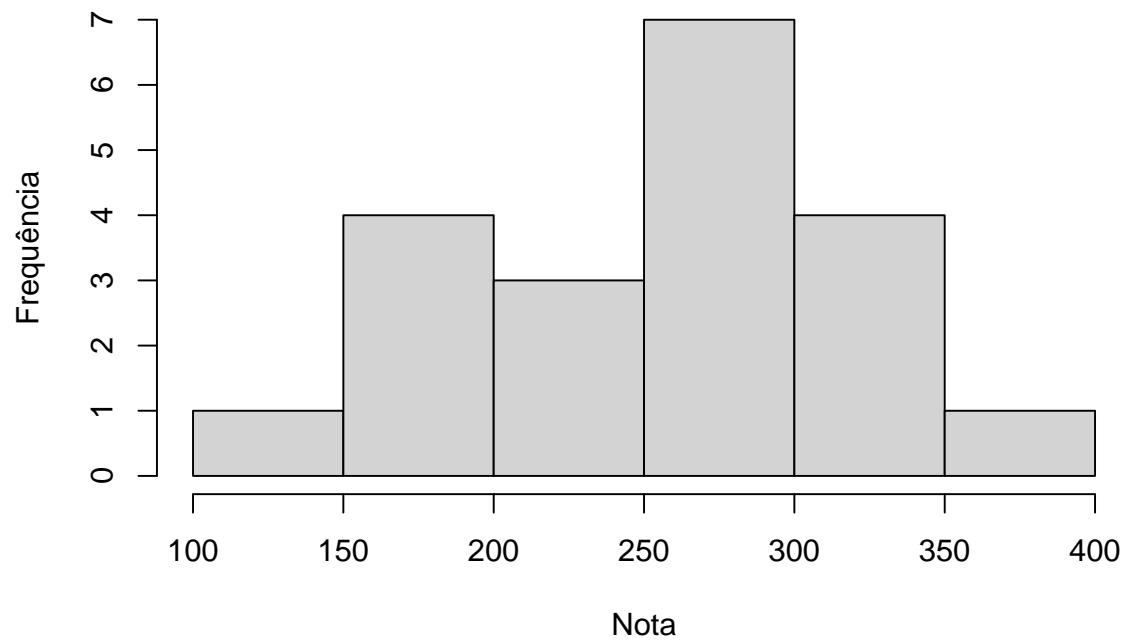
O teste de *Shapiro-Wilk* apresentou estatística de teste **W=0.9629786**, com **p-valor=0.604914**. O teste de *Anderson-Darling* apresentou estatística de teste **A=0.2660988**, com **p-valor=0.6525044**. O teste de *Kolmogorov (Lilliefors)* apresentou estatística de teste **D=0.1071565**, com **p-valor=0.7917618**.

Com esses resultados, devemos **aceitar a hipótese de normalidade da amostra**.

## Variável Nota em Língua Portuguesa

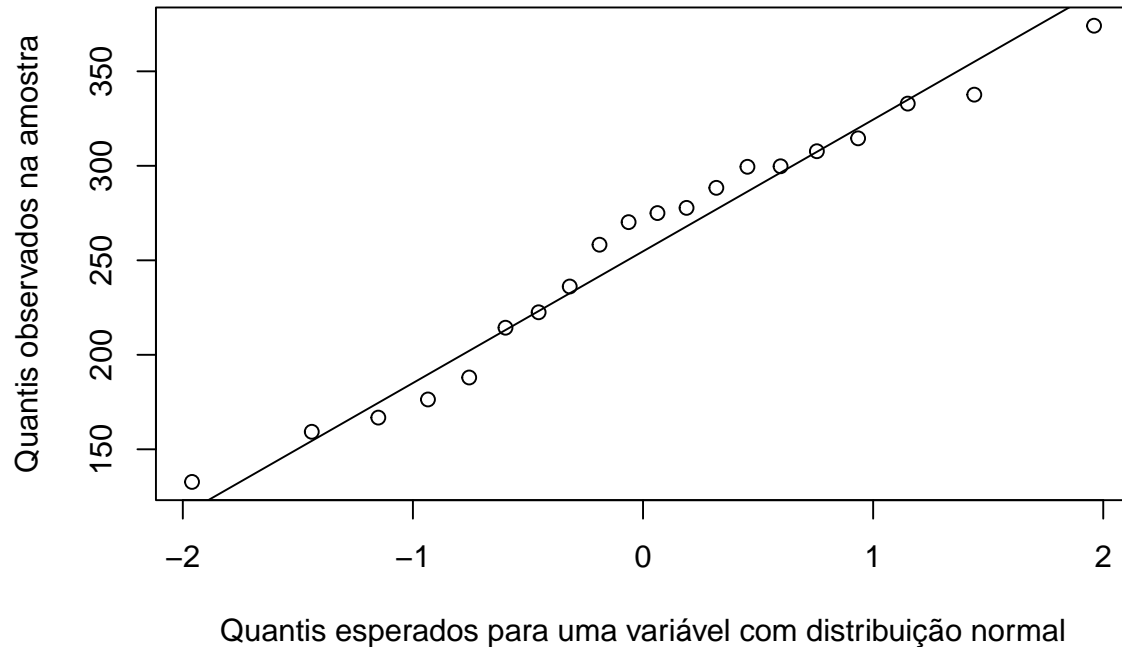
1º passo: Histograma

**Histograma Notas em Língua Portuguesa – Amostra n=20**



2º passo: Gráfico *Quantil-Quantil*

### Gráfico Q-Q da variável Nota em Língua Portuguesa – Amostra n=20



3º passo: Realizando testes de normalidade

O teste de *Shapiro-Wilk* apresentou estatística de teste **W=0.9657757**, com **p-valor=0.6643684**. O teste de *Anderson-Darling* apresentou estatística de teste **A=0.3020201**, com **p-valor=0.5435989**. O teste de *Kolmogorov (Lilliefors)* apresentou estatística de teste **D=0.1304634**, com **p-valor=0.4982526**.

Com esses resultados, devemos **aceitar a hipótese de normalidade da amostra**.

### 3 - Comente os resultados obtidos, à luz da análise descritiva realizada com a amostra com 2000 observações.

Na análise da atividade 1.3, os resultados obtidos **rejeitaram** a hipótese de normalidade para ambas as variáveis (Nota em Matemática e Nota em Língua Portuguesa) para a amostra  $n=2000$ . Nesta atividade (3.2) estou trabalhando com duas subamostras ( $n=20$  e  $n=200$ ), retirados dessa amostra maior. Apesar da amostra original eu ter **rejeitado** a hipótese de normalidade, esta se deu por conta do *p-valor* reduzido, o que é de se esperar em uma amostra grande. Como as amostras aqui trabalhadas são consideravelmente reduzidas, tanto as *estatísticas de teste* quanto os *p-valores* estão na margem de aceitação. Isso se dá por conta dos testes serem conservadores em relação à aceitação da hipótese nula em caso de  $n$  pequeno.