



Universidade de Brasília
IE- Departamento de Estatística
Métodos Estatísticos 2

Juliana Magalhães Rosa

Testes de Aderência
Amostra dos Resultados do Saeb

Brasília
2020

Sumário

1 Introdução	2
2 Distribuição de Notas	3
2.1 Notas em Língua Portuguesa	3
2.2 Notas em Matemática	5
3 Testes de Normalidade para NOTA_LP	7
3.1 Amostra de Tamanho 30	7
3.2 Amostra de Tamanho 100	8
4 Testes de Normalidade para NOTA_MT	9
4.1 Amostra de Tamanho 30	9
4.2 Amostra de Tamanho 100	10
5 Considerações Finais	11

1 Introdução

Testes de Aderência (ou Ajustamento) são testes de hipóteses que estudam as distribuições de probabilidade das variáveis. Geralmente indicam se uma variável segue ou não certa distribuição conhecida, a partir da análise de amostras. É comum existirem vários testes com o mesmo propósito, mas com métodos diferentes.

O foco deste trabalho estará em checar se as variáveis `NOTA_LP` e `NOTA_MT`, que indicam a proficiência dos alunos brasileiros em português e matemática na escala Saeb, seguem uma distribuição de probabilidade aproximadamente normal. Para isso serão utilizados os testes Qui-Quadrado, Kolmogorov, Shapiro-Wilk e Anderson-Darling.

Os dados a serem utilizados consistem em uma amostra de 2000 alunos do nono ano que participaram do Saeb 2017. Esse conjunto de dados fará o papel da população, e dele serão colhidas uma amostra de tamanho 30 e uma amostra de tamanho 100 para serem utilizadas nos testes.

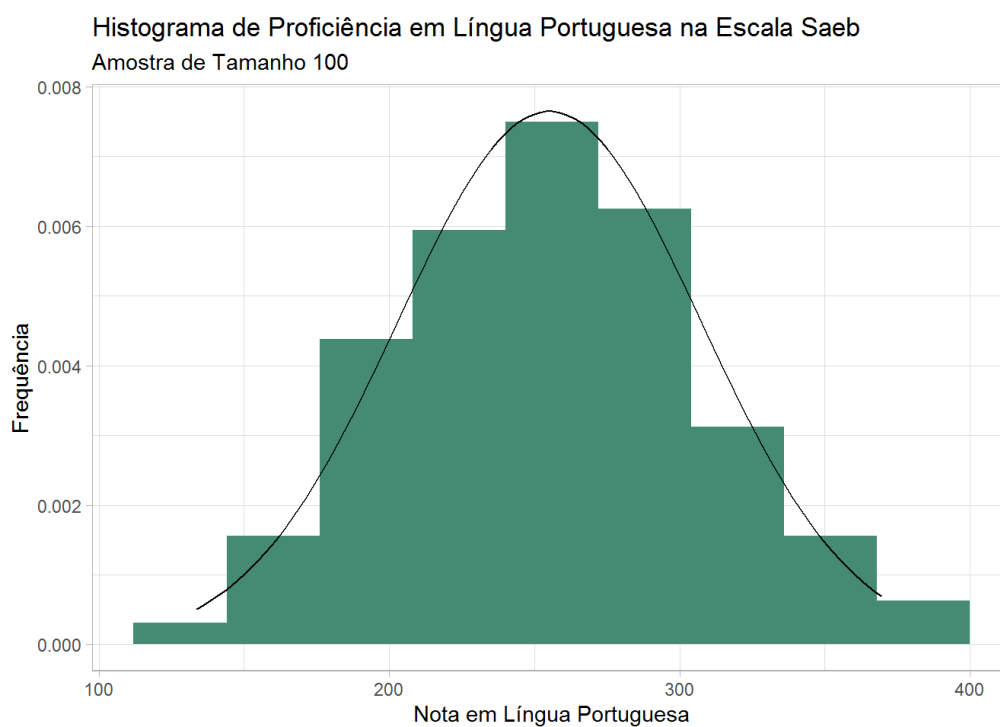
2 Distribuição de Notas

2.1 Notas em Português

Tabela 1- Distribuição de Frequências das Notas em Língua Portuguesa na Prova Saeb no Brasil em 2017

Intervalos de Classes das Notas de Língua Portuguesa	Frequência Absoluta
(133, 165]	4
(165, 197]	10
(197, 229]	18
(229, 261]	25
(261, 293]	20
(293, 325]	13
(325, 357]	7
(357, 329]	3

Fonte: <http://portal.inep.gov.br/educacao-basica/saeb>



Pela tabela de distribuição de frequências, observa-se que as classes centrais de valores são as que possuem mais observações. Isso é um primeiro indício de uma distribuição possivelmente próxima da normal.

O histograma mostra também uma maior concentração de dados na parte central da distribuição, em contraste com suas caldas. Utilizando o recurso da curva normal sobre o gráfico para facilitar a comparação, se torna evidente que a normal seria uma aproximação no mínimo razoável para a distribuição real das notas em português.

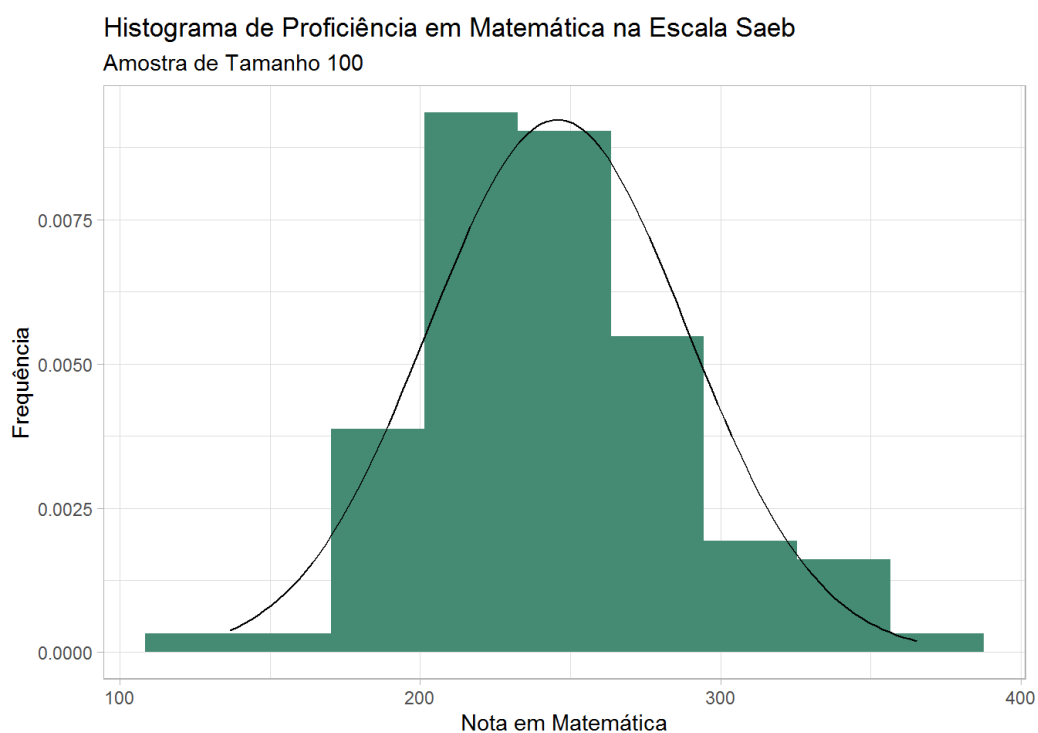
Para confirmar isso, foi realizado o teste Qui-Quadrado de aderência nas classes de notas. A estatística qui-quadrado do teste teve o valor de 0,96, que é um número pequeno. Como essa medida é baseada nas distâncias quadráticas proporcionais entre os valores observados e os que seriam esperados caso a distribuição fosse a normal, um resultado pequeno é evidência a favor da hipótese de normalidade. E, de fato, com um p-valor de 0,97, a hipótese nula é aceita, e conclui-se que as notas em língua portuguesa seguem uma distribuição aproximadamente normal.

2.2 Notas em Matemática

Tabela 2- Distribuição de Frequências das Notas em Matemática na Prova Saeb no Brasil em 2017

Intervalos de Classes das Notas de Matemática	Frequência Absoluta
(136, 167]	2
(167, 198]	12
(198, 229]	25
(229, 260]	25
(260, 291]	24
(291, 322]	6
(322, 353]	5
(353, 384]	1

Fonte: <http://portal.inep.gov.br/educacao-basica/saeb>



Para as notas em matemática, a tabela de frequência mostra uma diferença entre as primeiras e as últimas classes de valores. Os intervalos iniciais possuem uma concentração maior de frequências do que os finais. Isso aponta para uma distribuição assimétrica e, portanto, diferente da normal.

No histograma é possível enxergar essa assimetria à direita, indicando que as notas em matemática costumam ser mais baixas. Ao comparar a distribuição com a curva normal, percebe-se um deslocamento da primeira em relação ao centro, indicando que as duas distribuições são distintas.

De novo foi feito o teste Qui-Quadrado de ajustamento para checar a normalidade. Dessa vez a estatística do teste resultou em 4,8. Esse valor é maior do que o referente às notas em português, validando a impressão de que as pontuações em língua portuguesa seguem uma distribuição mais próxima da normal do que as de matemática. Porém, o qui-quadrado observado ainda é pequeno, indicando que, na verdade, há certa proximidade com a normal. O p-valor de 0,44 confirma
isso.

3 Testes de Normalidade para NOTA_LP

3.1 Amostra de Tamanho 30

O primeiro teste realizado foi o de Lilliefors para normalidade. Sendo uma adaptação do teste de Kolmogorov, sua estatística do teste representa a maior distância vertical entre a distribuição empírica acumulada e a função de distribuição do modelo teórico a ser testado. Nesse caso, a estatística assumiu um valor de 0,16. Apesar de não parecer um valor elevado, seu p-valor é de 0,04, permitindo a rejeição da hipótese de normalidade a um nível de significância de 0,05.

Também foi feito o teste de Shapiro-Wilk que é específico para o estudo da normalidade. A estatística do teste resultou em 0,94 que, por ser um valor próximo de 1, indica proximidade com a normal. Com um p-valor de 0,07, se aceita a normalidade, contradizendo o teste de Lilliefors com o mesmo nível de significância.

O teste de Anderson-Darling resultou em uma estatística de 0,67 e um p-valor de 0,07. Portanto, nesse teste aceita-se a hipótese de normalidade também. Esses resultados podem parecer contraditórios, mas na realidade os p-valores não são muito diferentes. Todos estão em torno de 0,05, então, talvez, se fosse escolhido outro nível de significância, os testes estariam em concordância. Com um nível crítico de 0,01, por exemplo, a hipótese nula seria aceita em todos os casos.

3.2 Amostra de Tamanho 100

Para a amostra maior, o teste de Lilliefors dá uma estatística que vale 0,04. Lembrando que esse valor representa a distância máxima entre as funções de distribuição empírica e normal, observa-se que a diferença entre as duas é pequena. De fato, o p-valor elevado de 0,96 é uma forte evidência a favor da hipótese de normalidade.

O teste de Shapiro-Wilk ratifica essa conclusão com sua estatística do teste muito próxima de 1 ($W=0,99$) e seu p-valor de 0,84 que não permite a rejeição da hipótese nula.

Por fim, o teste de Anderson-Darling com estatística valendo 0,15 e p-valor assumindo 0,96 também conclui normalidade. Dessa vez os três testes tiveram resultados parecidos, todos com p-valor muito alto. Essa consistência maior nos testes pode estar conectada com o aumento da amostra.

Quando foi feita a análise dos dados completos com as 2000 observações, a distribuição das notas em língua portuguesa apresentou uma assimetria à esquerda que não ficou tão clara no estudo atual das amostras. No histograma feito para a amostra de tamanho de 100, os dados pareceram encaixar-se bem com a curva normal, e os testes também apontaram para isso em ambas as amostras.

4 Testes de Normalidade para NOTA_MT

4.1 Amostra de Tamanho 30

O teste de Lilliefors apresentou uma distância vertical máxima de 0,09 entre as distribuições acumuladas, indicando proximidade com a normal. O p-valor de 0,78 não permite a rejeição da hipótese de normalidade.

Pelo teste de Shapiro-Wilk, obteve-se uma estatística do teste de 0,97, que de novo é próxima de 1. Isso é confirmado pelo p-valor de 0,45 que aceita a normalidade da distribuição.

Realizando o teste de Anderson-Darling, obteve-se uma estatística com valor de 0,29. O p-valor foi de 0,58, dando força às duas conclusões anteriores.

4.2 Amostra de Tamanho 100

Assim como para a amostra de tamanho 30, o teste de Lilliefors resultou em uma estatística de 0,09. Porém, pelo tamanho maior da amostra, o p-valor dessa vez foi de 0,03, permitindo, assim, a rejeição da hipótese de normalidade ao nível de significância de 0,05.

O teste de Shapiro-Wilk também apresentou valor similar com o da amostra anterior para a estatística do teste ($W=0,98$). E, novamente, seu p-valor foi mais reduzido, valendo 0,11. Mas dessa vez não há evidência suficiente para a rejeição da normalidade.

Com o teste de Anderson-Darling, é obtido um valor de 0,72 para a estatística do teste e de 0,06 para o p-valor. Por pouco não foi rejeitada a proximidade com a normal. Para esse tamanho de amostra foi mais fácil a rejeição da hipótese nula, como é de se esperar. Mas, de forma geral, as notas de matemática parecem tem comportamento suficientemente próximo da normal.

Novamente houve uma diferença entre o que foi observado na distribuição das 2000 notas e o que se concluiu ao aplicar os testes de aderência nas amostras. Tanto no histograma dos dados populacionais como no da amostra de tamanho 100, havia uma aparente assimetria à direita que servia de indicação contra a normal. Porém, os testes indicaram que essa assimetria não é acentuada ao ponto de descartar a distribuição normal como uma aproximação razoável.

5 Considerações Finais

Resultados de testes de hipóteses dependem de diversos fatores, como o nível descritivo escolhido, o tamanho da amostra, a amostra em si, entre outros. Mesmo em um processo de amostragem aleatória, pode acontecer de uma amostra não ser representativa da população, o que resulta em erros nas conclusões. Portanto, é simples perceber que diferentes testes resultam em diferentes evidências e, conseqüentemente, diferentes informações.

No caso dos testes de aderência, existem várias opções de procedimentos com propósitos e utilidades iguais, mas que ao serem aplicadas em uma mesma amostra podem ter resultados diferentes. Por isso é importante saber escolher o melhor teste para cada situação, e também ter a disposição de experimentar com diversos testes e ponderar as diferentes medidas.

Em suma, houve algumas diferenças entre população e amostra, entre amostras de tamanhos diferentes, e até entre testes diferentes para uma mesma amostra. Mas, de maneira geral, as notas tanto de matemática como de português parecem ser aproximadas de modo satisfatório pela distribuição normal.