

UNIVERSIDADE FEDERAL DA BAHIA

DEPARTAMENTO DE MATEMÁTICA E ESTATÍSTICA - IME

DISCIPLINA: MÉTODOS ESTATÍSTICOS - MAT236

PROF. Dr. Gilberto Pereira Sassi

ESTUDANTES:

Daija Diniz de Macedo

Djair Maykon de Novaes Miranda

Joanderson Gonçalves Santos

João Lucas Lima de Melo

Thalles Yan Santos Medrado

RESUMO

Este relatório traz resultados de uma análise estatística dos dados fornecidos pelo MEC a respeito dos vestibulandos e das provas do ENEM aplicadas em Salvador no ano de 2019. O material bruto foi previamente tratado de forma que analisamos apenas uma amostra com 1000 participantes. Nossa análise foi empenhada em estudar uma possível relação entre gênero, renda financeira e notas, a fim de compreender e contextualizar os resultados da análise na discussão sobre acesso à oportunidades e desigualdade social.

A análise nos trouxe resultados que atestaram primeiro uma relação de gêneros, onde mesmo com uma maior quantidade de participantes do sexo feminino, candidatos do sexo masculino tiveram melhor desempenho no exame independente da renda declarada na pesquisa de faixa salarial. Foi também observado que, geralmente, candidatos com renda superior a outros, independente do seu gênero, possuem melhor aproveitamento no exame.

Como esperado, portanto, foi atestada uma evidente correlação entre melhor condição financeira e melhor rendimento. Além disso, o desempenho melhor de pessoas do sexo masculino em todas as faixas de renda indicam possíveis questões de gênero atreladas à desigualdade.

INTRODUÇÃO

O exame nacional do ensino médio (ENEM), foi criado em 1998 para medir como estava o nível de alfabetização dos estudantes concluintes do período de educação básica. Foi em seguida adaptado para servir como ingresso dos estudantes na educação superior a partir de 2002, gerando grande interesse por parte da sociedade interessada no ingresso em uma faculdade e uma possível melhora na qualidade de vida, bem como dos

professores e pesquisadores, pois se trata de uma fotografia da realidade do ensino médio dos últimos 19 anos.

A nossa população-alvo são os participantes do ENEM 2019 na cidade de Salvador e nosso principal objetivo é constatar a existência e quais são as relações de sexo (masculino ou feminino), a nota obtida no exame (a média aritmética das quatro provas objetivas) e a renda per capita dos mesmos divulgada nos dados do Inep.

Tal análise pode prover dados suficientes para constatação de problemas sociais que afetam o rendimento dos participantes do ENEM e que, portanto, geram um impacto grande no futuro da educação no país.

MATERIAIS E MÉTODOS

Os dados analisados por nós são do INEP e se referem às informações dos participantes do ENEM 2019 em Salvador-BA. Esses dados foram previamente tratados de modo que houvesse informações apenas dos participantes que se enquadraram nos seguintes casos: esteve presente em todos os dias da prova, não fez prova adaptada, fez a prova de inglês, teve uma redação válida e não era treineiro. A partir desses dados, foi retirada uma amostra com 1000 participantes que se enquadraram nos filtros mencionados. Nossa análise se deu sobre essa amostra.

Para a análise dos dados, fizemos ajuste nos nomes das colunas para que fossem mais fáceis de manipular. Como pretendemos analisar os dados da renda como uma variável qualitativa nominal, renomeamos seus valores para letras de A-Q para representar as faixas de renda em ordem crescente.

Para a análise, utilizamos a ferramenta google colab e a linguagem Python em conjunto com frameworks de análise de dados e geração de gráficos, sendo eles: pandas, seaborn, numpy e pylab. O link para o colab produzido: <https://colab.research.google.com/drive/18OoIVSkYnhwQxOhozcw7hcDKbZgziFV?usp=sharing>.

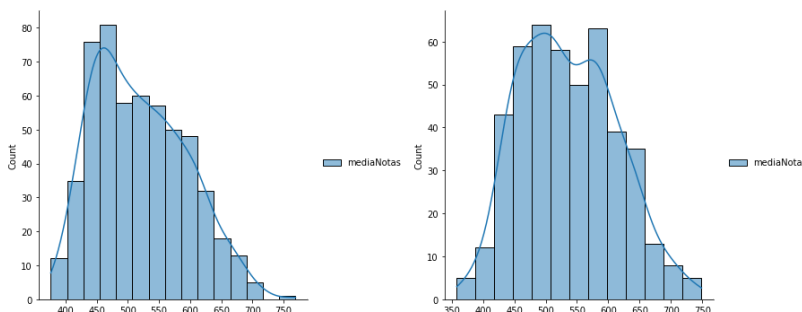
Começamos a análise utilizando medidas de posição e dispersão para fazer uma visão geral dos dados, criamos uma tabela agrupada por sexo e faixas de renda e verificamos a média, o desvio padrão e os quantis mínimo, 25%, mediana, 75%, e máximo (100%) da média das notas dos participantes. A partir desses dados, pudemos perceber qual caminho seguir para provar nossa hipótese da discrepância relacionada a questões de desigualdade social.

Para continuar e aprofundar nosso resultado, geramos box-plots e histogramas para a média das 4 notas de cada participante agrupados por sexo e por faixa de renda.

Após fazer essa análise mais geral e constatar alguns fatos dos dados, procuramos um estudo mais objetivo. Para isso, criamos box-plots e histogramas em 3 grupos de faixa de renda: A-B, C-O e P-Q. Calculamos algumas probabilidades para argumentar sobre nossas hipóteses de relação entre renda, gênero e desempenho. Por fim, utilizamos um intervalo de confiança para afirmar nossos resultados para todos os participantes do ENEM além da nossa amostra com um erro associado.

RESULTADOS

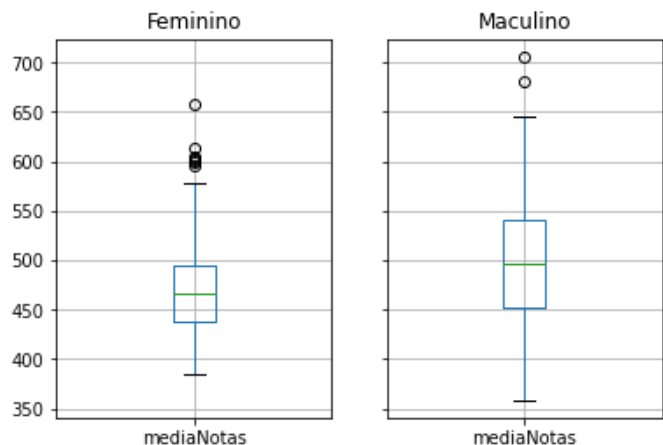
Os primeiros assuntos foram relacionados à construção de histogramas, mostrados na Fig. 1. À esquerda a imagem mostra o histograma da média das notas para o sexo feminino, e, à direita, a média para o sexo masculino. A observação dos gráficos sugere que a densidade de frequência de notas está mais concentrada nos valores abaixo de 600 pontos para ambos os sexos. Além disso, indicam que existe um maior número de pessoas do sexo feminino que não atingem a nota de corte. Vale destacar que essa nota não é a mesma para cada curso.



Em seguida, buscamos analisar se a diferença nas faixas de renda implica em algum tipo de variação nas médias por sexo. Para isso, construímos três pares de boxplots, um com as duas faixas de renda mais baixas, um com as treze faixas de renda intermediárias e um com as duas maiores faixas de renda. A média utilizada para os cálculos foi a média das 4 provas objetivas, e, com uma confiança de 95%, nas faixas de renda mais baixas foi apresentada uma variação de 6 pontos para mais ou para menos, um valor próximo do valor obtido para as faixas de renda intermediárias, de 5 pontos de variação.

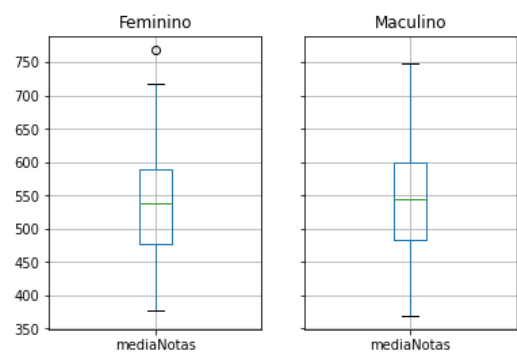
Ao contrário do que ocorreu com as outras faixas, os participantes com maiores rendas o intervalo de confiança foi substancialmente maior, apresentando uma variação de 20 pontos para mais ou para menos, o que pode ter sido causado pelo pequeno tamanho amostral de participantes nessa faixa de renda. Contudo, mesmo com uma variação de 20 pontos nos resultados das médias para as maiores rendas, as conclusões obtidas se mantêm prováveis de estarem corretas e possuem muita probabilidade de serem os valores observados para toda a população de estudantes, considerando os filtros citados, que realizaram a prova em Salvador no ano de 2019. Demais resultados obtidos são descritos a seguir.

Primeiro, analisamos os resultados para as faixas de renda mais baixas, até R\$ 998,00, apresentados na Fig 2. Grande parte das médias dos estudantes se encontram no primeiro quartil de dados, ou seja, abaixo da linha de mediana. Dessa forma, é gerada então uma leve assimetria à esquerda, que demonstra uma relação entre a baixa renda e um baixo desempenho para homens e mulheres, como já

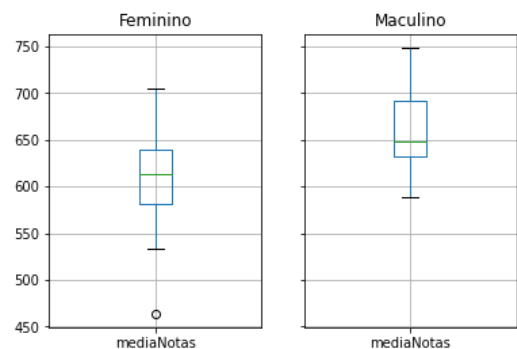


hávamos verificado no histograma anterior. Aqui, contudo, assinalamos a diferença entre o desempenho das pessoas que fizeram a prova com base em seu sexo. A mediana dos vestibulandos de sexo masculino é maior que o terceiro quartil do grupo Feminino. Assim, entre participantes das rendas mais baixas, fica claro o contraste do desempenho entre vestibulandos de sexos diferentes.

Em seguida, fizemos diagramas também para as faixas intermediárias de renda (Fig. 3), com participantes que declararam renda entre R\$ 998,00 e R\$ 14.970,00. Aqui, podemos observar um desempenho muito mais equivalente entre os diferentes sexos, ou seja, o sexo não foi determinante para estabelecer o desempenho do candidato.



Por fim, analisamos o desempenho dos participantes que pertencem à faixa de renda maior ou igual a 14.970,01 (Fig. 4). O encontrado sugere novamente um desempenho muito superior dos participantes de sexo masculino em detrimento dos participantes do sexo feminino.



A partir dos três conjuntos de boxplots, foi possível observar que, apesar de um indicativo de desempenho superior pelos participantes de sexo masculino ao analisar apenas as notas em todas as faixas de renda, ao separar os vestibulandos de diferentes faixas pudemos perceber que a discrepância entre seus desempenhos existe sobretudo nas faixas de renda mais extremas (maiores e menores), mas essa diferença não se mantém ao verificar os resultados obtidos nas faixas de renda intermediárias. Outra

observação possível é que em todos os boxplots há a presença de outliers, que podem ser produtos de erro de observação ou arredondamento.

Após verificar a diferença no desempenho nas faixas por sexo, calculamos a probabilidade de uma pessoa na faixa mais baixa e mais alta obter uma pontuação acima de 650 na média das quatro provas objetivas. O resultado obtido é que, entre os participantes nas faixas de renda menores ou iguais a R\$ 998,00, a probabilidade de obter essa pontuação foi de 1,06%. Enquanto isso, nas faixas de renda maiores ou iguais a 14.970,01, essa probabilidade foi de 34,29%. O dado fala por si só, mostrando claramente que o fator renda implica, e muito, no desempenho do participante.

A análise apresentada evidencia que o ENEM acaba tendo um caráter excludente quando observada a disparidade de renda dos participantes, o que pode ser explicado pela falta de oportunidades de educação de qualidade para essas pessoas. Por ser uma ferramenta extremamente poderosa para o ingresso em instituições de ensino, participantes de menor renda não possuem uma vantagem, ecoando em um processo segregatório de acesso à educação superior.

CONCLUSÃO E CONSIDERAÇÕES FINAIS:

Tendo como referência uma amostra dos dados fornecidos pelo INEP a respeito do ENEM de 2019 aplicado em Salvador, buscamos analisar e entender as relações de renda, gênero e desempenho na prova.

Desenvolvemos diagramas e usamos conceitos estatísticos para termos uma base mais sólida para manipulação, visualização e análise dos dados coletados. Pudemos perceber que, de forma geral, participantes com maior renda possuem desempenho consideravelmente superior aos participantes com menor renda, independente de gênero. Observamos ainda que, majoritariamente, as participantes femininas possuem desempenho inferior aos participantes masculinos em faixas de renda extremas.

Concluimos, portanto, que a condição financeira exerce uma influência diretamente proporcional à nota do candidato do ENEM. Podemos atribuir a isso melhores oportunidades de ensino, acesso à estruturas, plataformas e ambientes de melhor qualidade para a preparação para o vestibular. Ter o maior exame nacional para ingresso em instituições de ensino superior tão condicionado a fatores de renda em um país tão socialmente desigual não só é alarmante como também se apresenta como causa e sintoma dessa desigualdade. Nossas dificuldades foram encontrar uma plataforma que todos pudessem trabalhar com os dados, por isso usamos várias, além de que não tivemos tempo para executar os testes de hipótese devido ao assunto ser muito recente no curso, mas consideramos que dentro das limitações, o relatório foi bem sucedido.