

CANDIDATOS DO ENEM EM 10 ANOS DA LEI DE COTAS

Marília Magalhães Maia ¹
Silvio Fernandes ²
Luiza Helena Felix de Andrade³

RESUMO

Assumindo a desigualdade como uma realidade brasileira, em agosto de 2012, foi sancionada a Lei 12.711, a qual destina 50% das vagas de ingresso nas instituições federais de ensino público para estudantes oriundos das escolas públicas, com subdivisão relacionadas a renda familiar e a cor/raça do estudante. Esta foi criada como uma medida provisória para equiparar o acesso à educação entre aqueles afetados negativamente pelas desigualdades existentes no país, e em seu próprio texto exige a revisão de seus aspectos construtivos ao marco dos seus 10 anos de criação. Diante disso, se faz necessário estudos que avaliem o cenário atual e a efetividade desta política pública. Então, este trabalho propõe uma análise do perfil dos estudantes a partir dos microdados referentes ao Exame Nacional do Ensino Médio (Enem), fornecidos pelo Instituto Nacional de Pesquisa e Estatística Anísio Teixeira (INEP), ao longo de uma década, aplicados ao algoritmo de agrupamentos K-means, e os dados do Censo da Educação Superior 2020. Este trabalho avalia as mudanças no ensino superior após a implementação da Lei das Cotas e a necessidade de sua permanência diante do desempenho dos candidatos e das características socioeconômicas de cada agrupamento. Apesar do aumento significativo da população negra nos percentuais de matrículas no ensino superior, observou-se que os estudantes oriundos do ensino médio público e que sejam pretos, pardos e indígenas, ou tenham renda familiar per capita inferior ou igual a 1,5 salários-mínimos, apresentaram-se, expressamente, mais constantes nos grupos de menores desempenhos do Enem do que nos grupos de maiores, durante todo o período analisado. Dessa forma, entende-se a necessidade da permanência das cotas, a fim de fornecer oportunidade de acesso ao ensino superior para a população mais atingida pelas desigualdades.

Palavras-chave: Lei das cotas, desigualdade, Enem, Censo da Educação Superior.

INTRODUÇÃO

Diante das desigualdades que afetam os estudantes brasileiros, como medida provisória para diminuir os impactos causados por elas, em 2012, foi sancionada a Lei nº 12.711/2012, a Lei da Cotas. Esta prevê a reserva de 50% das vagas das instituições de ensino superior públicas federais para estudantes de escolas públicas, subdividindo-se entre os alunos quanto a sua renda *per capita* e cor/raça (Brasil, 2012).

¹ Graduando do Curso de Engenharia Elétrica da Universidade Federal Ruaral do Semi-Árido- Ufersa, marilia.maia@alunos.ufersa.edu.br;

² Doutor em Sistemas e Computação da Universidade Federal do Rio Grande do Norte - UFRN, silvio@ufersa.edu.br;

³ Doutora em Engenharia de Teleinformática da Universidade Federal do Ceará- UFC, luizafelix@ufersa.edu.br;

Diante desse debate, o presente trabalho tem como objetivo uma análise do perfil dos estudantes a partir dos microdados referentes ao Enem, ao longo de uma década, aplicados ao algoritmo de agrupamentos K-means, e os dados do Censo da Educação Superior 2020.

Dessa forma, será trabalhado no primeiro momento de uma revisão a literatura, apresentando um breve levantamento histórico do Enem e da Lei das Cotas, além de uma seção apresentando o algoritmo k-means. Em seguida, será apresentado o método de pesquisa, descrevendo quais os parâmetros utilizados no refinamento dos dados e suas justificativas. Como sequência, a apresentação dos resultados encontrados e a discussão sobre eles, em busca de compreender como se dividem os grupos e correlacionar ao contexto do debate sobre as cotas. E então, apresentar as considerações finais e temas para pesquisas futuras.

REFERENCIAL TEÓRICO

A reforma do ensino médio, guiada pelas Diretrizes Curriculares Nacionais do Ensino Médio ocorrida em 1998 e as criações, no mesmo ano, das avaliações padronizadas da educação, como o Exame Nacional do Ensino Médio (Enem) foram importantes mudanças ocorridas nesse período (Klitzke, 2014). O Enem foi desenvolvido como ferramenta para “avaliar o desempenho do estudante ao fim da escolaridade básica”, pelo Ministério da Educação (MEC), através do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) [5]. Porém, passou a ter, também, a finalidade de acesso à educação superior de instituições públicas e privadas, que fizeram uso das notas dos candidatos no processo seletivo. O exame passou por reformulação, em 2009, tornando-se o principal meio de ingresso as instituições de ensino superior (IES) públicas e o processo de seleção unificado das universidades públicas federais.

Diante as mudanças, o “novo Enem”, passou a possuir 180 questões distribuindo-se entre as áreas, linguagens e códigos, matemática, ciências da natureza e ciências humanas, além da redação. Além disso, passou a ser aplicado em dois dias, diferentemente de quando foi criado, quando era aplicado somente em um dia, mas possuía apenas 63 questões (Klitzke, 2014). Ainda mais, é válido ressaltar a implementação da Teoria de Resposta ao Item (TRI) como uma mudança ocorrida durante a reformulação, o que permitiu a comparação das notas com as aplicações de anos futuros, já que esta permite que o exame mantenha o grau de dificuldade (Rabelo, 2013). Dessa forma, passou a ser possível comparar o desempenho dos alunos que concluem o ensino básico ao passar dos anos e traçar os perfis dos estudantes brasileiros.

Ações afirmativas foram desenvolvidas em diversas IES, assumindo como realidade as desigualdades que afetam a população brasileira, quanto as diferenças de oportunidades entre raças e rendas e o desnível entre a educação privada e pública. Dentre as ações, o fornecimento de vagas de ingresso exclusivas para o público afetado (Anhaia, 2017). Diante do cenário de diferentes ações afirmativas, no dia 29 de agosto de 2012, foi sancionada a Lei das Cotas (Lei nº 12.711), a qual em seu Art. 1º diz:

“As instituições federais de educação superior vinculadas ao Ministério da Educação reservarão, em cada concurso seletivo para ingresso nos cursos de graduação, por curso e turno, no mínimo 50% (cinquenta por cento) de suas vagas para estudantes que tenham cursado integralmente o ensino médio em escolas públicas.

Parágrafo único. No preenchimento das vagas de que trata o caput deste artigo, 50% (cinquenta por cento) deverão ser reservados aos estudantes oriundos de famílias com renda igual ou inferior a 1,5 salário-mínimo (um salário-mínimo e meio) per capita” (Brasil, 2012)

Em seguida, no Art. 3º é expresso que dentre as vagas referidas no art. 1º dessa Lei, devem ser preenchidas por turno e curso, por autodeclarados pretos, pardos e indígenas e pessoas portadoras de deficiências seguindo a proposição de vagas às proporções de pretos, pardos e indígenas e pessoas portadoras de deficiências da população da unidade da Federação a qual se apresenta, seguindo o mais recente censo do Instituto Brasileiro de Geografia e Estatística (IBGE). A Figura 1 apresenta o esquema visual da divisão das vagas de ingresso nas IES estabelecida pela Lei.

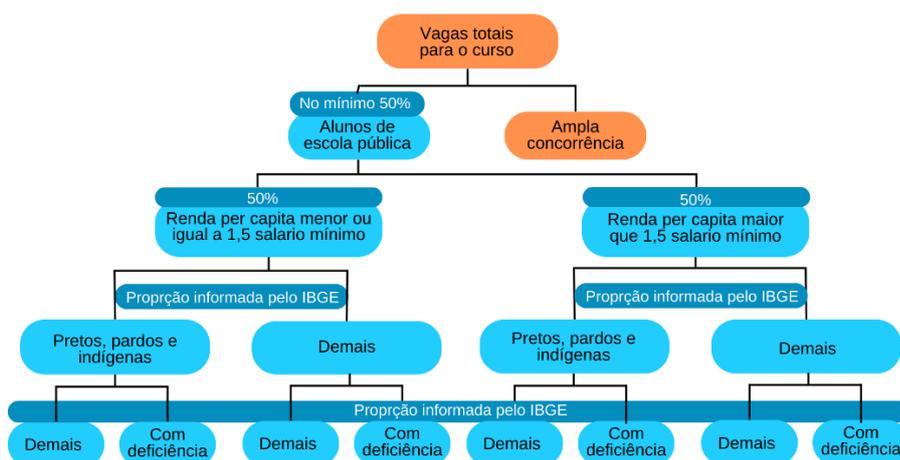


Figura 1 - Distribuição das vagas de ingresso no ensino superior. (Autoria própria)

- **K-means**

O k-means agrupa os dados que mais se aproximam de cada centroide e tem como parâmetros de funcionamento a inicialização aleatória, durante a determinação dos centroides iniciais, a quantidade de agrupamentos, nomeada de k , e a métrica, a qual a mais utilizada é a

distância Euclidiana ao quadrado, devido a existência e unicidade do ponto que a minimiza e este mesmo ponto, sendo o centroide, descrito matematicamente pela média aritmética (Jain, 2010; Lloyd, 1982). Diante disso, é possível descrever seu funcionamento pelos passos apresentados pela Figura 2.

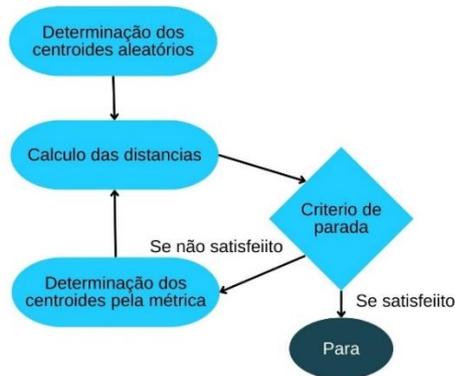


Figura 2 - Fluxograma do funcionamento do algoritmo *k-means*. (Autoria própria)

A determinação aleatória dos centroides é o passo inicial do algoritmo, sendo restrito à primeira iteração e se trata da determinação aleatória da localização dos k centroides. Sendo k definido, como forma de otimização, entre os dados de tal modo que não sejam próximos. Dessa forma, a partir da segunda iteração, os centroides passam a ser determinado pela métrica, sendo definidos pelos pontos que minimizam a soma das distancias a n pontos, sendo variável a métrica escolhida (Matte, 2020). Além disso, ainda segundo (Matte, 2020), no passo seguinte, são calculados as distancias de todos os dados a todos os centroides as quais são usadas no passo seguinte na seleção dos dados que apresentam as menores distancias a cada um dos centroides, de tal forma que um cluster (grupo) é composto por um centroide e todos os dados que se aproximam mais a ele. E por fim, o teste de parada realiza a paralização caso seja atingida a quantidade máxima de iterações ou não houver significativa mudança nos clusters entre uma iteração e a seguinte. Dessa forma, o resultado do algoritmo são os k clusters, sobre os quais é possível a realização de análises quanto as características que os agrupam.

METODOLOGIA

Com o objetivo de analisar a distribuição os alunos aptos as vagas de cotas, para o ensino superior, foram implementadas as suas proficiências no Enem, de cada aplicação desde 2009 a 2022, no algoritmo *k-means*. Os dados utilizados são disponibilizados pelo Instituto Nacional de Ensino e Pesquisa Anísio Teixeira (INEP) no endereço eletrônico (INEP, n.d.).

Diante do grande número de inscritos, se fez necessária a realização de uma amostragem, sobre todos os conjuntos, usando um nível de confiança de 99% e erro amostral de 1%, fornecendo valores proporcionais a quantidade de inscritos e computacionalmente viáveis. Em seguida, foi realizado dois processos de filtragem dos dados, selecionando inicialmente, os alunos tenham concluído o ensino médio no ano de aplicação do exame ou em anos anteriores, e que possuem um valor numérico para as proficiências das áreas avaliadas, ou seja, tenham presença e não tenham sido desclassificados de nenhuma das provas do exame. Após a filtragem inicial, foram criados os conjuntos de dados para as implementações no algoritmo, contendo apenas as proficiências das 5 áreas avaliadas de cada candidato. Estes conjuntos de dados foram aplicados no algoritmo *k*-means para a formação de 2 grupos, com a quantidade máxima de iterações igual a 10.

Em posse dos agrupamentos, foram resgatadas as informações referentes a características socioeconômicas e educacionais de cada indivíduo de cada grupo. Estas informações fazem parte do conjunto de dados disponibilizado pelo INEP e correspondem a respostas dos alunos aos questionários aplicados no momento da inscrição. Restringiu-se as informações referentes a renda, cor/raça e o tipo de escola que o aluno frequentou, visto que estas características são avaliadas para adesão às cotas de entrada nas IES. Entretanto, ocorreram diversas mudanças no questionário, desde a reformulação do exame, em 2009, até a mais recente aplicação. Dentre as mudanças, observa-se um aumento na quantidade de alternativas nas questões referentes a quantidade de pessoas que moram com o estudante e a renda bruta mensal familiar. Ambas as informações foram utilizadas para o cálculo da renda per capita. Para isso, em vista do uso de intervalos de valores proporcionais ao salário-mínimo do ano em vigor, como alternativas de resposta quanto a renda bruta mensal familiar, utilizou-se as médias dos intervalos das respostas de cada inscrito e a dividiu pela quantidade de moradores da residência do inscrito, ou a média do intervalo que informa a quantidade de moradores.

Uma outra mudança significativa nos formulários ocorreu sobre a questão “Em que tipo de escola você cursou o Ensino Médio?”, que surgiu a partir de 2010 e parou de ser aplicada em 2019. A resposta para tal pergunta informa se o aluno cursou todo o ensino médio em escolas públicas, requisito principal para a adesão às cotas. Entretanto, para nossa análise, foi utilizada, para os anos 2009 e 2019 a 2022, a pergunta referente a escola “Dependência administrativa” a qual o aluno deve informar se a escola que frequentou era federal, estadual, municipal ou privada, permitindo saber se o aluno era proveniente do ensino público ou privado. Ainda, ressalta-se que nos anos de 2019 a 2022, essa questão passou a não ser obrigatória, de modo

que candidatos às cotas que não responderam esta questão ficaram de fora da nossa análise. Ainda ocorreram mudanças nas questões relativas à cor/raça, mas não interferem na análise.

A implementação do pré-processamento dos dados, aplicação de métodos estatísticos, normatização das respostas dos candidatos e do algoritmo de *k*-means foi realizado na linguagem de programação Python, com auxílio de algumas bibliotecas padrões e de ciência de dados.

RESULTADOS E DISCUSSÃO

A partir das amostragens dos inscritos no exame, com nível de confiança de 99% e 1 % de erro amostral, a análise anual das características socioeconômicas e raciais é apresentada na Figura 3, de forma que os valores são expressos em percentuais para permitir a comparação entre os anos.

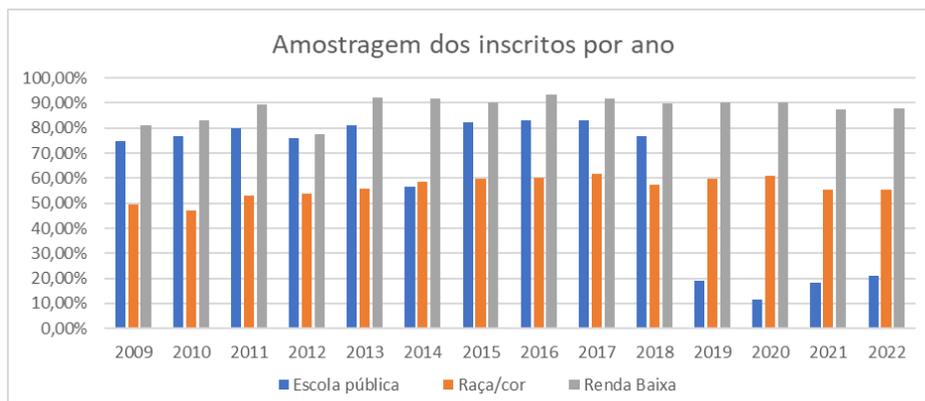


Figura 3 - Percentual de cada característica, a partir da amostra de cada ano. (Autoria própria)

Observa-se na Figura 3 que os valores referentes aos percentuais dos estudantes oriundos das escolas públicas se mantêm acima de 70%, com as exceções dos valores referentes ao ano de 2014, apresentando-se abaixo dos 60% e ao período de 2019 a 2022, devido a não obrigatoriedade de resposta da pergunta. Em relação aos percentuais de estudantes que possuem renda familiar per capita igual ou inferior a 1,5 salários-mínimos (destacados na legenda da Figura 1 como “Renda baixa”), se manteve praticamente constante e acima de 90% desde 2013. Os valores mais elevados sobre esta característica são atrelados ao fato de que, segundo (IBGE, 2022), desde 2012, a renda per capita mensal nacional é inferior a 1,5 salários-mínimos.

Ainda na Figura 3 são apresentados os valores referentes aos inscritos que se autodeclararam pretos, pardos e indígenas (destacados na legenda da Figura 1 como “Raça/cor”), sendo possível observar uma crescente até 2017, atingindo valores superiores a 60% e mantendo-se neste intervalo nos demais anos. Assim, demonstra-se que mais da metade dos

participantes do exame são estudantes pretos, pardos e indígenas, sendo assim uma representação fiel da população brasileira, já que está tem mais da metade composta pela população negra (MEC, 2018).

Apesar da maior participação no exame, segundo o Censo da Educação Superior de 2020 (Lévy, 1997), a população negra não é maioria nas matrículas no ensino superior, apesar de ter alcançado um significativo passando de 35% em 2011, para 46% em 2020, como apresentado na Figura 4.

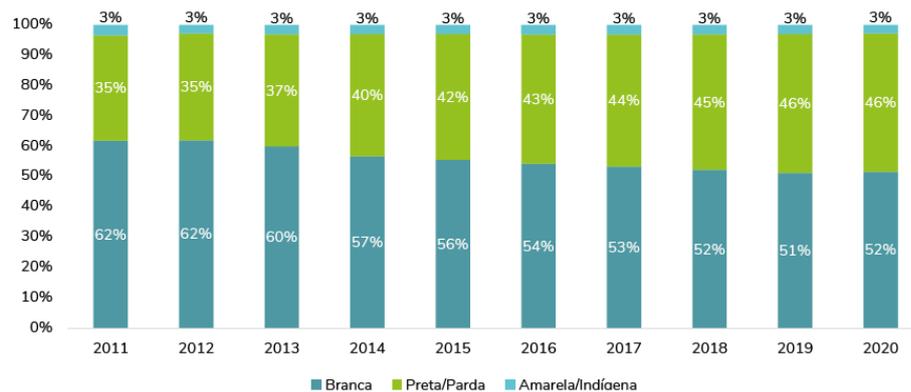


Figura 4 - evolução dos percentuais de declarações relativas à variável cor/raça do aluno (em relação à matrícula) – Brasil – 2011-2020

(Fonte: Elaborada por Deed/Inep com base nos dados do Censo da Educação Superior)

Após o pré-processamento dos dados e a sua implementação o algoritmo com o k igual a 2, são encontrados os 2 agrupamentos de semelhantes, baseados nas proficiências, para cada ano de aplicação do exame. Os tais grupos são representados graficamente e representados pela Figura 5.

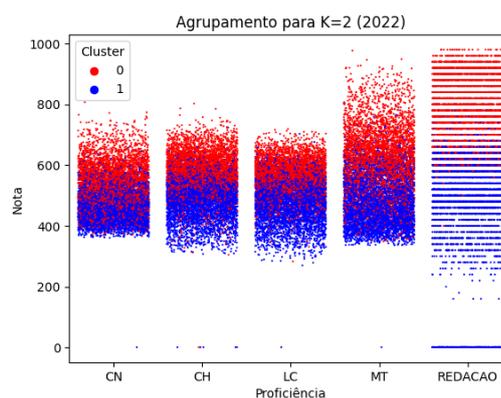


Figura 5 - Gráfico do agrupamento resultante do k -means, com $k=2$ sobre os dados de 2022. (Autoria própria)

No gráfico são expostas as médias das proficiências das provas do exame correspondentes a aplicação de 2022, servindo como demonstração da distribuição dos dados, uma vez que os gráficos dos demais anos analisados seguem o mesmo modelo e estão

disponíveis no link⁴. O algoritmo separou os indivíduos de maiores desempenhos, daqueles que tiveram um menor desempenho, como o esperado. Tornando visível esta separação a partir da observação da concentração dos pontos vermelhos na parte superior de todos os gráficos, indicando este grupo apresenta maiores proficiências em todas as provas, em relação aos indivíduos representados na cor azul. Nomeia-se o agrupamento em vermelho de Cluster 0 e de Cluster 1 o agrupamento em azul. Ocorrem variações nos tamanhos dos grupos de anos diferentes e por se tratar de elementos de 5 dimensões independentes (notas avaliadas), não há uma linha bem definida entre os dois grupos, mas sim uma separação um pouco difusa entre eles.

Com base nesses agrupamentos, as Figura 6 e Figura 7 apresentam gráficos com os percentuais das distribuições dos indivíduos nos grupos, por ano, em relação as características que são critério de adesão as cotas.

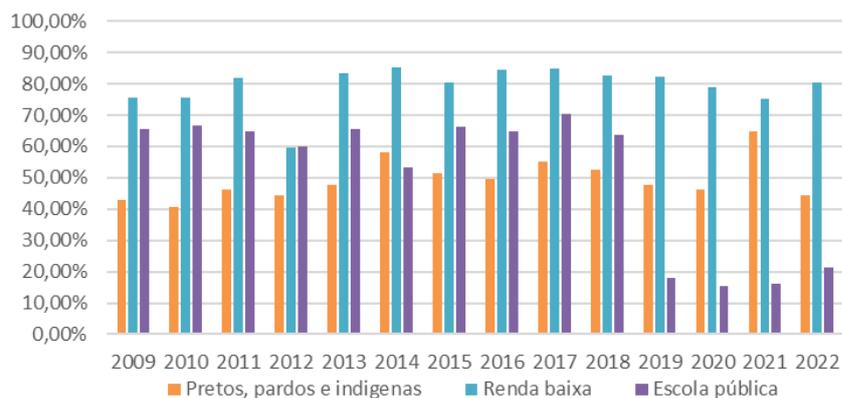


Figura 6 - Gráfico dos percentuais dos integrantes do Cluster 0 por característica, por ano. (Autoria própria)

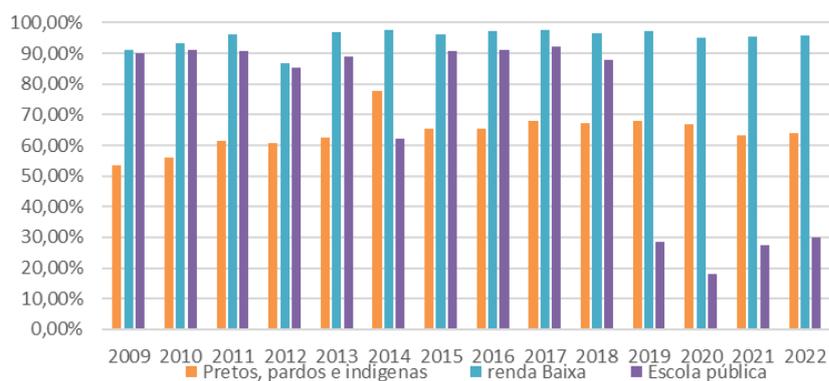


Figura 7 - Gráfico dos percentuais dos integrantes do Cluster 1 por característica, por ano. (Autoria própria)

Ao se analisar as figuras anteriores, quanto às características que diferenciam os grupos, observa-se uma disparidade quanto à questão racial. Estudantes que se autodeclaram pretos, pardos e indígenas, constituem, mais de 90% dos integrantes do grupo de menores notas (Figura

⁴ https://drive.google.com/drive/folders/1BNgY4_nllfok6b6xSxdE3WoGVaJIXNje?usp=drive_link

7), em todos os anos avaliados, exceto em 2014. Enquanto, no grupo de maiores notas (Figura 6) este percentual encontra-se, majoritariamente, abaixo dos 60%. A relação entre o desempenho no exame e a cor/raça dos candidatos se associam diretamente aos dados apresentados em (IBGE, 2018), especialmente quando se refere a taxa da frequência escolar da população preta ou parda entre 15 e 24 anos, decaindo de forma muito mais acentuada do que para a população branca. Além disso, como também exposto em (IBGE, 2018), o índice de analfabetismo é mais que o dobro na população preta e parda do que na população branca.

Analisando o aspecto se o candidato estudou em escola pública, em cada ano, há uma grande distinção entre os dois clusters. Desconsiderando os anos de 2019 a 2022, devido à pouca quantidade de alunos que responderam à pergunta, observa-se que o cluster de menores notas (Figura 7) apresenta uma variação entre os percentuais, mantendo-se entre 60% e 90%. No cluster de maiores notas (Figura 6) os percentuais variam entre 50 e 70%. Resultado semelhante foi encontrado por (Travitzki et al., 2016) e (Leoni & Sampaio, 2017), o qual apresenta como resultado a afirmação da existência do agrupamento natural entre os alunos por modalidade educacional frequentada.

Com relação à renda familiar per capita inferior ou igual a 1,5 salários-mínimos (legenda “renda baixa” na Figura 7), o grupo de melhor desempenho possui percentuais próximos a 80%, enquanto no outro grupo os percentuais superam 90%, sem variações significativas ao longo dos anos. Apesar de apresentar menores diferenças entre os percentuais entre os dois clusters, esta característica é significativa, tendo em vista que há uma relação entre a desigualdade educacional e a renda, como mostrado em (Helene & Mariano, 2020).

Para demonstrar a diferença da distribuição de estudantes aptos a concorrer a vagas de cotas em cada grupo encontrado pelo algoritmo, a Figura 8, expõem um gráfico com os valores em percentuais desta distribuição.

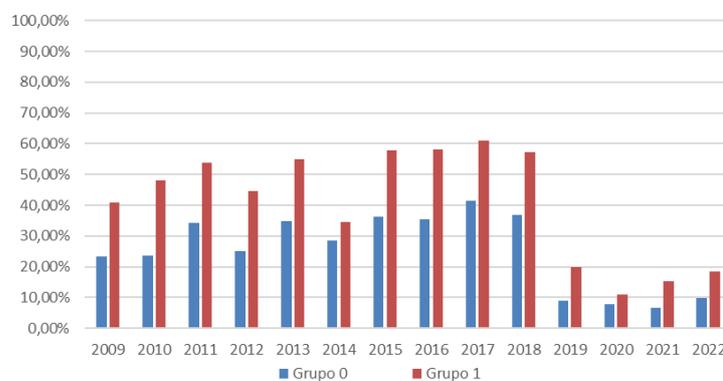


Figura 8 - Estudantes de escolas públicas, pretos pardos e indígenas e renda baixa. (Autoria própria)

Observa-se na Figura 8 o padrão de superioridade nos percentuais de alunos apitos as cotas no grupo de menores notas (grupo 1), em todos os anos analisados. É possível notar que na maioria dos anos, o percentual de alunos que compõem o grupo 1 é próximo ou superior a 50% e 20% maior que o valor correspondente ao grupo 0.

CONSIDERAÇÕES FINAIS

A Lei das Cotas, criada em 2012 para ser uma medida provisória para equiparar o acesso à educação entre os afetados negativamente pelas desigualdades existentes no país, tem o requerimento de revisão de seus aspectos construtivos ao marco dos seus 10 anos em seu próprio texto. Assim o presente trabalho propôs uma análise do perfil dos estudantes a partir dos microdados referentes ao Enem, ao longo de uma década, aplicados ao algoritmo de agrupamentos K-means, e os dados do Censo da Educação Superior 2020.

Os resultados apresentados demonstram que apesar do aumento significativo da população negra nos percentuais de matrículas no ensino superior, os estudantes oriundos do ensino médio público e que se autodeclaram pretos, pardos e indígenas, ou tenham renda familiar per capita inferior ou igual a 1,5 salários-mínimos, concentram-se, mais significativamente, nos grupos de menores desempenhos do Enem do que nos grupos de maiores, durante todo o período analisado. Dessa forma, entende-se a necessidade da permanência das cotas, a fim de fornecer oportunidade de acesso ao ensino superior para a população mais atingida pelas desigualdades.

REFERÊNCIAS

Anhaia, B. C. A Lei de Cotas e a universidade pública brasileira em debate. *Revista Contraponto*, Rio Grande do Sul. v.4, n.2, 2017. Disponível em: <<https://www.seer.ufrgs.br/contraponto/article/view/78915>>. Acesso em: 06 nov. 2023.

Brasil, Lei nº 12.711, de 29 de agosto de 2012. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12711.htm Acesso em: 06 nov. 2023

Helene, O., & Mariano, L. Education and inequality in income distribution. *Educacao e Sociedade*, 41, 2020. Disponível em: <https://doi.org/10.1590/es.223485>.

IBGE. *Desigualdades Sociais por Cor ou Raça no Brasil*. 2018.

IBGE. *Em 2021, rendimento domiciliar per capita cai ao menor nível desde 2012*. 2022.

INEP. Microdados - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

Jain, A. K. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666, 2010. Disponível em: <https://doi.org/10.1016/j.patrec.2009.09.011>

Klitzke, M. K. Exame nacional do ensino médio (Enem): uma tentativa de democratização do acesso ao ensino superior. *Res. Policy*, v. 9, n.2, p. 155–162, 2014. Disponível em: <<https://repositorio.ufsc.br/handle/123456789/196325>>. Acesso em: 04 nov 2023

Leoni, R. C., & Sampaio, N. A. de S. Desempenho das escolas públicas e privadas da região do vale do paraíba: uma aplicação da técnica de agrupamentos kmeans com base nas variáveis do Enem 2015. *Cadernos Do IME - Série Estatística*, 42(0), 2017. Disponível em: <https://doi.org/10.12957/cadest.2017.30347>

Lévy, P. *Programming Collective Intelligence Building Smart Web 2.0 Applications*, 1997.

Lloyd, S. P. Least Squares Quantization in PCM. In *IEEE TRANSACTIONS ON INFORMATION THEORY*. Vol. 28, 1982.

Matte, M. K. Impacto do Uso da Desigualdade Triangular para Acelerar o Algoritmo k-Means. 2020.

MEC. Enem - Ministério da Educação. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem>. Acesso em: 06 nov. 2023

MEC, “TRI - Ministério da Educação,” 2010. Disponível em: Acesso em: 06 nov. 2023

Rabelo, M. *Avaliação Educacional: fundamentos, metodologia e aplicações no contexto brasileiro*. Rio de Janeiro: SBM. 2013.

Travitzki, R., Ferrão, M. E., & Couto, A. P. Desigualdades educacionais e socioeconômicas na população Brasileira pré-universitária: Uma visão a partir da análise de dados do Enem. *Education Policy Analysis Archives*, 24, 2016. Disponível em: <https://doi.org/10.14507/epaa.24.2199>