

A importância das cotas do Sisu na redução das desigualdades no acesso ao ensino superior no Brasil: uma análise estatística dos dados do ENEM de 2012 a 2021

Ronaldo Costa da Silva ¹
Joelson da Cruz Campos ²

RESUMO

O Exame Nacional do Ensino Médio (ENEM) é uma das principais formas de ingresso no ensino superior no Brasil. O objetivo do Enem é avaliar o desempenho dos estudantes no final do ensino médio e avaliar a qualidade do sistema de educação brasileiro. Entretanto, sendo o Brasil um país de dimensões continentais é de se esperar que o sistema de ensino não seja igualitário, havendo locais com escolas melhores estruturadas do que outras, bem como pessoas com diferentes classes econômicas ou de diferentes grupos sociais. Visando tornar o acesso ao ensino superior igualitário, foi criado no Brasil o sistema de cotas para o Enem sendo este instituído pela Lei nº 12.711/2012, conhecida como Lei de Cotas, que determina que as instituições federais de ensino superior devem reservar, no mínimo, 50% das vagas de cada curso para estudantes que tenham cursado todo o ensino médio em escolas públicas. Dessas vagas, pelo menos metade devem ser destinadas a estudantes com renda familiar per capita igual ou inferior a um salário-mínimo e meio, e a outra metade destinada a estudantes de qualquer renda. Além disso, a Lei de Cotas prevê a reserva de vagas para pessoas com deficiência e para pessoas autodeclaradas pretas, pardas ou indígenas, seguindo as proporções no mínimo igual a sua presença na população. Essas cotas também são aplicadas em programas do governo federal, como o Programa Universidade para Todos (ProUni) e o Fundo de Financiamento Estudantil (FIES). No presente trabalho, realizamos uma análise estatística dos dados dos participantes do ENEM, abrangendo os exames de 2012 a 2021. Nosso objetivo foi investigar a evolução do desempenho dos participantes, comparando as notas dos cotistas e não cotistas. Nossos resultados reforçam a importância da existência de cotas, pois evidenciam as

¹Graduando do Curso de Estatística da Universidade Federal de Campina Grande - UFCG, ronaldo.costa@estudante.ufcg.edu.br;

²Professor orientador: Doutor, Universidade Federal de Campina Grande - UFCG, joelson.cruz@professor.ufcg.edu.br



desigualdades presentes no acesso ao ensino superior no Brasil.

INTRODUÇÃO

Considerada por Santos e Rocha (2007) como uma medida governamental, a palavra cotas na educação brasileira é vista como uma ação afirmativa, com a finalidade de criar reservas de vagas em instituições de ensino superior, públicas ou privadas com o objetivo de ampliar a inclusão social das classes menos favorecidas.

A Constituição Federal Brasileira de 1988 (BRASIL, 1988) em seu artigo 206, inciso I e a Lei de Diretrizes e Bases da Educação Brasileira Nacional (LDBEN) de 1996 (BRASIL, 1996), em seu artigo 3º, inciso I estabelece que o ensino seja ministrado respeitando-se o princípio da igualdade de condições para o acesso e permanência na escola. Em consonância com esse preceito constitucional, políticas de inclusão social passaram a ser formuladas e implementadas com o propósito de reduzir as desigualdades sociais que, ao longo dos séculos, têm predominado em nossa história, discriminando vários segmentos da sociedade. Neste contexto, inserem-se as ações afirmativas, que nada mais são do que políticas públicas ou medidas adotadas para promover a igualdade, reduzindo as injustiças sociais (Bayma, 2012).

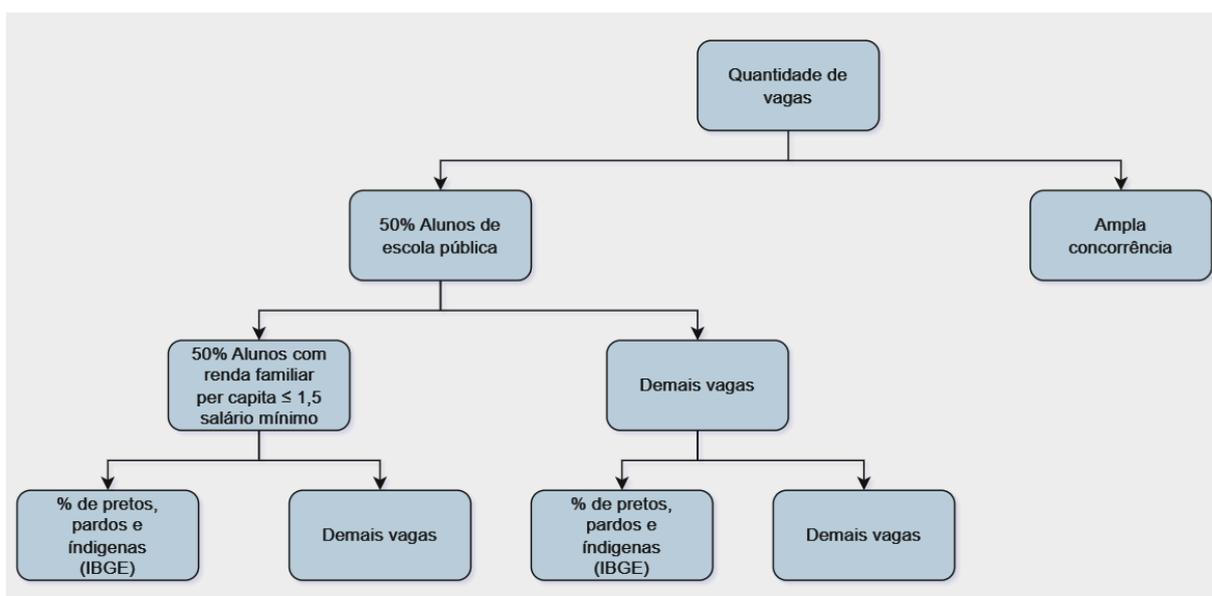
Essas políticas consistem em reservar vagas em universidades, concursos públicos e outras instituições para esses grupos, garantindo assim uma representação mais justa e equilibrada na sociedade. A escravidão, o racismo, o preconceito e a discriminação deixaram marcas profundas na sociedade, afetando negativamente a vida de muitas pessoas e impedindo o acesso a recursos e oportunidades. Logo essas políticas ajudam a corrigir essas desigualdades, promovendo uma sociedade mais justa e inclusiva. Visam restituir uma igualdade que foi rompida ou nunca existiu (Moehlecke, 2002)

Visando tornar o acesso ao ensino superior igualitário, a primeira proposta de cotas para negros em universidades públicas federais surgiu na UnB, em 1999. A proposta apresentada ao Conselho de Ensino e Pesquisa (CEPE), pelos professores José Jorge Carvalho e Rita Laura Segato (Carvalho & Segato, 1999), previa uma cota de 20% de vagas para estudantes negros. Em 2001, duas universidades já haviam adotado o sistema de cotas, a Universidade Estadual do Rio de Janeiro (UERJ) e a Universidade do Estado da Bahia (UNEB).

O Sistema de Seleção Unificada (SISU) é um programa do governo federal brasileiro

que utiliza as notas do Exame Nacional do Ensino Médio (ENEM) para selecionar estudantes para vagas em universidades públicas e institutos federais de educação superior em todo o país. No ano de 2012 foi criado no Brasil o sistema de cotas sendo este instituído pela Lei nº 12.711/2012, conhecida como Lei de Cotas, que determina que as instituições federais de ensino superior devem reservar, no mínimo, 50% das vagas de cada curso para estudantes que tenham cursado todo o ensino médio em escolas públicas. Dentro dessa categoria, há subdivisões para candidatos autodeclarados pretos, pardos e indígenas, assim como candidatos em situação de vulnerabilidade econômica, como podemos observar na Figura 1, onde o participante terá de escolher uma opção no momento da inscrição no SISU.

Figura 1: Fluxograma da distribuição de vagas do SisU



Fonte: Autoria própria

Existe também uma porcentagem de vagas destinadas a pessoas com deficiência, que são comprovadas mediante laudo médico, onde o objetivo é aumentar a inclusão de estudantes pertencentes a esses grupos.

Além disso, as cotas no SISU têm um efeito positivo na promoção da diversidade. A diversidade é importante porque enriquece as relações humanas e promove a troca de experiências e perspectivas diferentes. Ela possibilita a compreensão e a valorização das

diferentes culturas, tradições e modos de vida, e contribui para a construção de uma sociedade mais tolerante e democrática, promovendo a igualdade de oportunidades e garantindo o direito à educação, que corresponde a um direito fundamental garantido pela Constituição brasileira. Todas as pessoas têm direito a uma educação de qualidade, desde a educação infantil até o ensino superior, sem qualquer tipo de discriminação. O estado é o responsável por garantir o acesso à educação e por assegurar que a educação seja de qualidade e esteja disponível para todos, independentemente de sua origem social, raça, gênero ou condição financeira.

No entanto, as cotas também são controversas, e há críticas que questionam sua eficácia e legalidade. Alguns argumentam que as cotas podem criar novas formas de discriminação, favorecendo alguns grupos em detrimento de outros, e que a seleção baseada em critérios raciais ou étnicos é contrária aos princípios de igualdade e mérito. Além disso, algumas pessoas argumentam que as cotas podem gerar ressentimento e conflitos entre grupos, em vez de promover a união e a igualdade.

Este trabalho tem como pretensão realizar uma análise descritiva e inferencial dos dados do ENEM do ano 2012 até o 2021 no intuito de fazer um comparativo na qualidade do ensino fornecido por escolas públicas e privadas e grupos que fazem ou não o uso das cotas.

METODOLOGIA (OU MATERIAIS E MÉTODOS)

Esse trabalho foi desenvolvido através de uma iniciação científica dentro do programa de educação Tutorial (PET) de Matemática e Estatística da Universidade Federal de Campina Grande (UFCG) ao qual deixamos registrado nossos agradecimentos. Neste trabalho realizamos uma análise exploratória dos dados do Exame Nacional do Ensino Médio (ENEM) disponíveis no site do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) e fizemos uso dos microdados do ENEM de 2012 até 2021. A pesquisa utilizou os dados dos participantes que já haviam concluído o ensino médio e concluintes no ano corrente que preencheram o formulário socioeconômico, informando dados como o tipo de escola, situação de conclusão do ensino médio, entre outras informações. Além disso, a seleção limitou-se para participantes que estudaram em território nacional. Para realização da análise foi feito uso do *software R* como recurso computacional.

REFERENCIAL TEÓRICO

ENEM - Exame Nacional do Ensino Médio

O Exame Nacional do Ensino Médio é uma prova elaborada pelo Ministério da Educação do Brasil com o objetivo de avaliar o desempenho dos estudantes ao final do ensino médio e serve também como critério de acesso ao ensino superior em instituições públicas e privadas, através do Sistema de Seleção Unificada (Sisu), Programa Universidade para Todos (ProUni) e Fundo de Financiamento Estudantil (Fies). A prova é realizada anualmente em dois domingos subsequentes e é composta por 180 questões objetivas de múltipla escolha, distribuídas em quatro áreas do conhecimento: Linguagens, Códigos e suas Tecnologias; Ciências Humanas e suas Tecnologias; Ciências da Natureza e suas Tecnologias; e Matemática e suas Tecnologias. Além disso, a prova também inclui uma redação. No primeiro dia o exame tem duração de 5 horas e 30 minutos e no segundo dia tem duração de 5 horas. Entre 2020 e 2023, foi implementado o Enem Digital. Nesse período, tiveram provas impressas e digitais, em locais de aplicação pré-determinadas.

Linguagem escolhida para a análise

Foi selecionada a linguagem de programação R 4.1.2, juntamente com a IDE RStudio, para escrever o código que foi utilizado para fazer as análises descritivas e inferenciais. O R é uma linguagem de programação e ambiente de desenvolvimento muito utilizado na área de ciência de dados, estatística e análise de dados. Foi criada por Ross Ihaka e Robert Gentleman na Universidade de Auckland, Nova Zelândia, durante a década de 1990.

O R é uma linguagem interpretada, ou seja, o código fonte é executado diretamente, linha por linha, e possui uma sintaxe orientada a objetos. Ela se destaca pela sua extensa coleção de pacotes e bibliotecas, que fornecem funcionalidades adicionais para manipulação, visualização e análise de dados. Essa vasta biblioteca de pacotes faz com que o R seja uma das linguagens preferidas por cientistas de dados e estatísticos, pois oferece uma ampla gama de recursos e ferramentas específicas para essas áreas. Uma das características distintivas do R é a sua ênfase na análise estatística e construção de gráficos. Ela fornece uma ampla variedade

de funções estatísticas e métodos para a manipulação e transformação de dados, bem como para a criação de visualizações gráficas de alta qualidade. Outra vantagem do R é a sua comunidade ativa de usuários e desenvolvedores. Existem inúmeros recursos, fóruns e tutoriais disponíveis para ajudar os usuários a aprender e solucionar problemas na linguagem. A comunidade também contribui constantemente com novos pacotes e atualizações, enriquecendo ainda mais o ecossistema do R.

R é uma escolha popular para análise de dados em vários campos, incluindo ciências sociais, biologia, finanças, marketing e muitos outros. Seu uso se estende desde tarefas básicas de manipulação de dados até análises estatísticas complexas e modelagem preditiva. Além disso, possui integração com outras linguagens, como Python, o que permite a combinação de diferentes ferramentas e recursos.

Está disponível como *Software* Livre sob os termos da Licença Pública Geral GNU da Free Software Foundation em forma de código-fonte. Ele está disponível em uma ampla variedade de plataformas UNIX e sistemas similares (incluindo FreeBSD e Linux), Windows e MacOS.

Lógica da construção do script

Inicialmente foi realizado o download dos microdados do Enem do ano 2012 até o ano 2021 no site <https://www.gov.br/inep/pt-br>, em que o Inep disponibiliza de forma aberta, esses dados são o menor nível de desagregação de dados recolhidos por meio do exame e atendem à demanda por informações específicas ao disponibilizar as provas, os gabaritos, as informações sobre os itens, as notas e o questionário respondido pelos inscritos no Enem.

Em seguida, o *script* realiza a importação e leitura da base de dados, onde apenas as variáveis de interesse que são "NU_INSCRICAO", "TP_ESCOLA", "TP_ST_CONCLUSAO", "NU_NOTA_CN", "NU_NOTA_CH", "NU_NOTA_LC", "NU_NOTA_MT", "NU_NOTA_REDACAO", que respectivamente significam o número da inscrição, o tipo de escola, a situação de conclusão do ensino médio, a nota de ciências da natureza e suas tecnologias, a nota de ciências humanas e suas tecnologias, a nota de linguagens, códigos e suas tecnologias, a nota de matemática e suas tecnologias e a nota de redação estarão na base de dados.

Figura 2: Carregamento dos pacotes e importação dos dados no R (Exemplo com o ano 2013)

```
2 # Análise de dados ENEM 2013
3
4
5 library(data.table) # Fornece uma versão aprimorada de data.frames
6 library(tidyverse) # Metapacote
7
8 #***** Importação dos dados *****
9
10 notas_13 <- fread("MICRODADOS_ENEM_2013.csv",
11                  selec = c('NU_INSCRICAO', 'TP_ESCOLA', 'TP_ST_CONCLUSAO',
12                            'NU_NOTA_CN', 'NU_NOTA_CH', 'NU_NOTA_LC',
13                            'NU_NOTA_MT', 'NU_NOTA_REDACAO')) %>%
14                  rename("inscricao" = "NU_INSCRICAO",
15                          "tipo_escola" = "TP_ESCOLA",
16                          "sit_conclusao" = "TP_ST_CONCLUSAO",
17                          "ciencias_natureza" = "NU_NOTA_CN",
18                          "ciencias_humanas" = "NU_NOTA_CH",
19                          "linguagens_codigos" = "NU_NOTA_LC",
20                          "matematica" = "NU_NOTA_MT",
21                          "redacao" = "NU_NOTA_REDACAO")
```

Fonte: Autoria própria

A manipulação por sua vez tem como objetivo filtrar apenas os participantes que, preencheram no formulário de inscrição sua cor, o tipo de escola que estudou e a situação de conclusão do ensino médio, para então partir para o cálculo das notas médias de cada disciplina, separada por participantes que têm o direito de usar as cotas e os demais participantes.

Foram utilizados os seguintes pacotes na importação e manipulação dos dados:

- O "data.table" onde é uma biblioteca que oferece uma alternativa eficiente e rápida para a manipulação de conjuntos de dados grandes. Ele é projetado para trabalhar com tabelas de dados em memória e é particularmente útil para lidar com conjuntos de dados extensos.

Uma das principais características do pacote é a sua velocidade, pois ele utiliza uma sintaxe semelhante à do DataFrame (O DataFrame é uma estrutura de dados tabular bidimensional que organiza os dados em colunas e linhas, semelhante a uma tabela ou planilha.) padrão do R, mas é otimizado para oferecer desempenho superior em operações de manipulação de dados.

- O tidyverse que corresponde a um conjunto de pacotes que foram desenvolvidos para facilitar a manipulação, visualização e análise de dados de maneira consistente e eficiente. Ele é baseado em uma filosofia de "tidy data" (dados organizados), que

ênfatisa a padronização da estrutura dos conjuntos de dados para torná-los mais fáceis de trabalhar.

Figura 3: Manipulação dos dados e cálculos das notas médias no R

```
3
4  ##***** Manipulação e calculo da média*****
5
6  dados_filtrados_11 <- dados_2013 %>%
7    filter(tipo_escola == "Publica",
8          salario == "até 1,5 salario",
9          cor %in% c("Preta", "Parda", "Indígena"))
10
11
12 # Calcula as médias das linhas e cria um novo vetor
13 medias_linhas <- rowMeans(dados_filtrados_11[,5:9], na.rm = TRUE)
14
15 # Adiciona o novo vetor de médias ao DataFrame
16 dados_filtrados_11$media <- medias_linhas
17
18 dados_filtrados_11 <- na.omit(dados_filtrados_11)
19
```

Fonte: Autoria própria

Teste qui-quadrado de independência

De acordo com Bolfarine e Sandoval (2001) o teste qui-quadrado de independência é um teste estatístico usado para determinar se existe uma associação estatisticamente significativa entre duas variáveis categóricas. Ele é aplicado quando queremos verificar se a ocorrência de uma variável está relacionada à ocorrência de outra variável.

Baseia-se na comparação das frequências observadas em cada categoria das variáveis com as frequências esperadas, sob a suposição de que as variáveis são independentes uma da outra. A estatística de teste qui-quadrado de independência é calculada a partir da diferença entre as frequências observadas e esperadas, levando em consideração todas as categorias das variáveis.

Sua utilização nessa aplicação é dada da seguinte forma, inicialmente é feita a formulação da hipótese nula (H_0) em que dizemos que não existe uma associação significativa entre as variáveis, ou seja, afirmamos que o desempenho acadêmico independe do participante ser ou não cotista e a hipótese alternativa (H_a) em que dizemos que existe uma associação

significativa entre as variáveis, isto é, afirmamos que existe relação de dependência entre o desempenho acadêmico e o participante ser ou não cotista, em seguida é feita a construção da tabela de contingência onde é criada uma tabela cruzada para contar as frequências observadas de cada combinação de categorias das duas variáveis. Para assim utilizarmos os totais de cada categoria para realizarmos o cálculo das frequências esperadas, com base na hipótese nula, em que, assumimos que não há diferença significativa ou associação entre as variáveis em estudo. Em outras palavras, as frequências observadas nas categorias das variáveis são iguais às frequências esperadas, de acordo com a hipótese nula. Calculamos as frequências esperadas para cada célula da tabela de contingência. Por fim, calculamos a estatística de teste qui-quadrado de independência a partir das frequências observadas e esperadas.

A estatística de teste qui-quadrado de independência segue uma distribuição qui-quadrado. Com base na estatística de teste, é calculado o valor p , que indica a probabilidade de obter uma estatística de teste igual ou mais extrema do que a observada, sob a hipótese nula.

A tomada de decisão se dá com base no valor p , a hipótese nula é rejeitada se o valor p for menor que um nível de significância predefinido. Caso contrário, a hipótese nula não é rejeitada.

O teste qui-quadrado de independência é amplamente utilizado em diversas áreas, como ciências sociais, saúde, marketing e pesquisas de opinião, para investigar associações entre variáveis categóricas. Ele fornece uma medida estatística objetiva para determinar se as duas variáveis estão independentes ou se há alguma relação significativa entre elas.

RESULTADOS E DISCUSSÃO

Inicialmente fizemos uma nova coluna no conjunto de dados, onde essa é composta pela média das quatro áreas do conhecimento e redação, para então organizarmos em um rol do qual utilizamos para fazer a divisão em 3 categorias. Na escala de avaliação, classificamos as notas como baixas, médias e altas com base nos percentis. As notas no intervalo da nota mais baixa até o percentil 33 são categorizadas como baixas, aquelas que estão entre o percentil 33 e o percentil 66 são consideradas médias, e notas no intervalo do percentil superior ao 66 são classificadas como altas, por conseguinte realizamos a contagem de quantos participantes existem em cada categoria com base no seu direito em ser cotista e não ser cotista.

A sequência citada anteriormente é fundamental para a utilização do teste qui-quadrado de independência em questão.

O teste inicial teve como objetivo determinar se as notas médias dos participantes são afetadas pelo tipo de escola que frequentaram, em outras palavras, buscamos averiguar se o desempenho dos participantes que frequentaram escolas públicas era similar ao dos que frequentaram escolas privadas ou públicas (ampla concorrência), do qual concluímos que em todos os anos analisados existiam diferenças significativas entre o desempenho das categorias, em termos simples, as evidências sugerem que o desempenho depende de ser ou não cotista.

Logo se o participante é cotista realizamos outros testes com o objetivo de saber se as subcategorias do fluxograma da Figura 1 fazem sentido, iniciando o primeiro teste com participantes que estudaram em escolas públicas e participantes que estudaram em escolas públicas e possuem situação de vulnerabilidade econômica, em outros termos, determinar se as notas médias dos participantes cotistas que estudam em escolas públicas são afetadas ou não pela renda, do qual concluímos que em todos os anos analisados existiam diferenças significativas entre o desempenho dessas duas subcategorias, ou seja, as evidências sugerem que o desempenho depende se o participante estudou em escola pública ou estudou em escolas públicas e tem situação de vulnerabilidade econômica.

Em seguida realizamos o segundo teste entre as subcategorias das cotas, onde neste, o intuito foi verificar se dentro das subcategorias de estudar em escolas públicas e possuir situação de vulnerabilidade econômica a cor seria um fator dependente de diferença de desempenho, onde concluímos que existe uma relação de dependência deste fator.

Para finalizar realizamos o último teste qui-quadrado entre as subcategorias das cotas, onde neste, o intuito foi verificar se dentro das subcategorias de estudar em escolas públicas a cor seria um fator dependente de diferença de desempenho, onde concluímos que existe uma relação de dependência deste fator.

Realizamos também uma análise descritiva onde observamos aspectos relevantes como por exemplo que anualmente as notas médias dos cotistas são sempre menores do que a dos não cotistas, outra coisa importante é que as médias das disciplinas geralmente oscilam com o passar dos anos, porém as notas médias gerais têm uma tendência crescente com o passar dos anos.



CONSIDERAÇÕES FINAIS

Com base nos resultados apresentados, torna-se evidente a importância da manutenção de todos os grupos de cotas estabelecidos pela Lei de Cotas como um fator crucial para garantir a inclusão de seus membros no acesso ao ensino superior. Além disso, observamos que ao longo do tempo houve uma melhoria no desempenho geral dos participantes, embora o desempenho em áreas específicas tenha oscilado, mostrando melhorias em alguns anos e declínios em outros.

Portanto, torna-se imperativo direcionar investimentos adicionais para escolas e colégios públicos, especialmente em áreas como infraestrutura, capacitação de professores e adoção de tecnologias educacionais, entre outros fatores. Isso é essencial para mitigar as flutuações periódicas de desempenho em áreas específicas, como linguagens, códigos e suas tecnologias, bem como ciências humanas e suas tecnologias.

Palavras-chave: ENEM, Cotas, Estatística, Teste qui-quadrado.

REFERÊNCIAS

BAYMA, F. **Reflexões sobre a constitucionalidade das cotas raciais em universidades públicas no Brasil: referências internacionais e os desafios pós-julgamento das cotas.** aval. pol. públ. Educ., Rio de Janeiro, v. 20, n. 75, p. 325-346, abr./jun. 2012.

BRASIL. **Constituição da república federativa do Brasil de 1988.** Disponível em: <http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm> Acesso: 13 jul. 2023.

BRASIL. Lei nº 12.711, de 29 de agosto de 2012. **Dispõe sobre o ingresso nas universidades federais e nas instituições federais de ensino técnico de nível médio e dá outras providências.**

BRASIL. **Ministério da Educação.** Disponível em: <<https://www.gov.br/pt-br/servicos/fazer-o-exame-nacional-do-ensino-medio>>. Acesso em: 21 jul. 2023.



BOLFARINE, H.; SANDOVAL, M. C. **Introdução à Inferência Estatística**. Rio de Janeiro: SBM - Sociedade Brasileira de Matemática, 2001.

CARVALHO, J.J.; SEGATO, R.L. **Proposta para implementação de um sistema de cotas para minorias raciais na Universidade de Brasília**. Brasília, DF, 1999.

Diário Oficial da União, Brasília, 30 ago. 2012. Seção 1, p. 1. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/112711.htm>. Acesso em: 15 de julho de 2023.

DOWLE, M. et al. **data.table: Extension of 'data.frame'**. Versão: 1.14.8. Disponível em: <<https://cran.r-project.org/web/packages/data.table/index.html>>. Acesso em: 21 jul. 2023.

MOEHLECKE, S. **Ação afirmativa: história e debates no Brasil**, Cadernos de Pesquisa, São Paulo, n. 117, p. 197-217, nov. 2002.

R Core Team. (2019). **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing. Disponível em: <<https://www.R-project.org/>>. Acesso em: 21 jul. 2023.

SANTOS, I. ; ROCHA, J. G. **Diversidade & ações afirmativas**. Rio de Janeiro: CEAP, 2007.

WICKHAM, H. (2022). **tidyverse: Easily Install and Load the 'Tidyverse'**. Versão: 2.0.0. Disponível em: <<https://cran.r-project.org/web/packages/tidyverse/index.html>>. Acesso em: 21 jul. 2023.