



Thays Conceição da Silva

**Efeitos das Políticas de Cotas no Ensino
Superior Público Brasileiro: Uma Análise do
Acesso, Perfil Institucional e Distribuição das
Vagas Reservadas com Técnicas de
Agrupamento**

Recife

2026

Thays Conceição da Silva

Efeitos das Políticas de Cotas no Ensino Superior Público Brasileiro: Uma Análise do Acesso, Perfil Institucional e Distribuição das Vagas Reservadas com Técnicas de Agrupamento

Monografia apresentada ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Universidade Federal Rural de Pernambuco – UFRPE

Departamento de Estatística e Informática

Curso de Bacharelado em Sistemas de Informação

Orientador: Gabriel Alves

Recife

2026

Agradecimentos

Agradeço primeiramente a Deus e à minha família: meu pai, Antonio Rocha, minha mãe, Francisca Conceição e minha irmã, Thianny Cristina, por me proporcionarem educação e todos os recursos necessários para que eu pudesse cursar a universidade sem grandes dificuldades, permitindo que eu chegasse até este momento de conclusão de um curso que tanto almejei e pelo qual tanto me esforcei. Nada disso teria sido possível sem o apoio emocional e financeiro dos meus pais, e espero um dia conseguir retribuir, ainda que minimamente, tudo o que sempre fizeram por mim.

Agradeço também ao meu orientador, Gabriel Alves, que me incentivou a adotar um tema tão importante quanto as cotas e me guiou no desenvolvimento de um trabalho com sentido e coerência.

Aos meus amigos: Larissa, Eva, Ewellyn, Alexandro, Beatriz, Davi, Gustavo e Igor, que ao longo desse tempo me apoiaram e foram meus companheiros na jornada da vida, tornando esse período mais leve e divertido, e me acolhendo sempre com carinho.

Agradeço também a todos os professores que tive ao longo da minha trajetória, que sempre me incentivaram a continuar estudando e a me dedicar aos estudos, servindo de inspiração para mim. E a todas as outras pessoas que passaram pela minha vida nesse percurso: cada uma foi importante de algum modo, pois é também nesses encontros que minha história foi sendo construída.

Um agradecimento mais que especial ao meu professor de programação do ensino médio, Anderson Paulo da Silva, que foi meu primeiro mentor e um guia que sempre acreditou no meu potencial como programadora. Seu incentivo foi decisivo para que eu me desenvolvesse e escolhesse seguir nesta área. Você foi um professor e um amigo extraordinário, e espero que esteja descansando na paz que sempre mereceu.

Resumo

As políticas de cotas no ensino superior brasileiro, instituídas pela Lei nº 12.711/2012, ampliaram o acesso de grupos historicamente excluídos à graduação. Contudo, sua efetividade depende não apenas do ingresso, mas também da distribuição dos beneficiários entre cursos e instituições e da garantia de permanência desses estudantes. Este trabalho inicia com uma revisão histórica do acesso da população negra à graduação no país, a fim de compreender as origens das desigualdades estruturais de raça e classe, e tem como objetivo analisar os efeitos da reserva de vagas na rede pública por meio da integração das bases oficiais ENADE, Censo da Educação Superior (CES) e Censo Demográfico, aplicando técnicas de agrupamento para identificar padrões no perfil dos ingressantes, no preenchimento das vagas e em sua evolução temporal entre 2010 e 2023. Para isso, adotaram-se os algoritmos K-Means e de agrupamento hierárquico (Ward), com o número de grupos determinado pelo método do cotovelo e pelo índice de silhueta, e validação por meio do modelo de mistura gaussiana (GMM). Os resultados evidenciaram a presença do fenômeno *creamy layer*: os cotistas em cursos de maior retorno econômico apresentaram menor proporção de pretos, pardos e indígenas e de estudantes de escola pública do que os concentrados em cursos de menor retorno. No nível institucional, as universidades federais concentraram maior inclusão racial, porém em áreas de menor valorização, enquanto as estaduais apresentaram alta adesão numérica sem equidade racial proporcional. A dimensão temporal mostrou avanços expressivos após a implementação da Lei, mas sem alteração estrutural na distribuição dos cotistas entre áreas ao longo do período. Por fim, a contextualização com o Censo Demográfico de 2022 confirmou que as desigualdades observadas nas universidades reproduzem, em escala populacional, a estratificação racial por área de formação. Conclui-se que, embora a política tenha ampliado o acesso, o *creamy layer* e a concentração dos beneficiários em cursos de menor retorno contribuem para a persistência da estratificação horizontal, indicando que o ingresso, isoladamente, não é suficiente para romper o ciclo de reprodução das desigualdades estruturais da sociedade brasileira.

Palavras-chave: Ações afirmativas. Lei de Cotas. Ensino superior. Análise de agrupamento. Estratificação horizontal.

Abstract

The affirmative action policies in Brazilian higher education, established by Law No. 12,711/2012, expanded access to undergraduate education for historically excluded groups. However, their effectiveness depends not only on admission, but also on the distribution of beneficiaries across programs and institutions and on ensuring these students' retention. This work begins with a historical review of the Black population's access to undergraduate education in the country, in order to understand the origins of the structural inequalities of race and class, and aims to analyze the effects of the quota system in the public network through the integration of the official databases ENADE, the Higher Education Census (CES), and the Demographic Census, applying clustering techniques to identify patterns in the profile of incoming students, in the filling of reserved seats, and in their temporal evolution between 2010 and 2023. For this purpose, the K-Means and hierarchical clustering (Ward) algorithms were adopted, with the number of groups determined by the elbow method and the silhouette score, and validation through the Gaussian Mixture Model (GMM). The results revealed the presence of the *creamy layer* phenomenon: quota students in programs with higher economic returns showed a lower proportion of Black, mixed-race, and Indigenous students and of public-school graduates than those concentrated in programs with lower returns. At the institutional level, federal universities concentrated greater racial inclusion, but in less valued fields, while state universities showed high numerical adherence without proportional racial equity. The temporal dimension showed significant advances after the Law's implementation, but with no structural change in the distribution of quota students across fields throughout the period. Finally, the contextualization with the 2022 Demographic Census confirmed that the inequalities observed within universities reproduce, at the population scale, racial stratification by field of study. It is concluded that, although the policy expanded access, the *creamy layer* and the concentration of beneficiaries in programs with lower returns contribute to the persistence of horizontal stratification, indicating that admission alone is not sufficient to break the cycle of reproduction of the structural inequalities in Brazilian society.

Keywords: Affirmative action. Quota Law. Higher education. Cluster analysis. Horizontal stratification.

Lista de ilustrações

Figura 1 – Percentual de ingressantes pretos, pardos e indígenas por tipo de IES (2010–2023).	38
Figura 2 – Funil de inclusão: total vs cotistas por tipo de IES pública (2023). . .	38
Figura 3 – Método do cotovelo e índice de silhueta.	40
Figura 4 – Perfil normalizado dos clusters.	42
Figura 5 – Método do cotovelo e índice de silhueta.	46
Figura 6 – Dendrograma — agrupamento hierárquico Ward.	48
Figura 7 – Evolução dos indicadores por período	51
Figura 8 – Dendrograma dos períodos.	52
Figura 9 – Dendrograma por ano. Cores indicam o período de cada ano. . . .	53
Figura 10 – Evolução dos indicadores por tipo de IES e período.	55
Figura 11 – Variação dos indicadores entre períodos	56

Lista de tabelas

Tabela 1 – Áreas avaliadas no ENADE por ciclo trienal (2010–2023)	33
Tabela 2 – Adaptações realizadas em variáveis de informações da IES e do curso	34
Tabela 3 – Caracterização dos clusters	40
Tabela 4 – Comparação entre K-Means e GMM	43
Tabela 5 – Caracterização dos clusters	46
Tabela 6 – Distribuição por tipo de IES por cluster (%)	47
Tabela 7 – Comparação entre K-Means e Ward	48
Tabela 8 – Indicadores médios por período (CES, 2010–2023)	51
Tabela 9 – Classificação anual – K-Means e Ward	54

Lista de abreviaturas e siglas

CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CES	Censo da Educação Superior
CINE-F	Classificação Internacional Normalizada da Educação
EaD	Educação a Distância
ENADE	Exame Nacional de Desempenho dos Estudantes
ERE	Ensino Remoto Emergencial
GMM	Gaussian Mixture Models
IA	Inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
IES	Instituição de Ensino Superior
IFES	Instituições Federais de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
KDD	Knowledge Discovery in Databases
LGPD	Lei Geral de Proteção de Dados Pessoais
MEC	Ministério da Educação
ML	Machine Learning
PCA	Principal Component Analysis
PCD	Pessoa com Deficiência
PNAD	Pesquisa Nacional por Amostra de Domicílios
PNAES	Programa Nacional de Assistência Estudantil
PPI	Pretos, Pardos e Indígenas
RAIS	Relação Anual de Informações Sociais

REUNI	Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais
UERJ	Universidade do Estado do Rio de Janeiro
UENF	Universidade Estadual do Norte Fluminense
UnB	Universidade de Brasília
UNEB	Universidade do Estado da Bahia

Sumário

	Lista de ilustrações	5
1	INTRODUÇÃO	11
1.1	Justificativa e Objetivos da Pesquisa	12
2	REFERENCIAL TEÓRICO	14
2.1	Panorama Geral e Fundamentos das Ações Afirmativas	14
2.2	Inteligência Artificial e <i>Machine Learning</i>	15
2.2.1	Descoberta de Conhecimento em Bases de Dados (KDD)	16
2.2.2	Técnicas de Agrupamento	17
3	REVISÃO DA LITERATURA	19
4	METODOLOGIA	24
4.1	Base de Dados Utilizadas	24
4.2	Divisão Temporal	25
4.3	Ferramentas e Recursos Computacionais	25
4.4	Análises de Agrupamento	26
4.4.1	QP1 – Perfil dos Cotistas no Nível Institucional	26
4.4.2	QP2 – Padrão Institucional de Preenchimento de Vagas	28
4.4.3	QP3 – Evolução Temporal entre os Períodos	30
4.4.4	QP4 – Contextualização com o Censo Demográfico	31
5	BASE DE DADOS E PREPARAÇÃO	32
5.1	Visão Geral das Bases e Integração	32
5.2	ENADE	32
5.2.1	Estrutura e Variáveis Utilizadas	32
5.2.2	Harmonização e Preparação	33
5.3	CES	34
5.3.1	Estrutura e Variáveis Utilizadas	34
5.3.2	Harmonização e Preparação	35
5.4	Censo Demográfico	36
5.4.1	Tabelas Selecionadas e Recorte	36
5.4.2	Harmonização e Preparação	36
6	ANÁLISE DE AGRUPAMENTO E DISCUSSÃO DOS RESULTADOS	37
6.1	Caracterização Geral dos Dados	37

6.2	QP1 – Perfil dos Cotistas no Nível Institucional	38
6.2.1	Pré-processamento e definição dos clusters	39
6.2.2	Caracterização dos clusters	40
6.2.3	Validação por GMM	42
6.2.4	Discussão – <i>creamy layer</i> no nível institucional	43
6.3	QP2 – Padrão Institucional de Preenchimento de Vagas	45
6.3.1	Pré-processamento e definição dos clusters	45
6.3.2	Caracterização dos clusters institucionais	46
6.3.3	Validação por agrupamento hierárquico (Ward)	47
6.3.4	Discussão – <i>undermatching</i> e padrões institucionais	49
6.4	QP3 – Evolução Temporal entre os Períodos	50
6.4.1	Pré-processamento por período	50
6.4.2	Análise hierárquica e dendrograma	52
6.4.3	Validação por K-Means	54
6.4.4	Discussão – efeitos da Lei e da pandemia	54
6.5	QP4 – Contextualização com o Censo Demográfico	56
7	CONSIDERAÇÕES FINAIS	59
7.1	Limitações	61
7.2	Trabalhos Futuros	61
	REFERÊNCIAS	63

1 Introdução

Para compreender como o Brasil chegou ao ponto de necessitar de políticas de cotas e outras ações afirmativas voltadas para a população negra e pobre, é necessário olhar para o passado e entender os fatores que consolidaram as desigualdades estruturais de raça e classe no país.

O processo de colonização da América trouxe consigo práticas opressoras, que resultaram na exploração e escravização de povos indígenas brasileiros e de africanos trazidos forçadamente para o Brasil. A mão de obra escravizada gerou benefícios econômicos à Coroa Portuguesa e permitiu a grandes proprietários de terra e senhores de escravos acumular riqueza e poder entre os séculos XVI e XIX (PEREIRA; ZIENTARSKI, 2011). Essa estrutura não se encerrou com o fim da escravização: a Lei Áurea, promulgada em 1888, decorreu sobretudo de pressões econômicas internacionais que buscavam ampliar o mercado consumidor interno (BAYMA, 2012).

A abolição, contudo, não foi acompanhada de políticas de integração social. Sem apoio estatal, a população recém-liberta permaneceu marginalizada, sem acesso à educação, ao trabalho e à moradia, o que consolidou desigualdades sociais e econômicas que perduram até hoje, principalmente entre a população negra e parda (PIOVESAN, 2008). Somente nas últimas décadas o Estado começou a reconhecer essa realidade e a elaborar medidas de reparação voltadas à inclusão social e econômica (SILVA, 2018). É nesse cenário que as ações afirmativas surgem como instrumento estratégico para enfrentar desigualdades históricas no acesso educacional, promovendo a inclusão de grupos historicamente marginalizados, especialmente a população negra, em espaços predominantemente brancos.

As ações afirmativas podem ser definidas como “um conjunto de políticas públicas e privadas de caráter compulsório, facultativo ou voluntário, concebidas com vistas ao combate da discriminação de raça, gênero etc., bem como para corrigir os efeitos presentes da discriminação praticada no passado” (GOMES, 2001 apud DOMINGUES, 2005). Seu propósito central é ampliar o acesso de populações historicamente marginalizadas a espaços de formação e prestígio, tradicionalmente reservados a uma elite social e racial (GOMES, 2005). Trata-se de medidas temporárias, necessárias enquanto não se alcança a equidade educacional, e cuja efetividade depende de monitoramento sistemático, sob risco de se atribuir aos próprios beneficiários a responsabilidade por eventuais falhas da política (SENKEVICS, 2018).

Para que tais políticas se tornem desnecessárias no futuro, é fundamental que sejam acompanhadas de investimentos estruturais na educação básica de qualidade e de

mecanismos que assegurem a permanência e o sucesso acadêmico dos estudantes, já que o ingresso na universidade não implica, por si só, continuidade ou bom desempenho (SILVA, 2018). No Brasil, onde a educação superior oferece um dos maiores retornos econômicos do mundo (OLIVEIRA, 2021), e em um cenário de alta desigualdade, o acesso equitativo à universidade torna-se ainda mais urgente como fator de mobilidade social (ROZA, 2023).

1.1 Justificativa e Objetivos da Pesquisa

A implementação da Lei nº 12.711/2012 transformou o perfil de ingresso nas universidades brasileiras, ampliando o acesso de grupos historicamente sub-representados e alterando a composição do corpo discente ao longo da última década (SENKEVICS; MELLO, 2019). Essa transformação reafirma a importância de análises empíricas que mensurem a efetividade da política para além do aumento numérico de matrículas, avaliando a qualidade e a distribuição do acesso obtido.

Ainda são poucos os estudos que analisam de forma integrada os efeitos das ações afirmativas combinando múltiplas bases de dados públicas e aplicando técnicas de agrupamento para identificar padrões institucionais no acesso às cotas. Entre esses padrões, destacam-se dois fenômenos centrais para avaliar a efetividade da política: o *creamy layer*, que designa a tendência de os benefícios das cotas se concentrarem nos estratos menos vulneráveis dentro dos grupos elegíveis (JÚNIOR; DAFLON, 2015b); e o *undermatching*, que descreve a tendência de cotistas acessarem cursos de menor prestígio do que aqueles para os quais seriam elegíveis (MEDEIROS; NETO; GOMES, 2016).

As pesquisas existentes tendem a se concentrar nos dados do Exame Nacional de Desempenho dos Estudantes (ENADE) (INEP, 2023b), com foco na representação e na comparação do desempenho acadêmico entre cotistas e não cotistas, sem correlacioná-los com indicadores institucionais agregados ou com a evolução temporal do preenchimento das vagas reservadas. Há, portanto, espaço relevante para análises que combinem diferentes bases públicas — o ENADE, o Censo da Educação Superior (CES) (INEP, 2023a) e o Censo Demográfico (IBGE, 2022) — e que cubram períodos ainda pouco explorados, como o contexto pandêmico e pós-pandêmico, em que a suspensão das atividades presenciais pode ter afetado desproporcionalmente os grupos historicamente vulneráveis.

Diante disso, este trabalho tem como objetivo analisar os efeitos das políticas de cotas no ensino superior público brasileiro por meio da integração de bases de dados oficiais (ENADE, CES e Censo), identificando, por meio de técnicas de agrupamento, padrões no perfil racial e institucional dos ingressantes, na distribuição das vagas reser-

vadas entre tipos de instituição e áreas de conhecimento, e nas mudanças temporais produzidas pela *Lei de Cotas* ao longo do período 2010–2023. Para orientar o alcance do objetivo principal, o estudo busca responder às seguintes questões de pesquisa:

Questão de Pesquisa 1 (QP1): Os cotistas que acessam o ensino superior pertencem predominantemente aos grupos mais vulneráveis segundo os critérios de composição racial e origem socioeconômica dos ingressantes, ou há evidência de concentração nos estratos menos vulneráveis dentro dos grupos elegíveis (*creamy layer*)?

Esta questão é central porque a mera ampliação do número de cotistas não garante que a política alcance quem mais precisa dela: é preciso avaliar se a reserva cumpre sua finalidade redistributiva ou reproduz, internamente, as desigualdades que pretende corrigir.

Questão de Pesquisa 2 (QP2): As vagas reservadas estão sendo efetivamente preenchidas nos diferentes tipos de instituição e áreas de conhecimento, e os cotistas acessam cursos de maior prestígio econômico na mesma proporção que os não cotistas (*undermatching*)?

O preenchimento efetivo das vagas é condição básica para que a política produza efeitos, mas não suficiente: ainda que as vagas sejam ocupadas, os cotistas podem concentrar-se em cursos de menor retorno econômico, reproduzindo a estratificação horizontal. Avaliar conjuntamente onde as vagas são preenchidas e se há paridade de acesso às carreiras de maior retorno econômico permite distinguir o cumprimento formal da lei de sua efetividade real sobre a mobilidade social.

Questão de Pesquisa 3 (QP3): Como o perfil racial e institucional dos cotistas, medido pelos indicadores de inclusão de pretos, pardos e indígenas (PPI), origem escolar e prestígio dos cursos, se transformou entre os períodos pré-lei (P1), implementação (P2), consolidação (P3) e pandemia (P4)?

A dimensão temporal permite avaliar não apenas se a política produziu efeitos, mas quando e com que intensidade, identificando avanços, estagnações e retrocessos ao longo de mais de uma década, incluindo o período pandêmico.

Questão de Pesquisa 4 (QP4): Em que medida as mudanças observadas no perfil dos estudantes universitários refletem ou divergem da evolução das desigualdades estruturais de raça, renda e escolaridade na população brasileira geral?

Esta questão situa os resultados em um contexto mais amplo: ao confrontar os dados universitários com os do Censo Demográfico, avalia se as transformações internas às universidades acompanham a redução das desigualdades na sociedade ou permanecem restritas ao ambiente acadêmico.

2 Referencial Teórico

Este capítulo apresenta os conceitos teóricos e técnicos que embasam a metodologia adotada nesta pesquisa, com foco na aplicação de técnicas de inteligência artificial, especialmente o aprendizado de máquina não supervisionado, utilizando técnicas de agrupamento com base em dados educacionais para identificar padrões entre os pontos analisados.

2.1 Panorama Geral e Fundamentos das Ações Afirmativas

Internacionalmente, a Índia foi pioneira na implementação de ações afirmativas, estabelecidas legalmente na Constituição de 1950 para proteger grupos historicamente discriminados, como os *dalits* (“intocáveis”), por meio de cotas em legislaturas, empregos públicos e no ensino superior (JÚNIOR; DAFLON, 2015a). Nos Estados Unidos, surgiram na década de 1960 como resposta aos conflitos raciais provocados pela segregação institucionalizada — exemplificada pelo sistema *Jim Crow* — e pela força do movimento negro pelos direitos civis. Foi nesse contexto que o termo “ação afirmativa” foi introduzido oficialmente em 1963 pelo então presidente John F. Kennedy, designando políticas de natureza temporária e compensatória voltadas a corrigir desigualdades históricas (PEREIRA; ZIENTARSKI, 2011). Na África do Sul, as políticas afirmativas associam-se às ideias de “reparação” e “transformação”, voltadas a romper com as estruturas do *apartheid* após o fim do regime segregacionista em 1994 (FERREIRA et al., 2019).

No Brasil, o histórico de exclusão da população negra — marcado por quase 400 anos de escravidão sem políticas de integração subsequentes, exemplificado por medidas como a *Lei do Boi*, que durante o regime militar favorecia os filhos de proprietários rurais no acesso ao ensino superior — manteve a escolarização como privilégio das elites (PEREIRA; ZIENTARSKI, 2011). A educação só passou a ser tratada como prioridade de democratização nas últimas décadas, sobretudo a partir da III Conferência Mundial Contra o Racismo, realizada em Durban em 2001, cujo plano de ação recomendava medidas afirmativas às quais o Brasil aderiu (SILVA, 2018).

A partir dos anos 2000, algumas universidades estaduais anteciparam a adoção de cotas, com destaque para a Universidade do Estado do Rio de Janeiro (UERJ), que instituiu reservas de vagas para escola pública em 2000 e cotas raciais em 2001, seguida pela Universidade Estadual do Norte Fluminense (UENF) e pela Universidade do Estado da Bahia (UNEB) (TREVISOL; BELLO; NIEROTKA, 2023). No âmbito federal, a Universidade de Brasília (UnB) implantou seu plano de cotas em 2003 (FRANCO,

2008), e o programa REUNI, criado em 2007, incentivou a expansão dessas iniciativas (PINHEIRO; PEREIRA; XAVIER, 2021). Esse processo culminou na aprovação da Lei nº 12.711/2012, a *Lei de Cotas*, que estabeleceu a reserva de, no mínimo, 50% das vagas em universidades e institutos federais para estudantes egressos de escola pública, com recortes por renda e por autodeclaração de PPI — ampliada, pela Lei nº 13.409/2016, às pessoas com deficiência (PCD) (SENKEVICS; MELLO, 2019).

As ações afirmativas resultam de anos de mobilização social, especialmente do movimento negro (TREVISOL; BELLO; NIEROTKA, 2023), e respondem a desigualdades educacionais profundas e persistentes: em 2010, apenas 4,7% dos negros e 5,3% dos pardos com 25 anos ou mais possuíam diploma universitário, ante 15,0% dos brancos (FRIAS, 2014). Essas disparidades revelam como o racismo estrutural, aliado à pobreza, restringe as oportunidades educacionais da população negra e parda e se reflete tanto no acesso quanto na permanência no ensino superior (ROZA, 2023).

Dessa maneira, é fundamental considerar as desigualdades internas no ensino superior, especialmente a chamada *estratificação horizontal*, que diz respeito à distribuição desigual dos estudantes entre cursos e instituições (ROZA, 2023), influenciada por fatores de origem social, racial e de gênero. É nesse contexto que se inserem os fenômenos do *creamy layer* e do *undermatching*, apresentados na introdução: o primeiro revela que a reserva de vagas pode beneficiar preferencialmente os estratos menos vulneráveis dentro dos grupos elegíveis (JÚNIOR; DAFLON, 2015b), enquanto o segundo mostra que mesmo estudantes qualificados de grupos minoritários tendem a ser direcionados a cursos de menor prestígio (MEDEIROS; NETO; GOMES, 2016). Ambos os fenômenos operam como mecanismos que reproduzem, no interior do sistema, as hierarquias que as cotas pretendem enfrentar, e são investigados empiricamente nas análises QP1 e QP2 deste trabalho. Sem medidas complementares que enfrentem essas hierarquias, a inclusão promovida pelo acesso pode revelar-se frágil, razão pela qual somente a combinação entre ações afirmativas e políticas universais viabiliza um modelo educacional verdadeiramente democrático e inclusivo (TREVISOL; BELLO; NIEROTKA, 2023).

2.2 Inteligência Artificial e *Machine Learning*

A inteligência artificial (IA) é um campo tecnológico voltado para o desenvolvimento de sistemas que são capazes de aprender, analisar e processar grandes volumes de dados (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). No contexto deste trabalho, interessa especialmente sua aplicação à análise de bases de dados educacionais e socioeconômicas, com vistas à identificação de perfis dos ingressantes e agrupamentos de padrões institucionais.

Machine Learning (ML) ou Aprendizado de Máquina é um ramo da IA, que consiste em treinar algoritmos para que reconheçam padrões e tomem decisões a partir de conjuntos de dados, sem que as regras sejam explicitamente programadas. Esse aprendizado divide-se em três modalidades:

- **Aprendizado Supervisionado:** utiliza dados rotulados (entradas com saídas conhecidas) para treinar o algoritmo, sendo adequado a tarefas de classificação e predição;
- **Aprendizado Não Supervisionado:** trabalha com dados não rotulados, permitindo que o modelo identifique padrões e estruturas de uma base de dados por conta própria (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). É a modalidade adotada neste trabalho, por meio de técnicas de agrupamento (*clustering*), que particionam os dados em grupos de características semelhantes;
- **Aprendizado por Reforço:** baseia-se em tentativa e erro, ajustando as respostas do algoritmo com base em recompensas e penalidades até atingir o desempenho desejado.

2.2.1 Descoberta de Conhecimento em Bases de Dados (KDD)

A Descoberta de Conhecimento em Bases de Dados ou KDD (do inglês *Knowledge Discovery in Databases*) é um processo sistemático voltado à extração de conhecimento útil a partir de grandes volumes de dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). Seu objetivo é identificar padrões relevantes, relações não triviais e estruturas ocultas que subsidiem a tomada de decisões. O processo é composto por cinco etapas sequenciais:

- **Seleção:** identificação e escolha das fontes de dados e das variáveis relevantes para o problema de análise, definindo critérios de inclusão e exclusão, recorte temporal e populacional;
- **Pré-processamento:** etapa de preparação dos dados brutos para a análise, garantindo qualidade e confiabilidade das informações. Essa etapa envolve limpeza dos dados, como tratamento de valores ausentes, duplicidades, inconsistências de formato e valores atípicos (*outliers*), e normalização das variáveis para assegurar comparabilidade entre elas;
- **Transformação:** adequação dos dados ao formato exigido pelos algoritmos, podendo envolver agregação de registros, redução de dimensionalidade e codificação de variáveis categóricas em representações numéricas;

- **Mineração de dados:** etapa central do KDD, na qual algoritmos são aplicados sobre os dados transformados para descobrir padrões e agrupamentos. Neste trabalho, a técnica empregada é o agrupamento, descrito na Seção 2.2.2;
- **Interpretação e avaliação:** etapa de análise e validação dos padrões identificados, verificando sua relevância e utilidade para o problema de interesse. Os resultados são traduzidos em conhecimento aplicável, fundamentando conclusões e recomendações.

No contexto deste trabalho, o processo KDD orienta todas as etapas metodológicas conforme proposto por [Fayyad, Piatetsky-Shapiro e Smyth \(1996\)](#): a seleção das bases de dados (ENADE, CES e Censo), o pré-processamento e a harmonização dos registros, a transformação das variáveis para uso nos algoritmos de agrupamento e a interpretação dos perfis identificados à luz das políticas de cotas no ensino superior brasileiro.

2.2.2 Técnicas de Agrupamento

A técnica utilizada é o agrupamento (*clustering*), abordagem descritiva do aprendizado não supervisionado que organiza os registros em grupos internamente coesos com base na similaridade entre seus atributos, sem impor categorias predefinidas ([HASTIE; TIBSHIRANI; FRIEDMAN, 2009](#)). Os algoritmos distinguem-se entre agrupamento rígido (*hard clustering*), em que cada ponto pertence a um único cluster, e agrupamento suave (*soft clustering*), em que cada ponto recebe uma probabilidade de pertencer a cada grupo. Entre os principais métodos, destacam-se os adotados neste trabalho:

- **Modelos de centróide:** representados pelo algoritmo K-Means, organizam os dados em um número pré-definido de grupos k , alocando cada elemento ao centróide mais próximo e atualizando iterativamente suas posições até a convergência. Utiliza a distância euclidiana como medida de similaridade e é o método principal das análises QP1 e QP2, pela eficiência computacional e interpretabilidade dos resultados;
- **Modelos de conectividade:** constroem estruturas hierárquicas em forma de árvore (*dendrogramas*) com base na distância entre os elementos, sem exigir a definição prévia do número de clusters. O agrupamento hierárquico aglomerativo com critério de Ward ([JR., 1963](#)) escolhe a cada iteração a fusão de grupos que minimiza o aumento da variância intracluster, produzindo agrupamentos compactos. É o método adotado na análise QP3, por facilitar a comparação entre os quatro períodos temporais;

- **Modelos de distribuição:** representados pelos Modelos de Mistura Gaussiana ou GMM (*Gaussian Mixture Models*) (BISHOP, 2006), assumem que os dados seguem distribuições estatísticas específicas e estimam a probabilidade de cada elemento pertencer a cada cluster. Por ser uma abordagem de agrupamento suave, oferece maior flexibilidade que o K-Means ao lidar com grupos sobrepostos. É utilizado como método de validação na análise QP1, dada a tendência de sobreposição entre perfis institucionais.

Para determinar o número adequado de clusters, dois critérios são amplamente utilizados na literatura (HASTIE; TIBSHIRANI; FRIEDMAN, 2009):

- **Método do cotovelo (*elbow method*):** consiste em executar o algoritmo para diferentes valores de k e plotar a soma dos quadrados das distâncias intracluster (inércia) em função de k . O ponto de inflexão acentuada da curva, formando um cotovelo, indica o valor ótimo de k . Quando não há cotovelo claro, adota-se o critério de interpretabilidade;
- **Índice de silhueta (*silhouette score*):** proposto por Rousseeuw (1987), mede simultaneamente a coesão interna de cada cluster e sua separação em relação aos demais. Para cada ponto, calcula-se um valor entre -1 e 1 : valores próximos de 1 indicam boa alocação, próximos de 0 indicam sobreposição, e negativos sugerem alocação inadequada. O índice global é a média dos valores individuais.

Neste trabalho, o número ótimo de clusters em cada análise foi determinado pelo método do cotovelo e pelo índice de silhueta, garantindo a consistência e a comparabilidade entre os agrupamentos obtidos.

3 Revisão da Literatura

Como etapa inicial do trabalho, realizou-se uma revisão sistemática da literatura sobre o tema, utilizando *strings* (cadeia de caracteres) de busca elaboradas para identificar produções acadêmicas relevantes acerca das ações afirmativas no Brasil e em outros contextos internacionais. As buscas foram conduzidas em bases como Google Acadêmico, SciELO e o Portal de Periódicos da CAPES, a partir das seguintes strings de busca:

Polêmicas em torno das cotas

("ações afirmativas" OR "cotas raciais" OR "cotas sociais") AND ("polêmicas" OR "controvérsias" OR "debates" OR "críticas" OR "discriminação" OR "mérito" OR "oposição" OR "resistência")

Aplicação de cotas no ambiente acadêmico

("ações afirmativas" OR "cotas" OR "políticas públicas") AND ("ensino superior" OR "universidades" OR "ambiente acadêmico" OR "Brasil")

Desempenho de cotistas e não cotistas

("cotistas" OR "ações afirmativas") AND ("desempenho acadêmico" OR "rendimento" OR "notas" OR "evasão" OR "permanência") AND ("não cotistas" OR "comparação")

Impacto das cotas na sociedade

("cotas raciais" OR "cotas sociais") AND ("impacto social" OR "inclusão" OR "mobilidade social" OR "equidade" OR "desigualdades") AND ("Brasil")

Desigualdade no acesso e permanência ao ensino superior

("ações afirmativas" OR "cotas" OR "políticas públicas") AND ("desigualdade" OR "permanência" OR "acesso")

Coleta de dados da educação

("ENADE" OR "Censo da Educação Superior" OR "Censo educacional") AND ("dados socioeconômicos" OR "perfil dos estudantes" OR "indicadores educacionais" OR "cotas" OR "ações afirmativas")

Cotas em outros países

("affirmative action" OR "quota system") AND ("higher education" OR "universities" OR "institution") AND ("comparison" OR "effectiveness" OR "structural discrimination")

Ensino superior durante e pós-pandemia

(“ensino superior” OR “universidades” OR “ambiente acadêmico”) AND (“COVID” OR “pandemia”)

Como resultado, a análise do material coletado evidenciou não apenas os debates contemporâneos sobre as políticas de ação afirmativa, mas também a necessidade de compreender o processo histórico que moldou o ensino superior brasileiro. A literatura mostra que esse nível de ensino esteve, desde suas origens, vinculado a uma lógica elitista voltada à formação de quadros dominantes (TRAGTENBERG, 2002; ROMANELLI, 2007), o que contribuiu para a exclusão sistemática de negros e indígenas dos espaços acadêmicos e produziu efeitos duradouros sobre seus indicadores educacionais — analfabetismo, anos médios de estudo, rendimentos e acesso a cursos de maior prestígio (BAYMA, 2012).

É como resposta a essas desigualdades que as ações afirmativas se consolidam, não apenas como reparação histórica, mas como mecanismo de promoção da igualdade real (GOMES, 2005). Os debates que acompanharam a Lei nº 12.711/2012 refletem a tensão entre concepções meritocráticas e políticas redistributivas (MEDEIROS; NETO; GOMES, 2016), evidenciando que, embora polêmicas, as cotas foram fundamentais para romper com a exclusão histórica de determinados grupos no ambiente acadêmico.

Essa persistência das desigualdades raciais na educação brasileira é documentada ao longo das últimas décadas. Dados do IBGE mostram que, em 2001, cerca de 18% dos negros eram analfabetos (FRIAS, 2014), e que, em 2010, apenas 4,7% dos negros e 5,3% dos pardos com 25 anos ou mais haviam concluído o ensino superior, ante 15,0% dos brancos (FRIAS, 2014). O Censo de 2022 confirma a redução do gap, mas sua persistência: 23,8% dos brancos possuíam ensino superior completo, contra 10,7% dos pretos e 11,1% dos pardos (IBGE, 2022). Mesmo entre graduados, a diferença de rendimento entre brancos e não brancos indica que o diploma, por si só, não equaliza as condições de inserção no mercado de trabalho (IBGE, 2022). Análises da PNAD (Análises da Pesquisa Nacional por Amostra de Domicílios) entre 1973 e 2005 indicam ainda que o efeito da cor ou raça se intensifica à medida que se elevam os níveis educacionais: pouco pronunciado nos níveis básicos, torna-se decisivo no ensino superior e na pós-graduação, quando o acesso envolve bens escassos e competição (FRIAS, 2014).

Segundo Costa, Lima e Reis (2020 apud TREVISOL; BELLO; NIEROTKA, 2023), a análise dos dados referentes aos anos de 2017 e 2018 evidencia a persistência das desigualdades no acesso ao ensino superior. Os autores apontam que 48% dos jovens entre 18 e 24 anos pertencentes ao quartil mais rico da população estavam

matriculados em universidades, enquanto, entre os 25% mais pobres, essa taxa era de apenas 7,2%. Além disso, observou-se que a taxa de acesso entre jovens brancos era de 30,7%, reduzindo-se para 16,3% entre pardos e 15,1% entre pretos.

Entretanto, apesar das resistências iniciais à implantação da política de cotas, diversos estudos demonstram transformações significativas no perfil dos ingressantes após a adoção das cotas. Análises baseadas em dados da PNAD/IBGE e dos microdados do CES, presentes na literatura revisada ([BARBOSA, 2021](#); [TREVISOL](#); [BELLO](#); [NIE-ROTKA, 2023](#); [SENKEVICS, 2018](#)) mostram um aumento expressivo na presença de estudantes pretos, pardos e egressos de escolas públicas nas universidades. O percentual de ingressantes por reserva de vagas passou de 12,7% em 2012 para 39,6% em 2019. Nesse mesmo ano, 64,9% dos novos estudantes haviam concluído o ensino médio em escolas públicas e 47,4% se autodeclararam PPI, revelando um avanço importante na representatividade desses grupos.

De acordo com [Senkevics e Mello \(2019\)](#), a comparação entre a composição demográfica de jovens de 18 a 24 anos na população geral e os ingressantes nas Instituições Federais de Ensino Superior (Ifes) evidencia o impacto das cotas. Em 2012, a população brasileira dessa faixa etária era composta por 32% de brancos e amarelos de baixa renda, 50% de PPI de baixa renda e 18% de jovens com rendas familiares superiores a 1,5 salário mínimo per capita. Contudo, entre os ingressantes nas Ifes, esses percentuais eram de 37%, 34% e 29%, respectivamente, evidenciando a sub-representação dos PPI de baixa renda no acesso à universidade antes da implementação da Lei de Cotas. Já em 2016, após a consolidação da política, a participação dos PPI de baixa renda cresceu 8,8 pontos percentuais, enquanto, na população total, o crescimento desse grupo foi de apenas 8%. Assim, o crescimento de 26% entre os ingressantes das Ifes demonstra o impacto direto das cotas na correção de desigualdades históricas.

[Portes \(2007\)](#) também destaca essas mudanças, argumentando que a diversidade social e racial no ambiente acadêmico é um dos principais resultados positivos da política de cotas. A metodologia adotada por esses estudos — predominantemente quantitativa, com uso de estatísticas descritivas e análises comparativas entre cotistas e não cotistas — comprova que as ações afirmativas têm contribuído de maneira efetiva para ampliar o acesso de grupos historicamente excluídos do ensino superior.

Em relação ao desempenho acadêmico, os estudos de ([PINHEIRO](#); [PEREIRA](#); [XAVIER, 2021](#)), ([FRANCO, 2008](#)) e ([SOUZA, 2019](#)) buscam refutar o mito da inferioridade que recai sobre os cotistas. Utilizando bases como o ENADE e o CES, essas pesquisas comparam índices de rendimento, evasão e conclusão de curso entre cotistas e não cotistas. Os resultados demonstram que, uma vez garantido o acesso, os estudantes beneficiados pelas cotas tendem a apresentar desempenho equivalente — e, em alguns casos, superior — ao de seus colegas não cotistas. Tais achados reforçam

que a qualidade do ensino não é prejudicada pela adoção das ações afirmativas, e evidenciam a importância de se combater estigmas e preconceitos infundados no debate público.

Apesar dos avanços no acesso às universidades, a permanência dos estudantes cotistas no ensino superior ainda representa um desafio central. [Pena, Matos e Coutrim \(2020\)](#) e [Santos \(2009\)](#) alertam para a relevância do capital cultural e do apoio institucional na trajetória acadêmica dos estudantes de baixa renda, apontando que, sem políticas de permanência adequadas, muitos enfrentam dificuldades emocionais e econômicas que comprometem sua formação. Estudos mais recentes evidenciam a ausência de mecanismos eficazes de acompanhamento, demonstrando que, sem o devido suporte, a política de cotas pode perder parte de seu potencial transformador ([PEREIRA, 2017](#); [FERREIRA et al., 2020](#)). A garantia da permanência é, portanto, um passo indispensável para que o acesso se traduza em sucesso acadêmico e inclusão social efetiva.

Além dos desafios de acesso e permanência, é preciso considerar a desigualdade interna na distribuição dos estudantes entre cursos e instituições, fenômeno conhecido como estratificação horizontal. [Romanelli \(2007\)](#) aponta que estudantes de menor capital econômico e cultural tendem a se concentrar em cursos menos prestigiados, como licenciaturas e ciências humanas, enquanto os de maior renda ocupam cursos mais valorizados, como medicina, direito e engenharia. Essa dinâmica é reforçada por fatores institucionais e sociais, impactando diretamente as possibilidades de mobilidade social e empregabilidade. [Lahire \(1997\)](#) contribui com essa discussão ao destacar como a escolha do curso é mediada por disposições sociais herdadas. Assim, a democratização do acesso, sem o enfrentamento dessas hierarquias internas, pode reproduzir desigualdades no interior do próprio sistema educacional ([RISTOFF, 2014](#)).

Consequentemente, essa segmentação interna do sistema de ensino superior se reflete nas oportunidades disponíveis no mercado de trabalho. [Romanelli \(2007\)](#) e [Lahire \(1997\)](#) apontam que a trajetória universitária influencia fortemente a inserção profissional dos indivíduos, e que a origem social continua sendo um fator determinante mesmo após a obtenção do diploma. Alguns estudos analisam a inserção dos egressos cotistas, indicando que, embora as ações afirmativas ampliem o acesso ao ensino superior, os efeitos sobre a mobilidade social permanecem limitados diante da estratificação dos cursos e da seletividade do mercado ([OLIVEIRA, 2021](#); [RISTOFF, 2014](#)). Esses resultados indicam que a efetividade das cotas depende da articulação com outras políticas públicas voltadas à empregabilidade, à valorização de todos os cursos e ao acompanhamento dos egressos, de modo a garantir a justiça social para além do ambiente universitário.

A pandemia de COVID-19 provocou transformações profundas no ensino superior,

com a suspensão das atividades presenciais e a implementação do ensino remoto emergencial (ERE) (JOYE et al., 2020). Essa transição acentuou desigualdades já existentes: enquanto instituições privadas, com maior infraestrutura em EaD, adaptaram-se mais rápido (PEREIRA; DANTAS, 2021), as universidades públicas enfrentaram limitações de acesso à tecnologia e de formação para o ambiente virtual (CUNHA, 2021; CUNHA et al., 2022). A exclusão digital mostrou-se fortemente associada à cor e à renda: segundo dados de 2018, cerca de dois terços dos universitários sem acesso adequado à internet eram negros ou indígenas de baixa renda (CUNHA et al., 2022).

Apesar do volume crescente de produções sobre o período pandêmico, ainda são escassas as análises empíricas e quantitativas, baseadas em dados públicos, sobre os impactos reais da pandemia no acesso, permanência e desempenho dos estudantes (TREVISOL; BELLO; NIEROTKA, 2023) — lacuna que reforça a importância de pesquisas que integrem diferentes bases de dados educacionais e socioeconômicas.

Ademais, os estudos sobre ações afirmativas têm se apoiado em bases de dados públicas robustas. O ENADE é amplamente utilizado para examinar o perfil e o desempenho dos estudantes; o CES, para mensurar a evolução da presença de grupos historicamente excluídos; e o ENEM, para cruzar desempenho com cor/raça autodeclarada. No presente trabalho, concentra-se a análise no ENADE, no CES e no Censo, aplicando técnicas de agrupamento.

Apesar dos avanços proporcionados por esses estudos, persistem lacunas analíticas relevantes: a escassez de análises que integrem múltiplas bases de dados públicas, articulem indicadores institucionais e temporais, e cubram o período pandêmico com dados empíricos. É nesse espaço, somado à ainda pouco explorada trajetória dos egressos no mercado de trabalho pós-pandemia, que o presente trabalho se insere, conforme detalhado nos capítulos seguintes.

4 Metodologia

4.1 Base de Dados Utilizadas

As análises quantitativas deste trabalho foram conduzidas a partir de três bases de dados oficiais, selecionadas pela complementaridade entre as dimensões que cobrem: o perfil individual dos estudantes, a estrutura institucional do ensino superior e o contexto socioeconômico da população brasileira.

O Exame Nacional de Desempenho dos Estudantes (ENADE), aplicado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) em parceria com o Ministério da Educação (MEC) desde 2004 (INEP, 2023b), tem como finalidade mensurar o desempenho dos estudantes concluintes dos cursos de graduação em relação às competências, habilidades e conteúdos programáticos. O exame permite acompanhar a qualidade do ensino superior no Brasil e, a partir dos dados coletados, gerar diversos indicadores que contribuem para a consolidação de políticas públicas voltadas à qualidade da educação superior. Neste trabalho, no entanto, os microdados do ENADE são utilizados exclusivamente pelas variáveis disponíveis em arquivo único — modalidade de ingresso por cotas (POLITICAS_AFIRMATIVAS) e categoria administrativa da Instituição de Ensino Superior (IES) (CO_CATEGAD) —, em razão da reestruturação dos arquivos pelo INEP para adequação à *Lei Geral de Proteção de Dados Pessoais* (LGPD), conforme detalhado na Seção 5.2.2.

O Censo da Educação Superior (CES), levantamento anual realizado pelo INEP (INEP, 2023a), reúne informações detalhadas sobre as instituições de ensino superior, cursos de graduação, vagas, matrículas, concluintes e docentes. Seu objetivo é oferecer um panorama completo do sistema de educação superior no Brasil, possibilitando análises sobre o acesso, inclusão, desigualdades e qualidade do ensino. Neste trabalho, o CES constitui a principal fonte para mensurar a oferta e o preenchimento de vagas reservadas, desagregados por tipo de instituição, curso e modalidade de ensino.

O Censo Demográfico do IBGE (IBGE, 2022), realizado a cada dez anos por meio de visitas domiciliares em todos os municípios do país, é a principal referência sobre as condições de vida da população brasileira. Neste trabalho, o Censo é utilizado como base contextual comparativa, fornecendo o retrato da desigualdade estrutural da população por raça e renda para situar e interpretar os resultados das análises realizadas com ENADE e CES.

Em todas as análises de agrupamento, as variáveis utilizadas dividem-se em dois grupos. As *variáveis de entrada* (*features*) são aquelas efetivamente processadas pelo

algoritmo de agrupamento para calcular a similaridade entre os registros e formar os clusters. Já as *variáveis descritivas* não participam da formação dos grupos: são utilizadas apenas na etapa posterior de caracterização, para interpretar o perfil de cada cluster já definido. Essa separação permite que o agrupamento reflita as dimensões de interesse (perfil racial, origem escolar e prestígio) sem que atributos como o tipo de instituição dominem a partição, como detalhado na Seção 4.4.1.

A descrição técnica de cada base, incluindo estrutura dos arquivos, variáveis selecionadas e recorte temporal, está apresentada no próximo capítulo.

4.2 Divisão Temporal

Para evidenciar as transformações ocorridas ao longo do tempo, as análises são organizadas em quatro períodos, definidos com base nos marcos institucionais e conjunturais que caracterizam fases distintas da política de cotas no ensino superior brasileiro:

- **P1 – Fase Pré-Lei (2010–2012):** caracterização do ambiente universitário no período anterior à Lei nº 12.711/2012, servindo como linha de base para as comparações subsequentes;
- **P2 – Implementação da Lei (2013–2016):** período de introdução gradual das cotas nas instituições federais, com observação dos ajustes iniciais e impactos imediatos sobre o perfil dos ingressantes;
- **P3 – Consolidação (2017–2019):** estabilização das políticas após o prazo legal de implementação plena, com verificação dos efeitos de médio prazo sobre o acesso e o perfil dos estudantes;
- **P4 – Pandemia e pós (2020–2023):** análise das transformações recentes no contexto pandêmico e pós-pandêmico, considerando os impactos da suspensão de atividades presenciais e do ensino remoto emergencial sobre o acesso ao ensino superior.

Essa divisão é aplicada nas análises QP1, QP2 e QP3 (Seções 4.4.1, 4.4.2 e 4.4.3), permitindo identificar em quais períodos ocorreram as principais mudanças no perfil dos cotistas e no padrão de preenchimento das vagas reservadas.

4.3 Ferramentas e Recursos Computacionais

As análises realizadas neste trabalho foram desenvolvidas na linguagem Python ([Python Software Foundation, 2023](#)), escolhida pela ampla adoção em projetos de ciência de

dados, pela disponibilidade de bibliotecas especializadas em análise estatística e aprendizado de máquina e pela compatibilidade com o ambiente de execução utilizado. O processamento foi realizado no Google Colaboratory (Colab), plataforma de computação em nuvem que dispensa instalação local de software e permite a execução de código Python em notebooks interativos, com acesso direto aos arquivos armazenados no Google Drive.

Para o desenvolvimento e a depuração dos scripts, utilizou-se como ferramenta de apoio à programação o modelo de linguagem Claude, desenvolvido pela Anthropic. O acesso foi realizado por meio da interface conversacional disponível em <https://claude.ai>. O assistente foi empregado na geração, revisão e otimização de código Python, tendo todos os resultados verificados e validados pela autora antes de sua incorporação ao estudo.

4.4 Análises de Agrupamento

As análises conduzidas neste trabalho seguem uma estrutura metodológica padronizada, em que cada etapa é descrita de forma sequencial e replicável. Para cada análise, são especificados: o objetivo, a fonte de dados, as variáveis utilizadas, o pré-processamento aplicado, o método de agrupamento e o resultado esperado. Essa padronização visa garantir a transparência e a reprodutibilidade do processo analítico, além de facilitar a comparação entre as análises.

As três análises de agrupamento (QP1, QP2 e QP3) e a análise contextual (QP4) estão diretamente vinculadas às perguntas de pesquisa do trabalho, conforme descrito nas subseções a seguir.

4.4.1 QP1 – Perfil dos Cotistas no Nível Institucional

Objetivo: identificar padrões no perfil socioeconômico e racial dos estudantes que acessam as cotas, avaliando se os beneficiários pertencem predominantemente aos grupos mais vulneráveis dentro dos critérios elegíveis ou se há evidência de concentração nos estratos menos vulneráveis, no fenômeno denominado *creamy layer* (JÚNIOR; DAFLON, 2015b).

Definição operacional de vulnerabilidade: considera-se que um agrupamento de cursos apresenta perfil vulnerável quando concentra simultaneamente alta proporção de cotistas PPI, alto percentual de ingressantes cotistas e predominância em áreas de menor prestígio econômico — sendo o perfil oposto caracterizado como *creamy layer*.

Nota metodológica: a reestruturação dos microdados do ENADE para adequação à LGPD (INEP, 2022) impossibilitou a junção de informações no nível do estudante

individual. Por isso, a análise de perfil foi conduzida no nível institucional agregado via CES, sem uso de variáveis socioeconômicas individuais, conforme detalhado na Seção 5.2.2.

Fonte de dados: CES (2013–2023) como fonte principal para os indicadores de perfil agregado por curso e IES; e microdados do ENADE (2013–2023) para as variáveis de modalidade de ingresso por cotas, tipo de IES e período temporal. Os dados anteriores a 2013 foram excluídos das análises de perfil e clustering por corresponderem a sistemas voluntários de ação afirmativa de IES privadas, com perfil distinto dos cotistas da Lei nº 12.711/2012. Esses anos são utilizados apenas como linha de base contextual na análise QP3.

Variáveis de entrada (features):

- Percentual de cotistas ingressantes ($QT_ING_RESERVA_VAGA / QT_ING$);
- Percentual de ingressantes PPI;
- Percentual de ingressantes de escola pública;
- Prestígio médio dos cursos ($PRESTIGIO_NUM$), classificado conforme definição operacional apresentada a seguir.

Variáveis descritivas (pós-clustering):

- Tipo de categoria administrativa da IES;
- Modalidade de ingresso por cotas e tipo de cota (ENADE);
- Período temporal P1–P4.

Nota: o CES não disponibiliza o número de vagas reservadas ofertadas como campo separado, impossibilitando o cálculo direto da taxa de preenchimento. O percentual de cotistas ingressantes sobre o total ($QT_ING_RESERVA_VAGA / QT_ING$) foi utilizado como indicador substituto, sendo conceitualmente válido para mensurar o grau de adesão à política de reserva de vagas em cada curso e instituição.

Pré-processamento:

- Consolidação dos registros anuais do CES em base longitudinal única (2013–2023);
- Cálculo dos indicadores derivados por curso e IES: percentual de cotistas, percentual de PPI e percentual de ingressantes de escola pública;

- Exclusão de registros com zero ingressantes ou zero vagas reservadas;
- Integração dos indicadores do ENADE (tipo de cota e tipo de IES) no nível de agregação $\text{ANO} \times \text{TIPO_IES} \times \text{PERIODO}$;
- Padronização das variáveis numéricas pelo método *z-score*, garantindo que variáveis com escalas distintas — como percentuais (0–100) e categorias ordinais (0–2) — tenham pesos comparáveis no algoritmo de agrupamento (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Definição operacional de prestígio do curso: o prestígio foi classificado em três níveis ordinais (0 = Baixo, 1 = Médio, 2 = Alto), tomando como referência o rendimento médio dos graduados por área de formação apurado pelo Censo Demográfico de 2022 (IBGE, 2022) para definir quais áreas ocupam cada faixa. A atribuição foi operacionalizada por correspondência de termos no nome do curso (`NO_CURSO`): enquadraram-se como **alto** prestígio as áreas de maior rendimento e concorrência no acesso — Medicina, Direito, Engenharias, Computação e Sistemas, além de Farmácia, Odontologia e Enfermagem; como **baixo** prestígio, as licenciaturas e humanidades, de menor rendimento médio — Pedagogia, Letras, História, Geografia, Filosofia, Serviço Social, Educação Física, Música e Artes; e como prestígio **médio**, os demais cursos não enquadrados nos dois grupos anteriores. Essa hierarquia é coerente com o padrão de estratificação horizontal apontado por Romanelli (2007), segundo o qual estudantes de menor capital econômico e cultural tendem a se concentrar em cursos menos prestigiados, como licenciaturas e humanidades, enquanto os de maior renda ocupam carreiras mais valorizadas, como Medicina, Direito e Engenharia.

Método de agrupamento: K-Means como método principal, por sua eficiência computacional e interpretabilidade dos resultados. Como método de validação, utiliza-se o GMM, dado que os perfis institucionais tendem a apresentar sobreposição entre grupos. O número de clusters k será determinado pelo método do cotovelo e pelo índice de silhueta.

4.4.2 QP2 – Padrão Institucional de Preenchimento de Vagas

Objetivo: agrupar as instituições de ensino superior segundo o padrão de oferta e preenchimento das vagas reservadas, verificando se as vagas estão sendo efetivamente preenchidas nos diferentes tipos de instituição e áreas de conhecimento, e se os cotistas acessam cursos de maior prestígio econômico na mesma proporção que os não cotistas, ou se concentram em cursos de menor retorno econômico, fenômeno denominado *undermatching* (MEDEIROS; NETO; GOMES, 2016).

Fonte de dados: CES (2013–2023). Enquanto QP1 analisa o perfil dos cotistas no nível de curso, QP2 agrupa as instituições segundo o padrão de oferta e preenchimento das vagas reservadas, permitindo identificar quais tipos de IES cumprem ou não a política de cotas de forma sistemática.

Variáveis de entrada (features):

- Percentual de ingressantes cotistas sobre o total (PCT_COTISTAS);
- Percentual de ingressantes PPI (PCT_PPI);
- Percentual de ingressantes oriundos de escola pública (PCT_ESCOLA_PUB);
- Percentual de concluintes cotistas sobre o total de concluintes (PCT_COTISTAS_CONC);
- Prestígio médio dos cursos da IES (PRESTIGIO_NUM, em escala 0 = Baixo, 1 = Médio, 2 = Alto).

Variáveis descritivas (pós-clustering):

- Tipo de categoria administrativa da IES (federal, estadual, municipal, privada);
- Modalidade de ensino (presencial ou EaD);
- Composição racial dos ingressantes (branca, preta, parda, amarela, indígena).

Pré-processamento:

- Consolidação dos registros anuais do CES em uma única base, mantendo o histórico de cada instituição ao longo do período analisado;
- Cálculo de indicadores por instituição: percentual de cotistas entre os ingressantes, percentual de cotistas entre os concluintes e distribuição por área de conhecimento;
- Agrupamento das áreas de conhecimento nas categorias de prestígio definidas acima;
- Exclusão de registros com vagas reservadas iguais a zero;
- Padronização das variáveis numéricas pelo método *z-score*.

Método de agrupamento: K-Means como método principal, agrupando as instituições segundo o padrão de oferta e preenchimento das vagas reservadas. O agrupamento hierárquico aglomerativo (método de Ward) é utilizado como alternativa para

verificar a estrutura natural dos grupos institucionais e validar o número de clusters identificado pelo K-Means. O número de clusters k será determinado pelo método do cotovelo e pelo índice de silhueta.

4.4.3 QP3 – Evolução Temporal entre os Períodos

Objetivo: comparar o perfil dos cotistas e o padrão de preenchimento das vagas reservadas entre os quatro períodos definidos no estudo, identificando mudanças estruturais no acesso ao ensino superior ao longo do tempo e avaliando em que medida a política de cotas evoluiu em termos de alcance e equidade.

Fonte de dados: microdados do ENADE (2010–2023) para os indicadores de participação de cotistas por tipo de IES e tipo de cota; e CES (2013–2023) para os indicadores institucionais de preenchimento de vagas, perfil racial dos ingressantes e evolução dos concluintes. Para a análise temporal, o CES é utilizado com dois cortes complementares: ingressantes, para mensurar o acesso e o preenchimento das vagas reservadas em cada período; e concluintes, para avaliar a persistência dos cotistas ao longo do percurso formativo entre os períodos P1 e P4. Os dados do ENADE anteriores a 2013 são utilizados exclusivamente como linha de base do período pré-lei (P1), sem análise de perfil socioeconômico individual.

Variáveis utilizadas:

- Do ENADE: percentual de cotistas por período, utilizado como indicador complementar ao CES para validação da tendência temporal;
- Do CES: percentual de ingressantes PPI, percentual de cotistas ingressantes, percentual de concluintes cotistas e prestígio médio dos cursos, os mesmos indicadores de QP1 e QP2, calculados separadamente para cada período.

Pré-processamento: além das etapas descritas em QP1 e QP2, esta análise requer:

- Segmentação das bases pelo período temporal (P1–P4), garantindo que cada subconjunto seja tratado de forma independente;
- Verificação da comparabilidade das variáveis entre períodos, com atenção às alterações de categorias no CES ao longo dos anos.

Método de agrupamento: o agrupamento hierárquico aglomerativo é o método principal, por facilitar a comparação entre períodos por meio de dendrogramas e por preservar a estrutura temporal dos dados sem exigir a definição prévia do número

de clusters. O K-Means é aplicado como complemento para validar os agrupamentos identificados hierarquicamente.

4.4.4 QP4 – Contextualização com o Censo Demográfico

Objetivo: situar os resultados das análises QP1, QP2 e QP3 no contexto da desigualdade estrutural da população brasileira, verificando em que medida as transformações observadas dentro das universidades refletem ou divergem da evolução das desigualdades de raça, renda e escolaridade na população geral. Busca-se, com isso, uma leitura integrada dos dados universitários e populacionais que permita avaliar se as políticas de cotas contribuíram para reduzir disparidades históricas de raça e classe no acesso ao ensino superior (TREVISOL; BELLO; NIEROTKA, 2023), ou se as desigualdades estruturais da sociedade brasileira ainda se reproduzem dentro das universidades, mesmo após mais de uma década de implementação da Lei nº 12.711/2012.

Fonte de dados: tabelas do Censo referentes aos anos de 2000, 2010 e 2022, conforme descritas na seção 5.4. Esta análise não utiliza técnicas de clustering, pois os dados do Censo são apresentados como base contextual comparativa, e não como fonte de agrupamento.

Indicadores utilizados:

- Percentual da população com ensino superior completo, por cor ou raça (2000, 2010 e 2022);
- Número médio de anos de estudo por cor ou raça (2022);
- Distribuição da população ocupada por faixa de rendimento e cor ou raça (2022);
- Distribuição das pessoas com ensino superior por área de formação e cor ou raça (2022).

Procedimento analítico: os indicadores do Censo são apresentados em paralelo aos resultados de QP1, QP2 e QP3, comparando o perfil dos estudantes universitários com o da população geral — em especial quanto ao crescimento da participação PPI, à reprodução da estratificação horizontal por área de formação e à evolução do gap de escolaridade entre brancos e não brancos ao longo dos três censos.

5 Base de Dados e Preparação

5.1 Visão Geral das Bases e Integração

Os procedimentos descritos neste capítulo correspondem às três primeiras etapas do processo KDD apresentado na Seção 2.2.1: seleção das bases, pré-processamento e transformação (Seções 5.2, 5.3 e 5.4). As etapas de mineração e interpretação são apresentadas no Capítulo 6.

A escolha das três bases fundamenta-se na complementaridade entre as dimensões que cobrem: o perfil institucional e agregado dos ingressantes, a estrutura da oferta de vagas reservadas e o contexto socioeconômico da população brasileira. A integração entre ENADE, CES e Censo possibilita uma análise multiescalar que permite avaliar em que medida as transformações observadas dentro das IES refletem ou divergem das tendências de desigualdade estrutural da sociedade brasileira ao longo do período analisado.

Um dos principais desafios da preparação consistiu no fato de que as três bases apresentam formatos, estruturas e codificações distintas ao longo do tempo. Para enfrentá-lo, construiu-se um dicionário de equivalência de variáveis para cada base, sistematizando as particularidades de cada arquivo por ano de referência e assegurando a coerência dos dados integrados. Nas subseções a seguir, cada base é descrita quanto à estrutura, às variáveis selecionadas e aos procedimentos de harmonização.

5.2 ENADE

5.2.1 Estrutura e Variáveis Utilizadas

Para este estudo, foram selecionados os microdados do ENADE referentes aos anos de 2010 a 2023. A escolha desse período deve-se ao fato de que, a partir de 2010, o exame passou a ter caráter censitário, ampliando a cobertura e a representatividade dos dados. Embora a avaliação seja realizada anualmente, os cursos participam em ciclos trienais, o que gera uma alternância entre áreas avaliadas a cada ano. Cabe ressaltar que o ENADE não foi aplicado em 2020 devido à pandemia de COVID-19; o ciclo de Engenharias, previsto para aquele ano, foi realizado em 2021.

As áreas avaliadas em cada edição do ENADE ao longo do período analisado, organizadas em ciclos trienais, estão apresentadas na Tabela 1.

Tabela 1 – Áreas avaliadas no ENADE por ciclo trienal (2010–2023)

Área geral avaliada	Anos
Saúde e Ciências Agrárias	2010, 2013, 2016, 2019, 2023
Engenharias, Ciências Exatas e Licenciaturas	2011, 2014, 2017, 2021
Ciências Sociais Aplicadas e Gestão	2012, 2015, 2018, 2022

Fonte: INEP/MEC. Elaboração própria.

Embora o ENADE avalie conhecimentos específicos dos cursos com o objetivo de mensurar a qualidade do ensino superior, essa dimensão avaliativa não constitui o foco desta pesquisa. A análise concentra-se exclusivamente nas variáveis institucionais e de ingresso disponíveis nos microdados — a categoria administrativa da IES, a modalidade de ingresso por políticas de cotas e o tipo de cota utilizado. As variáveis socioeconômicas individuais não foram utilizadas, pela limitação LGPD detalhada na Seção 5.2.2.

5.2.2 Harmonização e Preparação

Os microdados do ENADE são disponibilizados em arquivos separados por ano, cada um correspondendo às respostas de uma questão específica dos questionários do estudante e da IES. Em 2022, o INEP reestruturou esses arquivos para adequação à LGPD, adotando uma estrutura em que cada arquivo é ordenado por variáveis distintas (INEP, 2022), impossibilitando a junção de informações no nível do estudante individual. As variáveis socioeconômicas individuais — renda familiar, cor ou raça, escolaridade dos pais e tipo de escola no ensino médio — portanto não são utilizadas neste estudo.

Das variáveis efetivamente utilizadas nas análises, o mapeamento identificou as seguintes equivalências entre ciclos: POLITICAS_AFIRMATIVAS disponível no arq18 (2010–2013) e arq21 (2014–2023); e CO_CATEGAD disponível no arq1 em todos os anos.

Realizou-se um mapeamento manual de todos os arquivos, identificando e validando as equivalências conceituais entre os diferentes períodos. Neste trabalho, os arquivos são referenciados pela posição que ocupam no conjunto de microdados de cada ano — arq1, arq2, e assim sucessivamente — conforme a nomenclatura original dos arquivos disponibilizados pelo INEP.

Nesse mapeamento, identificaram-se dois padrões relevantes na composição dos arquivos:

- **Arquivos estáveis:** os arquivos arq1 a arq6 mantiveram-se estruturalmente con-

sistentes em toda a série histórica, garantindo que a variável `CO_CATEGAD` continue confiável ao longo do período analisado;

- **Arquivos exclusivos por período:** a variável `POLITICAS_AFIRMATIVAS` foi disponibilizada no `arq18` entre 2010 e 2013 e no `arq21` a partir de 2014, com equivalência conceitual preservada entre os dois períodos.

Para além da renumeração, diversas variáveis sofreram alterações em suas categorias de resposta, exigindo transformações específicas para garantir a comparabilidade temporal. As duas principais adaptações, das variáveis utilizadas no trabalho, estão descritas na Tabela 2.

Tabela 2 – Adaptações realizadas em variáveis de informações da IES e do curso

Variável	Categorias originais por período	Estratégia de harmonização
Categoria Administrativa <code>CO_CATEGAD</code>	2010–2013 e 2014–2023: 1 Pública Federal 2 Pública Estadual 3 Pública Municipal 4 Privada com fins lucrativos 5 Privada sem fins lucrativos Acréscimos a partir de 2014: 7 Especial 8 Comunitária/ Confessional (2015+)	Agrupamento em três categorias amplas: 1–3: Pública 4–5: Privada 7–8: Outras Dicionários anuais específicos foram mantidos para análises detalhadas.
Organização Acadêmica <code>CO_ORGACAD</code>	2010–2013: Incluía o código 10027 para “Instituto Superior/Escola Superior”. 2014–2023: Código 10027 descontinuado.	Todos os registros com código 10027 foram reclassificados como “Outras” para preservar a comparabilidade ao longo da série histórica.

Após a harmonização individual de cada variável, os dados foram consolidados em um único *dataframe* com nomenclatura padronizada, codificação numérica consistente e tratamento uniforme de valores ausentes. Os dados foram armazenados no formato Parquet, escolhido pela eficiência de compressão e desempenho em análises longitudinais.

5.3 CES

5.3.1 Estrutura e Variáveis Utilizadas

O CES (INEP, 2023a) constitui a fonte principal para mensurar a oferta e o preenchimento de vagas reservadas, desagregados por tipo de instituição, curso e modalidade

de ensino, cobrindo o período de 2010 a 2023. Diferentemente do ENADE, os arquivos do CES apresentam estrutura padronizada ao longo dos anos, com os mesmos campos disponíveis em cada edição, o que facilita o tratamento e a análise longitudinal.

Em cada ano do CES, foram utilizados dois conjuntos principais de arquivos:

- MICRODADOS_CADASTRO_CURSOS_<ANO>.CSV
- MICRODADOS_CADASTRO_IES_<ANO>.CSV

Destaca-se que, a partir de 2022, o arquivo de IES passou a apresentar em alguns casos a nomenclatura MICRODADOS_ED_SUP_IES_<ANO>. Ambos os formatos foram tratados de forma equivalente.

Para os arquivos de IES, foram selecionadas as variáveis relativas à identificação da instituição, categoria administrativa e organização acadêmica. Para os arquivos de cursos, foram selecionadas variáveis relacionadas às características dos cursos, à modalidade de ensino e indicadores de ingresso e conclusão desagregados por cor ou raça e formas de ingresso, incluindo reserva de vagas. As demais colunas foram desconsideradas por não contribuírem diretamente para as análises propostas.

5.3.2 Harmonização e Preparação

O processamento dos dados do CES exigiu o desenvolvimento de uma metodologia específica para lidar com a heterogeneidade estrutural e técnica das bases ao longo do período de 2010 a 2023. As principais dificuldades estiveram associadas às variações nos padrões de nomenclatura dos arquivos, nos *encodings* (codificações de caracteres), nos separadores utilizados e na estrutura das colunas disponíveis em cada ano.

Para enfrentar esses desafios, foi implementado um *pipeline* (fluxo automatizado) de processamento composto por etapas sequenciais de detecção de arquivos, identificação de metadados, harmonização estrutural das variáveis, otimização de memória e validação de consistência interna. O processamento foi realizado de forma incremental, com carregamento em blocos, viabilizando o tratamento de arquivos de grande volume em ambiente computacional limitado.

Após o tratamento individual de cada ano, os dados foram consolidados em uma base longitudinal única e armazenados no formato Parquet, com cobertura completa do período de 2010 a 2023.

5.4 Censo Demográfico

5.4.1 Tabelas Seleccionadas e Recorte

O Censo Demográfico (IBGE, 2022) foi utilizado como base contextual e comparativa, permitindo situar os resultados das análises de agrupamento em um panorama mais amplo das desigualdades educacionais, raciais e socioeconômicas da população brasileira. Foram seleccionadas tabelas referentes aos censos de 2000, 2010 e 2022, com foco nos seguintes eixos analíticos:

- **Evolução do acesso ao ensino superior por raça:** análise da proporção da população com nível superior concluído segundo cor ou raça nos três censos, permitindo verificar se houve aumento da representação de pretos, pardos e indígenas na graduação em escala populacional no período de vigência da Lei nº 12.711/2012;
- **Renda e nível de instrução por raça:** análise dos rendimentos médios mensais e da distribuição por faixas salariais segundo cor ou raça e nível de instrução, verificando indícios de mobilidade econômica associada ao avanço educacional dos grupos historicamente sub-representados;
- **Concentração do ensino superior por área de formação e raça:** análise da distribuição de graduados por área geral de formação e cor ou raça, verificando em escala populacional se a concentração de não brancos em áreas de menor retorno econômico se reproduz no perfil geral da população com ensino superior concluído.

5.4.2 Harmonização e Preparação

As tabelas do Censo Demográfico apresentam estruturas hierárquicas complexas, o que dificulta sua leitura direta por bibliotecas usuais em Python. Diferentemente do ENADE e do CES, cujo processamento foi automatizado, as tabelas do Censo exigiram edição manual antes da importação, incluindo a remoção de linhas de título, a reorganização de áreas de formação e faixas etárias em abas e a separação de tabelas que agregavam múltiplos indicadores. Concluídas essas etapas, os dados das três bases foram consolidados em formatos adequados às análises de agrupamento, garantindo comparabilidade temporal e consistência estrutural entre as fontes.

6 Análise de Agrupamento e Discussão dos Resultados

Este capítulo apresenta os resultados das análises de agrupamento realizadas sobre as bases do CES e do ENADE, seguindo a estrutura metodológica definida no Capítulo 4. Para cada análise, descrevem-se o pré-processamento aplicado, os clusters identificados, sua caracterização e validação, e a discussão dos resultados à luz das perguntas de pesquisa. Cada uma das quatro análises (QP1 a QP4) responde a uma das perguntas, encerrando com a contextualização dos achados frente aos indicadores populacionais do Censo Demográfico de 2022 ([IBGE, 2022](#)).

6.1 Caracterização Geral dos Dados

Antes das análises de agrupamento, apresenta-se uma caracterização geral dos indicadores que fundamentam as discussões subsequentes, construídos a partir dos microdados do ENADE ([INEP, 2023b](#)) e do CES ([INEP, 2023a](#)).

No ENADE, o percentual de concluintes cotistas passou de 15,7% em 2010 para 24,4% em 2023, com participação maior nas IES públicas (40,9% em 2023) do que nas privadas (18,6%). No CES, o percentual de ingressantes autodeclarados pretos, pardos e indígenas (PPI) passou de 52,7% em 2010 para 61,5% em 2023, com crescimento mais acentuado nas instituições federais (Figura 1), onde a obrigatoriedade das cotas étnico-raciais exerce maior pressão inclusiva. Por área de conhecimento nas federais, a área de Saúde e bem-estar registrou o maior percentual de ingressantes cotistas (44,6%), e a de Educação, um dos menores (30,3%).

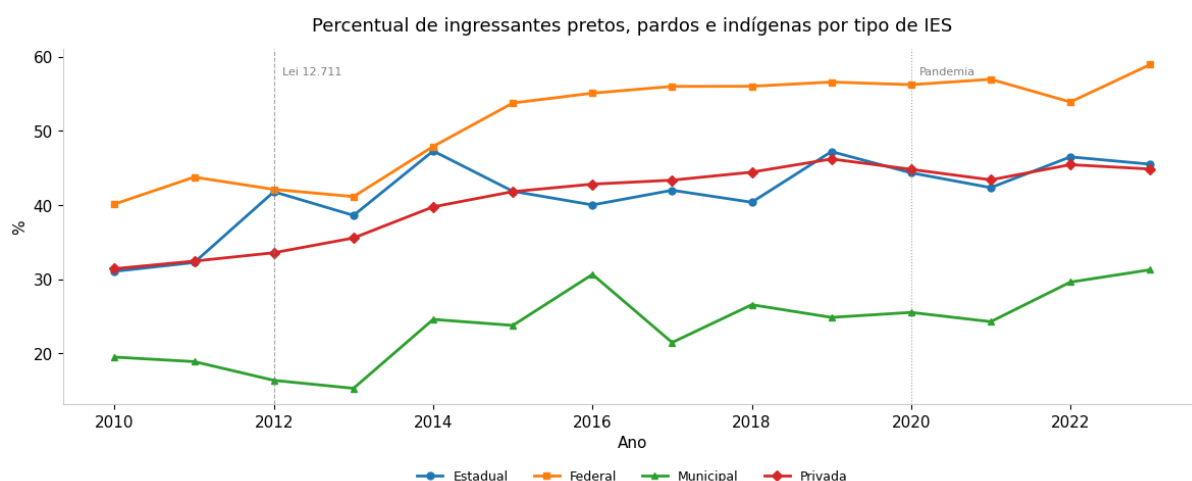


Figura 1 – Percentual de ingressantes pretos, pardos e indígenas por tipo de IES (2010–2023).

Fonte: Elaboração própria a partir dos dados do CES/INEP.

A comparação entre ingresso e conclusão revela um funil de inclusão (Figura 2): em 2023, os cotistas representavam 35,7% dos ingressantes e 37,2% dos concluintes nas IES federais, enquanto o percentual geral de concluintes cotistas no sistema passou de 0,8% em 2010 para 5,0% em 2023, mantendo-se abaixo do percentual de ingressantes ao longo de todo o período. Essa defasagem entre ingresso e conclusão evidencia o desafio da permanência estudantil, tema retomado nas discussões das análises a seguir.

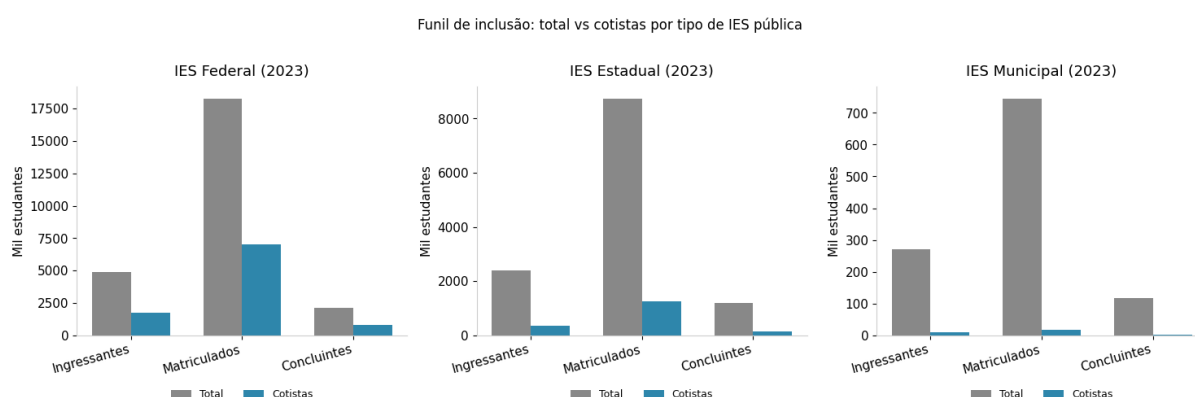


Figura 2 – Funil de inclusão: total vs cotistas por tipo de IES pública (2023).

Fonte: Elaboração própria a partir dos dados do CES/INEP.

6.2 QP1 – Perfil dos Cotistas no Nível Institucional

Esta seção responde à **QP1**: *os cotistas que acessam o ensino superior pertencem predominantemente aos grupos mais vulneráveis segundo os critérios de composição*

racial e origem socioeconômica dos ingressantes, ou há evidência de concentração nos estratos menos vulneráveis dentro dos grupos elegíveis (creamy layer)?

6.2.1 Pré-processamento e definição dos clusters

A base utilizada nesta análise foi construída a partir do CES consolidado (2013–2023), conforme descrito na Seção 5.3. Após o carregamento dos 2.903.652 registros disponíveis, foram computados quatro indicadores derivados por curso e IES: percentual de cotistas ingressantes sobre o total de ingressantes (PCT_COTISTAS), percentual de ingressantes autodeclarados pretos, pardos e indígenas (PCT_PPI), percentual de ingressantes oriundos de escola pública (PCT_ESCOLA_PUB) e prestígio do curso (PRESTIGIO_NUM), classificado em três categorias ordinais (0 = Baixo, 1 = Médio, 2 = Alto) com base no nome do curso, conforme definição operacional apresentada na Seção 4.4.1.

Em seguida, foram excluídos os registros de cursos sem nenhum ingressante cotista, resultando em 74.996 registros válidos, representando 2,6% do total original. Essa proporção reflete o fato de que a maioria dos cursos de IES privadas não registra ingressantes por reserva de vagas, concentrando os registros com cotistas majoritariamente em cursos de IES federais (75,4%) e estaduais (21,7%). As IES privadas representam 2,6% dos registros válidos, correspondendo às instituições que adotam sistemas voluntários de reserva de vagas.

As variáveis de tipo de IES foram mantidas apenas como descritivas para a caracterização pós-clustering — e não como *features* —, de modo que o agrupamento refletisse as diferenças de perfil dos cotistas, e não o tipo de instituição. As *features* finais foram padronizadas pelo método *z-score*, garantindo que variáveis com escalas distintas tivessem pesos comparáveis no agrupamento.

Para determinar o número ótimo de clusters, aplicou-se o método do cotovelo e o índice de silhueta para k de 2 a 10 (Figura 3). Os índices de silhueta variaram entre 0,217 ($k = 2$) e 0,262 ($k = 10$), com crescimento monotônico e sem cotovelo claro, padrão esperado para dados institucionais com distribuição contínua, em que os perfis variam de forma gradual, sem fronteiras naturais entre grupos. Diante disso, adotou-se $k = 3$ por critério de interpretabilidade: os três clusters gerados apresentam perfis distintos e nomeáveis, adequados para responder à pergunta de pesquisa sobre a concentração do acesso às cotas nos estratos menos vulneráveis.

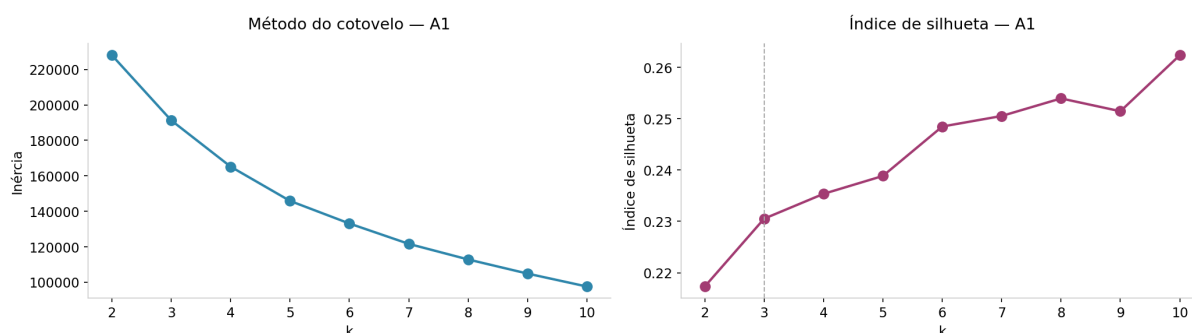


Figura 3 – Método do cotovelo e índice de silhueta.

Fonte: Elaboração própria.

6.2.2 Caracterização dos clusters

Os três clusters identificados pelo K-Means estão descritos na Tabela 3 e visualizados na Figura 4.

Tabela 3 – Caracterização dos clusters

Cluster	Registros	% Cotistas	% PPI	% Escola pública	Prestígio médio
C0 - Inclusão em cursos de prestígio	18.910	36,9%	38,2%	61,8%	2,0 (Alto)
C1 - Baixa inclusão racial	27.868	29,2%	27,6%	60,9%	0,8 (Médio/Baixo)
C2 - Alta inclusão PPI	28.218	45,7%	64,3%	81,9%	0,8 (Médio/Baixo)

Fonte: Elaboração própria a partir dos dados do CES/INEP (2013–2023).

Como os dados possuem quatro dimensões (PRESTIGIO_NUM, PCT_ESCOLA_PUB, PCT_COTISTAS e PCT_PPI), a visualização direta dos clusters não é possível. Para isso, aplicou-se a Análise de Componentes Principais (PCA), que comprime as quatro dimensões originais em dois eixos (PC1 e PC2) preservando o máximo de informação possível. A projeção capturou 63,4% da variância total dos dados (PC1 = 37,3%; PC2 = 26,1%), sendo considerada adequada para fins de visualização (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Os 36,6% restantes estão distribuídos em dimensões não visíveis no gráfico, razão pela qual a caracterização quantitativa dos clusters está apresentada na Tabela 3.

C0 – Inclusão em cursos de prestígio (18.910 registros): este cluster reúne exclusivamente cursos de alto prestígio (100% com PRESTIGIO_NUM = 2), majoritariamente em IES federais (77,6%). O percentual de cotistas é de 36,9% e o de PPI de 38,2%, os menores entre os três clusters para essas variáveis. O percentual de ingressantes de

escola pública (61,8%) é próximo ao de C1 (60,9%) e significativamente inferior ao de C2 (81,9%), indicando que os cotistas que acessam cursos de alto retorno econômico (como Medicina, Direito, Engenharia e Computação) apresentam composição racial e de origem escolar menos vulnerável do que os cotistas concentrados em cursos de menor prestígio.

C1 – Baixa inclusão racial (27.868 registros): cluster com o menor percentual de cotistas (29,2%) e o menor percentual de PPI (27,6%) de toda a análise. Os cursos são majoritariamente de prestígio Médio (78%) e Baixo (22%), com maior presença de IES estaduais (27,6%) em comparação aos demais clusters. O percentual de escola pública (60,9%) é próximo ao de C0, indicando que a baixa inclusão racial não está associada à origem escolar, mas a outros fatores institucionais ou de seleção.

C2 – Alta inclusão PPI (28.218 registros): cluster com os maiores valores em todas as dimensões de inclusão — maior percentual de cotistas (45,7%), maior percentual de PPI (64,3%) e maior percentual de ingressantes de escola pública (81,9%). Os cursos são predominantemente de prestígio Médio e Baixo, concentrados em IES federais (80,7%). Este cluster representa o funcionamento das cotas conforme previsto pela Lei nº 12.711/2012 — cotistas majoritariamente PPI, oriundos de escola pública, acessando cursos de nível médio e baixo prestígio econômico.

O gráfico de perfil normalizado (Figura 4) evidencia que C1 apresenta os menores valores em todas as dimensões simultaneamente após a normalização, posicionando-se próximo à origem do radar: menor percentual de cotistas, menor representação PPI, menor proporção de ingressantes de escola pública e menor prestígio médio dos cursos. C0 destaca-se pelo valor máximo na dimensão de prestígio, enquanto C2 apresenta os maiores valores nas dimensões de inclusão racial e de origem escolar pública.

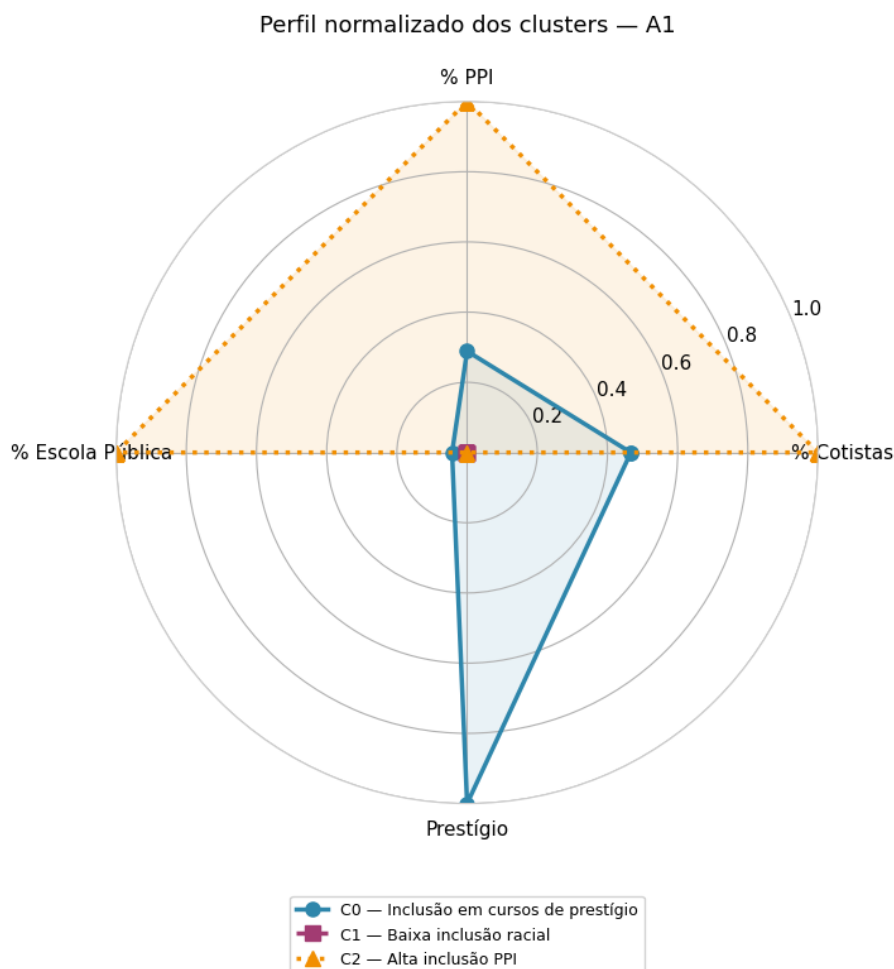


Figura 4 – Perfil normalizado dos clusters.

Fonte: Elaboração própria.

6.2.3 Validação por GMM

Para validar os agrupamentos obtidos pelo K-Means, aplicou-se o modelo GMM (*Gaussian Mixture Models*) com $k = 3$ sobre as mesmas *features* padronizadas. Como o GMM numera os clusters de forma independente, os rótulos foram realinhados pelo algoritmo húngaro — que encontra a correspondência de máxima concordância entre dois conjuntos de rótulos (HASTIE; TIBSHIRANI; FRIEDMAN, 2009) — antes do cálculo da concordância global, que resultou em 60,6%. Esse valor moderado é esperado para dados com distribuição contínua e fronteiras difusas entre grupos: por ser um método de agrupamento suave (BISHOP, 2006), o GMM atribui probabilidades de pertencimento em vez de rótulos fixos, distribuindo de forma distinta os registros nas zonas de sobreposição entre clusters, o que reduz a concordância sem invalidar a estrutura identificada.

A Tabela 4 compara as caracterizações dos clusters obtidos pelos dois métodos.

Tabela 4 – Comparação entre K-Means e GMM

Variável	C0 – Prestígio		C1 – Baixa inclusão		C2 – Alta PPI	
	K-Means	GMM	K-Means	GMM	K-Means	GMM
% Cotistas	36,9	39,9	29,2	26,4	45,7	41,4
% PPI	38,2	41,8	27,6	20,8	64,3	61,5
% Esc. pública	61,8	49,3	60,9	74,6	81,9	88,2
Prestígio médio	2,0	1,3	0,8	1,0	0,8	1,0

Fonte: Elaboração própria a partir dos dados do CES/INEP (2013–2023).

Os dois métodos convergem na identificação da mesma estrutura de três perfis. As maiores divergências ocorrem em C0, onde o GMM estima menor prestígio médio (1,3 versus 2,0) e menor percentual de escola pública (49,3% versus 61,8%), sugerindo que parte dos registros de prestígio médio migra para C0 no GMM por estarem nas zonas de fronteira com C1. Ainda assim, as diferenças não alteram a interpretação qualitativa dos clusters, e a concordância de 60,6% — equivalente a 3 em cada 5 registros classificados no mesmo grupo pelos dois métodos — é considerada aceitável para dados com sobreposição natural entre os perfis institucionais.

6.2.4 Discussão – *creamy layer* no nível institucional

Os resultados da análise QP1 evidenciam a presença do fenômeno denominado *creamy layer* no nível institucional do ensino superior público brasileiro. O conceito, discutido por Júnior e Daflon (2015b) no contexto das políticas de cotas, refere-se à tendência de que os benefícios das ações afirmativas se concentrem nos estratos menos vulneráveis dentro dos grupos elegíveis, isto é, estudantes que possuem maior capital social, cultural e econômico.

Na presente análise, esse fenômeno manifesta-se na comparação entre os clusters C0 e C2. Os cotistas inseridos em cursos de alto prestígio econômico (C0) apresentam percentual de PPI de 38,2% e percentual de ingressantes oriundos de escola pública de 61,8%, valores significativamente inferiores aos observados no cluster de alta inclusão PPI (C2), que reúne cotistas em cursos de menor retorno econômico, com 64,3% de PPI e 81,9% de estudantes oriundos da escola pública. Em outras palavras, entre os estudantes beneficiados pelas reservas de vagas, aqueles que conseguem acessar os cursos mais valorizados tendem a possuir perfil racial e trajetória escolar menos vulneráveis.

Esse resultado é reforçado pelo padrão de utilização dos tipos de cota: tanto nos dados do ENADE quanto nos do CES, a modalidade de escola pública é a mais acio-

nada, seguida pelas cotas de renda e étnico-racial. Isso indica que parte significativa dos cotistas inseridos em cursos de maior prestígio ingressa prioritariamente pela modalidade de escola pública — critério menos restritivo do ponto de vista racial — sem necessariamente pertencer aos segmentos mais vulneráveis dentre os elegíveis.

O padrão identificado é consistente com a literatura sobre estratificação horizontal no ensino superior brasileiro (TREVISOL; BELLO; NIEROTKA, 2023), segundo a qual a distribuição desigual dos estudantes entre cursos e instituições é condicionada por fatores sociais, raciais e econômicos. Embora a implementação da Lei nº 12.711/2012 tenha ampliado o acesso de grupos historicamente excluídos ao ensino superior, a distribuição desses grupos entre áreas de conhecimento e tipos de IES ainda reproduz desigualdades estruturais. Nota-se, por exemplo, o aumento da presença de cotistas em áreas antes majoritariamente ocupadas por brancos de alta renda (como Engenharia, Direito e Medicina) contribuindo para romper parte da estratificação histórica. Contudo, o setor privado continua concentrando ingressantes em cursos de maior retorno econômico sem a mesma proporção de cotistas PPI, fenômeno que pode ampliar desigualdades entre os próprios beneficiários das cotas (ROZA, 2023).

O cluster C1 revela um padrão adicional relevante. Mesmo entre os cotistas, há baixa participação de estudantes PPI (27,6%) em cursos concentrados majoritariamente em IES estaduais e em áreas de prestígio médio e baixo. Esse perfil indica que a sub-representação racial não se restringe aos cursos de maior prestígio econômico, manifestando-se também em cursos intermediários. Assim, as barreiras ao acesso de cotistas PPI demonstram envolver não apenas a seletividade dos cursos, mas também fatores institucionais e regionais.

A análise do funil de inclusão reforça essa leitura: embora o percentual de concluintes cotistas tenha crescido continuamente, de 0,8% em 2010 para 5,0% em 2023 (INEP, 2023a), a disparidade entre ingressantes e concluintes permanece ao longo de todo o período analisado. Esse resultado indica que a ampliação do acesso, embora fundamental, não é suficiente sem políticas efetivas de permanência estudantil, uma vez que as condições socioeconômicas influenciam diretamente o desempenho acadêmico e a conclusão dos cursos (PENA; MATOS; COUTRIM, 2020).

Em síntese, os três clusters identificados demonstram que os cotistas não constituem grupo homogêneo: há concentração dos benefícios nos estratos menos vulneráveis nos cursos de maior prestígio (C0), funcionamento conforme previsto nos de menor prestígio (C2) e um padrão intermediário de baixa inclusão racial (C1).

6.3 QP2 – Padrão Institucional de Preenchimento de Vagas

Esta seção responde à **QP2**: *as vagas reservadas estão sendo efetivamente preenchidas nos diferentes tipos de instituição e áreas de conhecimento, e os cotistas acessam cursos de maior prestígio econômico na mesma proporção que os não cotistas (under-matching)?*

6.3.1 Pré-processamento e definição dos clusters

A análise QP2 foi conduzida no nível da instituição, diferentemente de QP1 que operou no nível do curso. Cada registro corresponde a uma IES em um determinado ano, com indicadores calculados como média de todos os seus cursos com cotistas. A partir do CES consolidado (2013–2023), foram obtidos 2.220 registros, referentes a 492 IES únicas ao longo desses 11 anos.

Os indicadores utilizados como variáveis de entrada foram: percentual de ingressantes cotistas sobre o total (PCT_COTISTAS), percentual de ingressantes PPI (PCT_PPI), percentual de ingressantes oriundos de escola pública (PCT_ESCOLA_PUB), percentual de concluintes cotistas sobre o total de concluintes (PCT_COTISTAS_CONC) e prestígio médio dos cursos da IES (PRESTIGIO_NUM, em escala 0 = Baixo, 1 = Médio, 2 = Alto). O tipo de IES foi mantido apenas como variável descritiva, pelo mesmo motivo descrito em QP1: quando incluído como variável de entrada, domina o agrupamento por tipo de instituição em vez de por padrão de perfil. Todas as variáveis numéricas foram padronizadas pelo método *z-score*.

Para determinar o número ótimo de clusters, aplicou-se o método do cotovelo e o índice de silhueta para k de 2 a 10 (Figura 5). Os índices de silhueta variaram entre 0,188 ($k = 2$) e 0,244 ($k = 8$), com crescimento irregular e sem cotovelo claro — padrão semelhante ao observado em QP1, característico de dados institucionais com distribuição contínua. O valor de $k = 8$ foi sugerido pelo índice de silhueta, mas resultou em 8 clusters de difícil interpretação, com tamanhos muito desiguais (entre 96 e 582 registros). Diante disso, adotou-se $k = 3$ por critério de interpretabilidade, conforme justificado na Seção 6.2.

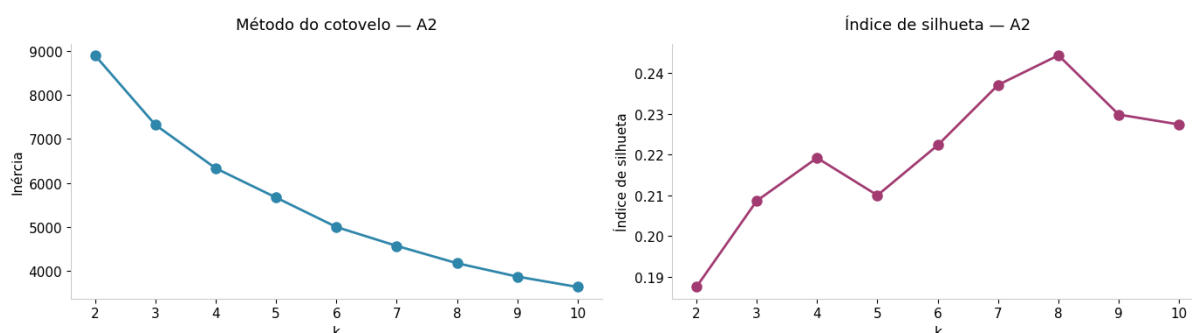


Figura 5 – Método do cotovelo e índice de silhueta.

Fonte: Elaboração própria.

6.3.2 Caracterização dos clusters institucionais

Os três clusters identificados pelo K-Means estão descritos na Tabela 5.

Tabela 5 – Caracterização dos clusters

Variável	C0 – Baixa inclusão		C1 – Alta inclusão PPI		C2 – Alta adesão	
	IES	%	IES	%	IES	%
IES (n)		836		939		445
% Cotistas		22,6		31,4		63,1
% PPI		25,8		61,1		33,7
% Esc. pública		59,0		82,5		79,6
% Conc. cotistas		9,8		11,7		41,5
Prestígio médio		1,3		1,0		1,1

Nota: prestígio em escala 0–2 (0 = Baixo, 1 = Médio, 2 = Alto).

Fonte: Elaboração própria a partir dos dados do CES/INEP (2013–2023).

Como em QP1, aplicou-se a Análise de Componentes Principais (PCA) para viabilizar a visualização, comprimindo as cinco dimensões em dois eixos. A projeção capturou 58,6% da variância total (PC1 = 33,2%; PC2 = 25,4%), valor ligeiramente inferior ao de QP1 pela dimensionalidade maior. Os valores reais estão na Tabela 5.

C0 – Baixa inclusão com prestígio médio (836 registros): cluster com baixo percentual de cotistas (22,6%) e o menor percentual de PPI (25,8%) entre os três grupos, combinados com prestígio médio relativamente alto (1,3). O percentual de ingressantes de escola pública (59,0%) é significativamente menor que nos demais clusters, sugerindo que essas IES atraem cotistas com perfil de origem escolar menos vulnerável. A distribuição por tipo de IES é equilibrada: 43,9% federais, 25,8% estaduais e 26,0%

privadas. O percentual de concluintes cotistas (9,8%) é o menor entre os clusters, indicando dificuldades de permanência ou baixa adesão histórica às cotas.

C1 – Alta inclusão PPI (939 registros): cluster com o maior percentual de PPI (61,1%) e o maior percentual de ingressantes de escola pública (82,5%), combinados com prestígio médio (1,0). Concentra majoritariamente IES federais (60,6%) e representa o funcionamento das cotas conforme previsto pela Lei nº 12.711/2012 — cotistas majoritariamente PPI, oriundos de escola pública. O percentual de concluintes cotistas (11,7%), embora modesto, é coerente com a defasagem temporal entre ingresso e conclusão.

C2 – Alta adesão com baixa inclusão racial (445 registros): cluster com o maior percentual de cotistas (63,1%) e o maior percentual de concluintes cotistas (41,5%), indicando alta adesão sistemática à política de cotas e boa taxa de permanência. Contudo, o percentual de PPI é de apenas 33,7%, o segundo menor entre os clusters, revelando que a alta adesão numérica às cotas não se traduz em inclusão racial proporcional. Esse padrão é predominante em IES estaduais (49,7%), sugerindo que essas instituições cumprem a reserva de vagas prioritariamente pela modalidade de escola pública e renda, sem atingir a mesma proporção de cotistas étnico-raciais.

A distribuição por tipo de IES dentro de cada cluster está consolidada na Tabela 6.

Tabela 6 – Distribuição por tipo de IES por cluster (%)

Cluster	Federal	Estadual	Municipal	Privada	Outras
C0 – Baixa inclusão	43,9	25,8	2,2	26,0	2,2
C1 – Alta PPI	60,6	18,0	0,3	21,0	0,1
C2 – Alta adesão	38,2	49,7	1,3	10,8	0,0

Fonte: Elaboração própria a partir dos dados do CES/INEP (2013–2023).

6.3.3 Validação por agrupamento hierárquico (Ward)

Para validar os agrupamentos obtidos pelo K-Means, aplicou-se o agrupamento hierárquico aglomerativo com critério de Ward (JR., 1963) sobre as mesmas *features* padronizadas. O dendrograma foi gerado a partir de uma amostra de 200 IES para viabilizar a visualização (Figura 6).

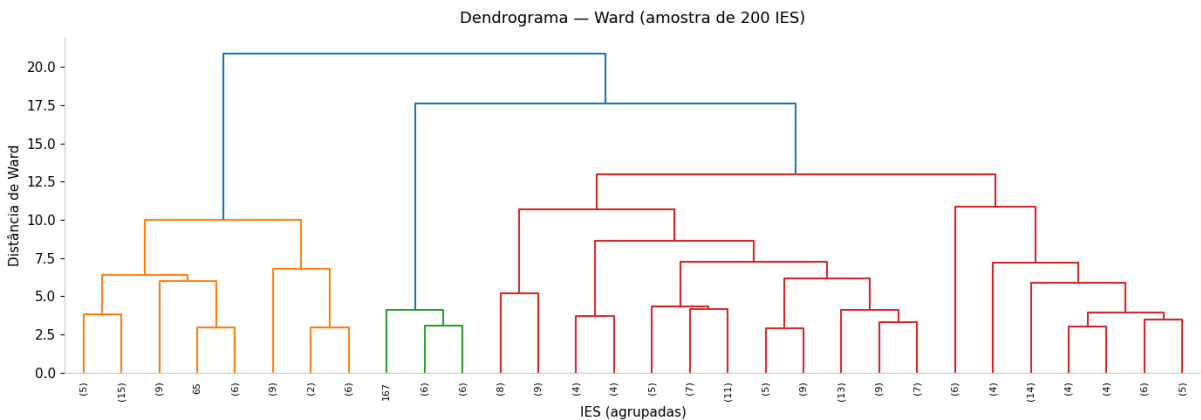


Figura 6 – Dendrograma — agrupamento hierárquico Ward.

Fonte: Elaboração própria.

A concordância global entre K-Means e Ward foi de 63,7%. Como os dois métodos numeraram os clusters de forma independente, os rótulos foram realinhados pelo algoritmo húngaro (HASTIE; TIBSHIRANI; FRIEDMAN, 2009) antes do cálculo. Após o realinhamento, 63,7% dos registros foram classificados no mesmo cluster pelos dois métodos, equivalente a aproximadamente 3 em cada 5 IES. Essa concordância moderada é esperada para dados com fronteiras difusas entre grupos institucionais.

Tabela 7 – Comparação entre K-Means e Ward

Variável	C0 – Baixa inclusão		C1 – Alta PPI		C2 – Alta adesão	
	K-Means	Ward	K-Means	Ward	K-Means	Ward
IES (n)	836	502	939	1578	445	140
% Cotistas	22,6	27,5	31,4	32,4	63,1	82,8
% PPI	25,8	33,9	61,1	46,1	33,7	29,9
% Esc. pública	59,0	44,5	82,5	81,1	79,6	85,1
Prestígio médio	1,3	1,3	1,0	1,1	1,1	1,1

Fonte: Elaboração própria a partir dos dados do CES/INEP (2013–2023).

A principal divergência entre os métodos ocorre em C1, onde o Ward agrupa 1.578 registros contra 939 do K-Means, incorporando parte dos registros de C0 e C2. Isso reduz o percentual de PPI de 61,1% para 46,1% no equivalente Ward, sugerindo que as fronteiras entre C1 e os demais clusters são difusas. Em C2, o Ward identifica um grupo menor (140 registros) mas mais homogêneo, com 82,8% de cotistas, padrão mais concentrado que o K-Means. Ainda assim, ambos os métodos convergem na identificação dos mesmos três perfis institucionais, e a concordância de 63,7% é considerada aceitável para dados com sobreposição natural entre instituições.

6.3.4 Discussão – *undermatching* e padrões institucionais

Os resultados da QP2 aprofundam os padrões identificados em QP1 no nível institucional.

O cluster C0 concentra IES com baixa inclusão de cotistas (22,6%) e baixa representação PPI (25,8%), mas com o maior prestígio médio entre os clusters (1,3). A composição institucional é equilibrada, sendo 43,9% federais e 26,0% privadas, indicando que a baixa inclusão não é exclusividade de um tipo de IES, mas um padrão comum em instituições com cursos de maior retorno econômico. Esse resultado converge com QP1: os cursos e IES de maior prestígio são justamente os que menos recebem e formam cotistas, evidenciando o fenômeno do *undermatching* (MEDEIROS; NETO; GOMES, 2016), em que cotistas elegíveis não acessam os cursos mais valorizados na mesma proporção que os não cotistas.

O cluster C1 reúne as IES com maior inclusão racial (61,1% de PPI) e maior proporção de ingressantes de escola pública (82,5%), sendo majoritariamente IES federais (60,6%). O prestígio médio (1,0) é o menor entre os clusters, confirmando o padrão identificado em QP1: as IES que mais recebem cotistas PPI são as que oferecem cursos de menor retorno financeiro. Embora esse cluster represente o funcionamento das cotas mais próximo do previsto pela Lei nº 12.711/2012, a concentração em áreas de menor prestígio reforça que a ampliação do acesso ainda não se traduziu em equidade na distribuição por áreas de conhecimento.

O cluster C2 apresenta o maior percentual de cotistas ingressantes (63,1%) e o maior percentual de concluintes cotistas (41,5%), indicando que o preenchimento das vagas reservadas ocorre conforme o esperado nessas IES, tanto no ingresso quanto na permanência até a conclusão. Contudo, o percentual de PPI é de apenas 33,7%, abaixo dos demais clusters de alta adesão, e a presença majoritária de IES estaduais (49,7%) sugere que essas instituições cumprem a lei prioritariamente pelas modalidades de escola pública e renda, sem atingir a representação étnico-racial proporcional prevista.

Em síntese, a QP2 evidencia que o cumprimento da política de cotas varia entre tipos de instituição: as IES federais concentram-se em C1 (maior inclusão racial, menor prestígio), as estaduais em C2 (alta adesão numérica, baixa equidade racial) e as privadas entre C0 e C1 (baixa inclusão geral). Isso sugere que a efetividade da Lei nº 12.711/2012 depende não apenas da obrigatoriedade legal, mas de fatores institucionais e regionais que condicionam a composição dos ingressantes por cotas (TREVISOLO; BELLO; NIEROTKA, 2023).

6.4 QP3 – Evolução Temporal entre os Períodos

Esta seção responde à **QP3**: *como o perfil racial e institucional dos cotistas, medido pelos indicadores de inclusão de pretos, pardos e indígenas (PPI), origem escolar e prestígio dos cursos, se transformou entre os períodos pré-lei (P1), implementação (P2), consolidação (P3) e pandemia (P4)?*

6.4.1 Pré-processamento por período

Para responder a essa questão, os dados foram organizados em quatro períodos sequenciais, permitindo avaliar o efeito da Lei nº 12.711/2012, sua consolidação e o impacto da pandemia.

As bases utilizadas foram o CES (2013–2023) e o ENADE (2010–2023), segmentadas nos quatro períodos definidos na Seção 4.2:

- **P1 – Pré-Lei (2010–2012)**: linha de base anterior à implementação da Lei nº 12.711/2012;
- **P2 – Implementação (2013–2016)**: introdução gradual das cotas nas instituições federais;
- **P3 – Consolidação (2017–2019)**: estabilização após o prazo legal de implementação plena;
- **P4 – Pandemia e pós (2020–2023)**: transformações no contexto pandêmico e pós-pandêmico.

Para cada período, foram calculados os indicadores médios de todos os registros do CES com cotistas: percentual de ingressantes cotistas (PCT_COTISTAS), percentual de ingressantes PPI (PCT_PPI), percentual de ingressantes oriundos de escola pública (PCT_ESCOLA_PUB), percentual de concluintes cotistas (PCT_COT_CONC) e prestígio médio dos cursos (PRESTIGIO). O resultado é uma matriz de quatro vetores (um por período) que representa a evolução estrutural do perfil dos cotistas ao longo do tempo. A Tabela 8 apresenta esses indicadores.

Tabela 8 – Indicadores médios por período (CES, 2010–2023)

Período	% Cot.	% PPI	% Esc. pub.	% Conc. cot.	Prestígio	Registros
P1 (2010–2012)	30,2	17,5	47,4	8,4	1,04	9.465
P2 (2013–2016)	35,3	36,0	65,2	11,0	1,09	24.439
P3 (2017–2019)	39,8	48,8	69,8	18,3	1,10	20.695
P4 (2020–2023)	37,3	47,4	71,6	27,3	1,11	29.862

Nota: prestígio em escala 0–2 (0 = Baixo, 1 = Médio, 2 = Alto).

Fonte: Elaboração própria a partir dos dados do CES/INEP (2010–2023).

A Figura 7 apresenta a evolução de cada indicador ao longo dos quatro períodos. Observa-se que o percentual de PPI saltou de 17,5% em P1 para 48,8% em P3, acompanhado pelo crescimento do percentual de ingressantes de escola pública (de 47,4% para 69,8%) e de cotistas ingressantes (de 30,2% para 39,8%). Em P4, houve leve retração em cotistas e PPI, enquanto o percentual de concluintes cotistas continuou crescendo — de 8,4% em P1 para 27,3% em P4 — reflexo da geração que ingressou em P2 e P3 concluindo seus cursos. O prestígio médio dos cursos permaneceu praticamente estável ao longo de todo o período (1,04 em P1; 1,11 em P4), indicando que, apesar do crescimento no acesso, os cotistas continuaram concentrados em cursos de prestígio médio, sem avanço estrutural rumo às áreas de maior retorno econômico, padrão coerente com os achados de QP1 e QP2.

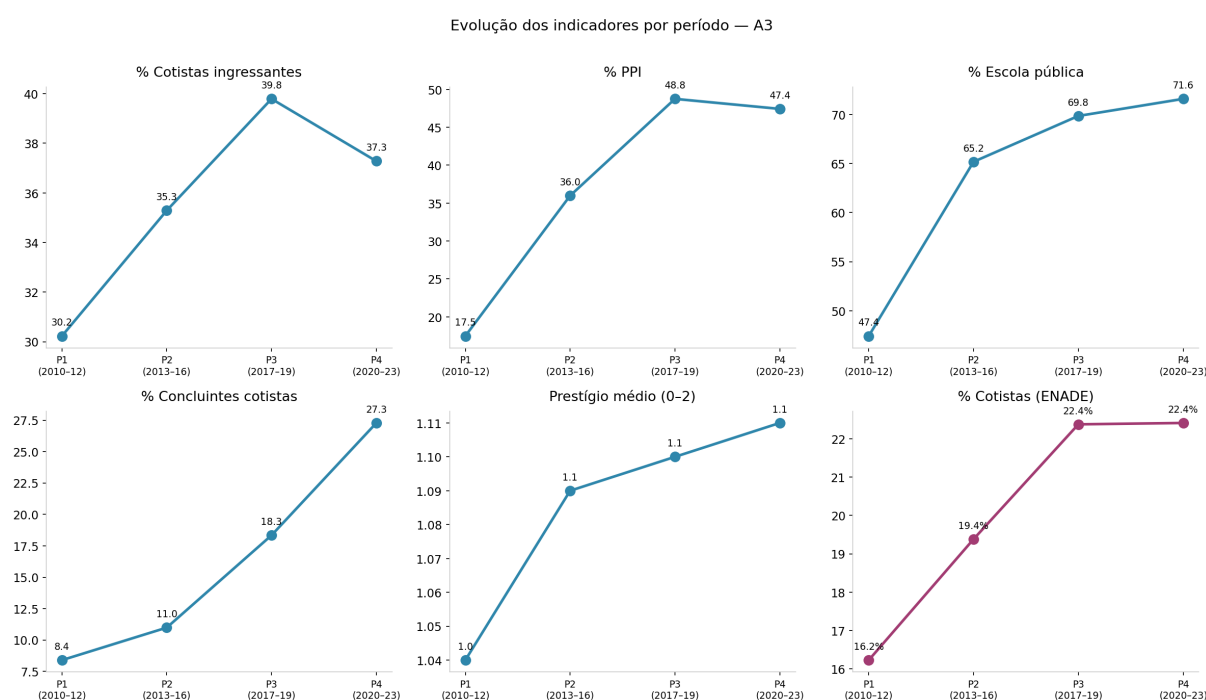


Figura 7 – Evolução dos indicadores por período

Fonte: Elaboração própria a partir dos dados do CES/INEP e ENADE/INEP (2010–2023).

6.4.2 Análise hierárquica e dendrograma

O agrupamento hierárquico aglomerativo com critério de Ward (JR., 1963) foi aplicado sobre os vetores padronizados dos quatro períodos, produzindo um dendrograma que evidencia a distância estrutural entre os perfis P1–P4 (Figura 8).

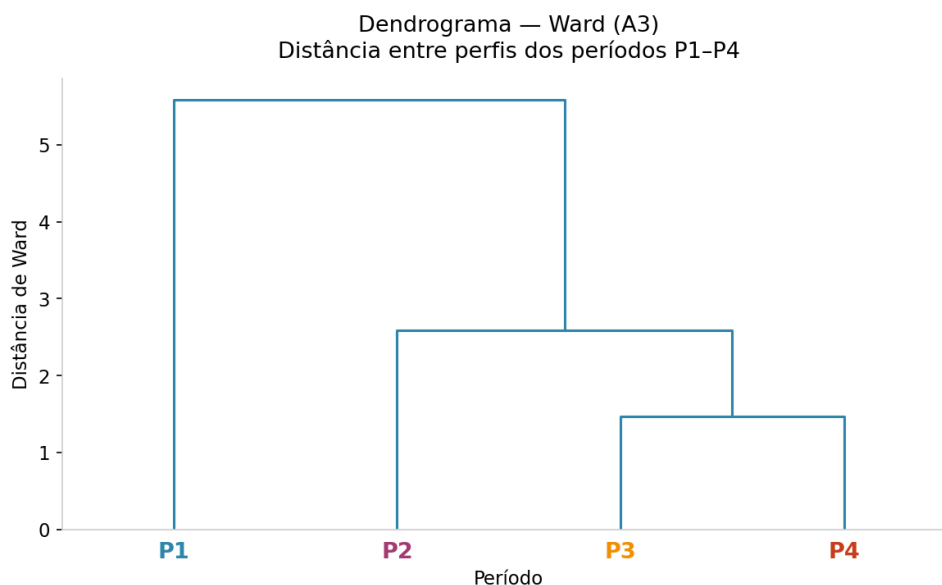


Figura 8 – Dendrograma dos períodos.

Fonte: Elaboração própria.

O dendrograma evidencia que P3 e P4 são os períodos mais similares entre si, fundindo-se na primeira ligação com distância de 1,476 (medida adimensional de variância intracluster). Em seguida, P2 se aproxima desse grupo com distância de 2,589. P1 permanece isolado até a última fusão (distância 5,578), confirmando que o período pré-lei apresenta perfil estruturalmente distinto de todos os demais.

Para aprofundar a análise, o agrupamento hierárquico foi também aplicado sobre os vetores anuais, permitindo visualizar a estrutura interna de cada período (Figura 9).

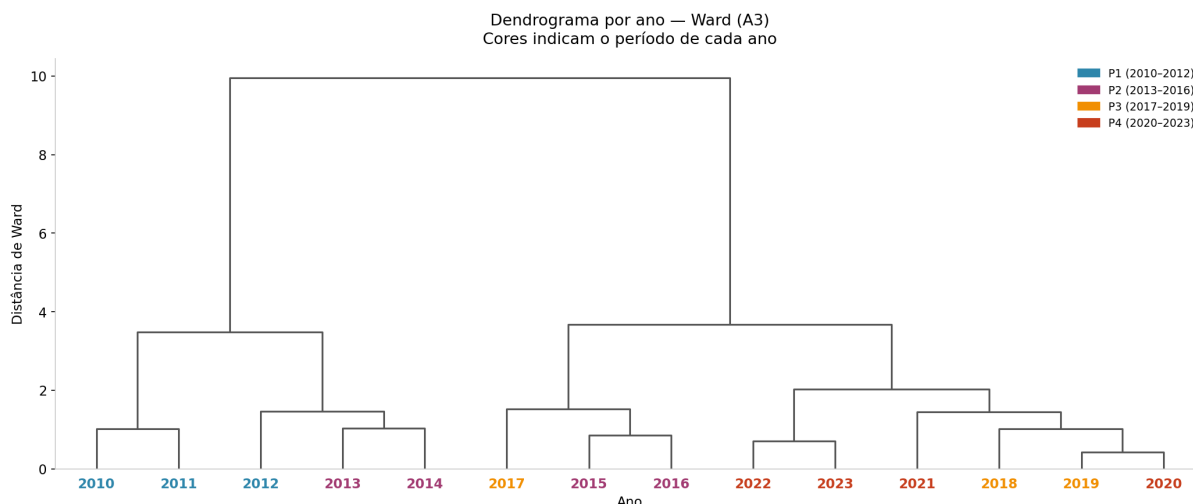


Figura 9 – Dendrograma por ano. Cores indicam o período de cada ano.

Fonte: Elaboração própria.

A análise por ano revela padrões de transição relevantes. Os anos de 2012 e 2013 agrupam-se no mesmo cluster, ainda que pertençam a períodos distintos (P1 e P2), reflexo da implementação gradual da Lei nº 12.711/2012, aprovada em agosto de 2012: seus efeitos sobre o perfil dos ingressantes já se manifestam ainda no mesmo ano, antes do início formal de P2. Esse resultado é coerente com o fato de que parte das IES federais antecipou a adequação ao novo sistema de reserva de vagas ainda em 2012 (SENKEVICS; MELLO, 2019).

Da mesma forma, os anos de 2015 e 2016 agrupam-se com 2017, cruzando a fronteira entre P2 e P3, indicando que a consolidação da política ocorreu de forma gradual, sem ruptura abrupta entre os períodos. O ano de 2018 foi o único com discordância entre K-Means e Ward, posicionando-se em uma zona de transição entre os anos de implementação plena (2015–2017) e os anos de estabilidade da política (2019–2023). Isso sugere que 2018 foi um ano de mudança gradual dentro do próprio período P3, em que os indicadores ainda estavam se ajustando antes de atingir os patamares que se manteriam estáveis até 2023.

Os anos de 2019 a 2023 agrupam-se consistentemente no mesmo cluster pelos dois métodos, formando o grupo mais coeso de toda a série histórica. Esse resultado indica que, a partir de 2019, o perfil dos cotistas atingiu uma estabilidade estrutural, tendo em vista que os indicadores de inclusão se mantiveram em patamares próximos mesmo durante a pandemia (2020–2021) e no período de retomada presencial (2022–2023). Embora P4 tenha registrado leve retração no percentual de cotistas ingressantes e de PPI em relação a P3, a magnitude da variação foi pequena o suficiente para não separar esses anos em clusters distintos, sugerindo que os efeitos da pandemia sobre o perfil dos ingressantes foram moderados e não reverteram os ganhos

estruturais acumulados nos períodos anteriores (CUNHA, 2021).

6.4.3 Validação por K-Means

O K-Means foi aplicado sobre os vetores anuais com $k = 4$ para validar os agrupamentos do dendrograma. A concordância global entre os dois métodos foi de 92,9%, calculada após realinhamento dos rótulos pelo algoritmo húngaro (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). O índice de silhueta de 0,426 é significativamente superior ao de QP1 (0,262) e QP2 (0,244), confirmando que a variação temporal é mais estruturada do que a variação entre cursos ou IES.

A Tabela 9 apresenta a classificação de cada ano pelos dois métodos, agrupada por período.

Tabela 9 – Classificação anual – K-Means e Ward

Período	Anos (cluster)	Concordância
P1	2010–2011 (C2), 2012 (C1)	Total
P2	2013–2014 (C1), 2015–2016 (C0)	Total
P3	2017 (C0), 2019 (C3)	2018 divergente (K-Means C0, Ward C3)
P4	2020–2023 (C3)	Total

Fonte: Elaboração própria a partir dos dados do CES/INEP (2010–2023).

6.4.4 Discussão – efeitos da Lei e da pandemia

Os resultados da QP3 evidenciam três movimentos estruturais distintos ao longo do período analisado.

P1→P2 – efeito da Lei (2012–2016): a transição entre os períodos pré e pós-lei produziu as maiores variações absolutas em todos os indicadores. O percentual de PPI cresceu 18,5 pontos percentuais (p.p.) e o percentual de ingressantes de escola pública cresceu 17,8 p.p., as maiores variações observadas em qualquer transição. Esse resultado confirma que a Lei nº 12.711/2012 produziu impacto imediato e expressivo sobre o perfil dos ingressantes, com efeito mais pronunciado nas IES federais, conforme evidenciado na desagregação por tipo de IES apresentada a seguir (SENKEVICS; MELLO, 2019), especialmente no que diz respeito à inclusão racial e à origem escolar.

P2→P3 – consolidação (2017–2019): o crescimento continuou, porém em ritmo menor. O percentual de cotistas ingressantes cresceu 4,5 p.p. e o de PPI 12,8 p.p. O percentual de concluintes cotistas registrou o maior salto relativo nessa transição (+7,3 p.p.), reflexo da geração que ingressou em P2 concluindo seus cursos em P3.

Esse padrão é consistente com a defasagem temporal entre ingresso e conclusão, geralmente de quatro a cinco anos (PENA; MATOS; COUTRIM, 2020).

P3→P4 – pandemia e pós (2020–2023): a pandemia de COVID-19 provocou leve retração no percentual de cotistas ingressantes (–2,5 p.p.) e de PPI (–1,3 p.p.), sugerindo que as condições adversas do ensino remoto emergencial afetaram desproporcionalmente os grupos mais vulneráveis (CUNHA, 2021). Contudo, o percentual de concluintes cotistas continuou crescendo de forma expressiva (+8,9 p.p.), resultado da geração que ingressou em P2 e P3 concluindo seus cursos em P4. O prestígio médio permaneceu estável (1,10–1,11), indicando que a distribuição dos cotistas entre áreas de conhecimento não sofreu alterações estruturais significativas no período pandêmico.

A desagregação por tipo de IES revela que o impacto da Lei nº 12.711/2012 foi heterogêneo entre as instituições (Figura 10). As IES federais apresentaram o crescimento mais expressivo em todos os indicadores: o percentual de PPI saltou de 20,0% em P1 para 51,5% em P3 (+31,5 p.p.), e o percentual de ingressantes de escola pública cresceu de 40,7% para 69,7% (+29,0 p.p.). Em P4, houve leve retração, de 51,5% para 49,3% em PPI e de 69,7% para 64,8% em cotistas, possivelmente associada aos impactos da pandemia sobre o acesso.

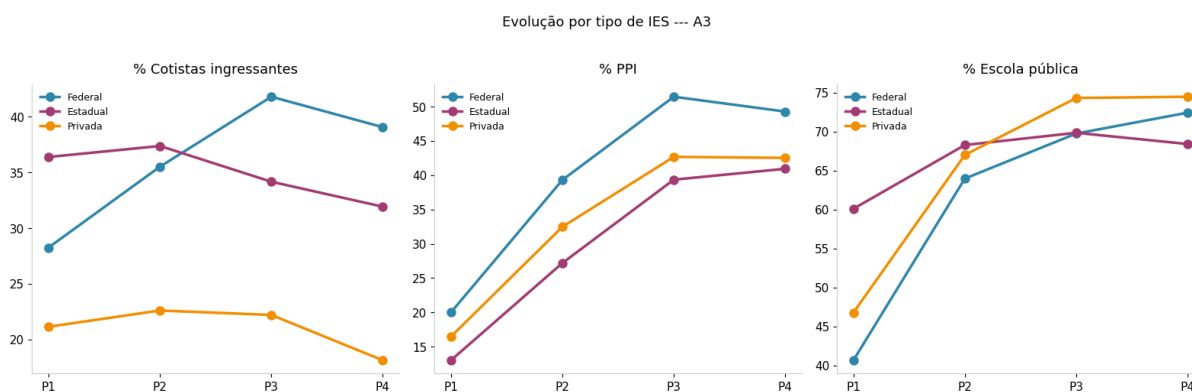


Figura 10 – Evolução dos indicadores por tipo de IES e período.

Fonte: Elaboração própria a partir dos dados do CES/INEP (2010–2023).

As IES estaduais apresentaram trajetória distinta: embora o percentual de cotistas já fosse relativamente alto em P1 (36,4%), o crescimento da inclusão racial foi mais lento. O percentual de PPI saiu de 13,0% em P1 para 40,9% em P4, crescimento expressivo mas concentrado nos períodos P2–P4, sem o salto imediato observado nas federais. Além disso, o percentual de cotistas ingressantes recuou de 37,4% em P2 para 31,9% em P4, sugerindo que as estaduais reduziram progressivamente a adesão às cotas após a consolidação da política, padrão coerente com o cluster C2 identificado em QP2, de alta adesão inicial com retração posterior.

As IES privadas mantiveram os menores percentuais de cotistas ao longo de todo o período (entre 18,1% e 22,6%), refletindo o caráter voluntário da reserva de vagas nesse segmento. O percentual de PPI, contudo, cresceu de 16,5% em P1 para 42,5% em P4, um crescimento proporcional semelhante ao das federais, indicando que, entre as privadas que adotaram cotas, houve avanço na inclusão racial mesmo sem obrigatoriedade legal.

O mapa de calor de variação (Figura 11) sintetiza essas três transições visualmente, destacando em verde os indicadores com crescimento e em vermelho os com retração.

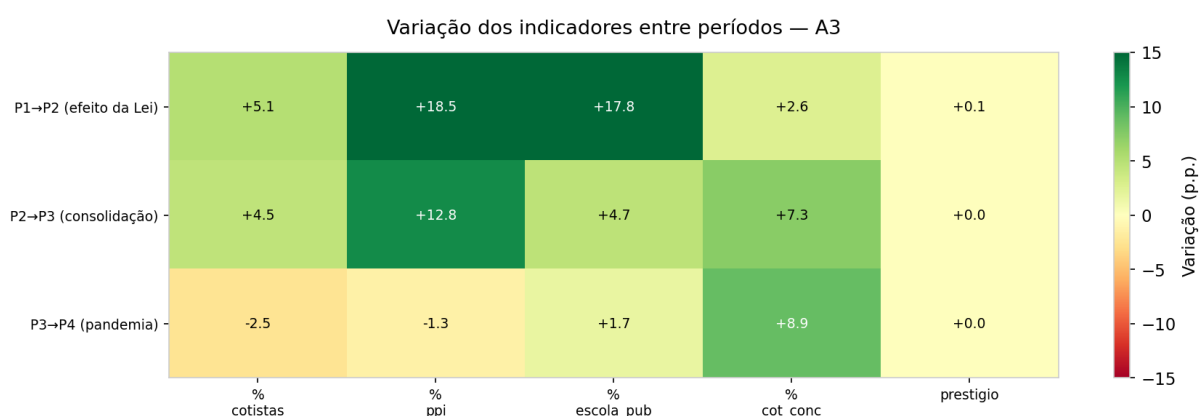


Figura 11 – Variação dos indicadores entre períodos

Fonte: Elaboração própria.

Em síntese, a análise QP3 evidencia que a política de cotas produziu avanços progressivos e consistentes entre P1 e P3, com impacto mais expressivo no momento de implementação da Lei (P1→P2). O período pandêmico (P4) interrompeu parcialmente essa trajetória no que diz respeito ao ingresso, mas não reverteu os ganhos de inclusão acumulados nos períodos anteriores. A leitura integrada com os resultados de QP1 e QP2 indica que, embora o acesso tenha crescido, a distribuição dos cotistas entre cursos de diferentes prestígios permaneceu relativamente estável ao longo dos quatro períodos, reforçando os padrões de estratificação horizontal identificados nas análises anteriores.

6.5 QP4 – Contextualização com o Censo Demográfico

Esta seção responde à **QP4**: *em que medida as mudanças observadas no perfil dos estudantes universitários refletem ou divergem da evolução das desigualdades estruturais de raça, renda e escolaridade na população brasileira geral?*

Para isso, os resultados de QP1, QP2 e QP3 são confrontados com os indicadores populacionais do Censo Demográfico, permitindo verificar se as transformações

internas às universidades acompanham a redução das desigualdades na sociedade brasileira.

Os dados do Censo de 2022 evidenciam a magnitude da desigualdade que as cotas buscam enfrentar. No que diz respeito ao acesso ao ensino superior, a diferença por cor ou raça é expressiva: 23,8% dos brancos com 25 anos ou mais haviam concluído o ensino superior, ante 10,7% dos pretos e 11,1% dos pardos (IBGE, 2022). Essa disparidade persiste mesmo após uma década de implementação da Lei nº 12.711/2012, indicando que as cotas contribuíram para reduzir o gap, mas não o eliminaram.

Essa desigualdade, contudo, não começa na universidade, ela se acumula desde a educação básica. Os anos médios de estudo por raça revelam trajetórias escolares desiguais: enquanto a população branca registra média de 10,3 anos de estudo, pretos e pardos chegam a 8,9 anos e indígenas a 7,5 anos (IBGE, 2022). Essa defasagem condiciona tanto o acesso quanto a permanência na graduação (SANTOS, 2009), uma vez que estudantes que chegam ao ensino superior com menor capital escolar acumulado enfrentam maiores dificuldades de adaptação e conclusão. Dessa forma, políticas que atuam exclusivamente no ingresso universitário enfrentam uma desvantagem estrutural já consolidada ao longo de toda a trajetória escolar anterior.

A distribuição dos graduados por área de formação e cor ou raça no Censo 2022 confirma os padrões de estratificação horizontal identificados em QP1 e QP2. As áreas de maior retorno econômico continuam sendo ocupadas desproporcionalmente por brancos, enquanto pretos e pardos se concentram em áreas de menor prestígio econômico, como Educação e Serviço Social (IBGE, 2022). Esse padrão se reproduz tanto nas IES públicas quanto nas privadas, com uma dimensão adicional relevante: a expansão acelerada do setor privado ao longo do período analisado (especialmente via EaD) ampliou significativamente a oferta de cursos de maior retorno econômico para estudantes com maior capacidade de pagamento, sem a exigência de reserva de vagas. Isso significa que, enquanto os cotistas das IES públicas concentram-se em cursos de prestígio médio e baixo, estudantes de maior renda acessam cursos de alto retorno no setor privado fora do sistema de cotas, aprofundando a estratificação entre os próprios graduados (ROZA, 2023). Esse mecanismo contribui para a manutenção da estrutura social: os grupos mais favorecidos continuam ocupando as posições de maior prestígio e retorno econômico no mercado de trabalho, independentemente da ampliação do acesso promovida pela Lei nº 12.711/2012. Esse padrão é coerente com o *under-matching* identificado em QP2 e com o *creamy layer* de QP1: mesmo entre os cotistas que acessam o ensino superior, a distribuição por áreas de conhecimento reproduz as desigualdades estruturais da população geral.

Os dados de rendimento médio mensal por nível de instrução e cor ou raça reforçam essa leitura. Mesmo entre os graduados, a diferença de rendimento entre brancos

e não brancos persiste, indicando que o diploma universitário não equaliza as condições de inserção no mercado de trabalho (IBGE, 2022). A análise dos clusters de QP1 e QP2 permite compreender parte desse mecanismo: os cotistas concentram-se majoritariamente em cursos de prestígio médio e baixo (como Educação, Serviço Social e Humanidades) que correspondem justamente às áreas de menor retorno salarial após a graduação. Assim, populações historicamente vulneráveis que ocupam essas posições no sistema universitário tendem a acessar empregos de menor remuneração após a conclusão, o que limita sua capacidade de mobilidade econômica e perpetua a desvantagem acumulada desde a educação básica. Esse ciclo de reprodução social — detalhado nas considerações finais (Capítulo 7) — indica que os efeitos das cotas sobre a mobilidade econômica dependem não apenas do acesso ao ensino superior, mas também do tipo de curso acessado e das condições de permanência e conclusão (PENA; MATOS; COUTRIM, 2020).

Em síntese, a leitura integrada dos dados universitários e populacionais revela que as políticas de cotas produziram avanços reais e mensuráveis, como o percentual de PPI no ensino superior que cresceu significativamente entre P1 e P3, e o aumento da presença de cotistas em áreas antes majoritariamente brancas. No entanto, as desigualdades estruturais da sociedade brasileira ainda se reproduzem dentro das universidades: a concentração de cotistas em cursos de menor prestígio, a baixa inclusão racial nos cursos de maior retorno econômico e a persistente disparidade entre ingressantes e concluintes indicam que o acesso, embora necessário, não é suficiente para romper o ciclo de reprodução social. Esses achados orientam as respostas às quatro perguntas de pesquisa, sistematizadas no capítulo seguinte.

7 Considerações Finais

As análises conduzidas neste trabalho permitem responder às quatro perguntas de pesquisa com base em evidências empíricas extraídas do ENADE, do CES e do Censo Demográfico.

A primeira pergunta, sobre o perfil dos cotistas e a presença do *creamy layer*, foi respondida pela análise QP1, que identificou três grupos distintos: C2, o maior cluster em número de registros, concentra cotistas com alto percentual de PPI (64,3%) e de ingressantes oriundos de escola pública (81,9%), acessando predominantemente cursos de prestígio médio e baixo — perfil que representa o funcionamento das cotas conforme previsto pela Lei nº 12.711/2012; C0 reúne cotistas em cursos exclusivamente de alto prestígio, com os menores percentuais de PPI (38,2%) e de escola pública (61,8%) entre os clusters, evidenciando a concentração dos benefícios nos estratos menos vulneráveis dentro dos elegíveis; C1 apresenta os menores valores em todas as dimensões de inclusão, em cursos de prestígio médio e baixo concentrados em IES estaduais, indicando que a sub-representação racial não se restringe aos cursos de maior prestígio. Esse padrão confirma que as cotas não atingem de forma homogênea os grupos mais vulneráveis, e que a estratificação horizontal persiste dentro do próprio sistema de reserva de vagas.

A segunda pergunta, sobre o padrão institucional de preenchimento de vagas e o *undermatching*, foi respondida pela análise QP2. As IES federais concentram-se em C1, com o maior percentual de PPI (61,1%) e de ingressantes de escola pública (82,5%), mas com baixo percentual de concluintes cotistas (11,7%), sugerindo dificuldades de permanência nesse segmento. As estaduais predominam em C2, com alta adesão às cotas (63,1% de cotistas) e boa taxa de permanência (41,5% de concluintes cotistas), mas com inclusão racial abaixo do esperado (33,7% de PPI), indicando cumprimento prioritário pelas modalidades de escola pública e renda. As privadas distribuem-se entre C0 e C1, com os menores percentuais de cotistas ao longo de todo o período, mas com presença crescente em cursos de maior prestígio e retorno econômico — padrão que aprofunda a estratificação entre os próprios graduados, conforme discutido na Seção 6.5. Em todos os clusters, os cotistas se concentram em cursos de prestígio médio (1,0 a 1,3 na escala 0–2), sem acesso proporcional às áreas de maior retorno econômico.

A terceira pergunta, sobre a evolução temporal da política, foi respondida pela análise QP3, que identificou três movimentos estruturais distintos. A transição P1→P2 produziu o maior impacto: +18,5 p.p. em PPI e +17,8 p.p. em ingressantes de escola

pública, reflexo direto da implementação da Lei nº 12.711/2012. A consolidação em P2→P3 manteve o crescimento, porém em ritmo menor, com o percentual de concluintes cotistas registrando o maior salto (+7,3 p.p.) pela geração que ingressou em P2 concluindo seus cursos. Em P3→P4, a pandemia provocou leve retração no ingresso (-2,5 p.p. em cotistas e -1,3 p.p. em PPI), sem reverter os ganhos acumulados. O prestígio médio dos cursos permaneceu praticamente estável ao longo de todo o período (1,04 em P1; 1,11 em P4), confirmando que a estratificação horizontal não foi alterada estruturalmente pela política de cotas ao longo dos 13 anos analisados.

A quarta pergunta, sobre a relação entre as transformações universitárias e as desigualdades populacionais, foi respondida pela análise QP4, que mostrou que os avanços no acesso universitário não eliminaram o gap racial registrado pelo Censo 2022: brancos ainda detêm 23,8% de graduados contra 10,7% de pretos e 11,1% de pardos. A distribuição dos graduados por área de formação reproduz os padrões de estratificação identificados em QP1 e QP2 — brancos concentrados em áreas de maior prestígio, pretos e pardos em áreas de menor retorno econômico —, e a persistência da diferença salarial entre graduados brancos e não brancos indica que o diploma universitário, por si só, não equaliza as condições de inserção no mercado de trabalho.

Transversalmente às quatro análises, os dados do ENADE e do CES revelam que a modalidade de escola pública é a mais acionada entre os cotistas, seguida pelas cotas de renda e étnico-racial. Isso significa que parte significativa dos beneficiários acessa as cotas por um critério menos restritivo do ponto de vista racial, o que contribui para o *creamy layer* observado em QP1.

Essa estrutura se retroalimenta: os mais privilegiados têm maior acesso à universidade e aos cursos de maior prestígio, melhores condições de permanência, maior probabilidade de conclusão, inserção em empregos de maior retorno econômico e capacidade de bancar educação básica de qualidade para as gerações seguintes. As cotas ajudam a alcançar a linha de largada, mas não equalizam os obstáculos até a linha de chegada — razão pela qual políticas de permanência são tão necessárias quanto as de acesso (PENA; MATOS; COUTRIM, 2020), (SILVA, 2018). Os resultados de QP1, QP2 e QP3 confirmam essa necessidade: em todos os clusters analisados, o percentual de concluintes cotistas é inferior ao de ingressantes, indicando que garantir o ingresso não é suficiente sem suporte efetivo à permanência.

Em síntese, a contribuição central desta pesquisa está em evidenciar, por meio da integração de múltiplas bases oficiais e de técnicas de agrupamento, que a efetividade das cotas não pode ser medida apenas pelo volume de ingressantes. Os padrões identificados nas quatro análises mostram que a política produziu avanços reais e mensuráveis no acesso, mas que a qualidade desse acesso e as condições de permanência permanecem como dimensões críticas, e é sobre elas que devem incidir as políticas

complementares para que o ingresso se converta em mobilidade social efetiva.

7.1 Limitações

Este trabalho apresenta limitações que devem ser consideradas na interpretação dos resultados. A principal delas é a reestruturação dos microdados do ENADE para adequação à LGPD (INEP, 2022), que impediu o uso de variáveis socioeconômicas individuais e restringiu a análise ao nível institucional agregado. Pesquisas que necessitem de dados individuais devem recorrer ao SEDAP/INEP.

A análise do perfil dos cotistas foi conduzida no nível institucional agregado, sem acompanhar individualmente a permanência, o desempenho e a inserção no mercado de trabalho de cada estudante após a conclusão. Os dados de concluintes do CES permitem inferências sobre permanência no nível agregado, mas não capturam as trajetórias individuais.

Além disso, a classificação de prestígio dos cursos foi operacionalizada com base no nome do curso e no rendimento médio de graduados do Censo 2022, o que pode não capturar nuances internas às áreas de conhecimento. O recorte temporal até 2023 também impede a avaliação dos efeitos das revisões recentes da Lei nº 12.711/2012, aprovadas após esse período.

7.2 Trabalhos Futuros

Os resultados deste trabalho apontam para quatro direções prioritárias de investigação. Primeiro, o uso dos dados individuais do ENADE via SEDAP/INEP permitiria analisar o perfil socioeconômico real dos cotistas e testar as hipóteses de *creamy layer* no nível do estudante, e não apenas institucional, oferecendo subsídios mais precisos para o aprimoramento da política de reserva de vagas.

Segundo, investigações sobre os determinantes da mobilidade social entre egressos cotistas — integrando dados de mercado de trabalho como a RAIS e a PNAD Contínua — permitiriam rastrear em que medida o acesso ao ensino superior se traduz em melhoria efetiva das condições econômicas dos beneficiários e de suas famílias, fechando o ciclo analítico que este trabalho não pôde completar.

Terceiro, estudos focados nos mecanismos de aceleração da inclusão nas IES são necessários. Os dados de QP3 mostram que as mudanças ocorrem de forma lenta: o prestígio médio dos cursos acessados por cotistas permaneceu praticamente estável ao longo de 13 anos. Pesquisas que identifiquem quais políticas institucionais e programas de ação afirmativa complementares foram mais eficazes em ampliar o acesso a cursos de maior prestígio poderiam orientar intervenções mais direcionadas.

Quarto, o aprimoramento das políticas de permanência estudantil merece atenção específica. Os resultados de QP1, QP2 e QP3 indicam persistência da disparidade entre ingressantes e concluintes cotistas, especialmente nas IES federais (C1 em QP2), onde o percentual de concluintes cotistas (11,7%) é significativamente inferior ao de ingressantes (31,4%). Pesquisas que avaliem a efetividade dos programas de assistência estudantil existentes (como o PNAES) e proponham modelos mais abrangentes de acompanhamento da trajetória acadêmica dos cotistas contribuiriam diretamente para que o acesso se traduza em conclusão e mobilidade social efetiva.

Referências

- BARBOSA, D. C. *Política de cotas e o desempenho entre cotistas e não-cotistas: uma análise para as instituições de ensino superior da rede pública no Brasil*. Dissertação (Dissertação (Mestrado em Economia Aplicada)) — Universidade Federal de Viçosa, Viçosa, 2021. Disponível em: <<https://locus.ufv.br/server/api/core/bitstreams/29212d39-7a7a-43f9-bfba-b7659c24cf95/content>>. Acesso em: 30 jul. 2025. Citado na página 21.
- BAYMA, F. Reflexões sobre a constitucionalidade das cotas raciais em universidades públicas no Brasil: referenciais internacionais e os desafios pós-julgamento das cotas. *Ensaio: Avaliação e Políticas Públicas em Educação*, v. 20, n. 75, p. 325–346, 2012. Disponível em: <<https://www.scielo.br/j/ensaio/a/xWK9mv8FbJ6NMnf53PvzQ9s>>. Citado 2 vezes nas páginas 11 e 20.
- BISHOP, C. M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006. Citado 2 vezes nas páginas 18 e 42.
- COSTA, M. d.; LIMA, M.; REIS, R. Entre o direito e o privilégio: a luta pelas oportunidades educacionais. In: *Desigualdade para inconformados: dimensões e enfrentamentos das desigualdades no Brasil*. Porto Alegre: Editora da UFRGS, 2020. Citado na página 20.
- CUNHA, L. H. A. *Universidades federais na pandemia da Covid-19: a falta de acesso à internet interdita mesmo o ensino?* [S.l.], 2021. Disponível em: <<https://www.econstor.eu/bitstream/10419/240831/1/td-2637.pdf>>. Acesso em: 30 jul. 2025. Citado 3 vezes nas páginas 23, 54 e 55.
- CUNHA, L. H. A. et al. Universidades federais na pandemia da covid-19: acesso discente à internet e ensino remoto emergencial. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 30, n. 115, p. 547–570, 2022. Disponível em: <<https://www.scielo.br/j/ensaio/a/53yPKgh7jK4sT8FGsYGn7cg>>. Acesso em: 30 jul. 2025. Citado na página 23.
- DOMINGUES, P. Ações afirmativas para negros no Brasil: o início de uma reparação histórica. *Revista Brasileira de Educação*, n. 29, p. 164–176, 2005. Disponível em: <<https://www.scielo.br/j/rbedu/a/g9K3wSLyhKn88LXn3GgJDvc/>>. Citado na página 11.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI Magazine*, v. 17, n. 3, p. 37–54, 1996. Disponível em: <<https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1230>>. Acesso em: 20 maio. 2026. Citado 2 vezes nas páginas 16 e 17.
- FERREIRA, A. et al. Ações afirmativas: análise comparativa de desempenho entre cotistas e não cotistas em uma universidade pública. *Revista Brasileira de Política e Administração da Educação*, v. 36, n. 3, p. 1297–1322, 2020. Citado na página 22.

FERREIRA, Z. L. et al. Ações afirmativas no Brasil e na África do Sul. *Tempo Social*, São Paulo, v. 31, n. 2, p. 251–273, 2019. Disponível em: <<https://www.scielo.br/j/ts/a/dWxr36GQ76SfDqs7gzYSCGy>>. Acesso em: 30 jul. 2025. Citado na página 14.

FRANCO, R. N. *Efeitos da política de cotas na Universidade de Brasília: uma análise do rendimento e da evasão*. [S.l.], 2008. Disponível em: <<https://biblioteca.flacso.org.br/files/2012/07/44.pdf>>. Acesso em: 30 jul. 2025. Citado 2 vezes nas páginas 15 e 21.

FRIAS, L. As cotas raciais e sociais em universidades públicas são injustas? *Revista Direito, Estado e Sociedade*, n. 41, p. 157–166, 2014. Disponível em: <<https://doi.org/10.17808/des.41.157>>. Citado 2 vezes nas páginas 15 e 20.

GOMES, J. B. *Ação afirmativa e princípio constitucional da igualdade: o direito como instrumento de transformação social*. Rio de Janeiro: Renovar, 2001. Citado na página 11.

GOMES, J. B. O uso simbólico da ação afirmativa. In: *Ações afirmativas: aspectos jurídicos e práticas sociais*. Brasília: OAB Editora, 2005. Citado 2 vezes nas páginas 11 e 20.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2. ed. New York: Springer, 2009. Disponível em: <<https://hastie.su.domains/ElemStatLearn/>>. Citado 9 vezes nas páginas 15, 16, 17, 18, 28, 40, 42, 48 e 54.

IBGE. *Censo Demográfico 2022*. 2022. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/populacao/22827-censo-demografico-2022.html>>. Acesso em: 20 maio. 2026. Citado 8 vezes nas páginas 12, 20, 24, 28, 36, 37, 57 e 58.

INEP. *Nota Técnica n.º 21/2022/DAES: Adequação dos microdados do ENADE às determinações da LGPD*. [S.l.], 2022. Disponível em: <https://download.inep.gov.br/microdados/nota_tecnica_21-2022-daes.pdf>. Citado 3 vezes nas páginas 26, 33 e 61.

INEP. *Censo da Educação Superior*. 2023. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-da-educacao-superior>>. Acesso em: 20 maio. 2026. Citado 5 vezes nas páginas 12, 24, 34, 37 e 44.

INEP. *ENADE – Exame Nacional de Desempenho dos Estudantes*. 2023. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade>>. Acesso em: 20 maio. 2026. Citado 3 vezes nas páginas 12, 24 e 37.

JOYE, C. R. et al. Educação a distância ou ensino remoto emergencial? impacto na educação durante a pandemia de COVID-19. *Research, Society and Development*, v. 9, n. 11, 2020. Citado na página 23.

JR., J. H. W. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, v. 58, n. 301, p. 236–244, 1963. Citado 3 vezes nas páginas 17, 47 e 52.

JÚNIOR, J. F.; DAFLON, V. T. Ação afirmativa na Índia e no Brasil: um estudo sobre a retórica acadêmica. *Sociologias*, v. 17, n. 40, p. 92–123, 2015. Citado na página 14.

JÚNIOR, J. F.; DAFLON, V. T. A nata e as cotas raciais: genealogia de um argumento público. *Opinião Pública*, v. 21, n. 2, p. 419–440, 2015. Citado 4 vezes nas páginas 12, 15, 26 e 43.

LAHIRE, B. *Sucesso escolar nos meios populares: as razões do improvável*. São Paulo: Ática, 1997. Citado na página 22.

MEDEIROS, H. A. V.; NETO, R. d. D. e. M.; GOMES, A. M. Limits of the quota law in federal public universities. *Education Policy Analysis Archives*, v. 24, n. 6, 2016. Citado 5 vezes nas páginas 12, 15, 20, 28 e 49.

OLIVEIRA, A. L. M. d. Perfil dos estudantes de graduação entre 2001 e 2015: uma revisão. *Avaliação: Revista da Avaliação da Educação Superior*, v. 26, n. 1, p. 157–177, 2021. Disponível em: <<https://www.scielo.br/j/aval/a/VmwKVwphJwPnKFyDdyZSgNq>>. Citado 2 vezes nas páginas 12 e 22.

PENA, M. A. C.; MATOS, D. A. S.; COUTRIM, R. M. d. E. Percurso de estudantes cotistas: ingresso, permanência e oportunidades no ensino superior. *Avaliação: Revista da Avaliação da Educação Superior*, Campinas; Sorocaba, v. 25, n. 1, p. 27–51, 2020. Disponível em: <http://educa.fcc.org.br/scielo.php?pid=S1414-40772020000100027&script=sci_arttext>. Acesso em: 20 maio. 2026. Citado 5 vezes nas páginas 22, 44, 55, 58 e 60.

PEREIRA, E. R.; DANTAS, M. L. No desligar das câmeras: experiências de estudantes de ensino superior com o ensino remoto no contexto da covid-19. *Civitas – Revista de Ciências Sociais*, v. 21, p. e40892, 2021. Disponível em: <<https://www.scielo.br/j/civitas/a/sh4bXpyhyNmPgSH4snCTrKx>>. Acesso em: 30 jul. 2025. Citado na página 23.

PEREIRA, S. M.; ZIENTARSKI, C. Políticas de ações afirmativas e pobreza no Brasil. *Revista Brasileira de Estudos Pedagógicos*, v. 92, n. 232, p. 493–515, 2011. Disponível em: <<https://www.scielo.br/j/rbep/a/D8mX7z9y9p7zG9z/>>. Citado 2 vezes nas páginas 11 e 14.

PEREIRA, S. R. d. S. *Determinantes do desempenho acadêmico: uma análise sobre as diferenças de cotistas e não cotistas*. Tese (Tese (Doutorado em Educação)) — Universidade Federal da Bahia, Salvador, 2017. Disponível em: <<https://repositorio.ufba.br/handle/ri/24148>>. Citado na página 22.

PINHEIRO, D. C.; PEREIRA, R. D.; XAVIER, W. S. Impactos das cotas no ensino superior: um balanço do desempenho dos cotistas nas universidades estaduais. *Revista Brasileira de Educação*, v. 26, 2021. Citado 2 vezes nas páginas 15 e 21.

PIOVESAN, F. Ações afirmativas no Brasil: desafios e perspectivas. *Revista Estudos Feministas*, v. 16, n. 3, p. 887–896, 2008. Disponível em: <<https://www.scielo.br/j/ref/a/JXPnmdcRhtfnv8FQsVZzFH>>. Citado na página 11.

PORTES, A. O trabalho escolar das famílias populares. In: CARVALHO, M. d. C. B. d. (Ed.). *A família contemporânea em debate*. São Paulo: Cortez, 2007. Citado na página 21.

Python Software Foundation. *Python Language Reference, version 3.11*. 2023. Disponível em: <<https://www.python.org>>. Acesso em: 20 maio. 2026. Citado na página 25.

RISTOFF, D. O novo perfil do campus brasileiro: uma análise do perfil socioeconômico do estudante de graduação. *Avaliação*, v. 19, n. 3, p. 723–747, 2014. Disponível em: <<http://educa.fcc.org.br/pdf/aval/v19n03/v19n03a10.pdf>>. Citado na página 22.

ROMANELLI, O. d. O. *História da Educação no Brasil (1930/1973)*. Petrópolis: Vozes, 2007. Citado 3 vezes nas páginas 20, 22 e 28.

ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, v. 20, p. 53–65, 1987. Citado na página 18.

ROZA, A. A. d. F. *As políticas de expansão e democratização da educação superior no Brasil reduziram desigualdades?: análise do perfil socioeconômico dos concluintes dos cursos de graduação no período entre 2011 e 2019*. Tese (Dissertação (Mestrado em Ciência Política)) — Universidade de São Paulo, São Paulo, 2023. Disponível em: <https://www.teses.usp.br/teses/disponiveis/8/8131/tde-30012024-174056/publico/2023_AliceAssisDeFigueiredoRoza_VOrig.pdf>. Acesso em: 30 jul. 2025. Citado 4 vezes nas páginas 12, 15, 44 e 57.

SANTOS, D. B. R. *Para além das cotas: a permanência de estudantes negros no ensino superior como política de ação afirmativa*. Tese (Tese (Doutorado em Educação)) — Universidade Federal da Bahia, Salvador, 2009. Disponível em: <<https://repositorio.ufba.br/bitstream/ri/11778/1/Tese%20%20Dyane%20Santos.pdf>>. Acesso em: 20 maio. 2026. Citado 2 vezes nas páginas 22 e 57.

SENKEVICS, A. S. Contra o silêncio racial nos dados universitários: desafios e propostas acerca da lei de cotas. *Educação e Pesquisa*, v. 44, 2018. Citado 2 vezes nas páginas 11 e 21.

SENKEVICS, A. S.; MELLO, U. M. O perfil discente das universidades federais mudou pós-lei de cotas? *Cadernos de Pesquisa*, v. 49, n. 172, p. 184–208, 2019. Citado 5 vezes nas páginas 12, 15, 21, 53 e 54.

SILVA, D. N. As propostas de políticas/ações afirmativas das universidades públicas e as políticas/ações de permanência nos ordenamentos legais. *Revista Brasileira de Estudos Pedagógicos*, v. 99, n. 251, p. 86–107, 2018. Disponível em: <<https://rbep.inep.gov.br/ojs3/index.php/rbep/article/view/3599/3334>>. Acesso em: 30 jul. 2025. Citado 4 vezes nas páginas 11, 12, 14 e 60.

SOUZA, E. M. d. Affirmative action and social stereotypes: Deconstructing the myth of minority inferiority. *Education Policy Analysis Archives*, v. 27, p. 75, 2019. Citado na página 21.

TRAGTENBERG, M. A universidade e a colonização. In: *Sobre educação, política e sindicalismo*. São Paulo: Editora UNESP, 2002. Citado na página 20.

TREVISOL, J. V.; BELLO, J. C. D.; NIEROTKA, R. L. A lei de cotas e as mudanças no perfil dos ingressantes das universidades federais brasileiras. *Série-Estudos*, v. 28, n. 64, p. 155–178, 2023. Disponível em: <<http://educa.fcc.org.br/pdf/sest/v28n64/>>

[1414-5138-sest-28-64-0155.pdf](#)>. Citado 8 vezes nas páginas 14, 15, 20, 21, 23, 31, 44 e 49.