



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE CIÊNCIAS AGRÁRIAS**  
**DEPARTAMENTO DE ECONOMIA AGRÍCOLA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA RURAL**

**ANTÔNIO CÉLIO FERREIRA DOS SANTOS**

**ENSAIOS SOBRE O ÍNDICE DE DESENVOLVIMENTO EDUCACIONAL DOS  
ALUNOS DA UNIVERSIDADE DA INTEGRAÇÃO INTERNACIONAL DA  
LUSOFONIA AFRO-BRASILEIRA (UNILAB)**

**FORTALEZA**

**2020**

ANTÔNIO CÉLIO FERREIRA DOS SANTOS

ENSAIOS SOBRE O ÍNDICE DE DESENVOLVIMENTO EDUCACIONAL DOS  
ALUNOS DA UNIVERSIDADE DA INTEGRAÇÃO INTERNACIONAL DA  
LUSOFONIA AFRO-BRASILEIRA (UNILAB)

Dissertação apresentada à Coordenação do Programa de Pós-Graduação em Economia Rural da Universidade Federal do Ceará (PPGER/UFC), como requisito parcial para obtenção do título de mestre em Economia Rural. Área de concentração: Políticas Públicas e Desenvolvimento Rural.

Orientador: Prof. Dr. Francisco José Silva Tabosa.

FORTALEZA

2020

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

S233e Santos, Antônio Célio Ferreira dos.  
Ensaio sobre o Índice de Desenvolvimento Educacional dos alunos da Universidade da Integração Internacional a Lusofonia Afro-Brasileira (UNILAB)/Antônio Célio Ferreira dos Santos. – 2020.  
64 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências Agrárias, Programa de Pós-Graduação em Economia Rural, Fortaleza, 2020.  
Orientação: Prof. Dr. Francisco José Silva Tabosa.

1. Sistema de cotas. 2. Propensity Score Matching. 3. Regressão Quantílica. 4. Desempenho acadêmico.

I. Título.  
CDD 338.1

---

ANTÔNIO CÉLIO FERREIRA DOS SANTOS

ENSAIOS SOBRE O ÍNDICE DE DESENVOLVIMENTO EDUCACIONAL DOS  
ALUNOS DA UNIVERSIDADE DA INTEGRAÇÃO INTERNACIONAL DA  
LUSOFONIA AFRO-BRASILEIRA (UNILAB)

Dissertação apresentada à Coordenação do Programa de Pós-Graduação em Economia Rural da Universidade Federal do Ceará (PPGER/UFC), como requisito parcial para obtenção do título de mestre em Economia Rural. Área de concentração: Políticas Públicas e Desenvolvimento Rural.

Orientador: Prof. Dr. Francisco José Silva Tabosa.

Aprovada em: \_\_\_ / \_\_\_ / \_\_\_\_.

BANCA EXAMINADORA

---

Prof. Dr. Francisco José Silva Tabosa (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. José Newton Pires Reis  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Pablo Urano de Carvalho Castelar  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Nicolino Trompieri Neto  
Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE)

À minha mãe, Francisca dos Santos Ferreira, e  
minha esposa Vanderli Ribeiro dos Santos, bem  
como meus filhos, Estéfane Ribeiro dos Santos,  
Luis Antônio Ribeiro dos Santos e Maria Ester  
Ribeiro dos Santos.

## AGRADECIMENTOS

A Deus, autor da minha vida, pelas infinitas graças alcançadas, força e determinação para seguir acreditando na realização dos meus sonhos.

À minha mãe, Francisca dos Santos Ferreira, por servir de base para a construção do meu caráter e da pessoa que sou hoje, e que mesmo sendo analfabeta teve a consciência de que a educação é a base para o crescimento profissional e engrandecimento como pessoa, e por isso sempre se esforçou para me dar educação.

À minha esposa, Vanderli Ribeiro dos Santos, com quem nos últimos 17 anos tenho compartilhado todas as minhas lutas e conquistas. Por acreditar e enaltecer sempre o meu potencial. E por acreditar que podemos crescer cada vez mais juntos.

Ao Prof. Dr. Francisco José Tabosa pelos conselhos, orientações e por me estender a mão do meu retorno ao mestrado, mesmo me conhecendo pouco, acreditou na minha área de estudo e aceitou me orientar.

Ao Prof. Robério Telmo Campos, por me incentivar a não desistir do curso, a retornar e concluir por acreditar no meu potencial.

Ao amigo de mestrado Domingos Isaias, por toda a colaboração e incentivo. Pessoas como você hoje são raras: que colocam o saber a serviço da coletividade.

Aos meus três filhos, Estéfane, Luis e Maria, razão pelo qual me levanto todos os dias com alegria, entusiasmo e perseverança, a eles dedico todo o meu esforço para o resultado deste trabalho. Destaco durante este meu retorno ao mestrado a minha primogênita Estéfane Ribeiro dos Santos, hoje já uma adolescente, que vibrou comigo por cada etapa conquistada.

Aos membros da Banca, Prof. Dr. José Newton Pires Reis, Prof. Dr. Pablo Urano de Carvalho Castelar e Prof. Dr. Nicolino Trompieri Neto, por me honrarem com a presença e pelas contribuições geradas. A todos os professores e funcionários que fazem parte do Programa de Pós-graduação em Economia Rural. A todos os colegas de curso com quem compartilhei momentos de alegria e de ansiedade.

“Enquanto houver vida, se realmente for seu objetivo, levante a cabeça, arregace as mangas e lute para conquistá-lo”. (Autoria própria).

## RESUMO

A presente dissertação é composta por dois ensaios que analisam o índice de desenvolvimento educacional dos alunos da Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB) referente ao primeiro semestre letivo de 2018. No primeiro ensaio, intitulado “Diferenciais de desempenho acadêmico dos alunos da UNILAB: Uma análise sob o prisma da Regressão Quantílica” buscou-se detectar efeitos nos quantis (ou percentis) da distribuição do índice de desenvolvimento educacional (IDE – Índice de Desenvolvimento Educacional), permitindo identificar e verificar se determinado fator tem maior efeito em estudantes de alto, médio ou baixo desempenho. Dentre os principais resultados, constatou-se que a modalidade presencial tem um efeito positivo no IDE, tal que os melhores desempenhos são nos alunos com maior IDE, também foi possível evidenciar que o melhor desempenho deve-se às mulheres em detrimento dos homens. Por outro lado, em relação a variável “reside em Redenção” o resultado apresentado neste estudo foi o inverso do esperado, fato este explicado por características peculiares dos alunos estrangeiros. No segundo ensaio, “Disparidades de Rendimento Acadêmico entre Cotistas e não Cotistas: Uma análise para a Universidade da Integração da Lusofonia Afro-Brasileira (UNILAB)”, objetivou-se verificar a existência (ou não) de disparidades de rendimento acadêmico (IDE) entre cotistas e não cotistas na UNILAB, associando para tanto a técnica de Propensity Score Matching (PSM) balanceado por entropia. Dentre os principais resultados encontrados, destaca-se que as cotas têm uma relação causal positiva, indicando um melhor desempenho para os beneficiários do sistema de cotas; os resultados apontaram que o IDE dos alunos cotistas foi de 77,152, valor este superior ao dos alunos não cotistas cujo resultado obtido foi de 76,075.

**Palavras-chave:** Sistema de Cotas. Propensity Score Matching. Regressão Quantílica. Desempenho acadêmico.



## ABSTRACT

The main purpose of this dissertation is to analyze the so-called academic achievement index of students who entered the *Universidade da Integração Internacional da Lusofonia Afro-Brasileira* (UNILAB) in 2018. The dissertation is divided into two parts, namely two essays. A major merit of the first essay, **Academic standard differentials from UNILAB students measured by means of the Quantum Regression**, lies in identifying what kind of particular (due to quantil index effect) factor contributes to the proper ranked grades performed by the students. As a remarkable result, the on-campus modality showed to be a positive parameter as for the academic achievement index (highest index applying to students with best academic achievements), having the better performance women students as compared with men students. In contrast, a result related to the explanatory variable “lives in Redenção” showed to be unexpected, this fact basically rests on the individual characteristics to foreign students. The second essay, **Academic success disparities from UNILAB quota and non-quota students**, aims at verifying by means of the Propensity Score Matching (PSM), entropy balancing included, whether exists (or not) conflictual academic achievement disparities relatively to quota and non-quota students. It is in it highlighted that quota students were better ranked, given their index was approximately 77,152 as compared with non-quota students, whose academic index counted 76.075.

**Keywords:** Quota system. Propensity Score Matching. Quantile regression. Academic achievement.

## LISTA DE FIGURAS

Figura 1 – Reserva de vagas de acordo com a Lei de Cotas 2012 .....	38
Figura 2 – Curva ROC: Logit e Probit .....	51
Figura 3 – Histograma dos escores de propensão, beneficiários e não beneficiários das Cotas na UNILAB 2018.1.....	54

## LISTA DE QUADROS

Quadro 1 – Definição das variáveis do modelo de regressão quantílica .....	25
Quadro 2 – Definição de variáveis do modelo PSM .....	48

## LISTA DE TABELAS

Tabela 1 – Estatística descritiva do modelo de regressão quantílica.....	27
Tabela 2 – Coeficientes das variáveis do modelo de regressão quantílica.....	29
Tabela 3 – Estatística descritiva do modelo PSM.....	49
Tabela 4 – Estatística descritiva entre cotistas e não cotistas.....	49
Tabela 5 – Estatística descritiva: Comparação Cotistas/Não cotistas.....	50
Tabela 6 – Critérios de seleção do modelo.....	51
Tabela 7 – Coeficiente do modelo Probit e seus efeitos marginais.....	52
Tabela 8 – Diferenças nas características observáveis antes e depois do pareamento.....	53
Tabela 9 – Resultado das estimativas de PSM balanceado por entropia.....	54
Tabela 10 – Análise de sensibilidade dos Limites de Rosenbaum.....	55

## LISTA DE ABREVIATURAS E SIGLAS

ATT	Efeito Médio do Tratamento no Tratado
CPLP	Comunidade dos Países de Língua Portuguesa
CV	Coeficiente de Variação
DTI	Diretoria de Tecnologia da Informação
EAD	Educação a Distância
ENADE	Exame Nacional de Desempenho dos Estudantes
ENEM	Exame Nacional do Ensino Médio
EUA	Estados Unidos da América
FEC	Função Expectativa Condicional
FQC	Função Quantílica Condicional
HIC	Hipótese de Independência Condicional
IBGE	Instituto Brasileiro de Geografia e Estatística
IDE	Índice de Desempenho do Estudante
IFES	Instituições Federais de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
MEC	Ministério da Educação
MQO	Mínimos Quadrados Ordinários
PARU	Programa de Avaliação da Reforma Universitária
PAIUB	Programa de Avaliação Institucional nas Universidades Brasileiras
PDI	Plano de Desenvolvimento Institucional
PSM	<i>Propensity Score Matching</i>
RLM	Regressão Linear Múltipla
ROC	<i>Receiver Operating Characteristic</i>
RQ	Regressão Quantílica
SIAC	Sistema Acadêmico
SINAES	Sistema Nacional de Avaliação da Educação Superior
UERJ	Universidade Estadual do Rio de Janeiro
UFBA	Universidade Federal da Bahia
UnB	Universidade de Brasília
UNILAB	Universidade da Integração da Lusofonia Afro-Brasileira

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO GERAL</b> .....	16
<b>2</b>	<b>DIFERENCIAIS DE DESEMPENHO ACADÊMICO DOS ALUNOS DA UNILAB: UMA ANÁLISE SOB REGRESSÕES QUANTÍLICAS</b> .....	18
<b>2.1</b>	<b>Introdução</b> .....	18
<b>2.2</b>	<b>Revisão da Literatura</b> .....	20
<b>2.3</b>	<b>Metodologia</b> .....	23
2.3.1	Bases de dados.....	23
2.3.2	Regressões quantílicas.....	25
<b>2.4</b>	<b>Resultados e Discussão</b> .....	27
<b>2.5</b>	<b>Considerações Finais</b> .....	30
<b>3</b>	<b>DISPARIDADES DE RENDIMENTO ACADÊMICO ENTRE COTISTAS E NÃO COTISTAS: UMA ANÁLISE PARA A UNIVERSIDADE DA INTEGRAÇÃO DA LUSOFONIA AFRO-BRASILEIRA (UNILAB)</b> .....	32
<b>3.1</b>	<b>Introdução</b> .....	32
<b>3.2</b>	<b>Revisão da Literatura</b> .....	35
3.2.1	Ações afirmativas no Brasil e no mundo.....	35
3.2.2	Ações afirmativas: uma breve discussão sobre a implantação de cotas no Brasil .....	37
<b>3.3</b>	<b>Metodologia</b> .....	40
3.3.1	Bases de dados.....	40
3.3.2	Método do Propensity Score Matching (PSM).....	41
3.3.2.1	<i>Balanceamento por entropia</i> .....	45
3.3.2.2	<i>Análise de sensibilidade</i> .....	46
<b>3.4</b>	<b>Resultados e Discussão</b> .....	48
3.4.1	Estatística descritiva .....	48
3.4.2	Resultados do Propensity Score Matching (PSM) .....	50
<b>3.5</b>	<b>Considerações Finais</b> .....	56
<b>4</b>	<b>CONCLUSÃO GERAL</b> .....	58
	<b>Referências</b> .....	59
	<b>APÊNDICE A</b> .....	63
	<b>APÊNDICE B</b> .....	65

## 1 INTRODUÇÃO GERAL

As desigualdades de oportunidades educacionais produzem grandes diferenças na qualidade de vida das pessoas por ocasionar mão de obra menos qualificada, disparidades de rendimento, e diminuem as chances de jovens mais pobres chegarem ao nível médio e superior. Neste contexto, a avaliação dos fatores que influenciam o desempenho educacional se torna muito importante, já que auxilia na formulação de estratégias para redução de desigualdades socioeconômicas e culturais.

Nos últimos anos, tem-se observado o esforço dos governos em aumentar tanto o acesso ao ensino quanto a sua qualidade. Aumentar a qualidade do ensino é um objetivo que tem sido amplamente buscado, em sendo tal finalidade alcançada, tem-se decerto importantes consequências positivas para a sociedade.

Em políticas públicas, as questões educacionais têm ganhado cada vez mais destaque. Há um olhar mais atento na análise econômica a este tema, o que vem gerando uma vasta literatura sobre o tema, que conta inclusive com periódicos especializados, e um número crescente de pesquisadores dedicados ao assunto (AMARAL, MENEZES-FILHO, 2011).

Ao longo das décadas, houve uma expansão da educação superior no Brasil a partir do aumento do número de Universidades Federais e crescimento do financiamento estudantil – ampliação dos programas como Prouni, FIES, entre outros. Houve claramente uma significativa preocupação com a universalização do ensino superior que é fundamental para que a nação consiga atingir um elevado grau de desenvolvimento no longo prazo. E partindo do pressuposto de que quantidade não significa necessariamente qualidade, torna-se importante que sejam feitas avaliações das políticas públicas na área da educação.

Neste trabalho faz-se uso de um estudo de caso da Universidade da Integração da Lusofonia Afro-Brasileira (UNILAB), que tem por características, desde sua criação, o caráter de implementação de políticas públicas voltadas a questão racial e a consequente política afirmativa, internacionalização, bem como atender às demandas de interiorização do ensino superior.

O foco principal desta dissertação é analisar o desempenho educacional dos alunos da UNILAB a partir do IDE – Índice de Desenvolvimento Educacional. De modo a cumprir tal desiderato, o trabalho, escrito na forma de ensaio, está estruturado em dois capítulos, a saber: no primeiro capítulo, sob o prisma das regressões quantílicas, objetiva-se detectar efeitos nos quantis (ou percentis) da distribuição do IDE, a fim de identificar se determinado fator tem maior efeito em estudantes de alto, médio ou baixo desempenho.

Partindo-se do pressuposto segundo o qual dentre as atuais políticas públicas desenvolvidas no Brasil relativas à área da educação, destaca-se o sistema de cotistas, procurou-se, então, no segundo capítulo, verificar se os desempenhos dos alunos cotistas são semelhantes aos alunos não cotistas, a partir da técnica de *Propensity Score Matching*, o qual espera-se ser capaz de demonstrar a existência (ou não) de disparidades de rendimento acadêmico entre cotistas e não cotistas na UNILAB.



## **2 DIFERENCIAIS DE DESEMPENHO ACADÊMICO DOS ALUNOS DA UNILAB: UMA ANÁLISE SOB REGRESSÕES QUANTÍLICAS**

O presente capítulo visa compreender quais fatores afetam o desempenho acadêmico dos estudantes universitários da UNILAB, a partir dos resultados oriundos do IDE, com uso do modelo de regressões quantílicas.

### **2.1 Introdução**

Ao longo das últimas décadas no Brasil, nota-se que a avaliação em larga escala vem se consolidando como um instrumento das políticas públicas em educação, em seus diferentes âmbitos de gestão, seja em nível federal, estadual ou municipal. No entanto, ainda se observam inúmeros desafios quanto ao uso efetivo dessas avaliações na formulação, reformulação e no monitoramento de ações, projetos e programas educacionais, e, notadamente, no que diz respeito ao trabalho no interior das unidades escolares (GIMENES, 2015).

Além de serem avaliados oficialmente pelos órgãos federais, dentre os quais se destaca a avaliação dos alunos no ENADE (Exame Nacional de Desempenho de Estudantes), as IFES (Instituições Federais de Ensino Superior) criam suas próprias formas para avaliar seus discentes a partir de critérios pré-estabelecidos e aprovados em seus conselhos universitários. Estas formas de avaliação podem ajudar a entender os diferenciais de desempenhos entre os alunos e, conseqüentemente, podem ser utilizadas para a formulação de políticas públicas (pelos governos) e na gestão interna das universidades com a finalidade de melhorar a eficiência em seus resultados.

Nos estudos acerca dos fatores associados ao desempenho escolar, seja nos níveis fundamental, médio ou superior, tem destaque a metodologia de modelos hierárquicos, uma vez que existem fatores conjuntamente associados ao estudante e à escola onde estuda, ou seja, há um nível de hierarquia primário – os estudantes –, e um nível agregado, nomeadamente as escolas (ALVES; SOARES, 2008; NASCIMENTO, 2003; PONTES; SOARES, 2016).

Tendo em vista que a maioria das formas de analisar o desempenho de um estudante numa avaliação de larga escala é, geralmente, medida por uma escala contínua, a maioria dos estudos usa os denominados modelos hierárquicos lineares. Os modelos lineares hierárquicos podem ser considerados extensão da regressão linear múltipla para acomodar uma estrutura hierárquica. Por intermédio do modelo adotado nesse estudo, procura-se avaliar a parcela do desempenho devida a cada fator associado ao estudante, nomeadamente a modalidade de ensino (presencial ou à distância), renda familiar, tamanho da família, raça, sexo, local de residência e nacionalidade).

Tradicionalmente, os modelos relacionam o valor médio do desempenho dos estudantes com os diversos fatores extra e intraescolares. Mas muitos desses fatores podem de maneira diferente afetar estudantes que apresentam baixo e alto desempenhos. Notadamente, nas políticas educacionais que buscam melhorar o desempenho dos alunos, há interesse em encontrar fatores intraescolares com maior efeito em alunos que registram baixo desempenho.

Neste estudo, procura-se analisar o caso específico da Universidade da Integração da Lusofonia Afro-Brasileira (UNILAB), que tem por características, desde sua criação, levar a bom termo a implementação de políticas públicas direcionadas a questão racial (e a consequente política afirmativa), internacionalização, e atender as demandas de interiorização do ensino superior.

A partir de informações relativas aos estudantes da UNILAB, busca-se verificar quais fatores afetam o desempenho acadêmico desses universitários, recorrendo-se aos seus resultados registrados pelo IDE disponibilizados institucionalmente pelas Diretoria de Tecnologia da Informação (DTI) e o Centro de Processamento de Dados, responsável pelo Sistema Acadêmico (SIAC) da Instituição.

Dado o exposto, avalia-se o efeito dos diferentes fatores associados não só na média do desempenho dos estudantes da UNILAB (medidos pelo IDE), mas também em diversos quantis da distribuição, por meio de Regressões Quantílicas, permitindo assim verificar se determinado fator tem maior efeito para o correspondente desempenho (alto, médio e baixo) do estudante.

Esta pesquisa, além desta introdução, é composta por mais quatro seções. Na seção dois se encontra a revisão da literatura sobre avaliação, desempenho acadêmico, na qual está especificada a evolução da experiência brasileira. A metodologia desta pesquisa está na terceira seção, na qual consta a descrição do modelo proposto (regressões quantílicas) e a análise do problema aqui destacado, bem como a apresentação da estatística. Os resultados e as discussões estão na quarta seção, seguidas pelas considerações finais do estudo.

## 2.2 Revisão da Literatura

A avaliação é uma tarefa didática necessária e permanente do trabalho docente, que deve acompanhar passo a passo o processo de ensino e aprendizagem. Através dela, os resultados obtidos no decorrer do trabalho conjunto do professor e dos alunos são comparados com os objetivos propostos a fim de constatar progressos, dificuldades, e reorientar o trabalho para as correções necessárias (LIBÂNEO, 1994).

Espera-se que, pelo ato de avaliar, possa ser analisada a qualidade do desempenho dos alunos. A literatura internacional costuma elencar fatores determinantes ao desempenho acadêmico. A título de ilustração, os fatores familiares e as características pessoais – idade, renda domiciliar, escolaridade dos pais, sexo, raça, entre outros constituem um desses determinantes. Questões relacionadas à qualidade do Ensino Médio são bastantes evidenciadas pela literatura internacional, a saber: recursos da escola, qualificação dos professores, passando pelo próprio desempenho do aluno nessa época que antecede a sua entrada na Universidade. Ou seja, o exame de entrada, comumente conhecido no Brasil como vestibular, também surge como fator explicativo à performance acadêmica dos alunos universitários (FARIAS, 2013).

Autores como Furtado (1954), Holanda (1956) e Freire (1951) já traziam o debate sobre o tema da qualidade da educação brasileira. Porém, o tema ocupava espaço marginal nas obras mais gerais acerca do pensamento social brasileiro. Vale ressaltar que, mais recentemente, diversos trabalhos que analisavam a qualidade levavam em consideração os recursos investidos na área.

O processo de avaliação do Ensino Superior no Brasil é bastante recente: remonta ao início da década de 1960, coincidindo com o primeiro boom da educação superior no país (ZAINKO, 2008). Em 1983, foi criado o Programa de Avaliação da Reforma Universitária (PARU), que vigorou até 1984 e que tinha como intuito gerar dados para fomentar estudos e discussões sobre tal questão. Em razão do PARU, observou-se a ausência de parâmetros necessários para a avaliação de desempenho, a qual, por sua vez, contribuiu para determinar de forma efetiva alocação dos recursos públicos no setor.

Na década de 1990, o governo brasileiro passou a enfatizar os processos de avaliação de ensino em bases que visasse a qualidade. Em 1993, no governo Itamar Franco, foi criada pela Secretaria da Educação Superior (SEC) a Comissão Nacional de Avaliação (CNA), cuja responsabilidade repousava na implantação de um Programa de Avaliação Institucional (PAIUB) nas universidades brasileiras. Assim, o MEC passava a articular, viabilizar e financiar a avaliação do ensino superior (ZAINKO, 2008).

Pela Lei nº. 9.131, de 24 de novembro de 1995, cria-se o Exame Nacional de Cursos, popularmente conhecido por Provão, numa tentativa pioneira de objetivamente avaliar a performance das IES e detectar a aptidão dos estudantes em fase de conclusão dos seus respectivos cursos de graduação.

Não obstante, a partir da Lei nº 10.861, de 14 de abril de 2004, criou-se o Sistema Nacional de Avaliação da Educação Superior (SINAES) que, dentre seus instrumentos para medir a qualidade da educação superior, conta com o Exame Nacional de Desempenho de Estudantes – ENADE (INEP, 2013). A partir dessa data, o ENADE substituiu o Provão.

Nos últimos anos, com mais frequência tem sido a discussão acerca da natureza do desempenho educacional no Brasil. Em geral, tem-se enfatizado apenas alguns determinantes em detrimento de outros como evidencia a literatura empírica nesta área. Por exemplo, segundo Barros e Mendonça (1996), o ambiente familiar tem seguramente impacto sobre o desempenho educacional dos estudantes brasileiros. Fatores como este são frequentemente abordados pelos autores que constituem a literatura internacional acerca do tema.

Com efeito, apesar da melhoria educacional verificada no Brasil nos últimos 20 anos, a evolução da educação foi mais lenta do que a dos demais países da América Latina (MENEZES-FILHO, 2001). De fato, mesmo com a ampliação, nos últimos anos, do acesso à educação básica, no Brasil, para grande parte das classes de renda mais baixa, a demanda pelo ensino superior cresceu intensamente.

Sendo assim, apesar do considerável crescimento das vagas universitárias, os jovens pobres e negros seguem com baixa representação entre os ingressantes na universidade pública – em geral, a mais concorrida (MENEZES-FILHO *et al.*, 2019).

Atualmente, a avaliação é um dos pilares da política educacional do Ministério da Educação (MEC). As três principais inovações inseridas no Plano de Desenvolvimento da Educação (PDE), criado em 2007 pelo MEC, foram: i) incorporação dos objetivos de accountability; ii) criação de um indicador sintético da qualidade da educação básica, que considera tanto o desempenho dos estudantes em exames padronizados quanto a progressão desses alunos no sistema; e iii) definição de metas seja para o país, seja para cada sistema e escola em particular (FERNANDES; GREMAUD, 2009).

Amaral e Menezes-Filho (2008) estudaram a relação entre os gastos educacionais e o desempenho escolar no Brasil, e neste trabalho desenvolvido por eles, fica clara a grande dispersão existente nos gastos municipais no ensino fundamental: enquanto a média dos gastos é de R\$1.015,34, seu desvio-padrão é de R\$1.165,42. Além disso, o artigo atenta para o fato segundo o qual a relação entre gastos educacionais e qualidade do ensino é significativa apenas

para os municípios com notas mais altas na Prova Brasil.

Do ponto de vista geral, o governo brasileiro intensificou, de fato, a avaliação em larga escala a partir do início da década de 1990 e encontrou contexto particularmente fértil para sua consolidação a partir de 1995 (BONAMINO, 2002). Seguindo os desenvolvimentos já em curso, em 1996 a LDB explicitou que “compete à União assegurar o processo nacional de avaliação do rendimento escolar no ensino fundamental, médio e superior, em colaboração com os sistemas de ensino, objetivando a definição de prioridades e a melhoria do ensino” (BRASIL, 1996, p. 67).

Outro aspecto que deve ser levado em consideração ao se analisar os desempenhos dos alunos, apontados por diversos autores, é a renda, tomada como fator determinante, ou seja, quanto maior a renda melhor o desempenho dos alunos. Como exemplo, pode ser citado o estudo longitudinal, com a base de dados do Saeb, elaborado por Leon e Menezes-Filho (2003), em que consta a classificação dos níveis socioeconômicos em quantis e o efeito da seleção na progressão escolar a partir dos níveis de renda.

As simulações mostraram que a renda se tornou mais determinante na probabilidade de reprovação ao longo do período, revelando que, para a 8ª série e o 3º ano do ensino médio, os estudantes com maior renda apresentam menor chance de reprovação sob os efeitos marginais de 1996-1997, do que se deparassem com os efeitos marginais de 1984-1985.

Vale destacar que no Brasil as universidades criam suas próprias formas de avaliação de seus discentes com base em critérios pré-estabelecidos e aprovados em seus conselhos universitários, e que nos últimos anos o Brasil vem democratizando o acesso ao ensino superior, objetivando amenizar a dívida histórica com os afrodescendentes e diminuir o hiato existente entre alunos que concluem o ensino médio nas redes públicas e privadas.

Neste estudo, a qualidade a ser analisada corresponde ao desenvolvimento dos alunos da UNILAB medidos por um índice cuja fórmula parte de uma definição interna da própria universidade, o IDE – Índice de Desenvolvimento do Estudante.

De fato, conhecer os principais fatores que afetam o desempenho estudantil auxilia os gestores e os professores a orientarem suas ações de forma mais objetiva e pragmática, com a conseqüente melhoria no desempenho acadêmico e na diminuição das diferenças entre eles.

## 2.3 Metodologia

A metodologia será apresentada em duas subseções: Base de dados e o modelo de regressões quantílicas.

### 2.3.1 Bases de dados

A metodologia proposta é aplicada no sistema de avaliação do desempenho do discente dos alunos da graduação da UNILAB, o qual é calculado pelo Índice de Desempenho do Estudante (IDE). A avaliação envolveu uma amostra de estudantes ativos nos cursos de graduação presencial e à distância da UNILAB, no ano de 2018 que fizeram matrícula no semestre 2018.1. A escolha do semestre 2018.1 se deu por ser o período de consolidação de uma base de dados mais sólida e confiável feita a partir de estudos, aperfeiçoamento de coletas de dados e discussões internas na universidade, e que proporcionaram a partir de então, obter informações de variáveis importantes para as análises dos modelos propostos nesta pesquisa.

Para cada elemento da amostra, existem informações socioeconômicas declaradas durante a realização da inscrição no processo seletivo e algumas de quando ingressado na universidade. Os dados foram coletados junto à Diretoria de Tecnologia da Informação (DTI), no Sistema Acadêmico (SIAC) e pelo Centro de Processamento de Dados. Os dados obtidos apresentam, além das habituais informações socioeconômicas dos ingressantes, informações sobre número de matrícula e acompanhamento acadêmico do estudante no decorrer do curso. A amostra abrange 2967 observações, entre alunos que ingressaram por meio do sistema de reserva de vagas, vestibular de ampla concorrência e processos específicos para alunos estrangeiros e EAD.

As variáveis socioeconômicas utilizadas na análise, com exceção das variáveis *renda familiar*, IDE e *tamanho da família*, as demais variáveis são dicotômicas, com valor igual à unidade, indicando que aquele aspecto pode ser atribuído ao indivíduo, caso contrário, atribui-se o valor zero.

A UNILAB acompanha o desempenho de seus alunos baseado numa avaliação de aprendizagem, aprovada no Conselho universitário na resolução 27 de 11 de novembro de 2014. Esta avaliação foi constituída como:

- Parte integrante do ensino aprendizagem;
- Processo que subsidia a reflexão e prática docente;
- Elemento constitutivo do processo formativo do estudante.

A avaliação de aprendizagem na UNILAB visa o êxito acadêmico do discente de graduação calculado por um componente curricular, em que se combina o aproveitamento

acadêmico e a assiduidade. O aproveitamento acadêmico consiste no êxito discente de cada componente em cada disciplina e a assiduidade se refere a frequência regular às atividades acadêmicas de cada componente curricular.

O desempenho do discente da graduação é calculado pelo IDE, isto é, trata-se da média das notas finais das disciplinas ponderada pela carga horária respectiva, multiplicada por um redutor por reprovação (falta e/ou trancamento) de disciplina, conforme exposto abaixo:

$$IDE = \left( \frac{\sum_{i=1}^n NO_i \cdot CH_i}{\sum_{i=1}^n CH_i} \right) \left( 1 - \frac{0,3 \cdot NTR + 0,7 \cdot NRF}{NT} \right) \quad (1)$$

Em que:

- IDE – Índice de Desempenho do estudante;
- $i$  –  $i$ -ésima disciplina que o estudante concluiu, com ou sem reprovação (com exceção daquelas em que houve reprovação por falta);
- $n$  – total de disciplinas concluídas, com ou sem reprovação (com exceção daquelas em que houve reprovação por falta);
- NTR – número de disciplinas trancadas;
- NRF – número de disciplinas com reprovação por falta;
- NT – número total de disciplinas que o aluno teve matrícula efetivada;
- $NO_i$  – nota em que o aluno obteve em uma dada disciplina  $i$ ; e,
- $CH_i$  – carga horária da disciplina  $i$ .

OBS1: O primeiro fator,  $\left( \frac{\sum_{i=1}^n NO_i \cdot CH_i}{\sum_{i=1}^n CH_i} \right)$  representa a fórmula básica, que é a média ponderada das notas obtida nas disciplinas pela carga horária. Exclui-se desta fórmula as notas e carga horária da disciplina em que o estudante obteve reprovação por falta.

OBS2: O segundo fator,  $\left( 1 - \frac{0,3 \cdot NTR + 0,7 \cdot NRF}{NT} \right)$  é um fator de redução. Caso o estudante não tenha apresentado trancamentos ou reprovações por falta, a segunda parcela do fator de redução se anula e a fórmula é multiplicada por 1, não havendo qualquer perda para o estudante.

Estima-se o modelo a partir de variáveis socioeconômicas, as quais encontram-se elencadas no Quadro 1. Com exceção da variável *renda familiar*, *tamanho da família* e IDE, as demais variáveis são dicotômicas, com valor igual à unidade indicando que aquele aspecto pode ser atribuído ao indivíduo (caso contrário, atribui-se o valor zero).

Quadro 1 – Definição das variáveis do modelo de regressão quantílica

Variável	Definição
Modalidade	Dummy 1 se presencial, 0 se a distância
Tamanho da família	Número de pessoas na família
Renda familiar	Rendimento mensal da família (R\$)
Sexo	Dummy 1 se do sexo masculino, 0 se do sexo feminino
Estrangeiro	Dummy 1 se Estrangeiro, 0 se brasileiro
Reside em Redenção	Dummy 1 se reside em Redenção, 0 se não reside em Redenção
Cotista	Dummy 1 se for Cotista, 0 se não for Cotista
IDE	Nota obtida em uma escala de 0 a 100

Fonte: Elaboração própria a partir dos dados da pesquisa (2018).

Salienta-se que foi considerada a *renda familiar*, o valor monetário declarado pelos estudantes no preenchimento das informações quando da inscrição; e, neste estudo, foram retirados outliers, ao considerar erros de preenchimento dos formulários e ou valores muito discrepantes do restante da população.

### 2.3.2 Regressões quantílicas

Regressão quantílica é uma técnica estatística baseada em uma generalização do conceito de regressão por MQO, recorrendo à estimação de vários quantis da distribuição condicional associada ao modelo (KOENKER; HALLOK, 2000). Enquanto nas técnicas dos modelos de regressão usuais estimam o valor médio da distribuição condicional da variável dependente, na metodologia da regressão quantílica é possível estimar toda uma família de quantis, fornecendo informações mais completas sobre a relação existente entre a variável resposta e as variáveis explicativas do modelo.

Segundo Koenker e Bassett (1978), a função quantílica condicional (FQC) no quantil  $p$  para uma variável aleatória contínua  $y_i$  com uma função densidade bem comportada, dado um vetor de regressores  $X_i$ , pode ser definida como:

$$Q^{(p)}(y_i|X_i) = F^{-1}(p|X_i) \quad (2)$$

Tem-se que a FQC da RQ é a versão quantílica condicional da função expectativa condicional (FEC), a qual é derivada da solução do problema de minimização do erro médio ao quadrado, base da regressão linear múltipla (RLM). Neste caso, a média é o valor que minimiza o erro médio ao quadrado. Para a regressão quantílica, tem-se que o quantil desejado minimiza a esperança dos desvios absolutos.

Com relação à regressão linear, os estimadores de mínimos quadrados dão os valores dos coeficientes da regressão,  $\hat{\beta}s$ , que minimizam a soma dos quadrados dos resíduos,



da seguinte forma:

$$\text{mim} \sum_i (y_i - x' \beta)^2 \quad (3)$$

Por outro lado, os coeficientes da RQ,  $\beta^{(p)}$ s, solucionam o seguinte problema de minimização:

$$\text{mim} \sum_i |y_i - x' \beta| \quad (4)$$

Detalhadamente, os  $\beta^{(p)}$ s estão envolvidos na minimização da seguinte expressão:

$$\sum_{i=1}^n (d_p(y_i, \hat{y}_i)) = p \sum_{y_i \geq x' \beta^{(p)}} |y_i - x' \beta^{(p)}| + (1-p) \sum_{y_i < x' \beta^{(p)}} |y_i - x' \beta^{(p)}| \quad (5)$$

Em que  $d_p$  é distância do valor de  $Y$  com relação à sua média estimada  $\hat{y}_i$ , e  $p$  e  $(1-p)$  são os pesos atribuídos aos desvios.

Se estes desvios fossem iguais ou superiores à  $\hat{y}_i$ , receberão peso  $p$ , caso sejam inferior, receberão peso  $(1-p)$ . Logo, os valores  $\hat{\beta}^{(p)}$ s são aqueles que minimizam a soma ponderada dos desvios absolutos entre  $y_i$  e o valor ajustado  $\hat{y}_i = x' \hat{\beta}^{(p)}$  (SPDA).

Assim,

$$\beta^{(p)} = \text{arg}_b \text{mim} E |y_i x' b^{(p)}| \quad (6)$$

Como a função objetivo (05) não é diferenciável, há a necessidade de ser otimizada pelo método simplex, que promove interações entre os números, a fim de minimizar as somas dos desvios na expressão. Os estimadores  $\hat{\beta}^{(p)}$ s que conduzem ao valor mínimo dessas somas possuem boas propriedades assintóticas. Sob condições gerais, eles são assintoticamente normais (CAMERON; TRIVEDI, 2009).

Segundo Fonseca e Martins (1987), para avaliar o grau de afastamento de uma distribuição em relação à unidade de referência, deve-se usar medidas de assimetria; e se numa distribuição, a mediana, a moda e a média forem iguais, tem-se uma distribuição simétrica. Mas, se a mediana for maior do que a média e menor do que a moda, então a distribuição é assimétrica à esquerda (assimetria negativa). E, se a mediana for maior do que a moda e menor do que a média, tem-se assimetria à direita (distribuição assimétrica positiva).

## 2.4 Resultados e Discussão

Nessa seção, encontram-se os resultados das estatísticas descritivas, bem como a discussão dos resultados de regressões quantílicas hierárquicas aplicadas aos desempenhos dos alunos da UNILAB medidos pelo IDE. A obtenção das estimativas teve por base um conjunto de 2967 estudantes, que correspondem às unidades observacionais com respostas nas variáveis em estudo.

Para os dados da amostra, a Tabela 1 apresenta uma análise exploratória dos dados. Observa-se que 89,2% dos alunos estudam presencialmente com uma renda média em torno de R\$1.205,00 reais e quantidade média de membros da família de 4,69 membros. Observa-se, ainda, que a maioria dos alunos da amostra são mulheres, sendo que do total 47,6 são homens, 34,7 são negros.

Tabela 1 – Estatística descritiva do modelo de regressão quantílica

Variáveis	Média	Desvio Padrão	Min	Máx	CV
Renda familiar	1205,821	1273,289	100	20000	1,055
Tamanho da família	4,699	2,847	1	26	0,606
Sexo	0,476	0,499	0	1	1,047
Estrangeiro	0,209	0,407	0	1	1,942
Reside em Redenção	0,258	0,438	0	1	1,691
Cotista	0,178	0,382	0	1	2,146
IDE	76,267	14,329	2,58	98,42	0,188

Fonte: Elaboração própria a partir dos dados da pesquisa (2018).

Tratando-se das regressões quantílicas, as quais destacam a quantidade nas diferenças de valores das variáveis explicativas – ou da presença de fatores que costumam afetar o desempenho – alteram os valores dos quantis e, mais ainda, se o efeito é maior (ou menor) nos quantis que separam os estudantes de desempenho mais baixo ou nos quantis que separam aqueles de desempenho mais alto.

Os resultados dos modelos de regressão quantílica foram estimados para os quantis 0,10, 0,25, 0,50, 0,75 e 0,90, para as variáveis observadas, as quais usou-se como base: modalidade (EAD); sexo (mulher); negro (não negro); cotista (não cotista); reside em Redenção (não residente em Redenção); estrangeiro (ser nativo).

Pela Tabela 2, as variáveis *modalidade*, *renda familiar* e *reside em Redenção* mostraram-se estatisticamente significantes ao nível de significância de (0,10) para quase todos os quantis analisados, com exceção do quantil ( $q_{50}$ ) para a variável *renda* e ( $q_{75}$ ) para *reside em Redenção*. O fato segundo o qual o estudante estuda presencialmente impacta positivamente a nota dele no IDE, sendo que o maior impacto se relaciona a alunos com menores desempenhos ( $q_{10}$ ), cujo aumento é da ordem de 33,86 pontos.

A partir deste resultado, pode-se evidenciar que os alunos que participam de aulas

presenciais têm como consequências melhores resultados relativamente aos alunos de menor desempenho acadêmico na UNILAB. Ressalta-se ainda que o menor efeito positivamente foi nos alunos de melhor desempenho  $q_{90}$ , cujo aumento foi de 8,6 pontos nas notas dos alunos. Estes fatos podem ser observados também no Apêndice B (Mudança nos Coeficientes Quantílicos–Modalidade), os quais mantêm um padrão negativamente inclinado na medida em que aumenta o quantil.

Em relação a renda familiar, tem-se efeitos antagônicos, o qual se observa que para os alunos de menor desempenho ( $q_{10}$  e  $q_{25}$ ) até a mediana ( $q_{50}$ ) o efeito foi negativamente, enquanto para os alunos de melhor desempenho ( $q_{75}$ ,  $q_{90}$ ) o efeito foi positivo, conforme também se pode verificar no Apêndice B (Mudança nos Coeficientes Quantílicos – Renda).

Quando se analisa os resultados da Tabela 3, observados os coeficientes das variáveis do modelo de regressão quantílica, pode-se perceber algumas peculiaridades para os grupos de alunos divididos segundo os resultados do IDE por quantis.

O fato de o aluno residir em Redenção evidenciou para os alunos da UNILAB efeitos negativos, sendo que tais efeitos foram maiores nos alunos de menor desempenho ( $q_{10}$ ) e os de menores efeitos nos de maior desempenho. Para aqueles, a redução foi de (4,7) pontos e para estes (-0,7) pontos. Isto pode ser explicado pelo fato de boa parte dos alunos estrangeiros morarem em Redenção; e sabe-se que há uma deficiência ao entrar na universidade na língua portuguesa e na matemática, o que reduz, pelo menos no início, as notas do IDE.

Para a variável *cotista*, em que se analisa o aluno é ou não cotista não se revelou estatisticamente significante para nenhum dos quantis considerados. Este fato pode ser explicado pela grande diferença entre o número de não cotistas comparativamente ao de cotistas, algo em torno de 4,5 vezes a mais. Quando se divide em grupos através de quantis pode gerar resultados insignificantes. Já o sexo mostrou-se significante para os alunos de menor desempenho ( $q_{10}$  e  $q_{25}$ ), portanto, o fato de ser homem implica um efeito negativo no desempenho das notas que compõem o IDE. Todavia, observa-se que o efeito negativo é maior nos alunos de menor desempenho ( $q_{10}$ ), cuja redução na nota é de 3,8 pontos.

O fato de o aluno ser negro apresentou um efeito positivo nas notas do IDE para o grupo de alunos pertencentes a mediana e o quantil ( $q_{25}$ ), sendo que para a mediana, a nota relativa ao IDE é aumentada em 1,6 pontos e para o quantil ( $q_{25}$ ) 2,7 pontos.

A variável *estrangeiro* é levada em consideração se os alunos vêm de outros países ou naturalizados no Brasil; e para os dados em análise, foi verificado que apenas o quantil ( $q_{10}$ ), que ilustra os menores desempenhos, manifestou-se estatisticamente significante. O efeito apresentado foi positivo e para este grupo, com menores desempenhos, o fato de o aluno ser

estrangeiro aumentou em 4,4 pontos a nota do IDE, o que sugere que caso seja analisado apenas o grupo de alunos de menor desempenho, os estrangeiros, para o caso específico da UNILAB, obtiveram melhores resultados.

Em relação ao *tamanho da família*, esta variável também se mostrou significante em apenas um quantil, no caso em questão, positivo o efeito para o quantil ( $q_{25}$ ), aumentando o resultado do IDE em 0,34 pontos.

Tabela 2 – Coeficientes das variáveis do modelo de regressão quantílica

Variável	QUANTIL				
	q <sub>10</sub>	q <sub>25</sub>	q <sub>50</sub>	q <sub>75</sub>	q <sub>90</sub>
Constante	28,75*	45,79*	64,35*	74,79*	80,95*
	3,43	2,186	2,198	1,168	-1,101
Modalidade	33,86*	26,85*	16,10*	10,88*	8,628*
	3,19	2,034	2,097	1,062	(0,999)
Tamanho da família	0,197	0,347*	0,099	-0,035	1,80e-02
	0,291	0,158	0,12	0,0864	-8,30e-02
Renda familiar	-8,21e-3*	-4,16e-3*	-2,20E-03	3,17e-4*	3,1e-4*
	-5.e-3	-2,34e-3	2,144	2,29e-04	-1,40e-03
Sexo	-3,839*	-2,474*	-0,435	-0,0204	-0,231
	1,38	0,739	0,469	0,406	-0,37
Negro	2,289	2,703*	1,601*	1,833	1,023
	2,032	0,981	0,637	0,585	(0,439)
Cotista	-0,046	0,375	0,374	0,304	0,181
	2,078	0,893	0,565	0,523	-0,545
Reside	-4,715*	-3,860*	-2,484*	-1,455	-0,742*
	1,559	0,845	0,596	0,462	0,343
Estrangeiro	4,465*	-0,919	0,196	0,599	-0,151
	2,521	1,386	1,179	0,832	0,690
Pseudo R <sup>2</sup>	0.172	0.125	0.073	0.049	0.037

Fonte: Elaboração própria a partir dos dados da pesquisa (2018).

Nota: Usou-se como base: modalidade (EAD); sexo (mulher); negro (não negro); cotista (não cotista); reside em Redenção (não residente em Redenção); estrangeiro (ser nativo). Sendo: P < 0,10\*.

## 2.5 Considerações Finais

Neste trabalho, discutiu-se os fatores por trás do desempenho dos estudantes da UNILAB no semestre 2018.1. Utilizou-se os dados disponibilizados pela DTI, para estimar o modelo econométrico de regressões quantílicas para os 2967 alunos avaliados naquele ano.

Nessa base de dados constam características pessoais e socioeconômicas dos alunos. Observou-se que a variável referente a modalidade de estudo presencial se revelou estatisticamente significativa em todos os quantis, visto que foram observados os efeitos positivos em cada quantil, sendo o maior o que diz respeito aos alunos com menor desempenho acadêmico.

Em decorrência disso, resta evidente que para estes alunos certamente é importante a presença física do professor, fato este que corrobora com outros estudos já feitos, onde se observa que alunos com melhor desempenho, são mais autodidatas e aproveitam melhor as tecnologias EAD.

Em relação a *renda familiar*, destaca-se um antagonismo em relação aos efeitos apresentados nos extremos dos quantis; e, em relação ao sexo para os alunos de menor desempenho o fato de o aluno ser homem reduziu a nota em -3,8 para o quantil ( $q_{10}$ ) e -2,4. É recorrente na literatura evidenciar que as mulheres obtêm, em média, melhores resultados que os homens.

Espera-se que quanto menor for a distância da casa dos estudantes para o seu local de ensino os resultados acadêmicos sejam melhores, haja vista perda em horas de estudo com deslocamento. Entretanto, o resultado apresentado neste estudo contrariou isso. Uma provável explicação está associada ao fato segundo o qual a maioria dos alunos residentes em Redenção são estrangeiros. Por sua vez, estes ingressam na universidade com déficit de aprendizagem, sobretudo na Língua Portuguesa e Matemática.

Em detrimento dessa questão, cursos de nivelamento têm sido criados pela Universidade. Com um tempo, podem ser analisadas as diferenças de notas entre alunos que fizeram o curso de nivelamento e os que não fizeram ou não o concluíram.

Os resultados com respeito aos alunos da UNILAB vêm encorajar novos estudos, em outras populações, que possam corroborar várias relações discutidas neste trabalho. Sabe-se que há políticas públicas realizadas com o objetivo de atenuar as diferenças de desempenho entre os estudantes, na tentativa de melhorá-lo, especialmente, daqueles estudantes com maior dificuldade de aprendizado.

Para uma correta análise deste tipo de problemática, a regressão quantílica

hierárquica torna-se um bom instrumento na avaliação do efeito dessas políticas, pois ela mede o efeito dos fatores de interesse em diferentes níveis do desempenho educacional, não só em seu valor esperado.

### **3 DISPARIDADES DE RENDIMENTO ACADÊMICO ENTRE COTISTAS E NÃO COTISTAS: UMA ANÁLISE PARA A UNIVERSIDADE DA INTEGRAÇÃO DA LUSOFONIA AFRO-BRASILEIRA (UNILAB)**

O presente capítulo visa verificar a existência (ou não) de disparidades no rendimento acadêmico do grupo de cotistas e não cotistas da UNILAB, a partir da metodologia do *Propensity Score Matching* (PSM) balanceado por entropia.

#### **3.1 Introdução**

As questões educacionais têm ocupado cada vez mais a análise econômica, sendo recorrente o tema da educação como propulsora do crescimento e redutora das desigualdades sociais. Esse efeito é discutido e confirmado por diversos pesquisadores, tais como Alon e Tienda (2007), Acemoglu e Angrist (2000) e Krishna e Tarasov (2013) que mostram ser a educação promotora do crescimento, além de ser (uma das) responsável pela incorporação de tecnologias, a geração de externalidades positivas e aumento dos ganhos salariais.

Em relação às desigualdades sociais, o Brasil é um país em que mais da metade da sua população se declara negro, pardo ou indígena. Segundo o IBGE (2010), em conjunto a maioria dessa população é classificada como pessoas de baixa renda. Com efeito, uma das medidas que buscou diminuir a desigualdade entre brancos e negros no país foi a criação, em 2012, da Lei de Cotas 12.711 (BRASIL, 2012).

Esta Lei surgiu como uma forma eficaz de democratizar o acesso ao ensino superior, na tentativa de se amenizar a dívida histórica com os afrodescendentes e diminuir o hiato existente entre alunos que concluem o ensino médio nas redes públicas e privadas. Este tipo de instrumento já foi implantado em alguns países para mitigar desigualdades sociais, econômicas e educacionais entre raças, onde as cotas raciais foram apresentadas com um modelo de ação afirmativa, e que, no caso brasileiro, vem sendo implantado desde o início dos anos 2000, quando a Universidade de Brasília (UnB) decidiu fazer reserva de vagas para alguns candidatos em seu processo seletivo.

A Lei de Cotas brasileira prevê, em seu primeiro artigo, 50% das vagas a estudantes que tenham cursado, integralmente, o ensino médio em escolas públicas; em seu terceiro artigo, descreve que em cada instituição federal de ensino superior, as vagas de que trata o art. 1º serão preenchidas, por curso e turno, por autodeclarados pretos, pardos e indígenas e por pessoas com deficiência, nos termos da legislação, em proporção ao total de vagas no mínimo igual à proporção respectiva de pretos, pardos, indígenas e pessoas com deficiência na população da unidade da Federação onde está instalada a instituição (IBGE, 2010).

As instituições federais brasileiras, em conformidade à Lei das Cotas (8º artigo), foram obrigadas a cumprir até 2016 a reserva de 50% das vagas para estudantes que cursaram o ensino médio em escolas públicas. Deste percentual, metade deveria ser destinada a candidatos que possuíam renda mensal *per capita* igual ou menor a 1,5 salários mínimos e a outra metade para os estudantes com renda maior que 1,5 salários mínimos. De fato, vem aumentando o número de universidades que adotam a política de ação afirmativa baseada em raças, e hoje já representa a maioria das universidades federais.

Nesse sentido, o regime de cotas nas universidades públicas vem sendo tratado nos últimos anos como um tema polêmico na sociedade brasileira, onde claramente se verificam duas posições antagônicas: de um lado, há os que defendem as cotas, afirmando e demonstrando os sofrimentos e humilhações sofridos pelos negros na história do Brasil, e, portanto, defendem a reparação histórica e justiça social derivadas dessa política (BITTAR; ALMEIDA, 2005; MOEHLECKE, 2002; GUARNIERI; MELO SILVA, 2007); e, por outro lado, as críticas às ações afirmativas estão relacionadas à hipótese de incompatibilidade ou *mismatch hypothesis* (SANDER, 2004; SOWELL, 2004). De acordo com esta hipótese, a alocação de indivíduos em ambientes no qual o nível educacional é incompatível ao aprendizado acumulado, particularmente em cursos mais seletivos, pode gerar um efeito de desânimo, o que impacta negativamente sobre o desempenho acadêmico e a permanência do estudante na universidade.

Partindo-se do pressuposto segundo o qual muitos alunos que concluíram o ensino médio em escolas públicas acabavam não conseguindo entrar em uma universidade pública, muitas vezes por deficiência na qualidade de ensino que eles tiveram, além da superioridade da qualidade das escolas particulares, as cotas apresentam-se como a tentativa de facilitar a entrada para estes alunos nas universidades federais.

A crítica contra as cotas aponta que muitos alunos com desempenho melhor no exame nacional (ENEM) deixam de entrar em uma universidade em razão da política de afirmação. Alguns afirmam que ao entrar na universidade os alunos cotistas teriam pior desempenho comparativamente aos não cotistas, tendo em vista deficiências em etapas anteriores de sua trajetória escolar.

Espera-se que a inserção dos estudantes cotistas em um mesmo ambiente de ensino com alunos ingressantes pelo sistema de ampla concorrência, oriundos de estruturas de ensino normalmente de melhor qualidade, pode ocasionar diferenciais de desempenho acadêmico, com aumento nas taxas de reprovação e evasão dos estudantes cotistas.

Ao investigar esta temática, buscou-se analisar o caso específico da Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB), tendo em vista que desde



a sua criação, conforme se pode verificar no PDI da própria universidade, que o seu programa da assistência estudantil tem buscado ampliar e sedimentar suas ações afirmativas, aqui entendidas como diretrizes, dispositivos normativos, medidas, iniciativas, atividades que procurem compensar déficits culturais, formativos e psicossociais que representem perdas históricas cumulativas, de grupos sociais e culturais alvo de preconceitos e discriminações devidas aos fatores de sexo, gênero, etnia e raça.

Para além disso, vale ressaltar que, de acordo com a sua lei de criação (Lei Nº 12.289 de 20 de julho de 2010), a Instituição tem como objetivo ministrar ensino superior, desenvolver pesquisas nas diversas áreas de conhecimento e promover a extensão universitária, e tendo como missão institucional específica formar profissionais e cidadãos para contribuir com a integração entre o Brasil e os demais estados-membros da Comunidade dos Países de Língua Portuguesa (CPLP), especialmente os países africanos (Angola, Cabo Verde, Guiné-Bissau, Moçambique, Portugal, São Tomé e Príncipe) e Timor Leste, bem como promover o desenvolvimento regional e o intercâmbio cultural, científico e educacional.

Diante disso, cogita-se o seguinte problema: ao entrar na universidade, os alunos cotistas obtêm desempenhos semelhantes aos alunos não cotistas? Desse modo, o objetivo desse ensaio é verificar a existência (ou não) de disparidades de rendimento acadêmico entre cotistas e não cotistas na UNILAB.

A despeito da diversidade de estudos sobre a política de cotas no Brasil, este trabalho diferencia-se pelo uso de micro dados de uma instituição de ensino superior, associado à adoção de estratégias de pareamento amplamente utilizadas em diferentes áreas do conhecimento. Fez-se uso da técnica de *Propensity Score Matching* (PSM) balanceada por entropia, tendo recorrido também ao teste de sensibilidade de Rosenbaum (1983).

Esta pesquisa, além desta introdução, é composta por mais quatro seções. A seção dois tem-se a revisão de literatura sobre os sistemas de cotas; os procedimentos metodológicos adotados estão apresentados na terceira seção, nomeadamente a estatística descritiva e o modelo proposto para análise do problema em questão, ou seja, *Propensity Score Matching* (PSM) balanceado por entropia e análise de sensibilidade. Na quarta seção analisa-se os resultados e discussões encontrados, e, por fim, são apresentadas as considerações finais do estudo.

## 3.2 Revisão da Literatura

Nesta seção, será feito um mapeamento dos principais estudos concernentes ao tema da pesquisa.

### 3.2.1 Ações afirmativas no Brasil e no mundo

O termo ação afirmativa data da década de 1960, nos Estados Unidos da América, com o Presidente John F. Kennedy, como forma de se referir a atos que promovem a igualdade entre os negros e brancos norte-americanos (MOEHLECKE, 2002). De forma ampla, a literatura define as ações afirmativas como conjunto de medidas que buscam reduzir desigualdades e ou diferentes tipos de discriminações.

Não obstante, as ações afirmativas são também definidas como o conjunto de medidas públicas e privadas que atribuem privilégios a grupos sociais historicamente inferiorizados e marginalizados, em razão de discriminações raciais, religiosas, étnicas, de gênero e por apresentarem algum tipo de deficiência física (PAULA, 2004; PISCINO, 2006).

Tomei (2005) acredita ser preciso atacar tanto a discriminação direta quanto a indireta, mesmo que seja necessário o tratamento de pessoas diferentes de maneiras diferentes. Vale ressaltar que na literatura especializada as ações afirmativas não podem ser confundidas com discriminação positiva: esta tem como finalidade selecionar pessoas que estejam em situação de desvantagem tratando-as desigualmente e favorecendo-as com alguma medida que as tornem menos desiguais; aquelas tratam de ações de incentivo e suporte a grupos de pessoas a que se destinam, tais como a criação de cursinhos pré-vestibulares para afrodescendentes e pessoas oriundas de escolas públicas, ou a criação de horários de reuniões (em partidos políticos, sindicatos etc.) que permitam a participação de mulheres com filhos.

Uma outra questão merecedora de atenção é a dos grupos focais ou grupos-alvos, para os quais se consideram características estigmatizantes desses grupos, tais como cor/raça, gênero, etnia, religião, idade, renda e deficiências físicas. Supõe-se que o sistema de cotas assegura uma falta de competitividade entre os grupos focais, pois o número de vagas é definido para cada grupo focal. Em muitos casos, existe apenas um grupo focal, representado por estudantes oriundos de escolas públicas. Estes grupos podem ser específicos ou direcionados aos negros/pardos, ocorrendo uma discriminação positiva, ou no caso de políticas universais, beneficiando todos que sofrem as deficiências do ensino médio público, independente da cor. Nesse caso, pode-se falar de uma política redistributiva universal, pois existe apenas um grupo focal, o que assegura igual tratamento para todos que estão nesse grupo.

Segundo Brandão Júnior e Amaral (2007), a identificação dos grupos sub-

representados pode gerar conflitos com indivíduos pertencentes a esses grupos-alvos optando por não serem assim classificados e indivíduos fora dos critérios de classificação indicando possuir algumas características semelhantes ao grupo beneficiado. Outra dificuldade é a possibilidade de as ações beneficiarem integrantes mais favorecidos dos grupos beneficiados, ampliando dessa forma a desigualdade interna dos grupos-alvos, mesmo que reduzindo as desigualdades de renda entre os diversos segmentos da sociedade.

No Brasil, percebe-se esta problemática, razão pela qual surgiram as subdivisões internas dentro do grupo de beneficiários, expandindo as modalidades de classificação dos alunos que se propõem fazer o Enem, os quais aprecem alunos com deficiência classificados em subgrupo a partir de declarações de renda e estudos em escolas públicas e ou privadas.

Menezes-Filho *et al* (2019), em um estudo feito sobre a mobilidade intergeracional de educação no Brasil, descreveram que negros continuam a pertencer a famílias com menor nível de instrução: 64,6% dos filhos de pais sem escolaridade são negros, embora esse grupo corresponda a apenas 49,9% da amostra do trabalho. Isto demanda uma atenção especial para aqueles que objetivam diminuir a desigualdade social.

Segundo Moehlecke (2002), as primeiras iniciativas das ações afirmativas referem-se a mudanças na legislação trabalhista, adotadas nos Estados Unidos, em 1935, mas que somente se consolidaram com o Movimento dos Direitos Civis na década de 1960. Países como Índia, Paquistão, Austrália, África do Sul, Malásia e Brasil, antes mesmo do uso do termo ações afirmativas, nesse mesmo período, já praticavam algum tipo de política pública direcionada a grupos em desvantagem, objetivando reduzir os efeitos dos processos de discriminação (TOMEI, 2005).

Su (2005), ao analisar o efeito de políticas de ação afirmativa em uma estrutura hierárquica de educação superior nos EUA, verificou o impacto desta política quanto ao acesso ao ensino superior. Tendo como hipótese que a desigualdade racial implique, à primeira vista, em oportunidade de desenvolvimento inferior a alunos no grupo desfavorecido, a admissão destes na educação superior, com base apenas no ranking de pontuação do teste de admissão, prejudica esses grupos, tendo em vista a discriminação histórica.

Velloso (2009) utilizou a análise de corte para testar a existência de diferenças entre as médias das notas obtidas nos cursos da Universidade de Brasília (UnB), a partir de dados dos alunos cotistas e não cotistas da UnB, entre 2004 e 2006. O autor verificou que, em 11% dos cursos da área de Humanidades, os estudantes cotistas apresentaram um desempenho superior aos estudantes não cotistas; na área de Ciências, em 2004, os cotistas apresentaram um desempenho inferior aos não cotistas, mas para os demais anos essa relação se inverte, tendo os

cotistas desempenhos superiores em cerca de 5% dos cursos. Na área da Saúde, os estudantes cotistas apresentaram desempenho inferior ao dos não cotistas na maioria dos cursos.

No Brasil, estudos feitos sobre o sistema de cotas implantado nas universidades brasileiras têm demonstrado não haver diferença significativa de desempenho entre os estudantes cotistas e não cotistas. Das pesquisas realizadas, destacam-se institucionalmente as universidades UERJ, a UnB e a UFBA, pioneiras que foram na adoção das ações afirmativas, ainda em meados dos anos 2000.

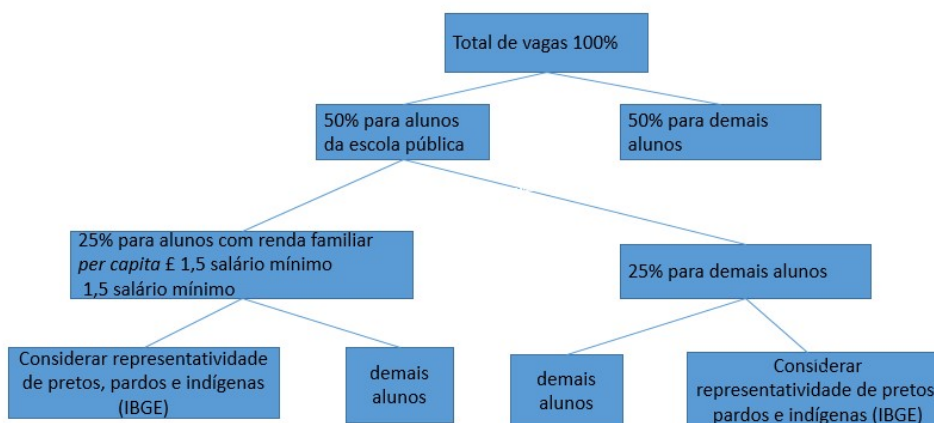
### 3.2.2 Ações afirmativas: uma breve discussão sobre a implantação de cotas no Brasil

No Brasil, as ações afirmativas ganharam mais destaque após 1995, com a Marcha Zumbi dos Palmares – contra o racismo pela cidadania e a vida, cujo objetivo era a luta contra o racismo e a favor da igualdade social (NASCIMENTO, 2003). Iniciativas na área da educação foram adotadas a partir de 2012, quando o Governo Federal sancionou a Lei nº 12.711, obrigando todas as universidades federais a adotarem o Sistema de Reserva de Vagas em seus vestibulares.

Regulamentada pelo Decreto nº 7.824/2012, a Lei definiu as condições gerais de reservas de vagas e estabeleceu o acompanhamento e a transição das reservas de vagas nas Instituições Federais de Ensino Superior (IFES). Com a Portaria Normativa nº 18/2012 do Ministério da Educação, foram estabelecidos os conceitos básicos para a aplicação da Lei. A legislação determina que 50% de todas as vagas sejam destinadas a estudantes oriundos de colégios públicos, alocadas entre negros, pardos, indígenas e de baixa renda, e as 50% restantes destinadas à ampla concorrência.

As cotas são divididas em quatro modalidades: estudantes não elegíveis (ou seja, que não estudaram em escola pública), estudantes de escolas públicas, estudantes com renda familiar inferior a um salário mínimo e meio, e estudantes pretos, pardos ou indígenas. As vagas destinadas às cotas são divididas em duas: (i) metade para estudantes de escolas públicas com renda familiar bruta igual ou inferior a um salário mínimo e meio *per capita*; e (ii) metade para estudantes de escolas públicas com renda familiar superior a um salário mínimo e meio. Em ambos os casos, será reservado um percentual mínimo correspondente ao da soma de pretos, pardos e indígenas no estado, de acordo com o último Censo Demográfico do Instituto Brasileiro de Geografia e Estatística (IBGE).

Figura 1 – Reserva de vagas de acordo com a Lei de Cotas 2012



Fonte: Elaboração própria.

Foi definido um intervalo de quatro anos para que as instituições federais de ensino superior se adequassem à referida Lei, sendo que, obrigatoriamente, 25% da reserva de vagas prevista deveria ser implementada a cada ano. Assim, as IFES passaram a aplicar a Lei de Cotas em seus vestibulares gradualmente, desde que, no vestibular de 2013, fosse reservado ao menos 12,5% do número de vagas ofertadas. E, até o ano de 2016, a implantação das cotas deveria ser feita em sua totalidade conforme determinado em lei, com metade das vagas sendo destinadas aos alunos oriundos de escolas públicas, divididas entre alunos com renda abaixo e acima de 1,5 salário mínimo *per capita* (BRASIL, 2015).

A alteração promovida pela promulgação da Lei de Cotas, em 2012, teve impactos nos incentivos dos indivíduos contemplados pela política. Com a reserva de vagas, criou-se uma situação em que as possibilidades de acesso ao ensino superior, antes percebidas como fora de alcance pelos elegíveis às cotas, tornaram-se alcançáveis (FRYER; LOURY, 2005).

Alguns autores apontam que o caráter unificado do ENEM e a Lei de Cotas, provavelmente impactaram a composição dos participantes do Enem. É razoável supor que, além disso, também impactaram o esforço desses alunos, haja vista o caráter competitivo de exames do tipo vestibular (VILELA *et al.*, 2017).

No período de 2003 a 2009 já se observou grande expansão de ações afirmativas, com o crescimento do acesso tanto em universidades federais, como em universidades estaduais. Nesta época, destacam-se os primeiros estudos acadêmicos feitos sobre o sistema de cotas implantado nas universidades brasileiras, os quais demonstraram não haver diferença significativa de desempenho entre os estudantes cotistas e não cotistas. Entretanto, os melhores desempenhos geralmente estão correlacionados a alunos não cotistas. Os resultados destas pesquisas tornaram-se importante por tentar dar uma resposta às críticas que se faz a política de cotas, notadamente pela afirmação segundo a qual esta tende a rebaixar o nível acadêmico.

Gomes (2017) realizou um outro estudo por meio do qual procurou analisar o impacto das políticas de cotas no ensino superior brasileiro: uma análise a partir do ENADE para os cursos de engenharia nos anos de 2008, 2011 e 2014. Observou-se que os resultados obtidos por meio dos métodos de Regressão múltipla e *Propensity Score Matching* indicaram que houve algumas diferenças significativas entre o desempenho dos alunos cotistas e não cotistas. Os alunos optantes pelas cotas sociais apresentaram desempenho superior aos não cotistas, enquanto os alunos optantes pelas cotas raciais apresentaram desempenho inferior aos não cotistas.

Entretanto a maioria dos estudos realizados demonstram uma superioridade dos alunos não cotistas. Como exemplo, vale o estudo, feito pela universidade da Paraíba, que recorreu a diferentes instrumentos de pareamento não-experimental e tendo como variável de resultado o nível de esforço medido pelo CRA relativo, que evidenciou que os alunos cotistas da UFPB apresentam rendimento médio significativamente inferior quando comparados aos alunos ingressantes por ampla concorrência. O impacto é ainda maior quando a análise contempla os quantis mais elevados da distribuição do CRA relativo (SILVA *et al.*, 2019).

Por sua vez, o estudo referente a ação afirmativa em universidades públicas desenvolvido por Santos (2015), que objetivava investigar a performance de estudantes universitários, constatou que de 15 cursos, em 9 deles, os estudantes beneficiados tiveram performance inferior aos dos não beneficiados.

Ainda ao analisar os diferenciais de desempenho entre estudantes cotistas e não cotistas da UFBA pelo PSM, os cotistas apresentam um desempenho inferior quando comparado aos não cotistas. Há níveis de conhecimento diferentes entre os estudantes, entretanto, as diferenças tendem a reduzir durante o decorrer do curso (CAVALCANTI *et al.*, 1999).

### 3.3 Metodologia

A metodologia apresentada será vista em duas subseções: Base de dados e métodos do *Propensity Score Matching*, balanceado por entropia e análise de sensibilidade.

#### 3.3.1 Bases de dados

A amostra utilizada neste artigo compreende os estudantes ativos nos cursos de graduação presencial da UNILAB, no ano de 2018 que fizeram matrícula no semestre 2018.1. Para cada elemento da amostra, existem as informações socioeconômicas, declaradas durante a realização da inscrição no processo seletivo e algumas de quando ingressado na universidade. Os dados foram coletados junto à Diretoria de Tecnologia da Informação (DTI), no Sistema Acadêmico (SIAC) e pelo Centro de Processamento de Dados. Os dados obtidos apresentam, além das informações socioeconômicas dos ingressantes, esclarecimentos acerca do número de matrícula e acompanhamento acadêmico do estudante no decorrer do curso. A amostra abrange 2967 observações, entre alunos que ingressaram por meio do sistema de reserva de vagas, vestibular de ampla concorrência e processos específicos para alunos estrangeiros.

A UNILAB acompanha o desempenho de seus alunos a partir de uma avaliação de aprendizagem que foi aprovada no Conselho Universitário na resolução 27 de 11 de novembro de 2014. Esta avaliação foi constituída como:

- Parte integrante do ensino aprendizagem;
- Processo que subsidia a reflexão e prática docente;
- Elemento constitutivo do processo formativo do estudante.

A avaliação de aprendizagem na UNILAB visa o êxito acadêmico do discente de graduação calculado por um componente curricular, em que se combina o aproveitamento acadêmico e a assiduidade. O aproveitamento acadêmico consiste no êxito discente de cada componente em cada disciplina e a assiduidade se refere a frequência regular às atividades acadêmicas de cada componente curricular.

O desempenho discente de graduação é calculado pelo IDE, que é a média das notas finais das disciplinas, ponderada pela carga horária respectiva, multiplicada por um redutor por reprovação por falta e por trancamento de disciplina, conforme exposto abaixo.

$$IDE = \left( \frac{\sum_{i=1}^n NO_i \cdot CH_i}{\sum_{i=1}^n CH_i} \right) \cdot \left( 1 - \frac{0,3 \cdot NTR + 0,7 \cdot NRF}{NT} \right) \quad (7)$$

Em que:

- IDE – Índice de Desempenho do Estudante;

- $i$  –  $i$ -ésima disciplina que o estudante concluiu, com ou sem reprovação (com exceção daquelas em que houve reprovação por falta);
- $n$  – total de disciplinas concluídas, com ou sem reprovação (com exceção daquelas em que houve reprovação por falta);
- NTR – número de disciplinas trancadas;
- NRF – número de disciplinas com reprovação por falta;
- NT – número total de disciplinas que o aluno teve matrícula efetivada;
- $NO_i$  – nota em que o aluno obteve em uma dada disciplina  $i$ ; e,
- $CH_i$  – carga horária da disciplina  $i$ .

OBS1: O primeiro fator  $\left(\frac{\sum_{i=1}^n NO_i \cdot CH_i}{\sum_{i=1}^n CH_i}\right)$  representa a fórmula básica, que é a média ponderada das notas obtidas nas disciplinas pela carga horária. Exclui-se desta fórmula as notas e carga horária da disciplina que o estudante obteve reprovação por falta.

OBS2: O segundo fator  $\left(1 - \frac{0,3 \cdot NTR + 0,7 \cdot NRF}{NT}\right)$  é um fator de redução. Tem-se que caso o estudante não tenha trancamentos ou reprovações por falta, a segunda parcela do fator de redução se anula e a fórmula é multiplicada por 1, não havendo qualquer perda para o estudante.

Do acompanhamento acadêmico tem-se a análise do coeficiente de rendimento, o IDE facilita acompanhar o desempenho de cada aluno da universidade, e que a partir dele, espera-se neste estudo identificar se, para o período cursado, os alunos que entraram com pontuações diferentes conseguiram ou não eliminar as diferenças de desempenho. Com o acesso ao banco de dados, foi possível identificar os estudantes que entraram pela política de cotas, pois uma das informações contidas na base de dados era a categoria de entrada.

### 3.3.2 Método do Propensity Score Matching (PSM)

Analisou-se o efeito que o sistema de cotas causa no rendimento escolar dos alunos matriculados na UNILAB, a partir da utilização da metodologia *Propensity Score Matching*, ou pareamento por escore de propensão. Este método trabalha com um grupo contrafactual ou de controle que seja tão semelhante ao grupo de tratamento quanto possível em termos de características observáveis. A determinação do grupo-controle é um dos principais aspectos da avaliação do impacto (BACKER, 2000).

Este grupo-controle deve ser análogo ao grupo de tratamento, e ser retirado a partir de um grande grupo de não participante, indivíduos que sejam observacionais semelhantes aos participantes em termos de características não afetadas pelo programa, onde cada observação



no grupo de tratamento teve um par no grupo-controle, onde esse par reuniu características comuns, de modo que o par associado seria o resultado da observação caso ela não fosse tratada.

Portanto, cada participante foi combinado com um participante não observacional similar e, a partir daí, observou-se a diferença média nos resultados entre os dois grupos, podendo assim, ser possível comparar o efeito do tratamento do programa.

No PSM, cada participante é comparado a um não participante com base em um único escore de propensão, que é a probabilidade condicional de receber um tratamento, dadas diferentes características observáveis  $X$  (ROSENBAUM; RUBIN, 1983).

Esse método, portanto, estima o efeito médio do tratamento sobre os tratados. Para tanto, tem como pressuposto primário a seleção nos observáveis, ou ignorabilidade (ROSENBAUM; RUBIN, 1983). Este requer que seja satisfeita a hipótese de independência condicional (HIC), para que os estimadores do pareamento apresentem uma interpretação causal.

Supõe-se que a designação do tratamento às unidades de estudo depende apenas das variáveis observáveis, o vetor  $X$ , o qual, contém informações capazes de analisar com a devida completude o resultado potencial  $Y(0)$  na ausência de tratamento  $T$ . Assim, a variável  $Y(0)$  tende a cair independente de  $T$ , assim como o outro resultado potencial  $Y(1)$  também.

$$0 < P(D_i = 1|X_i) < 1 \quad (8)$$

Essa condição assegura que as observações do grupo de tratamento tenham observações comparáveis do grupo de controle quanto às características  $X_i$ . Como é apenas na área em que existe suporte comum que é possível realizar inferências causais, em contextos em que essa condição é fracamente satisfeita pode ser necessário eliminar as observações que não pertencem a esse conjunto. De uma forma geral, o pareamento se torna inviável quando são muitas as variáveis que devem ser correspondidas entre o grupo de tratamento e o grupo de controle.

O Teorema do Escore de Propensão, proposto por Rosenbaum e Rubin (1983), torna prático o pareamento ao reduzir o número de variáveis do vetor  $X_i$  a um único escalar, a probabilidade de recebimento do tratamento, dadas as características observadas. Como mostra Angrist e Pischke (2009), o Teorema do escore de propensão pode ser enunciado da seguinte forma:

Teorema do Escore de Propensão:

A suposição da Hipótese de Independência Condicional (HIC) é satisfeita tal que

$$\{Y_{0i}, Y_{1i}\} \perp d_i | X_i \quad (9)$$

Então,

$$\{Y_{0i}, Y_{1i}\} \perp d_i | P(X_i) \quad (10)$$

Esse teorema afirma que se os resultados potenciais são independentes da variável de tratamento condicionada a um vetor multivariado  $X_i$ , então os resultados potenciais são independentes da variável de tratamento condicionada a uma função escalar desse mesmo vetor, que é o escore de propensão, definido como:

$$P(X_i) \equiv E[d_i | X_i] = P[d_i = 1 | X_i] \quad (11)$$

Quando se supõe a hipótese da independência condicional (HIC),  $E[Y_{1i} - Y_{0i} | X_i]$ . Pelo teorema do escore de propensão o *ATT* resultante do pareamento direto dos valores de propensão entre tratados e não tratados, aplicando a lei de expectativas iteradas sobre  $X_i$ , é:

$$ATT = E[Y_{1i} - Y_{0i} | D_i = 1] = E\{E[Y_i | P(X_i), D_i = 1] - E[Y_i | P(X_i), D_i = 0] | D_i = 1\} \quad (12)$$

O estimador típico do pareamento por escore de propensão tem a seguinte forma:

$$ATT_{PSM} = \frac{1}{N_T} \left[ \sum_{i \in T} Y_{1,i} - \sum_{j \in C} \omega(i, j) Y_{0,j} \right] \quad (13)$$

Em que:

$N_T$  é o número de indivíduos tratados pertencentes à região de suporte comum;  $\omega(i, j)$  é o esquema de pesos utilizado para agregar o resultado potencial dos indivíduos do grupo de controle e depende do escore de propensão do participante  $i$ ,  $P(X_i)$ , e do escore de propensão do não-participante  $j$ ,  $P(X_j)$ .

O método, portanto, ao objetivar calcular o efeito do tratamento do programa, deve-se primeiro calcular o escore de propensão  $P(X)$  com base em todas as covariadas observáveis  $X$  que afetam conjuntamente a participação e a variável de interesse. Para isto, foi encontrado um grupo de comparação mais próximo a partir de uma amostra de não tratados, com relação a amostra de tratados do programa. Distintos métodos podem ser utilizados para fazer o pareamento, dentre os quais se destacam os que seguem: Pareamento por vizinho mais próximo (Nearest Neighbor Matching), Pareamento por alcance (Radius Matching), Pareamento Kernel (Kernel Matching).

Sendo válidas todas as hipóteses apresentadas, para a estimação do efeito causal da política em estudo é necessário apenas comparar as unidades do grupo de controle e do grupo de tratamento que apresentam a mesma probabilidade de recebimento do tratamento. Angrist e Pischke (2009) apontam que a estimação por escore de propensão se dá em dois passos:

primeiro estima-se  $P(X_i)$  com algum modelo paramétrico, como Probit ou Logit. Em seguida, a estimação do efeito do tratamento pode ser computada ou pelo pareamento do valor encontrado no primeiro passo ou utilizando algum esquema de pesos.

Neste estudo, a escolha entre os modelos Probit ou Logit se deu seguindo os seguintes passos: Primeiro foram calculados os casos corretamente especificados, e em seguida calculou-se os Critérios de Akaike's e Bayesiano (BiC), além da Curva ROC para os dois modelos.

No critério de Akaike (1974), este utilizou a informação de Kullback-Leibler para testar se um dado modelo é adequado. Porém, seu uso é limitado, pois depende da distribuição  $g$  (modelo verdadeiro), que é desconhecida. Em seus estudos, mostrou que o viés é dado assintoticamente por  $p$ , em que  $p$  é o número de parâmetros a ser estimados no modelo, e definiu seu critério de informação como:

$$AIC = -2\log L(\hat{\theta}) + 2(p) \quad (14)$$

Já o critério de Informação Bayesiano (BIC), proposto por Schwarz (1978) é dado por:

$$BIC = 2\log f(x_n|\theta) + p\log n \quad (15)$$

Em que  $f(x_n|\theta)$  é o modelo escolhido,  $p$  é o número de parâmetros a serem estimados e  $n$  é o número de observações da amostra. AIC e BIC podem ser vistos como medidas que combinam ajuste e complexidade. O ajuste é medido negativamente por  $-2 * \ln$  (probabilidade), e quanto maior o valor, pior o ajuste. Complexidade é medido positivamente, quer por  $2 * k$  (AIC) ou  $\ln(N) * k$  (BIC). Dado que dois modelos se encaixam nos mesmos dados, o modelo com o menor valor do critério de informação é considerado melhor.

Quanto a Curva ROC (*Receiver Operating Characteristic*), esta é uma medida sobre a capacidade de o modelo discriminar as categorias da variável dependente. Geometricamente, a curva ROC é um gráfico de pares “ $x$ ” e “ $y$ ” (que correspondem, a 1-especificidade e a sensibilidade, respectivamente) num plano designado por plano ROC unitário. A designação de plano ROC unitário, deve-se ao fato das coordenadas deste gráfico representarem medidas de probabilidade, e, por conseguinte variarem entre zero e um.

Caso a área sob a curva seja menor ou igual a 0,5, o modelo não consegue discriminar as categorias, e se a área alcançar valores acima de 0,8, o modelo possui poder discriminatório excelente; enquanto, nos demais casos, o poder discriminatório é apenas aceitável.

### 3.3.2.1 Balanceamento por entropia

O critério de balanceamento é um método de pareamento, e neste estudo é utilizado como uma outra hipótese necessária para a estimação dos efeitos de tratamento. O critério de balanceamento é satisfeito quando para cada valor do escore de propensão,  $X$  tem a distribuição similar para os grupos de tratamento e controle. Formalmente,  $D \perp X | P(X)$ . Para que o PSM funcione, os grupos de tratamento e comparação devem estar equilibrados, pois os escores de propensão semelhantes são baseados em  $X$  similares observados.

Os testes de balanceamento também podem ser realizados para verificar se, dentro de cada quantil da distribuição de escore de propensão, o escore médio de propensão e a média de  $X$  são os mesmos.

Partindo do pressuposto de uma amostra com  $n_1$  observações pertencentes ao grupo dos tratados e  $n_0$  unidades de controle, os quais foram selecionados aleatoriamente de uma população de tamanho  $N_1$  e  $N_0$ , respectivamente ( $n_1 \leq N_1$  e  $n_0 \leq N_0$ ). Seja  $D_i \in \{1,0\}$  uma variável de tratamento binária, onde irá assumir o valor igual a 1 se a unidade  $i$  pertence ao tratamento, e 0 se pertencer ao grupo de controle. Seja  $X$  uma matriz que contém as observações de  $J$  variáveis exógenas de pré-tratamento;  $X_{ij}$  corresponde o valor da  $j$ -ésima covariada da unidade  $i$ , tais que,  $X_i = [X_{i1}, X_{i2}, \dots, X_{ij}]$  refere-se ao vetor de características da unidade  $i$  e  $X_j$  refere-se ao vetor coluna com  $j$ -ésima covariada.

Assim, a densidade das covariadas nas populações de tratamento e controle são obtidas a partir da expressão:

$$f_{X|D=1} = f_{X|D=0} \quad (16)$$

O resultado potencial  $Y_i(D_i)$  corresponde ao par de resultados para a unidade  $i$  dadas as condições de tratado e controle, assim, o resultado observado é dado por:

$$Y = Y(1)D + (1 - D)Y(0) \quad (17)$$

O balanceamento por entropia generaliza a abordagem de ponderação do escore de propensão ao estimar os pesos diretamente de um conjunto de restrições de equilíbrio que exploram o conhecimento sobre os momentos da amostra. Considere  $w_i$  o peso do balanceamento por entropia escolhido para cada unidade de controle, os quais foram encontrados pelo seguinte esquema de ponderação que minimiza a distância métrica de entropia:

$$\min H(w) = \sum w_i \log(w_i/q_i) \quad (18)$$

Sujeito à normalização:

$$\sum w_i cr_i(X_i) = mr \quad \{i|D = 0\} \quad (19)$$

$$\sum\{i|D = 0\}w_i = w_i \geq 0 \text{ para todo } i, \text{ tal que } D = 0 \text{ com } r \in 1, \dots, R \quad (20)$$

Sendo que,  $q_i = 1/n_0$  é um peso base usual e  $cr_i(X_i) = mr$  descreve um conjunto de  $R$  restrições impostas aos momentos das covariáveis no grupo de controle reponderados. Uma restrição típica do balanceamento é formulada de tal forma que  $mr$  contenha o momento de uma covariável específica  $X_j$  para o grupo de tratamento e a função do momento para o grupo de controle é especificada como:  $cr_i(X_{ij}) = Xr$  ou  $cr_i(X_{ij}) = (X_{ij} - \mu_j)r$  com média  $\mu_j$ .

Dessa forma, o balanceamento por entropia procura para um conjunto de unidades, pesos  $W = [w_i, \dots, w_{n_0}]$  que minimizam a equação (11), a distância de entropia entre  $W$  e o vetor base de pesos  $Q = [q_i, \dots, q_{n_0}]$ , sujeita as restrições de balanceamento (equação 12), restrição de normalização, e restrição de não negatividade.

Sendo assim, o Balanceamento por Entropia permite que os grupos sejam mais homogêneos em termos de características observáveis.

### 3.3.2.2 Análise de sensibilidade

Quando as variáveis não observadas afetam os resultados, os estimadores do escore de propensão podem não ser mais considerados consistentes. Consequentemente, a omissão de variáveis causa viés nos resultados por via do efeito médio do tratamento sobre os tratados. Para contornar esse problema, o método dos Limites de Rosenbaum, determina a força da influência de uma variável omitida sobre a seleção na participação (ROSENBAUM, 2002).

A probabilidade de participação  $i$  é dada por:

$$\pi_i = \Pr(D_i = 1|x_i) = f(\beta x_i + \gamma x_i) \quad (21)$$

Espera-se que caso não haja viés de seleção no escore de propensão, a probabilidade de participação se dará exclusivamente pelas características observáveis. Por outro lado, quando existe viés de seleção, duas pessoas com as mesmas covariáveis observadas  $x$  terão diferentes chances de receber tratamento.

Portanto, na análise de sensibilidade é avaliado o quanto do efeito do Programa é alterado pela mudança nos valores de  $\gamma$  e de  $u_i - u_j$ . Isso significa examinar os limites “ $\Gamma$ ” da odds ratio de participação. Rosenbaum (2002) mostra que 16 implica os seguintes limites para a odds ratio: se  $\Gamma=1$ , os pareados possuem a mesma probabilidade de participação. Se, entretanto,  $\Gamma= 2$ , então indivíduos aparentemente semelhantes em termos de  $x$  diferem nas

probabilidades de receberem tratamento por um fator de até 2.

### 3.4 Resultados e Discussão

Nesta seção, os principais resultados encontrados serão fruto de uma discussão.

#### 3.4.1 Estatística descritiva

As variáveis socioeconômicas utilizadas na análise estão elencadas no Quadro 2. Com exceção da variável *renda familiar*, *tamanho da família* e IDE, as demais variáveis são dicotômicas, com valor igual à unidade indicando que aquele aspecto pode ser atribuído ao indivíduo (caso contrário, atribui-se o valor zero).

Quadro 2 – Definição de variáveis do modelo PSM

Variável	Definição
Tamanho da família	Número de pessoas na família
Renda familiar	Rendimento mensal da família (R\$)
Sexo	Dummy 1 se do sexo masculino, 0 se do sexo feminino
Estrangeiro	Dummy 1 se Estrangeiro, 0 se brasileiro
Reside em Redenção	Dummy 1 se reside em Redenção, 0 se não reside em Redenção
Cotista <sup>1</sup>	Dummy 1 se for Cotista, 0 se não for Cotista
IDE	Nota obtida em uma escala de 0 a 100

Fonte: Elaboração própria a partir dos dados da pesquisa (2018).

Salienta-se que foi considerada a *renda familiar*, o valor monetário declarado pelos estudantes no preenchimento das informações quando da inscrição, e neste estudo foram retirados *outliers*, ao considerar erros de preenchimento dos formulários e/ou valores muito discrepantes do restante da população.

Na Tabela 3 está expressa a estatística descritiva dos alunos da UNILAB, onde foram analisados 2967 no semestre 2018.1. Observou-se uma média da *renda familiar* de R\$1.205,821 e 4,6 membros por família.

Quanto às questões de gênero e nacionalidade, 47,6% dos alunos são do sexo masculino, ou seja, as mulheres ocupam um espaço maior frente aos homens no ingresso como alunas da UNILAB e 20,9% dos alunos da amostra são estrangeiros, valor este inferior ao número previsto na lei de criação da universidade que é de 50%. Ainda vale ressaltar no que diz respeito ao local de residência, 25,8% dos alunos afirmaram que residem em Redenção, o

<sup>1</sup> Considera-se os alunos ingressados nas seguintes categorias: L1 – com renda familiar bruta *per capita* igual ou inferior a 1,5 salários mínimos (Lei nº 12.711/2012); L2 – autodeclarados PPI, com renda familiar bruta *per capita* igual ou inferior a 1,5 salários mínimos (Lei nº 12.711/2012); L3 – independentemente da renda (Lei nº 12.711/2012); L5 – independentemente da renda (Lei nº 12.711/2012); L6 – autodeclarados PPI, independentemente da renda (Lei nº 12.711/2012).

que facilita a locomoção e permanência na universidade, e conseqüentemente pode gerar melhor rendimento acadêmico.

Tabela 3 – Estatística descritiva do modelo PSM

Variáveis	Média	Desvio-Padrão	Min	Máx	CV
Renda familiar	1205,821	1273,289	100	20000	1,055
Tamanho da família	4,699	2,847	1	26	0,606
Sexo	0,476	0,499	0	1	1,047
Estrangeiro	0,209	0,407	0	1	1,942
Reside em Redenção	0,258	0,438	0	1	1,691
Cotista	0,178	0,382	0	1	2,146
IDE	76,267	14,329	2,58	98,42	0,188

Fonte: Elaboração própria a partir dos dados da pesquisa (2018).

Na Tabela 4 descreve-se a estatística descritiva entre cotistas e não cotistas, e nela observa-se que a média dos alunos cotistas, o qual obteve-se 77,15 foi superior aos não cotistas com 76,07. O resultado geral, em que se considera o valor global combinado entre cotistas e não cotistas, foi de 76,26.

Já na tabela 5, descreve-se a comparação entre cotistas e não cotistas, porém analisa-se cada variável envolvida no modelo. Nela observa-se que do total de alunos da amostra, 531 são cotistas e 2436 são não cotistas. Em seguida, descreve-se um pouco dos resultados desta análise.

Em relação ao *tamanho da família* os alunos não cotistas são maiores, 4,930; enquanto os cotistas tiveram média de 3,635. Isto pode ser parcialmente explicado por estar dentro do grupo de alunos não cotistas, os alunos estrangeiros, e estes, conforme se pode verificar na tabela do apêndice deste trabalho, apresentam um tamanho médio bem superior, 8,77, enquanto os nativos apresentaram 3,619.

Em relação ao gênero, as mulheres são maioria tanto no grupo de cotistas como não cotistas e praticamente todos os alunos cotistas são brasileiros. Vale ainda ressaltar que os alunos não cotistas afirmaram em percentual maior que os cotistas que moram em Redenção, fato este explicado pela alta procura de alunos da região do maciço de Baturité onde se encontra a universidade.

Tabela 4 – Estatística descritiva entre cotistas e não cotistas

Grupo	Observações	Média	Erro Padrão	Desvio padrão	Intervalo de confiança (95%)	
					Mínimo	Máximo
Cotista	531	77,15251	0,5772636	13,27706	76,0185	78,28653
Não Cotista	2.436	76,07559	0,2947361	14,54396	75,49763	76,65354
Combinado	2.967	76,26779	0,2632102	14,32985	75,7517	76,78388

Fonte: Elaboração própria a partir dos dados da pesquisa (2018).



Tabela 5 – Estatística descritiva: Comparação Cotistas/Não cotistas

Variáveis	Cotistas: 531 alunos				
	Média	Desvio Padrão	Min	Máx	CV
Renda familiar	1235,252	1117,36	100	13731,3	0,904
Tamanho da família	3,635	1,614	1	15	0,444
Sexo	0,425	0,494	0	1	1,163
Estrangeiro	0,056	0,751	0	1	13,253
Reside em Redenção	0,181	0,385	0	1	2,125
Variáveis	Não Cotistas: 2436 alunos				
	Média	Desvio Padrão	Min	Máx	CV
Renda familiar	1119,428	1304,824	100	20.000	1,087
Tamanho da família	4,930	3,001	1	26	0,608
Sexo	0,487	4,999	0	1	1,024
Estrangeiro	0,253	0,435	0	1	1,715
Reside em Redenção	0,275	0,447	0	1	1,620

Fonte: Elaboração própria a partir dos dados da pesquisa (2018).

### 3.4.2 Resultados do Propensity Score Matching (PSM)

Nesta seção, serão apresentados os resultados do *Propensity Score Matching* para analisar se o sistema de cotas causa efeito no IDE. A priori, conforme especificado na metodologia, foram utilizados critérios de seleção do modelo.

Conforme exposto na Tabela 6, os dois modelos Probit e Logit tiveram o mesmo resultado no que diz respeito aos casos corretamente classificados, obtendo 82,15%, e dado que os dois modelos se encaixam nos mesmos dados, o modelo com o menor valor do critério de informação é considerado melhor. Neste estudo, o modelo escolhido por este critério foi o Probit, por apresentar valores menores tanto do ponto de vista de ajuste (2529,323) quanto de complexidade (2565,289) frente aos mesmos resultados para o Logit (2529,743 e 2565,709).

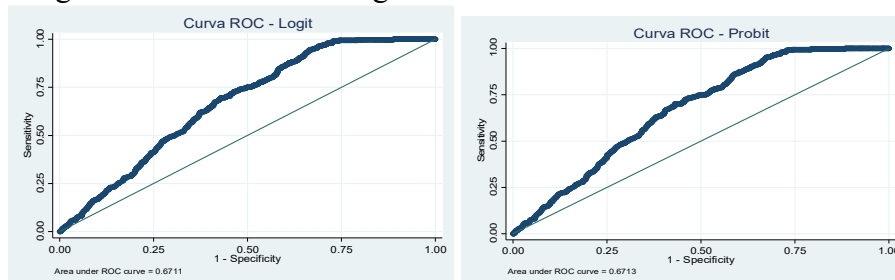
Tabela 6 – Critérios de seleção do modelo

TESTES	MODELO LOGIT	MODELO PROBIT
Casos corretamente classificados	82,15%	82,15%
Critério de Akaike's (AIC)	2529,743	2529,323
Critério de Bayesiano (BiC)	2565,709	2565,289
Goodness-of-fit	0,0606	0,1279
ROC curve	0,6610	0,6604

Fonte: Elaboração própria a partir dos dados da pesquisa (2018).

Foi utilizado como medida sobre a capacidade de o modelo discriminar as categorias da variável dependente, a Curva ROC (*Receiver Operating Characteristic*), tendo no modelo Probit o resultado de 0,6604, portanto possui poder discriminatório aceitável, conforme se pode verificar na Figura 2.

Figura 2 – Curva ROC: Logit e Probit



Fonte: Elaboração própria a partir dos dados da pesquisa (2018).

A Tabela 7 apresenta os resultados do modelo Probit e seus efeitos marginais. Entre as variáveis testadas, apenas os coeficientes do *tamanho da família e sexo* não se mostraram significantes do ponto de vista estatístico ao nível de significância de 5%. Em relação a covariada *estrangeiro*, o fato de o aluno ser estrangeiro diminui em 39,04% a probabilidade de ser cotista. Já em relação a covariada *reside em Redenção*, o fato de o aluno residir em Redenção diminui em 4,91% a probabilidade de ser cotista. Isto pode ser explicado, em parte, pelo fato de que 37,36% (ver UNILAB em números – site da universidade) dos alunos negros são estrangeiros, e como os alunos estrangeiros não entram no sistema de cotas brasileiras, e quando estes estudantes chegam ao Brasil recebem bolsa instalação para residirem em Redenção.

Para fazer uma comparação dos resultados entre cotistas e não cotistas foi necessário selecionar dois grupos, um de tratamento e outro de controle, de tal forma que fossem o mais homogêneo possível. Para isto, o passo seguinte foi realizado o balanceamento com entropia, em que as variáveis utilizadas para o balanceamento devem captar fatores que influenciam a probabilidade de o indivíduo ser selecionado para pertencer ao grupo de tratamento.

Tabela 7 – Coeficiente do modelo Probit e seus efeitos marginais

Covariadas	Coeficiente	Efeitos marginais
Renda familiar	-0,000* (0,0000)	-0,0000* (5,14e-06)
Tamanho da família	0,0075 (0,0190)	0,0015 (0,0039)
Sexo	0,0090 (0,0580)	0,0018 0,0119
Estrangeiro	-1,8966* (0,2291)	-0,3904* (0,0349)
Residente em Redenção	-0,2385* (0,0711)	-0,0491* (0,0146)
Constante	-0,6591* (0,0810)	-
Nº obs.: 2967	Prob>chi2 = 0,0000	Log likelihood = -1258,66

Fonte: Elaboração própria a partir dos dados da pesquisa (2018).

Nota: foram utilizadas como base para as *dummies*, sexo (mulher); estrangeiro (nativo); residente em Redenção (não residente).  $P < 0,10^*$ .

A Tabela 8 mostra as médias amostrais de todas as covariadas divididas entre os dois grupos: grupo de tratamento (cotistas) e outro de controle (não cotistas). Para que os dois grupos pudessem ser comparáveis, era necessário que eles fossem o mais parecido possível, diante disso, foi calculado o balanceamento por entropia. Os resultados apresentam a comparação entre as médias das covariáveis que influenciam a probabilidade de ser cotista, revelam que os alunos cotistas antes do balanceamento diferem dos não cotistas em relação a todas as características. Pode-se perceber que depois do balanceamento por entropia, os três momentos, média, variância e assimetria ficam praticamente iguais para os grupos de tratamento e controle, indicando que agora os grupos são comparáveis entre si.

Tem-se que as médias das características de pré-tratamento do grupo de tratamento com as do grupo de controle revelam a eficácia do balanceamento por entropia. Todas as covariáveis são praticamente balanceadas e não há diferença estatisticamente significativa, ou seja, o balanceamento está perfeitamente ajustado para os três primeiros momentos da distribuição das covariáveis, fato este também observado no histograma dos escores de

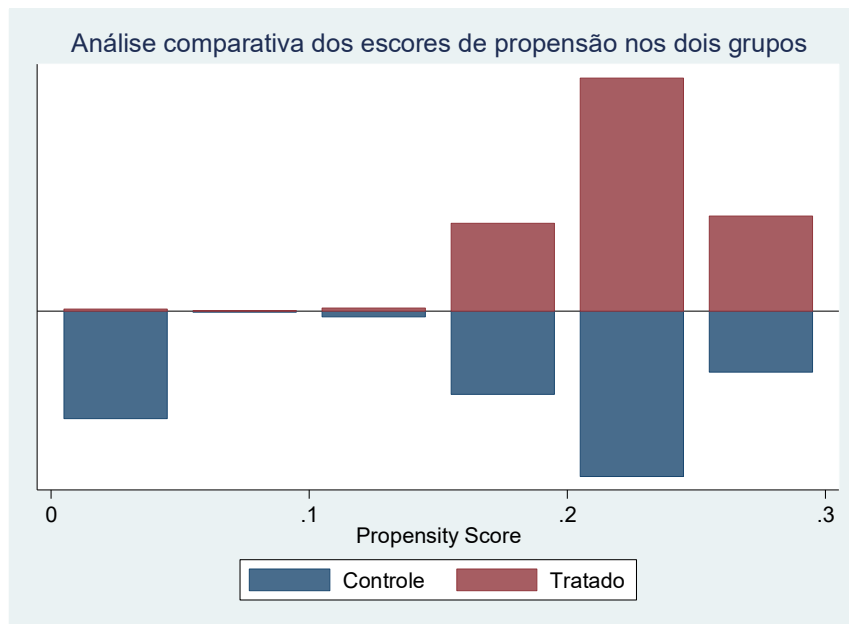
propensão dos beneficiários e não beneficiários das cotas na UNILAB (Verificar na Figura 3).

Tabela 8 – Diferenças nas características observáveis antes e depois do pareamento

	<b>Pré-balanceamento</b>					
	<b>Tratado</b>			<b>Controle</b>		
	<b>Média</b>	<b>Variância</b>	<b>Assimetria</b>	<b>Média</b>	<b>Variância</b>	<b>Assimetria</b>
Renda familiar	1235	1248,499	4,641	1199	1702,566	5,744
Tamanho da família	3,635	2,607	1,362	4,93	9,008	1,510
Sexo	0,425	0,244	0,302	0,487	0,25	0,048
Estrangeiro	0,005	0,005	13,17	0,253	0,189	1,131
Residente em Redenção	0,181	0,148	1,653	0,276	0,199	1,002
	<b>Pós balanceamento</b>					
	<b>Tratado</b>			<b>Controle</b>		
	<b>Média</b>	<b>Variância</b>	<b>Assimetria</b>	<b>Média</b>	<b>Variância</b>	<b>Assimetria</b>
Renda familiar	1235	1248,499	4.621	1235	1102032	3.783
Tamanho da família	3,635	2.607	1.362	3.635	2.261	1.197
Sexo	0,425	0.244	0.302	0.425	0.244	0.302
Estrangeiro	0,005	0.005	13.17	0.005	0.005	13.16
Residente em Redenção	0,181	0.148	1.653	0.181	0.148	1.653

Fonte: Elaboração própria a partir dos dados da pesquisa (2018).

Figura 3 – Histograma dos Escores de Propensão. Beneficiários e não beneficiários das Cotas na UNILAB 2018.1



Fonte: Elaboração própria a partir dos dados da pesquisa (2018).

Assim, conseqüentemente, segue-se com a confiança de que o grupo de controle na análise empírica subsequente é composto por contrafactuais confiáveis para a amostra de observações de indivíduos que são beneficiários cotistas, ou seja, o grupo de tratamento.

Conforme se pode verificar na Tabela 9, o IDE dos alunos cotistas foi de 77,152, valor este superior ao dos alunos não cotistas, cujo resultado obtido foi de 76,075. Logo, verifica-se uma relação causal positiva, indicando um desempenho no IDE positivo para os beneficiários do sistema de cotas exatamente de 1,076, comparado com os não beneficiários do sistema de cotas. Esses resultados evidenciam que os alunos cotistas da UNILAB têm um desempenho do IDE melhor do que os não cotistas.

Tabela 9 – Resultado das estimativas de Propensity Score Matching balanceado por entropia

Variável	Tratado	Controle	ATT	Erro padrão	Test <i>t</i>
<b>IDE</b>	77,152	76,075	1,076	0,87	3,48

Fonte: Elaboração própria a partir dos dados da pesquisa (2018).

A análise do viés de seleção é feita com amparo nos resultados na Tabela 10, onde se testou a robustez dos resultados com base na hipótese de que variáveis não observáveis impactam nos resultados estimados. Observa-se que os valores ( $\Gamma=1$  até  $\Gamma= 1,3$ ) foram estatisticamente significantes, corroborando os resultados encontrados no PSM. Logo, não existem variáveis omitidas, que afetam a seletividade do grupo de tratamento e, daí, nos resultados, não sendo necessário aplicar a técnica de correção pelo viés mínimo e viés corrigido.

Tabela 10– Análise de sensibilidade dos Limites de Rosenbaum

$\Gamma$	<b>Sig+</b>	<b>Sig-</b>	<b>t-hat+</b>	<b>t-hat-</b>
1	0	0	78,58	78,58
1,05	0	0	78,32	78,84
1,1	0	0	78,065	79,085
1,15	0	0	77,815	79,315
1,2	0	0	77,575	79,53
1,25	0	0	77,34	79,735
1,3	0	0	77,11	79,925

Fonte: Elaboração própria a partir dos dados da pesquisa (2018).

### 3.5 Considerações Finais

O sistema de cotas no Brasil surgiu como uma forma de democratizar o acesso ao ensino superior de maneira mais eficaz na tentativa de se amenizar a dívida histórica com os afrodescendentes e diminuir o hiato existente entre alunos que concluem o ensino médio nas redes públicas e privadas. Espera-se que ao longo do tempo sejam diminuídas as desigualdades sociais, econômicas e educacionais entre raças.

Uma vez ingressando na universidade é importante que os alunos que entraram por cotas obtenham resultados semelhantes aos de ampla concorrência. Diante disso, neste estudo o objetivo foi analisar o efeito que o sistema de cotas causa no rendimento escolar (IDE – Índice de desenvolvimento educacional) dos alunos matriculados na UNILAB, a partir da utilização da metodologia *Propensity Score Matching*, balanceado por entropia.

Os principais resultados indicam que as cotas têm uma relação causal positiva, indicando um melhor desempenho para os beneficiários do sistema de cotas, exatamente de 1,076, comparado com os não beneficiários dos sistemas de cotas. De fato, os resultados apontaram que o IDE dos alunos cotistas foi de 77,152, valor este superior ao dos alunos não cotistas cujo resultado obtido foi de 76,075.

Acredita-se que este resultado pode ser explicado pela característica peculiar da UNILAB, tal como o grande número de alunos estrangeiros, que acabam sendo registrados no estudo como não cotistas. Esses alunos vêm apontando dificuldades de língua portuguesa e matemática, além da baixa qualidade do ensino médio em seus países de origem (Angola, Cabo Verde, Guiné-Bissau, Moçambique, Portugal, São Tomé e Príncipe e Timor-Leste). Para amenizar esse problema, surgiram projetos de extensão visando o atendimento a estas demandas com criação de cursos de nivelamento.

Além dos alunos estrangeiros, outras formas de acesso à universidade que não são contabilizadas no sistema de cotas merecem atenção especial, tais como seleção de quilombolas ou indígenas, portador de diploma, transferência e mobilidade interna, e que somadas superam a proporção de alunos de ampla concorrência. Por algumas destas se tratarem de um outro tipo de cotas com características semelhantes aos alunos cotistas, possam ter contribuído para a nota do IDE dos não cotistas terem sido inferior.

Outra possível explicação é devido a UNILAB ser uma universidade localizada em uma cidade do interior, e que ingressam bastante alunos do não somente de Redenção, mas de outras cidades interioranas vizinhas, principalmente do maciço de Baturité. Conforme Rodrigues (2017), ao analisar os fatores que determinam o diferencial de desempenho escolar

entre estudantes da zona rural e urbana, concluiu que o desempenho entre alunos de escolas urbanas e rurais, apontaram resultados inferiores para os rurais decorrente ao que chamaram de efeito-característica.

Outro resultado encontrado neste estudo é que apenas os coeficientes do tamanho da família e sexo não se mostraram significantes do ponto de vista estatístico; e em relação a covariada *estrangeiro*, o fato de o aluno ser estrangeiro diminui em 39,01% a probabilidade de ser cotista; assim como a covariada *reside em Redenção*, onde o fato de o aluno residir em Redenção diminui em 4,91% a probabilidade de ser cotista.

Assim, sugere-se que o Programa Sistema de Cotas para o caso específico da UNILAB se mostra relevante na seleção dos alunos beneficiários, visto que com o resultado do ATT de 1,076, está demonstrado que cotistas e não cotistas não possuem o mesmo desempenho, sendo que os cotistas obtiveram um resultado no IDE superior aos não cotistas. Esses resultados evidenciam a necessidade de se fortalecer o Programa para o caso específico da UNILAB.

Sugerem-se aqui, outros estudos futuros para a UNILAB, acrescentando às demais cotas e retirando-os da ampla concorrência, bem como para as demais universidades, haja vista que com a autonomia administrativa vem surgindo outras formas de acesso nas universidades que ultrapassa os tipos de cotas expressa na lei de cotas.



#### 4 CONCLUSÃO GERAL

Essa dissertação teve dois objetivos principais, um dos quais se propôs avaliar o efeito dos diferentes fatores associados ao desempenho dos estudantes da UNILAB medidos pelo modelo de regressões quantílicas, permitindo verificar se determinado fator tem maior efeito em estudantes de alto, médio ou de baixo desempenho. Já o outro, tratado no segundo capítulo, objetivou verificar a existência (ou não) de disparidades de rendimento acadêmico entre cotistas e não cotistas na UNILAB.

No primeiro capítulo, constatou-se que a modalidade presencial tem um efeito positivo no IDE, a saber, os melhores desempenhos são nos alunos com maior IDE, e que diante disso mostra-se que os alunos de menor desempenho necessitam mais da presença física do professor para melhorarem seu aprendizado.

Também foi evidenciado melhor desempenho das mulheres em relação aos homens, o que corrobora com a literatura existente sobre o assunto, que apresentam que as mulheres em média obtêm melhores desempenhos que os homens. Por outro lado, em relação a variável *reside em Redenção*, o resultado apresentado neste estudo foi o inverso do esperado, fato este explicado por características peculiares dos alunos estrangeiros, que em sua grande maioria chegam com um déficit de aprendizado sobretudo nas disciplinas de português e matemática.

Quanto ao segundo capítulo, observa-se que as cotas têm uma relação causal positiva, indicando um melhor desempenho para os beneficiários do sistema de cotas, onde os resultados apontaram que o IDE dos alunos cotistas foi de 77,152, valor este superior ao dos alunos não cotistas cujo resultado obtido foi de 76,075.

Conclui-se, portanto, que o Sistema de cotas se mostra relevante na seleção dos alunos beneficiários, visto que com o resultado do ATT de 1,076, demonstra-se que os cotistas e não cotistas não possuem o mesmo desempenho, sendo que os cotistas obtiveram um resultado no IDE superior aos não cotistas. (Esses resultados evidenciam a necessidade de se fortalecer o Programa).

## REFERÊNCIAS

- ACEMOGLU, D.; ANGRIST, J. How large are the social returns to education? Evidence from compulsory schooling laws. **NBER Working papers**, n. 7444, p. 9-59, 2000.
- ALVES, M. T. G.; SOARES, J. F. O efeito das escolas no aprendizado dos alunos: um estudo com dados longitudinais no ensino fundamental. **Educação e Pesquisa**. São Paulo, v. 34, n. 3, p. 527-544, 2008.
- AMARAL, L. F. L. E. do.; MENEZES-FILHO, N. A. A relação entre gastos educacionais e desempenho escolar, RJ. *In*: ASSOCIAÇÃO NACIONAL DOS CURSOS DE PÓS-GRADUAÇÃO EM ECONOMIA, 10, 2008, Rio de Janeiro. **Anais [...]**. Rio de Janeiro: ANPEC, 2008. p.300-325.
- AMARAL, L. F. L. E. do.; MENEZES-FILHO, N.A. **Os determinantes dos gastos educacionais e seus impactos sobre a qualidade do ensino**. 44f. 2011. Dissertação (Mestrado em Economia) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2011.
- ANGRIST, J.; PISCHKE, J.S. **Most Harmless Econometrics**: an empiricist's companion. Princeton, New Jersey: Princeton University Press, 2009.
- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**. Boston, v.19, n.6, p.716-723, 1974.
- ALON, S.; TIENDA, M. Diversity, opportunity and the shifting meritocracy in higher education. **American Sociological Review**. California, v. 72, n. 4, p. 487-511, 2007.
- BACKER, J. Evaluating the impact of development projects on poverty: a handbook for practitioners. Washington: **The World Bank**, p. 203, 2000.
- BARROS, R. P. de.; MENDONÇA, R. **O impacto do ambiente comunitário sobre o desempenho educacional**. Rio de Janeiro: IPEA, 1996. Mimeografado.
- BECKER, S.; ICHINO, A. Estimation of average treatment effects based on propensity scores. **The Stata Journal**, v. 2, n. 4, p. 358-377, 2002.
- BITTAR, E.C.B.; ALMEIDA, G. A. de. **Curso de Filosofia do Direito**. 4. ed. São Paulo: Atlas, 2005.
- BONAMINO, A.C. de. **Tempos de avaliação educacional**: o SAEB, seus agentes, referências e tendências. Rio de Janeiro: Quartet, 2002.
- BRANDÃO JUNIOR, J.; AMARAL, S.T. Ações afirmativas: aspectos gerais, SP. *In*: ETIC-ENCONTRO DE INICIAÇÃO CIENTÍFICA, 5, 2007, São Paulo. **Anais [...]**, Presidente Prudente, v. 3, n. 3, p. 1-8, 2007.
- BRASIL. Congresso Nacional. **Lei nº 12.711, de 29 de agosto de 2012**. Dispõe sobre o ingresso nas universidades federais e nas instituições federais de ensino técnico de nível médio e dá outras providências. Brasília, DF, 2012.

BRASIL. LDB – Leis de Diretrizes e Bases. Lei nº 9.394. 1996.

BRASIL. Ministério da Educação. **Entenda as cotas para quem estudou todo o ensino médio em escolas públicas.** Brasília, 2015.

CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics using stata.** Texas: Stata press, 2009.

CAVALCANTI, I.T.N.; ANDRADE, C.S.M.; TIRYAKI, G.F.; COSTA, L.C.C. Desempenho acadêmico e o sistema de cotas no ensino superior: evidência empírica com dados da Universidade Federal da Bahia. **Avaliação: Revista da Avaliação da Educação Superior.** São Paulo. v. 24, p. 305-327, 2019.

FARIAS, P. P. **Os determinantes do desempenho acadêmico dos estudantes de ciências econômicas da UFC.** 48f. 2013. Monografia de Graduação – Curso de Ciências Econômicas, Universidade Federal do Ceará, Fortaleza, 2013.

FERNANDES, R.; GREMAUD, A.P. **Qualidade da educação: avaliação, indicadores e metas.** In: Educação básica no Brasil: construindo o país do futuro [S.l: s.n.], 2009.

FONSECA, J.S.da.; MARTINS, G.A. **Curso de estatística.** 3 ed. São Paulo: Atlas, 1987.

FRYER, R. G.; LOURY, G. C. Affirmative Action and its Mythology. **Journal of Economic Perspectives,** Washington D.C, v. 19, n. 3, p. 147-162, 2005.

FREIRE, G. **Sobrados e mucambos: decadência do patriarcado rural e desenvolvimento do urbano.** 2. ed. ampl. Rio de Janeiro: José Olympio, 1951.

FURTADO, C. **A economia brasileira: contribuição a análise do seu desenvolvimento.** Rio de Janeiro: Noite, 1954.

GIMENES, N. Avaliação em larga escala no Brasil: tensões e desafios. **Estudos em Avaliação Educacional.** São Paulo, v. 26, n. 62, p. 254-261, 2015.

GOMES, V.S. **O impacto das políticas de cotas no ensino superior brasileiro: uma análise a partir do ENADE para os cursos de engenharia nos anos de 2008, 2011 e 2014.** 2017. Trabalho de Conclusão de Curso (Graduação em Ciências Econômicas) – Universidade Federal de Uberlândia, Uberlândia, 2017

GUARNIERI, F.V.; MELO-SILVA, L.L. Ações afirmativas na educação. **Psicol. Soc.** v.9, n.2, p.70-78, 2007.

HOLANDA, S.B de. **Raízes do Brasil.** 3. ed. Rio de Janeiro: José Olympio, 1956.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Estatísticas do cadastro central de empresas 2011.** Rio de Janeiro: IBGE, p.191, 2013.

INEP - INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. Censo da Educação Básica: 2012 – **Resumo técnico.** Brasília: INEP, p.41, 2013.

INEP - INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Manual do ENADE 2013**. 2013.

KOENKER, R.; BASSETT, G. Regression quantiles. **Econometrica**, New York, v. 46, n. 1, p. 33-50, 1978.

KOENKER, R.; HALLOCK, K. **Quantile Regression: An Introduction**. Paper. May 31, 2000.

LEON, F.L.L de.; MENEZES-FILHO, N.A. Reprovação, avanço e evasão escolar no Brasil. **Pesquisa e Planejamento Econômico**. v. 32, n.3, p.417-452, 2003.

LIBÂNIO, J.C. **Didática**. 4 ed. São Paulo: Cortez, 1994.

MENEZES-FILHO, N. A. **Educação e desigualdade**. In: LISBOA, M., MENEZES-FILHO, N. A. (orgs.). Microeconomia e sociedade no Brasil. Rio de Janeiro: EPGE, 2001.

MENEZES-FILHO, N.A.; KOMATSU, B.K. Revisitando a Mobilidade Intergeracional de Educação no Brasil. **Revista Brasileira de Economia**. v. 73, p. 159-180, 2019.

MOEHLECKE, S. Ação afirmativa: história e debates no Brasil. **Cad. Pesqui**. São Paulo, n. 117, p. 4, 2002.

NASCIMENTO, A. do. As políticas de ação afirmativa como instrumentos de universalização de direitos. **Revista Lugar Comum – Estudos de Mídia, Cultura e Democracia**. Rio de Janeiro, n. 18, 2003.

PAULA, A.S. de. Ação afirmativa: instrumento de cidadania ou discriminação reversa? **Semina: Ciências Sociais e Humanas, Londrina**. Paraná, v. 25, p. 3-17, 2004.

PISCINO, M.R.P. **Teoria da ação afirmativa**. 2006. Trabalho de Conclusão de Curso - UniFMU, São Paulo, 2006.

PONTES; L. A. F.; SOARES, T. M. As metas escolares do Ideb: uma proposta alternativa de cálculo. **Estudos em Avaliação Educacional**. São Paulo, v. 27, n. 66, p. 690-715, 2016.

RODRIGUES, L.O. **Ensaio sobre diferencial de desempenho escolar entre alunos de escolas rurais e urbanas no Brasil**. 109f. 2017. Dissertação de mestrado da Universidade Federal do Ceará, centro de Ciências Agrárias, Programa de Pós-Graduação em Economia Rural, Fortaleza, 2017.

ROSENBAUM, P.; RUBIN, R. The Central Role of the Propensity Score in Observational Studies for Causal Effects. **Biometrika**, v. 70, n.1, p. 41-55, 1983.

SANDER, R.H. A systemic analysis of affirmative action in American law schools. **Stanford Law review**. Stanford, v. 57, n. 57, p. 367, 2004.

SANTOS, L.G.S. **Ação afirmativa em universidades públicas brasileiras: diagnóstico da performance de beneficiados**. 37f. 2015. Monografia (Bacharelado em Ciências Econômicas), Universidade de Brasília, Brasília, 2015.

SILVA, A. F.; ALMEIDA, A. T. C.; LOMBARDI FILHO, S. C.; RAMALHO, H. M. B. Efeitos de Políticas Afirmativas sobre Esforço e Abandono: Evidências a partir da Universidade Federal da Paraíba, CE. *In: ENCONTRO REGIONAL DE ECONOMIA*, 24, 2019, Fortaleza. **Anais [...]**. Fortaleza: Banco do Nordeste do Brasil, 2019.p. 413-488.

SOWELL, T. **Affirmative action around the world: an empirical study**. New Haven: Yale University Press, 2004.

SU, X. **Education hierarchy: within-group competition and affirmative action**. SSRN, United States, 2005.

TOMEI, M. Ação afirmativa para a igualdade racial: características, impactos e desafios. Brasília: OIT, 2005. **Documento de trabalho elaborado no âmbito do Projeto Igualdade Racial OIT/Brasil**. Tradução: Hélio Guimarães. 2005.

VELLOSO, J. Cotistas e não-cotistas: rendimento de alunos da Universidade de Brasília. **Cad. Pesqui.**, São Paulo, v. 39, n. 137, p. 621-644, 2009.

VILELA, L.; TACHIBANA, T.Y.; MENEZES-FILHO, N.A.; KOMATSU, B. As cotas nas universidades públicas diminuem a qualidade dos ingressantes?. **Estudos em Avaliação Educacional**. São Paulo. v. 28, p. 652, 2017.

ZAINKO, M. A. S. Avaliação da educação superior no brasil: processo de construção histórica. **Avaliação: Revista da Avaliação da Educação Superior**. São Paulo. v. 13, n. 3, p. 827-831, 2008.

**APÊNDICE A – REGRESSÕES QUANTILICAS ESTIMADAS**

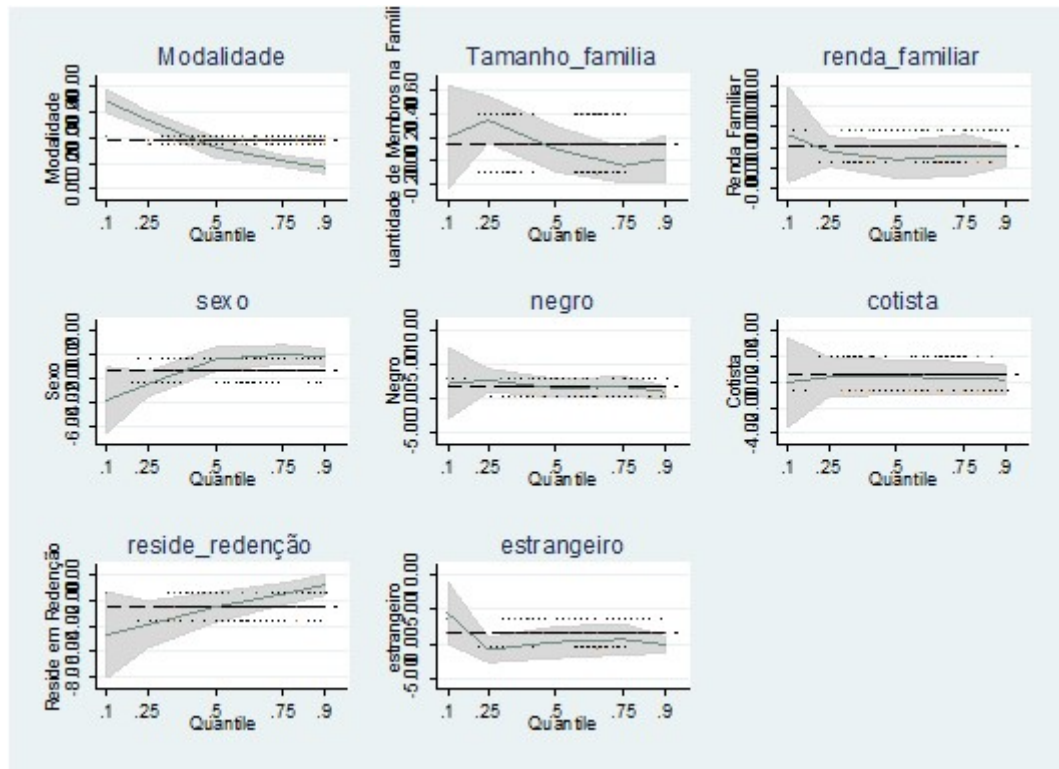
Regressão quantílica simultânea			Número de obs			
			2.967			
Bootstrap (1000)	SEs	0.100	Pseudo R2	=	0.172	
		0.250	Pseudo R2	=	0.125	
		0.500	Pseudo R2	=	0.073	
		0.750	Pseudo R2	=	0.049	
		0.900	Pseudo R2	=	0.037	
Covariadas	Coef.	Std.Err.	t	P> t	Intervalo de confiança (95%)	
<b>QUANTIL 10</b>						
Modalidade	33.86	3.190	10.61	0	27.60	40.11
Tamanho família	0.197	0.291	0.680	0.498	-0.373	0.767
Renda familiar	-8.21e-3	-5.e-3	1.640	0.101	-0.000	0.001
Sexo	-3.839	1.380	-2.780	0.005	-6.544	-1.134
Negro	2.289	2.032	1.130	0.260	-1.696	6.273
Cotista	-0.046	2.078	-0.020	0.982	-4.121	4.028
reside Redenção	-4.715	1.559	-3.030	0.003	-7.771	-1.659
Estrangeiro	4.465	2.521	1.770	0.077	-0.477	9.408
Cons	28.75	3.430	8.380	0	22.03	35.48
<b>QUANTIL 25</b>						
Modalidade	26.85	2.034	13.20	0	22.86	30.83
Tamanho família	0.347	0.158	2.190	0.028	0.036	0.657
Renda familiar	-4.16e-3	-2.34e-3	1.780	0.076	-4.30e-5	-8.75e-3
Sexo	-2.474	0.739	-3.350	0.001	-3.923	-1.025
Negro	2.703	0.981	2.760	0.006	0.780	4.627
Cotista	0.375	0.893	0.420	0.675	-1.377	2.126
reside Redenção	-3.860	0.845	-4.570	0	-5.518	-2.202
Estrangeiro	-0.919	1.386	-0.660	0.507	-3.637	1.798
Cons	45.79	2.186	20.95	0	41.51	50.08
<b>QUANTIL 50</b>						
Modalidade	16.10	2.097	7.680	0	11.98	20.21
Tamanho família	0.099	0.120	0.820	0.412	-0.137	0.334
Renda familiar	-2.20e-3	2.14e-4	1.030	0.303	-1.99e-4	6.4e-4
Sexo	-0.435	0.469	-0.930	0.354	-1.353	0.484
Negro	1.601	0.637	2.510	0.0120	0.352	2.850
Cotista	0.374	0.565	0.660	0.508	-0.734	1.482
reside Redenção	-2.484	0.596	-4.170	0	-3.652	-1.315
Estrangeiro	0.196	1.179	0.170	0.868	-2.117	2.508
Cons	64.35	2.198	29.27	0	60.04	68.66
<b>QUANTIL 75</b>						
Modalidade	10.88	1.062	10.24	0	8.798	12.96
Tamanho família	-0.035	0.0864	-0.410	0.682	-0.205	0.134
Renda familiar	3.17e-4	2.29e-4	1.390	0.165	-1.3e-4	7.7e-4
Sexo	-0.0204	0.406	-5e-3	0.960	-0.816	0.775
Negro	1.833	0.585	3.130	2e-4	0.685	2.981
Cotista	0.304	0.523	0.580	0.562	-0.722	1.330
reside Redenção	-1.455	0.462	-3.150	2e-4	-2.360	-0.549
Estrangeiro	0.599	0.832	0.720	0.471	-1.033	2.231

Cons	74.79	1.168	64.04	0	72.50	77.08
QUANTIL 90						
Modalidade	8.628	0.999	8.640	0	6.670	10.59
Tamanho família	1.8e-2	8.3e-2	0.230	0.821	-0.146	0.183
Renda familiar	3.1e-4	1.4e-3	2.240	0.025	3.9e-05	5.8e-5
Sexo	-0.231	0.370	-0.620	0.533	-0.958	0.495
Negro	1.023	0.439	2.330	0.0200	0.163	1.883
Cotista	0.181	0.545	0.330	0.740	-0.887	1.248
Reside Redenção	-0.742	0.343	-2.170	0.0300	-1.414	-0.070
Estrangeiro	-0.151	0.690	-0.220	0.827	-1.505	1.203
Cons	80.95	1.101	73.52	0	78.79	83.11

Fonte: Elaborado pelo autor a partir de dados da UNILAB.

Nota: Usou-se como base: modalidade (EAD); sexo (mulher); negro (não negro); cotista (não cotista); reside em Redenção (não residente em Redenção); estrangeiro (ser nativo).

## APÊNDICE B – MUDANÇAS NOS COEFICIENTES QUANTÍLICOS



Fonte: Elaboração própria.