



Análise do desempenho acadêmico de alunos estrangeiros no Brasil: o caso da Universidade da Integração da Lusofia Afrobrasileira (UNILAB)

Analysis of the academic performance of foreign students in Brazil: the case of the Universidade da Integração da Lusofia Afrobrasileira (UNILAB)

DOI: 10.55905/oelv21n7-039

Recebimento dos originais: 05/06/2023

Aceitação para publicação: 05/07/2023

Antonio Célio Ferreira dos Santos

Mestre em Economia Rural

Instituição: Universidade Federal do Ceará - Programa de Pós-Graduação em Economia Rural (UFC - PPGER)

Endereço: Av. da Universidade, 2853, Benfica, Fortaleza - CE, CEP: 60020-181

E-mail: acfsantos@unilab.edu.br

Francisco José Silva Tabosa

Doutor em Economia

Instituição: Universidade Federal do Ceará - Programa de Pós-Graduação em Economia Rural (UFC - PPGER)

Endereço: Av. da Universidade, 2853, Benfica, Fortaleza - CE, CEP: 60020-181

E-mail: franzetabosa@ufc.br

Pablo Urano de Carvalho Castelar

Doutor em Economia

Instituição: Universidade Federal do Ceará (UFC)

Endereço: Av. da Universidade, 2853, Benfica, Fortaleza - CE, CEP: 60020-181

E-mail: pcastelar@ufc.br

Domingos Isaias Maia Amorim

Doutorando em Economia

Instituição: Escola Superior de Agricultura "Luiz de Queiroz" da Universidade de São Paulo (ESALQ- USP)

Endereço: Av. Pádua Dias, 11, Agronomia, Piracicaba - SP, CEP: 13418-900

E-mail: domingos_isaias@usp.br

Maria Josiell Nascimento da Silva

Doutoranda em Economia

Instituição: Escola Superior de Agricultura "Luiz de Queiroz" da Universidade de São Paulo (ESALQ- USP)

Endereço: Av. Pádua Dias, 11, Agronomia, Piracicaba - SP, CEP: 13418-900

E-mail: nascimentojosiell@hotmail.com

Francisca Ingrid Gouveia Ferreira

Mestranda em Economia Rural

Instituição: Universidade Federal do Ceará - Programa de Pós-Graduação em Economia Rural (UFC - PPGER)

Endereço: Av. da Universidade, 2853, Benfica, Fortaleza - CE, CEP: 60020-181

E-mail: ingridgouve@gmail.com

RESUMO

O objetivo deste estudo foi analisar o desempenho acadêmico dos alunos estrangeiros no Brasil no semestre 2018.1, com foco na UNILAB. Para isso, utilizou-se o método do propensity score matching com balanceamento por entropia. Os resultados mostraram que os alunos brasileiros possuem um desempenho acadêmico melhor do que os alunos estrangeiros. Grande parte desse resultado deve-se a baixa qualidade do ensino médio dos alunos estrangeiros, além da dificuldade na língua portuguesa e matemática.

Palavras-chave: desempenho acadêmico, educação, UNILAB.

ABSTRACT

The objective of this study was to analyze the academic performance of foreign students in Brazil in the 2018.1 semester, focusing on UNILAB. For this, the propensity score matching method with entropy balancing was used. The results showed that Brazilian students have a better academic performance than foreign students. A large part of this result is due to the low quality of secondary education for foreign students, in addition to the difficulty in the Portuguese language and mathematics.

Keywords: academic performance, education, UNILAB.

1 INTRODUÇÃO

As questões educacionais têm tomado dimensão cada vez maior na análise econômica, sendo recorrente o tema da educação como propulsora do crescimento e redutora da desigualdade. Esse efeito é discutido e confirmado que a educação promove o crescimento, a incorporação de tecnologias, a geração de externalidades positivas e

aumento dos ganhos salariais (ALAN E TIENDA, 2007; ARCIDIANO, AUCEJO E SPENNER, 2012; KRISHNA E TARASOV, 2013).

O Brasil caracteriza-se por ser um país onde existe um ingresso de alunos estrangeiros, principalmente oriundo de países do continente africano, que tem a língua portuguesa como predominante. Esses alunos ingressam no país em busca de uma melhor qualificação profissional e/ou emprego que lhes proporcionem melhores salários e uma melhor qualidade de vida.

Ao analisar esta temática, buscou-se analisar o caso específico da Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB), que de acordo com a sua Lei de criação (Lei No 12.289 de 20 de julho de 2010), 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 e Timor Leste, bem como promover o desenvolvimento regional e o intercâmbio cultural, científico e educacional.

Ao fomentar a Cooperação Sul-Sul, a UNILAB atende à diretrizes internacionais de ampliação da oferta de cursos superiores em regiões carentes, das relações de cooperação com o continente africano. A universidade busca construir uma ponte histórica e cultural entre o Brasil e os países de língua portuguesa, especialmente os da África, compartilhando soluções inovadoras para processos históricos similares. E, ainda deseja auxiliar no fortalecimento de uma rede internacional que, com respeito à soberania dos países parceiros, permitirá a realização de ações e intervenções de apoio técnico, acadêmico, científico, cultural e humanitário. Assim, a UNILAB oferece 50% das vagas para alunos vindos da África.

Os candidatos estrangeiros são submetidos a uma avaliação do histórico escolar do Ensino Médio (Secundário) e prova de redação, realizadas nos próprios países de origem. Os interessados devem se inscrever nas Missões Diplomáticas brasileiras dos países parceiros (Angola, Cabo Verde, Guiné-Bissau, Moçambique, São Tomé e Príncipe e Timor Leste). O calendário de seleção é divulgado através de editais.

Vale ressaltar que os 50% previstos em sua lei de criação para alunos estrangeiros lusófonos de origem africana em seus oito anos de existência ainda não foi atingido, e destes alunos vindos da África serem em sua grande maioria negra e pobre, neste trabalho, procuramos analisar os alunos previstos no sistema de cotas brasileiro.

Diante disso cogita-se o seguinte problema: Ao entrar na universidade os alunos estrangeiros obtêm desempenhos semelhantes aos alunos brasileiros? Nesse sentido, o objetivo desse estudo é analisar o desempenho dos alunos estrangeiros no rendimento escolar (IDE – Índice de desenvolvimento educacional) dos alunos da UNILAB.

A despeito da diversidade de estudos sobre a política de alunos estrangeiros 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, onde foi utilizada como metodologia a técnica de Propensity Score Matching (PSM) balanceado por entropia, tendo recorrido, também, ao teste de sensibilidade de Rosebaum e Rubin (1983).

Nesse estudo, além desta introdução, é composto por mais quatro seções. A seção dois se refere a metodologia, elaborou-se a estatística descritiva e descreve o modelo proposto para análise do problema em questão, Propensity Score Matching (PSM) balanceado por entropia e análise de sensibilidade. Na quarta seção analisam-se os resultados encontrados, e por fim são apresentadas as conclusões do estudo.

2 METODOLOGIA

A metodologia será vista em duas subseções: Base de dados e métodos do propensity score matching, balanceado por entropia e análise de sensibilidade.

2.1 BASES DE DADOS

A base de dados neste trabalho será analisada em duas subseções: Dados da análise e estatística descritiva.

2.1.1 Dados da análise

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 a DTI (Diretoria de Tecnologia da informação), no Sistema Acadêmico (SIAC) e pelo Centro de Processamento de Dados. Os dados obtidos apresentam, além das informações socioeconômicas dos ingressantes, informações sobre, número de matrícula e acompanhamento acadêmico do estudante no decorrer do curso. A amostra abrange 2.964 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 aprovado no Conselho universitário a 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, onde 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 dos alunos da graduação é calculado pelo Índice de Desempenho do estudante (IDE). O IDE é a média das notas finais das disciplinas, ponderadas pela carga horária respectiva, multiplicada por um redutor por reprovação por falta e por trancamento de disciplina, conforme exposto abaixo:

$$IDE = \left(\frac{\sum_{n=1}^n NOi.CH}{\sum_{n=1}^n CHi} \right) \cdot \left(1 - \frac{0,3.NTR+0,7.NRF}{NT} \right), \quad (1)$$

Onde:

IDE – Índice de Desempenho do estudante;

I – i-ésimadisciplina 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;

NOi – nota em que o aluno obteve em uma dada disciplina i; e,

CHi carga horária da disciplina i.

OBS1: O primeiro fator:

$\left(\frac{\sum_{i=1}^n NOi \cdot CHi}{\sum_{i=1}^n CHi}\right)$, 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. Caso o estudante não teve 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 que permite 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.

2.1.2 Estatística descritiva

As variáveis socioeconômicas utilizadas na análise são: estrangeiro, renda familiar tamanho da família, gênero e reside em Redenção. A variável estrangeiro é uma variável dicotômica, onde 1 representa o aluno brasileiro e 0 o aluno estrangeiro. A variável tamanho da família é o número de indivíduos que residem na residência do aluno. Gênero corresponde a 1 se o aluno for do sexo masculino e 0 caso contrário. E a reside em Redenção, 1 se o aluno reside em Redenção e 0 caso contrário.

Salientamos 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 considerarmos erros de preenchimento dos formulários e ou valores muito discrepantes do restante da população.

Na tabela 1, está expressa a estatística descritiva dos alunos da Unilab no semestre 2018.1, o qual apresentou uma média da renda familiar de R\$1.205,82. Há alunos que, em sua resposta, chegou a declarar renda familiar de R\$ 100,00, em uma situação ainda mais extrema, outros declararam uma renda de R\$ 20.000,00.

A média foi de 4,699 membros por família e além disso, outros resultados encontrados referem-se valem apenas ressaltar, tais como às questões de gênero e nacionalidade, onde 47,6 % dos alunos são do sexo masculino, ou seja, as mulheres ocupando um espaço maior frente aos homens no ingresso como aluno 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%.

No que diz respeito ao local de residência 25,8 % dos alunos afirmaram que residem em Redenção, mas vale ressaltar que com o crescimento da universidade expandindo para os municípios de Acarape (CE) e São Francisco do Conde (BA), cabe a universidade atualizar seu banco de dados de modo que estes municípios podem ter recebido migração de estudantes, de tal forma que possa descrever a correlação entre os cursos ofertados em cada município e local de residência dos alunos, objetivando para uma melhor análise futura sobre a correlação entre os desempenhos dos alunos e o local onde residem.

Ao comparar os níveis de renda entre estrangeiros e brasileiros, observa-se que a renda dos brasileiros, em média, é maior do que dos estrangeiros (R\$1.342,58 contra R\$689,82). Em relação ao gênero, o percentual masculino de estrangeiros são maiores do que dos brasileiros; assim como o tamanho da família.

Tabela 1 – Estatística descritiva – Total, Estrangeiros e Brasileiros.

Variáveis	Obs	Média	Std.Dev	Min	Max
Estrangeiro	2,964	0,209514	0,407031	0	1
Tamanho da família	2,964	4,699055	2,847903	1	26
Renda familiar	2,964	1.205,82	1.273,28	100	20.000
Gênero	2,964	0,476721	0,499542	0	1
Reside em Redenção	2,964	0,259109	0,43822	0	1
Estrangeiro					
Tamanho da família	621	8,771337	3,127281	1	26
Renda familiar	621	689,82	889,26	100	20.000
Gênero	621	0,677939	0,467643	0	1
Reside em Redenção	621	0,37037	0,483293	0	1

Brasileiro					
Tamanho da família	2,343	3,619718	1,452371	1	14
Renda familiar	2,343	1.342,58	1323,82	100	18.500
Gênero	2,343	0,423389	0,494201	0	1
Reside em Redenção	2,343	0,22962	0,420678	0	1

Fonte: Elaborado pelos autores a partir de dados da UNILAB.

2.2 MÉTODO DO PROPENSITY SCORE MATCHING (PSM)

Analisou-se 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, 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, p. 10).

O Propensity Score Matching (PSM) é um dos métodos existentes para análise de avaliação de impacto, no qual consistem na seleção de um grupo de controle, com base em um modelo de probabilidade (Logit/Probit) de participar do tratamento, comparáveis com um grupo de tratamento por meio do pareamento (matching) dos grupos, utilizando características observáveis. A partir disso, podem-se comparar os resultados dos tratados e não tratados com escores de propensão semelhantes para obter o efeito dos alunos estrangeiros. O escore de propensão, desenvolvido por Rosenbaum e Rubin (1983), é definido como a probabilidade condicional de receber um tratamento, dado por características do pré-tratamento observáveis, isto é:

$$p(X) = \Pr(T = 1|X) = E(T|X) \quad (2)$$

Onde

T indica a posição do tratamento para o aluno (1 – para o aluno brasileiro, 0 – aluno estrangeiro) e X é o vetor de características. Assim, o efeito médio do tratamento sobre os tratados (ATT), é dado por:

$$\tau = E_{P(X)|T=1}\{E[Y_{1i}|T_i = 1, p(X_i)] - E[Y_{0i}|T_i = 0, p(X_i)]|T_i = 1\} \quad (3)$$

Para que se possa calcular o ATT, são necessárias que sejam atendidas duas hipóteses. A primeira hipótese diz a respeito à independência condicional, que implica que os fatores não observados não afetam a participação. Posteriormente é realizado um sorteio aleatório para a concessão do programa.

Hipótese 1: $Y_i(1), Y_i(0) \perp T_i | X_i$

A segunda hipótese trata do suporte comum, que conforme, este pressuposto garante que as observações de tratamento têm observações de comparação "nas proximidades" na distribuição de escore de propensão. Para atender essa hipótese, as variáveis escolhidas para o matching foram aquelas com mais similaridades entre os alunos brasileiros e estrangeiros.

Hipótese 2: Para algum $c > 0$, $c < p(x) < 1 - c$.

De acordo com Rosenbaum e Rubin (1983), essas duas hipóteses combinadas são conhecidas como ignorabilidade forte. Além disso, ressalta-se que a estimação do escore de propensão não é suficiente para estimar o ATT, sendo necessário aliá-la ao matching. Com relação a isso, diferentes métodos podem ser utilizados para fazer o pareamento, dentre eles, destacam-se os que seguem: Pareamento por vizinho mais próximo (Nearest Neighbor Matching), Pareamento por alcance (Radius Matching), Pareamento Kernel (Kernel Matching).

2.2.1 Análise de sensibilidade

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

Dessa forma, dado que a probabilidade de participação de um indivíduo i seja:

$$\pi_i = \Pr(D_i = 1 | x_i) = F(\beta x_i + \gamma x_i) \quad (4)$$



Caso não haja viés de seleção no escore de propensão, então γ será igual a zero e a probabilidade de participação será exclusivamente dada pelas características observáveis. Por outro lado, quando existe de viés de seleção, dois indivíduos com as mesmas covariáveis observadas x terão diferentes chances de receber tratamento $\frac{\theta_i}{1-\theta_i}$ e $\frac{\theta_j}{1-\theta_j}$. A odds ratio será dada por:

$$\frac{\frac{\theta_i}{1-\theta_i}}{\frac{\theta_j}{1-\theta_j}} = \frac{\theta_i(1-\theta_j)}{\theta_j(1-\theta_i)} = \exp[\gamma(u_i - u_j)] \quad (5)$$

Quando os indivíduos possuem as mesmas características observáveis, o vetor x se cancela. Então, se não houver diferenças nas variáveis não observadas ($u_i = u_j$) e se estas variáveis não influenciarem a probabilidade de participação ($\gamma = 0$), a odds ratio será igual a 1, denotando que não existe viés de seleção. Deste modo, quando há presença de não-observáveis, suas odds de participação são diferentes, ou seja, a odds ratio é diferente de 1. Portanto, na análise de sensibilidade é avaliado o quanto do efeito do programa é alterado pela mudança nos valores de γ e de $u_i - u_j$. Isso significa examinar os limites da odds ratio de participação. Rosenbaum (2002) mostra que (4) implica os seguintes limites para a odds ratio):

$$\frac{1}{e^\gamma} \leq \frac{\alpha_i(1-\theta_j)}{\theta_i(1-\theta_i)} \leq e^\gamma \quad (6)$$

Onde:

Se $e\gamma=1$ os indivíduos pareados possuem a mesma probabilidade de participação. Entretanto, se $e\gamma = 2$, então indivíduos aparentemente semelhantes em termos de x diferem nas probabilidades de receberem tratamento por um fator de até 2.

2.3 BALANCEAMENTO POR ENTROPIA

O método de balanceamento por entropia consiste em um esquema de ponderação que permite o ajustamento dos pesos que satisfazem um conjunto potencialmente grande



de restrições de momentos. Esse método permite que o pesquisador especifique um nível de equilíbrio desejável para as covariadas, usando um conjunto de condições associadas aos momentos da distribuição.

Esse procedimento foi desenvolvido por Hainmueller (2012), que garante haver algumas vantagens nas etapas de pré-processamento dos dados para estimação posterior do efeito tratamento em relação aos métodos de avaliação de impactos baseados em scores de propensão.

Considere 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 o aluno 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 do município i , tais que, $X_i = [X_{i1}, X_{i2}, \dots, X_{ij}]$ refere-se ao vetor de características do aluno i e X_j refere-se ao vetor coluna com j -th covariada. A densidade das covariadas nas populações de tratamento e controle é dada por $f_{X|D=1}$ e $f_{X|D=0}$, respectivamente. O resultado potencial $Y_i(D_i)$ corresponde ao par de resultados para o aluno i dadas as condições de tratado e controle, assim, o resultado observado é dado por $Y = Y(1)D + (1 - D)Y(0)$.

O Efeito Médio Tratamento sobre os Tratados (ATT) é dado por

$$\tau = E[Y(1)|D = 1] - E[Y(0)|D = 1] \quad (7)$$

A primeira esperança pode ser diretamente identificada do grupo de tratados, mas a segunda corresponde ao contrafactual, o qual não é observada. Rosenbaum e Rubin (1983) mostra que, assumindo seleção nos observáveis, $Y(0) \perp D|X$, e sobreposição, $\Pr(D = 1|X = x) < 1$ para todo x no suporte de $f_{X|D=1}$, o ATT é identificado como:

$$\tau = E[Y|D = 1] - \int E[Y|X = x, D = 0] f_{X|D=1}(x) dx \quad (8)$$



Para estimar o último termo da equação 8, a distribuição da covariável no grupo de controle necessita ser ajustada para torná-la semelhante à distribuição no grupo de tratados, tal que o indicador de tratamento D se torne mais perto de ser ortogonal em relação às covariáveis. Uma variedade de métodos de pré-processamento de dados, tais como pareamento pelo vizinho mais próximo, escore de propensão têm sido propostos para reduzir o desequilíbrio na distribuição de variáveis independentes. Uma vez que as distribuições de variáveis independentes são ajustadas, métodos de análise padrão, tais como a regressão pode ser posteriormente utilizada para estimar o tratamento com menor erro e modelo de dependência (IMBENS, 2004; RUBIN, 2006; HO et al., 2007; SEKHON, 2009).

Considere o caso mais simples onde o efeito tratamento nos dados pré-processados é estimado usando a diferença nos resultados médios entre os grupos de tratados e controle ajustado. Um método de pré-processamento popular é usar escore de propensão ponderado (HIRANO; IMBENS, 2001; HIRANO et al, 2003), onde a média contrafactual é estimada como:

$$E[Y(0)|\widehat{D} = 1] = \frac{\sum_{\{i|D=0\}} Y_i d_i}{\sum_{\{i|D=0\}} d_i} \quad (9)$$

Onde:

as unidades de controle recebem um peso dado por $d_i = \frac{\hat{p}(x_i)}{1-\hat{p}(x_i)}$. $\hat{p}(x_i)$ na equação 8 é o escore de propensão, o qual é comumente estimado através de uma regressão probit ou logit. Se este modelo estiver corretamente especificado, então o peso estimado d_i vai assegurar que a distribuição do covariável das unidades de controle reponderadas irá corresponder a distribuição no grupo de tratamento. No entanto, na prática, essa abordagem muitas vezes não consegue equilibrar conjuntamente todas as covariáveis.

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 do pesquisador sobre os momentos de amostra. Considere w_i o peso do balanceamento por entropia escolhido para cada unidade de controle, os



quais foram encontrados pelo seguinte esquema de reponderação que minimiza a distancia métrica de entropia:

$$\min_{w_i} H(w) = \sum_{\{i|D=0\}} w_i \log(w_i/q_i) \quad (10)$$

Sujeito as restrições de equilíbrio e normalização

$$\sum_{\{i|D=0\}} w_i c_{ri}(X_i) = m_r \quad \text{com } r \in 1, \dots, R \quad (11)$$

$$\sum_{\{i|D=0\}} w_i = 1 \quad (12)$$

E $w_i \geq 0$ para todo i , tal que

$$D = 0 \quad (13)$$

Onde

$q_i = 1/n_0$ é um peso base e $c_{ri}(X_i) = m_r$ descreve um conjunto de R restrições impostas aos momentos das covariadas no grupo de controle reponderados. Inicialmente, escolhe-se a covariada que será incluída na reponderação. Para cada covariada, especifica-se um conjunto de restrições de balanceamento para equiparar os momentos das distribuições das covariadas entre os grupos de tratamento e controles reponderados. As restrições de momentos podem ser a média (primeiro momento), a variância (segundo momento), e a assimetria (terceiro momento). Uma restrição típica do balanceamento é formulada de tal forma que m_r contenha o momento de uma covariada específica X_j para o grupo de tratamento e a função de momento para o grupo de controle é especificada como: $c_{ri}(X_{ij}) = X_{ij}^r$ ou $c_{ri}(X_{ij}) = (X_{ij} - \mu_j)^r$ com média μ_j .

Dessa forma, o balanceamento por entropia procura, para um conjunto de unidades, pesos $W = [w_1, \dots, w_{n_0}]'$ no qual minimiza a equação (8), distancia de entropia entre W e o vetor base de pesos $Q = [q_1, \dots, q_{n_0}]'$, sujeita as restrições de balanceamento

na equação (11), restrição de normalização (Eq 12), e restrição de não-negatividade (Eq 13).

3 RESULTADOS

Nesta seção, serão apresentados os resultados do propensity score matching para analisar o desempenho dos alunos estrangeiros na UNILAB no IDE. A priori, conforme especificado na metodologia, foram utilizados critérios de seleção do modelo.

Conforme exposto na tabela 2, os dois modelos Probit e Logit tiveram o mesmo resultado no que diz respeito ao casos corretamente classificados, obtendo 82,15% e dado que 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 2 – 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 de dados da UNILAB.

A tabela 3 apresentam os resultados do modelo probit e seus efeitos marginais. Todas as variáveis testadas se mostraram ser estatisticamente significantes no modelo. Apenas a variável renda familiar aumenta as chances de ser aluno estrangeiro. As demais variáveis reduzem as chances do aluno ser estrangeiro.

Tabela 3 – Efeitos marginais do Modelo Probit.

Estrangeiro	Coefficientes	Limite inferior	Limite Superior
Tamanho da família	0,5409 * (0,0205)	0,5008	0,5810
Renda familiar	-0,7323 * (0,0608)	-0,8514	-0,6132
Gênero	0,4652 * (0,0831)	0,3024	0,6280

Reside em Redenção	0,1777 * (0,0868)	0,0075	0,3479
Constante	0,8275 * (0,3969)	0,0494	1,6055

Fonte: Elaborado pelos autores a partir de dados da UNILAB.
Obs: $P < 0,10^*$.

A Tabela 4 mostra as médias amostrais de todas as covariadas divididas entre os dois grupos: grupo de tratamento e controle: brasileiro, ou seja, o grupo de tratamento e estrangeiro, que é o grupo de controle potencial. 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 desta tabela apresentam a comparação entre as médias das covariáveis que influenciam a probabilidade de ser brasileiro, revelam que os alunos brasileiros antes do balanceamento diferem dos estrangeiros 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.

A comparação das médias das características de pré-tratamento do grupo de tratamento com as do grupo de controle revela 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 alunos brasileiros e estrangeiros.

Tabela 4 – Diferenças nas características observáveis antes e depois do pareamento

Pré-balanceamento	Tratado			Controle		
	Média	Variância	Assimetria	Média	Variância	Assimetria
Tamanho da família	8,771	9,78	0,5085	3,62	2,109	0,8183
Ln da renda familiar	6,385	0,2107	1,153	6,904	0,5922	-0,1196
Gênero	0,6779	0,2187	-0,7616	0,4234	0,2442	0,3101
Reside em Redenção	0,3704	0,2336	0,5369	0,2296	0,177	1,286
Pós balanceamento	Tratado			Controle		



	Média	Variância	Assimetria	Média	Variância	Assimetria
Tamanho da família	8,771	9,78	0,5085	8,771	16,13	0,1144
Ln da renda familiar	6,385	0,2107	1,153	6,385	1,017	-0,5877
Gênero	0,6779	0,2187	-0,7616	0,6779	0,2184	-0,7614
Reside em Redenção	0,3704	0,2336	0,5369	0,3703	0,2333	0,537

Fonte: Elaborado pelos autores a partir de dados da UNILAB.

Conforme se pode verificar na tabela 5 abaixo, o IDE dos alunos brasileirosna de foi de 79,98, valor este superior ao dos alunos estrangeiros, cujo resultado obtido foi de 70,71. Assim, foi apresentada uma diferença positiva e estatisticamente significativa de 9,27. Isso implica dizer que os alunos brasileiros possuem um desempenho médio acadêmico melhor do que os alunos estrangeiros.

Dentre outros fatores que afetam o rendimento dos alunos estrangeiros, pode-se destacar o baixo nível de escolaridade no ensino médio dos alunos estrangeiros, sendo que todos eles são oriundos do continente africano. Para amenizar esse problema, a UNILAB adotou, nos últimos semestres, um projeto de extensão do o intuito de proporcionar aos alunos estrangeiros um curso de nivelamento de português e matemática. Espera-se que essa ação proporcione melhores resultados nos semestres seguintes.

Tabela 5 – Resultado das estimativas de Propensity Score Matching balanceado por entropia

Variável	Amostra	Tratados	Controle	Diferença	Erro padrão	t-stat
Estrangeiro	ATT	79,9835	70,7105	9,2729	2,6919	3,44

Fonte: Elaborado pelos autores a partir de dados da UNILAB.

A análise do viés de seleção é feita com amparo nos resultados na tabela 6. Deste modo, testou-se a robustez dos resultados com base na hipótese de que variáveis não observáveis impactam nos resultados estimados. Como pode ser observado, os valores ($\Gamma=1$ até $\Gamma= 1,3$) foram estatisticamente significantes, corroborando os resultados encontrados no PSM. Logo, não existem variáveis omitidas, afetando a seletividade do grupo de tratamento e, conseqüentemente, nos resultados, não sendo necessário aplicar a técnica de correção pelo viés mínimo e viés corrigido.

Tabela 6 – Análise de sensibilidade dos Limites de Rosenbaum

Γ	sig+	sig-	t-hat+	t-hat-	CI+	CI-
1	0	0	78.58	78.58	78.13	79.02
1.05	0	0	78.32	78.84	77.855	79.275
1.1	0	0	78.065	79.085	77.595	79.515
1.15	0	0	77.815	79.315	77.34	79.735
1.2	0	0	77.575	79.53	77.09	79.945
1.25	0	0	77.34	79.735	76.845	80.145
1.3	0	0	77.11	79.925	76.605	80.335
1.35	0	0	76.89	80.11	76.375	80.515
1.4	0	0	76.67	80.285	76.145	80.685
1.45	0	0	76.455	80.455	75.925	80.85
1.5	0	0	76.245	80.61	75.71	81.005
1.55	0	0	76.04	80.765	75.495	81.155
1.6	0	0	75.84	80.91	75.29	81.3

Fonte: Elaborado pelos autores a partir de dados da UNILAB.

4 CONCLUSÕES

O objetivo deste estudo foi analisar o desempenho acadêmico dos alunos estrangeiros no Brasil no semestre 2018.1, com foco na UNILAB. Para isso, utilizou-se o método do propensity score matching com balanceamento por entropia.

Os principais resultados indicam que todas as variáveis explicativas aumentam as chances do aluno ser brasileiro, exceto a variável renda domiciliar. Ao analisar o resultado das estimativas do propensity score Matching balanceado por entropia, observa-se que os alunos brasileiros possuem um desempenho acadêmico melhor do que os alunos estrangeiros.

Os alunos estrangeiros quando ingressam na universidade são recepcionados pelo PAIE (Programa de acolhimento e Integração de estudantes estrangeiros), o qual além de hospedarem em hotel e ou pousada, são feitas reuniões sistemáticas para o monitoramento, o que oportuniza correções no processo seletivo e são apontadas dificuldades do aluno ingresso bem como criadas estratégias para solucionar dificuldades, como de língua portuguesa e matemática, os quais surgiram projetos de extensão visando o atendimento a estas demandas com criação de cursos de nivelamento.

Sugere-se a realização de futuros estudos para o caso específico da UNILAB, bem como para as demais universidades, visando analisar políticas que equilibrem esse desempenho acadêmico entre alunos brasileiros e estrangeiros.

REFERÊNCIAS

ALAN, Sigal; TIENDA, Marta. Diversity, opportunity and the shifting meritocracy in higher education. **American Sociological Review, Califórnia**, v. 72, n. 4, 2007. Disponível em: <https://journals.sagepub.com/doi/10.1177/000312240707200401>. Acesso em 16 de ago de 2019.

ARCIDIACONO, Peter; AUCEJO, Esteban; SPENNER, Ken. What happens after enrollment? An analysis of the time path of racial differences in gpa and major choice. **IZA. Journal of Labor Economics, Chicago**, v. 1, n. 5, 2012. Disponível em: <https://izajole.springeropen.com/articles/10.1186/2193-8997-1-5>. Acesso em 19 de ago de 2019.

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. Disponível em: . Acesso em: 6 mai. 2019.

CORBUCCI, P. R. Desafios da educação superior e desenvolvimento no Brasil. Brasília: Ipea, 2007. (**Texto para Discussão**, n. 1287).

DINIZ, C. C.; VIEIRA, D. J. Ensino superior e desigualdades regionais: notas sobre a experiência recente do Brasil. **Revista Paranaense de Desenvolvimento**, v. 36, n. 129, p. 99-115, 2015.

KRISHNA, Kala; TARASOV, Alexander. **Affirmative action: one size does not fit all.**, Cambridge: National Bureau of Economic Research , oct. 2013.

RAVALLION, M. (2007). **Evaluating Anti-Poverty Programs**. In: Handbook of Development Economics (Vol. 4).

RODRIGUES, Luciana de Oliveira: Ensaio sobre diferencial de desempenho escolar entre alunos de escolas rurais e urbanas no Brasil, **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. (1983). The Central Role of the Propensity Score in Observational Studies for **Causal Effects Biometrika**, 70(1), pp. 41-55.

HAINMUELLER, J. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. **Political Analysis**, v. 20 n.1, p. 25-46, 2012

HIRANO K; IMBENS G. Estimation of Causal Effects Using Propensity Score Weighting: An Application of Data on Right Heart Catheterization. **Health Services and Outcomes Research Methodology**. 2001, p. 259–278

HIRANO K, IMBENS G, RIDDER G. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. **Econometrica**, v. 71 n.4, p.1161-1189, 2003

HO, D.E.; IMAI, K.; KING, G.; STUART, E.A. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. **Political Analysis**, v.15 n.3, p.199. 2007

IMBENS, G.W. Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review. **Review of Economics and Statistics**, v.86 n.1, p.4-29, 2004

ROSENBAUM, P. **Observational Studies**. 2nd edition. New York: **Springer**, 2002.

ROSENBAUM, P. R.; RUBIN, D. **Observational Studies**. New York: Springer, 2002

RUBIN, D.B. **Matched Sampling for Causal Effects**. Cambridge University Press. 2006

SEKHON, J.S. Opiates for the Matches: Matching Methods for Causal Inference. **Annual Review of Political Science**, 2009.