

**João Pedro Fernandes Martins**

**Avaliação da desigualdade no acesso ao  
ensino superior no Brasil em 2019**

Niterói - RJ, Brasil

20 de Julho de 2022

**João Pedro Fernandes Martins**

**Avaliação da desigualdade no acesso  
ao ensino superior no Brasil em 2019**

**Trabalho de Conclusão de Curso**

Monografia apresentada para obtenção do grau de Bacharel em  
Estatística pela Universidade Federal Fluminense.

Orientador(a): Profa. Dra. Ana Beatriz Monteiro Fonseca

Co-Orientador(a): Prof. Dr. Rafael Santos Erbisti

Niterói - RJ, Brasil

20 de Julho de 2022

**João Pedro Fernandes Martins**

**Avaliação da desigualdade no acesso ao ensino superior no Brasil em 2019**

Monografia de Projeto Final de Graduação sob o título “*Avaliação da desigualdade no acesso ao ensino superior no Brasil em 2019*”, defendida por João Pedro Fernandes Martins e aprovada em 20 de Julho de 2022, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

---

**Profa. Dra. Ana Beatriz Monteiro Fonseca**  
Departamento de Estatística – UFF

---

**Prof. Dr. Rafael Santos Erbisti**  
Departamento de Estatística – UFF

---

**Profa. Dra. Márcia Marques de Carvalho**  
Departamento de Estatística – UFF

---

**Profa. Dra. Mariana Albi de Oliveira Souza**  
Departamento de Estatística – UFF

Ficha catalográfica automática - SDC/BIME  
Gerada com informações fornecidas pelo autor

M379a Martins, João Pedro Fernandes  
Avaliação da desigualdade no acesso ao ensino superior no Brasil em 2019 / João Pedro Fernandes Martins ; Ana Beatriz Monteiro Fonseca, orientadora ; Rafael Santos Erbisti, coorientador. Niterói, 2022.  
61 f. : il.

Trabalho de Conclusão de Curso (Graduação em Estatística)-Universidade Federal Fluminense, Instituto de Matemática e Estatística, Niterói, 2022.

1. Acesso ao ensino superior. 2. Regressão logística. 3. Índice de dissimilaridade. 4. Índice de Oportunidades Humanas. 5. Produção intelectual. I. Fonseca, Ana Beatriz Monteiro, orientadora. II. Erbisti, Rafael Santos, coorientador. III. Universidade Federal Fluminense. Instituto de Matemática e Estatística. IV. Título.

CDD -

# Resumo

Atualmente, a oportunidade de começar a estudar em uma universidade não parece chegar da mesma forma para todos aqueles que concluem o ensino médio. Em vista disso, muito se discute sobre quais fatores poderiam determinar uma maior chance de ingresso no ensino superior. O objetivo do presente trabalho foi realizar uma análise exploratória sobre os microdados levantados na Pesquisa Nacional por Amostra por Domicílios Contínua (PNADc), realizada em 2019, especialmente aqueles relacionados ao tema de Educação e, a partir deles, construir um modelo de regressão logística para relacionar a chance de iniciar uma graduação a partir de características de cada indivíduo, como sexo, cor/raça, região geográfica, grau de escolaridade do chefe do domicílio e renda domiciliar. Com base neste modelo, foi estimado o Índice de Oportunidades Humanas (IOH), proposto inicialmente por Barros et al. (2009), para o Brasil no ano de 2019, sob a perspectiva de avaliar as desigualdades com relação ao ingresso em instituições de ensino superior (IES), possibilitando, também, comparar o acesso às IES públicas e privadas, em separado. Ao ser desconsiderado o tipo de instituição de ensino, o IOH encontrado foi igual a 0,331. Investigando-se somente o acesso às IES públicas, o IOH para 2019 foi de 0,172 e já para o ingresso em instituições particulares, o valor do IOH em 2019 foi de 0,154. Assim, foi percebida uma maior desigualdade de oportunidades em relação ao ingresso na educação superior nas IES privadas do que nas públicas. Destaca-se, também, que, nos três casos estudados, os índices calculados se encontraram distantes de 1, valor que indicaria uma situação onde o acesso à educação superior seria universal, ou seja, possível a todos.

Palavras-chave: Acesso ao ensino superior. Regressão logística. Índice de dissimilaridade. Índice de Oportunidades Humanas.

# Dedicatória

Dedico este trabalho aos meus pais, por todo o apoio nesse caminho.

# Agradecimentos

Aos meus pais, Martha e Ricardo, por todo o amor, pelo suporte em todos os momentos e por terem feito possível que eu chegasse até aqui. Especialmente à minha mãe, que me deu a oportunidade de conhecer a UFF desde cedo, o que me fez sentir muito mais ambientado quando dei início à minha graduação, por sempre me dar forças para seguir em frente e nunca desistir dos meus sonhos.

À minha família, que sempre me incentivou e torceu pelo meu melhor.

À minha orientadora, Dra. Ana Beatriz Monteiro Fonseca, por ter aceitado participar do projeto, pela dedicação e disponibilidade, pela paciência, pelos incentivos e pela verdadeira orientação dada para que eu conseguisse fazer este trabalho desde o zero.

Ao meu coorientador, Dr. Rafael Santos Erbisti, também por ter aceitado o convite de participação no projeto, por todo o empenho, atenção e ensinamentos passados durante este trabalho e durante todo o curso, sobretudo em períodos de ensino remoto.

Às professoras Dra. Márcia Marques de Carvalho e Dra. Mariana Albi de Oliveira Souza, que aceitaram participar da banca examinadora e pelas sugestões que possam dar para este trabalho. Em especial à professora Márcia, que me sugeriu o tema desta monografia, tornando viável que este projeto fosse realizado.

Aos demais professores da Estatística da UFF e de outros departamentos, pela colaboração e pelos conhecimentos transmitidos.

À equipe da secretaria da coordenação do curso e à coordenadora Dra. Ana Maria Lima de Farias, por todo o auxílio prestado e conselhos passados durante o curso.

Aos amigos e colegas que conheci na UFF, pelas vivências compartilhadas, pelos momentos de descontração e, principalmente, pelas trocas de conhecimentos, ajudas sempre necessárias e participações em trabalhos em grupo.

E finalmente, a Deus, por ter me dado a oportunidade de trilhar esta jornada.

# Sumário

## Lista de Figuras

## Lista de Tabelas

<b>1</b>	<b>Introdução</b>	p. 11
1.1	A Desigualdade Social . . . . .	p. 11
1.1.1	Acesso ao ensino superior no Brasil . . . . .	p. 13
1.2	Objetivos . . . . .	p. 16
1.2.1	Objetivo Geral . . . . .	p. 16
1.2.2	Objetivos Específicos . . . . .	p. 16
1.3	Organização . . . . .	p. 17
<b>2</b>	<b>Materiais e Métodos</b>	p. 18
2.1	Pesquisa Nacional por Amostra de Domicílios Contínua . . . . .	p. 18
2.1.1	Demanda Potencial ao Ensino Superior . . . . .	p. 19
2.2	Estimação pontual . . . . .	p. 20
2.3	Modelos Lineares Generalizados . . . . .	p. 22
2.3.1	Família Exponencial de distribuições . . . . .	p. 22
2.3.2	Regressão Logística para dados binários . . . . .	p. 23
2.3.2.1	Estimação dos parâmetros . . . . .	p. 25
2.3.2.2	Inclusão do plano amostral na estimação dos parâmetros	p. 26
2.3.2.3	Razão de chances . . . . .	p. 28

2.3.2.4	Avaliação da qualidade de ajuste do modelo . . . . .	p. 30
2.4	Índice de Oportunidades Humanas . . . . .	p. 32
<b>3</b>	<b>Análise dos Resultados</b>	<b>p. 36</b>
3.1	Análise Exploratória . . . . .	p. 36
3.1.1	Demanda e Ingresso no Ensino Superior . . . . .	p. 36
3.1.2	Outros descritores da demanda potencial ao ensino superior matriculada em alguma IES . . . . .	p. 39
3.2	Estimação e Ajuste do Modelo . . . . .	p. 42
3.3	Estimação do IOH . . . . .	p. 47
<b>4</b>	<b>Conclusões</b>	<b>p. 51</b>
	<b>Referências</b>	<b>p. 53</b>
	<b>Apêndice A</b>	<b>p. 57</b>

# Lista de Figuras

1	Perfil dos indivíduos da demanda potencial ao ensino superior que frequentavam alguma IES de acordo com o sexo e a etnia ( $n = 4.379.783$ )	p. 39
2	Perfil dos indivíduos da demanda potencial ao ensino superior que frequentavam alguma IES de acordo com a região de residência e a renda domiciliar <i>per capita</i> ( $n = 4.379.783$ )	p. 40
3	Perfil dos indivíduos da demanda potencial ao ensino superior de acordo com o tipo de IES que frequentava e o sexo ( $n = 4.379.783$ )	p. 40
4	Perfil dos indivíduos da demanda potencial ao ensino superior de acordo com o tipo de IES que frequentava e a etnia ( $n = 4.379.783$ )	p. 41
5	Perfil dos indivíduos da demanda potencial ao ensino superior de acordo com o tipo de IES que frequentava e a escolaridade do chefe do domicílio ( $n = 4.379.783$ )	p. 41
6	Perfil dos indivíduos da demanda potencial ao ensino superior de acordo com o tipo de IES que frequentava e a renda domiciliar <i>per capita</i> ( $n = 4.379.783$ )	p. 42
7	Perfil dos indivíduos da demanda potencial ao ensino superior de acordo com a região de residência e o tipo de IES que frequentava ( $n = 4.379.783$ )	p. 42
8	Probabilidades estimadas de ingresso em instituições de ensino superior segundo os grupos	p. 58
9	Probabilidades estimadas de ingresso em instituições de ensino superior públicas segundo os grupos	p. 59
10	Probabilidades estimadas de ingresso em instituições de ensino superior privadas segundo os grupos	p. 60

# Lista de Tabelas

1	Descrição das variáveis . . . . .	p. 20
2	Perfil da demanda potencial ao ensino superior (em mil pessoas) — Brasil (2019) . . . . .	p. 37
3	Perfil dos ingressantes e não ingressantes do ensino superior entre a demanda potencial (em mil pessoas) — Brasil (2019) . . . . .	p. 38
4	Resultados da estimação do Modelo Logit geral - Brasil (2019) . . . . .	p. 45
5	Resultados da estimação do Modelo Logit, para a categoria administrativa de IES pública - Brasil (2019) . . . . .	p. 46
6	Resultados da estimação do Modelo Logit, para a categoria administrativa de IES privada - Brasil (2019) . . . . .	p. 46
7	Medidas de qualidade do ajuste para os Modelos Logit - Brasil (2019) . . . . .	p. 47
8	Taxas de cobertura e índices D e IOH para o acesso ao ensino superior - Brasil (2019) . . . . .	p. 48

# 1 Introdução

Na introdução, será apresentado um panorama geral do mundo e do Brasil a respeito da desigualdade social e, em especial, do acesso ao ensino superior no Brasil, além de abordar estudos que concentraram seus esforços sobre o Índice de Oportunidades Humanas (IOH), que mede a desigualdade de oportunidades presente na sociedade para acesso a determinado bem ou serviço. Especificamente, pretende-se nesta pesquisa iniciar uma reflexão sobre o acesso ao ensino superior e como ele é influenciado por fatores sociodemográficos, estabelecendo uma leitura dialética entre educação e desigualdade.

## 1.1 A Desigualdade Social

Ao longo da história contemporânea, a preocupação com a desigualdade social passou a suscitar teorias que visavam reduzir ou eliminar as diferenças econômicas entre ricos e pobres (BARROS; FOGUEL; ULYSSEA, 2006). Roemer (1998), por exemplo, teorizou a respeito do conceito de igualdade de oportunidades, dizendo que uma situação de desigualdade ocorreria se houvesse condições que interferissem negativamente sobre o acesso de um grupo socioeconômico a determinados serviços ou vantagens.

As desigualdades, em um âmbito mundial, são de grande preocupação para a sociedade, bem como também o são para vários países (NASCIMENTO, 2016). Na América Latina e Caribe, por exemplo, apesar das altas taxas de cobertura em relação ao alcance de serviços como energia elétrica e educação (mais de 90,0% da população), estão presentes grandes defasagens quanto a serviços essenciais, como o acesso ao saneamento básico e à água encanada (GROUP, 2016).

Segundo dados de 2016 do Fórum Econômico Mundial - FEM (2016), a desigualdade social existente na América Latina se assemelha à encontrada na África Subsaariana. Destacando-se o indicador “distância da paridade de gênero”, que mede a diferença média remanescente entre gêneros com base no rendimento ganho, a região africana apresentava

o valor desse índice de 32,0%, representando uma das piores situações no mundo. Já na América do Sul e Caribe, o valor encontrado foi bem próximo (30,0%).

A fim de diminuir problemas como estes, países integrantes da cooperação Sul-Sul - tipo de colaboração entre nações subdesenvolvidas ou em desenvolvimento (Sul Global), criada através da Conferência de Bandung, na Indonésia, em 1955 - têm como desafio primordial investir na luta contra a fome e a extrema pobreza, enfrentadas por alguns de seus países membros, além de lidar com questões como as mudanças climáticas, doenças graves, infraestruturas precárias e conflitos internos (KONE, 2019; NAÇÕES UNIDAS, 2019).

Para Figueiredo e Silva (2012), a desigualdade de acesso a serviços e oportunidades no país não está atrelada somente a fatores econômicos, mas, também, a um *background* familiar, através de variáveis que traçam um perfil sociodemográfico dos indivíduos, tais como gênero, etnia, local de nascimento e educação dos pais, e que afetam as chances de acesso a melhores condições de vida.

Conforme o trabalho de Brito et al. (2017), as desigualdades de oportunidades vieram à tona no Brasil com as históricas diferenças existentes entre localidades. Pode-se destacar o limitado acesso aos serviços fundamentais, tais como, eletricidade, saneamento, recursos hídricos e educação. Em contrapartida, segundo Suliano e Filho (2017), a década de 2000 teve como característica principal a maciça diminuição do grau de desigualdade, levando a uma melhor situação em relação às quatro décadas anteriores e permitindo um cenário de maior isonomia na sociedade civil.

No Brasil, as grandes diferenças de capital entre as famílias são duradouras e implicam em pequena mobilidade educacional (REIS; RAMOS, 2011; FERREIRA; VELOSO, 2015). Para entender a forma como ocorre o acesso ao ensino superior no Brasil deve-se analisar, historicamente, como o seu percurso foi traçado e de que maneira ele se relaciona com as políticas públicas, em que o Estado atua na modificação da forma em que as instituições de ensino superior se organizam (FACHINETTI; CANDIDO; CARNEIRO, 2020).

Um exemplo desse fenômeno pode ser observado no estudo de Carvalho e Waltenberg (2015), em que a taxa de cobertura média, que conforme os pesquisadores, pode ser interpretada como a porcentagem disponível da educação superior para a sociedade no Brasil, cresceu de 0,285 para 0,316 no período de 2003 a 2013, considerando informações da Pesquisa Nacional por Amostra de Domicílios (PNAD). Ainda assim, pelos resultados obtidos no estudo, os indivíduos da etnia branca ou amarela estavam mais presentes no

ambiente universitário do que os que não pertenciam a essa etnia e da mesma forma, aqueles cujo rendimento domiciliar *per capita* era de mais de 1,5 salário mínimo estavam em maior quantidade do que os que tinham renda inferior ou igual a esse ponto de corte (OLIVEIRA et al., 2018).

### 1.1.1 Acesso ao ensino superior no Brasil

Através dos dados da PNAD estudados por Andrade (2012), a parcela da população jovem (com idade entre 18 e 24 anos) que, em 2009, possuía acesso à educação superior, ou seja, estava matriculada em algum curso de graduação em instituição de ensino superior, era de apenas 19,0%. Esse fato foi um dos que motivou a realização de investigações que objetivam evidenciar as razões que afetam as chances de ingresso nesse nível educacional e, da mesma forma, instigou o surgimento de políticas de ação com interesse de incrementar essas mesmas chances e a qualidade do ensino, além de buscar se sobrepor a restrições ocasionadas por disparidades socioeconômicas exibidas por todo o sistema de educação.

Nesta mesma vertente, uma das leis que influenciou bastante a admissão de jovens à educação superior é a Lei n° 12.711/2012 (mais conhecida como “Lei de Cotas”), e regulamentada pelo Decreto n° 7.824/2012 que, dentre outras atribuições, destina metade das vagas de institutos e universidades federais para pessoas que cursaram integralmente o ensino médio público, e o restante das vagas para a ampla concorrência (demais candidatos). A mesma lei também leva em conta a cor ou raça e a renda familiar bruta do candidato que deseja se matricular em um curso de graduação, dividindo as vagas de estudantes cotistas entre oito categorias de acesso (MEC, 2012).

Os Planos Nacionais de Educação (BRASIL, 2001; BRASIL, 2014) estabeleceram metas e diretrizes em relação à educação brasileira para os decênios de 2001-2010 e 2014-2024, respectivamente, afirmando que o país não pode ser considerado desenvolvido se não possuir um sistema universitário bem estruturado. Dessa forma, torna-se essencial a oferta de um ensino de qualidade e de políticas encarregadas de reduzir as desigualdades no ingresso à graduação (NOGUEIRA, 2018).

Para que sejam tiradas conclusões acerca da situação em que o país se encontra em relação à desigualdade social, é possível sintetizar em um indicador as medições de como as oportunidades básicas estão distribuídas. Nesse sentido, primeiramente, pode-se utilizar dados de pesquisas domiciliares e, posteriormente, estimar um índice a fim de, por exemplo, traçar um panorama de como a sociedade se encontra distante de um acesso mais abrangente a determinados bens e serviços (BARROS et al., 2009; DILL; GONÇALVES,

2011).

O Índice de Oportunidades Humanas (IOH), em especial, é um indicador capaz de sintetizar como os fatores que não são relacionados aos esforços pessoais, e sim às circunstâncias que se encontram fora da capacidade de alteração por parte dos indivíduos (como cor ou raça, sexo, renda da família, área de residência e grau de escolaridade dos pais), interferem no acesso a oportunidades, tais como bens e serviços básicos. Esse índice pode apresentar valores entre 0 e 1, em uma escala que vai de uma situação de completa desigualdade ( $IOH = 0$ ) até um cenário de acesso universal a oportunidade estudada ( $IOH = 1$ ) (BARROS et al., 2009; DILL; GONÇALVES, 2011; SULIANO; MIRO, 2014).

Serão abordados, resumidamente, alguns estudos apontados na literatura que mencionam o uso do IOH em diferentes temas (como a avaliação da desigualdade social quanto ao acesso a bens e serviços básicos e, em especial, à educação no nível superior) e localidades brasileiras, ressaltando a vasta gama de possibilidades de sua aplicação. Os três primeiros levam em consideração as taxas de cobertura relacionadas à infraestrutura doméstica, tais como energia elétrica, saneamento básico e recursos hídricos. Já os demais são voltados diretamente à investigação das oportunidades de acesso ao ensino superior, em que as variáveis consideradas para as análises estavam condicionadas a características sociodemográficas dos indivíduos, como gênero, cor/raça, estado civil, escolaridade, rendimento familiar e situação do domicílio (urbano/rural).

No estudo com abrangência nacional que levou em conta os dados da PNAD, de Dill e Gonçalves (2011), verificou-se que o IOH geral, calculado considerando a média do índice para as oportunidades de acesso ao saneamento básico, água, energia elétrica e educação, aumentou de 0,633 para 0,729, aproximadamente, no período entre 1999 e 2009. Assim, houve melhora nas condições de acesso da sociedade aos componentes analisados, fazendo o Brasil se aproximar mais de um contexto de igualdade ( $IOH = 1$ ) na cobertura destes bens e serviços. No mesmo trabalho, percebeu-se que, em 2009, o acesso à eletricidade teve um dos melhores índices, com IOH de 0,971, bem próximo ao ideal, indicando um cenário de ampla cobertura a esse bem. Por outro lado, o saneamento apresentou o pior valor do IOH, sendo esse igual a 0,393, ou seja, o acesso a esse serviço era o mais desigual dentre os pesquisados.

Suliano e Miro (2014), com base nos dados do estado do Ceará, compararam as informações das PNADs de 2001 e de 2011, e chegaram à conclusão que houve melhora nos resultados do IOH ao se investigar dados referentes a condições habitacionais, onde o

índice de oportunidades humanas para o saneamento básico passou de 0,226 para 0,356, e o IOH do abastecimento de água, de 0,468 para 0,685. Desse modo, o acesso aos recursos hídricos ficou mais perto de ser considerado universal no Ceará, mas, para o saneamento, apesar do incremento no valor do IOH, o mesmo cenário desejado ainda parecia estar mais distante.

Gonçalves et al. (2015) pesquisaram acerca das diferenças de acesso às seguintes oportunidades/serviços: eletricidade, saneamento, água e escola no Distrito Federal (DF). Os dados foram extraídos da Pesquisa Distrital por Amostra de Domicílios (PDAD) de 2013, a partir da qual foi calculado o IOH através de um modelo de regressão logística. Foi observado que a oportunidade com melhores resultados foi aquela relacionada ao acesso à energia elétrica ( $IOH = 0,990$ ), indicando uma situação de quase universalização de acesso a esse serviço. Para todas as oportunidades estudadas, exceto o saneamento básico ( $IOH = 0,751$ ), o índice calculado era superior a 0,900, até mesmo quando considerado o IOH global, evidenciando um cenário favorável para a região. Também se detectou que a região administrativa - são 31 diferentes no DF - associada ao domicílio dos respondentes influenciava significativamente nas chances de acesso aos bens e serviços estudados.

A pesquisa de Carvalho e Waltenberg (2015) investigou a situação do Brasil nos anos de 2003 e 2013, com o objetivo de avaliar o índice de oportunidades humanas com enfoque no acesso ao ensino superior, utilizando os microdados das PNADs de 2003 e 2013. Para este último ano, o IOH encontrado para universidades públicas era de 0,102 e para as particulares era de 0,230. Ao se considerar ambos os tipos de instituições, o IOH foi igual a 0,281. De forma semelhante, Ferreira et al. (2015) também voltaram seu estudo para a desigualdade no acesso às instituições públicas e privadas de ensino superior no Brasil, com base nas informações da PNAD 2013. O IOH no acesso a universidades públicas e privadas foi, respectivamente, igual a 0,123 e 0,246 (valores maiores que aqueles encontrados em Carvalho e Waltenberg (2015), citado anteriormente). Vale ressaltar que o IOH é um índice estimado e, sendo assim, a metodologia utilizada para o seu cálculo pode variar entre trabalhos. Os resultados apresentados pelos dois estudos identificam que ter acesso à educação superior era um fenômeno que não acontecia de forma universal no ano de 2013, já que os valores de IOH eram bem distantes do valor ideal, 1.

Oliveira et al. (2018) tomaram por base o estado do Mato Grosso do Sul (MS) e embasaram seu trabalho nas informações fornecidas pelas PNADs de 2005 e de 2015. Fundamentando-se nelas, realizaram uma análise através de um modelo de regressão logística para que fosse construído o IOH voltado a avaliar o acesso ao ensino supe-

rior no referido estado. Dentre as principais constatações, identificou-se a variação no índice de 0,306 para 0,374 no período de observação, evidenciando uma melhora nos resultados quanto ao acesso a esse nível de educação no MS, principalmente devido à maior disponibilidade de vagas e sua forma de distribuição.

Então, neste contexto faz-se relevante avaliar como evoluiu (até 2019) o acesso dos estudantes ao ensino superior. A partir disso, os objetivos desse trabalho são tais como os descritos na seção 1.2.

## **1.2 Objetivos**

### **1.2.1 Objetivo Geral**

O objetivo geral deste estudo é avaliar a desigualdade no acesso à educação superior no Brasil, em 2019, também investigando, separadamente, o ingresso em universidades públicas e privadas.

### **1.2.2 Objetivos Específicos**

Entre os objetivos específicos deste trabalho, estão:

- Verificar o perfil de indivíduos que potencialmente poderiam ingressar na educação superior;
- Fazer uma análise exploratória sobre as variáveis estudadas de forma uni e bivariada;
- Construir modelos de regressão logística para relacionar a chance de iniciar uma graduação com fatores sociodemográficos indicados pela revisão bibliográfica;
- Agrupar os indivíduos de acordo com as categorias das variáveis sociodemográficas consideradas pelos modelos;
- Calcular a probabilidade esperada de cada grupo de indivíduos ingressar em uma instituição de ensino superior (IES), desconsiderando o tipo de universidade ou avaliando o acesso às IES públicas ou privadas, de forma separada;
- Estimar os Índices de dissimilaridade (D) e de Oportunidades Humanas (IOH) para o Brasil em 2019 para o ensino superior.

## 1.3 Organização

O presente trabalho está organizado conforme descrito nesta seção. O capítulo 2 apresentará a descrição da base de dados e a metodologia empregada. No capítulo 3, serão apresentados e comentados os resultados obtidos. Por fim, no capítulo 4, serão expostas conclusões obtidas a partir dos resultados encontrados.

## 2 Materiais e Métodos

Este capítulo expõe detalhadamente, na seção 2.1, as informações relativas ao banco de dados utilizado, oriundo da Pesquisa Nacional por Amostra de Domicílios Contínua (PNADc) e define a Demanda Potencial ao Ensino Superior (subseção 2.1.1). Além disso, aborda questões metodológicas nas seções que tratam das Estimativas Pontuais (2.2), dos Modelos Lineares Generalizados (2.3) e do Índice de Oportunidades Humanas (2.4).

### 2.1 Pesquisa Nacional por Amostra de Domicílios Contínua

A criação da Pesquisa Nacional por Amostra de Domicílios (PNAD) pelo Instituto Brasileiro de Geografia e Estatística (IBGE), em 1967, foi o marco para a implantação do sistema de pesquisas domiciliares no país e tinha como objetivo produzir dados que tornassem possível avaliar o desenvolvimento socioeconômico do país (NERI, 2012).

No primeiro trimestre de 2012, o IBGE uniu as pesquisas PNAD e a Pesquisa Mensal de Emprego (PME), dando origem à Pesquisa Nacional por Amostra de Domicílios Contínua (PNADc) (IBGE, 2021a). A PNADc considera uma amostra que contempla domicílios investigados por 5 trimestres consecutivos, em visitas que acontecem uma vez a cada trimestre (IBGE, 2020).

Ainda segundo as notas técnicas do IBGE (2021a), a PNADc adota um plano de amostragem por conglomerado em dois estágios, de modo que o primeiro deles é estratificar as Unidades Primárias de Amostragem (UPAs), formadas por setores censitários ou grupo de setores de ao menos 60 domicílios particulares permanentes, e selecioná-las para compor a amostra mestra trimestral. Para o segundo estágio, 14 domicílios ocupados e cadastrados no Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE) são selecionados dentro de cada UPA, através de amostragem aleatória simples.

O objetivo fundamental da PNADc é verificar como a sociedade brasileira evoluiu

em relação a diferentes dimensões: trabalho, educação, demografia, habitação, economia, entre outras temáticas (ALMEIDA et al., 2018). Conforme a definição dada pelo IBGE (2021b), a pesquisa tem suas informações divulgadas de forma mensal, trimestral, anual ou variável, dependendo dos indicadores estudados, e possui como unidade de investigação o domicílio.

Os dados apresentados neste trabalho são relacionados a informações amostrais que totalizam 433.535 observações de 363 variáveis, baseados na PNADc de 2019, relativa ao ano que antecedeu o início da pandemia de COVID-19 no mundo, e tomam como referência a 5<sup>a</sup> visita aos domicílios, já que somente por meio dela foi possível ter acesso à variável “rendimento domiciliar *per capita*”, necessária para o trabalho, de forma a considerar os temas “Outras formas de trabalho (afazeres domésticos, cuidados de pessoas, produção para o próprio consumo e trabalho voluntário)” e “Trabalho de crianças e adolescentes” (IBGE, 2021b).

### 2.1.1 Demanda Potencial ao Ensino Superior

A base de dados estudada contém apenas indivíduos que estariam aptos a ingressar no ensino superior, grupo denominado “Demanda Potencial ao Ensino Superior”. As características comuns a esses indivíduos são:

- situação domiciliar ligada a áreas urbanas;
- idade entre 17 e 29 anos, pois esse intervalo abrange a faixa etária que o Ministério da Educação (MEC) considera apropriada para um indivíduo estar cursando alguma graduação (de 18 a 24 anos), levando em consideração os fenômenos de atraso nos níveis educacionais e o alto contingente de pessoas fora do ambiente universitário;
- ensino médio completo como mais alto grau de instrução, com período de tempo de estudo compreendido entre 11 e 14 anos;
- condição no domicílio igual à categoria “filhos”.

Uma vez feita a filtragem dos indivíduos com esse perfil, foram analisados os indivíduos que de fato acessaram o ensino superior, os comparando com os que não acessaram. Portanto, com a finalidade de avaliar a probabilidade de acesso à educação superior para o grupo da demanda potencial ao ensino superior, seguindo Carvalho e Waltenberg (2015),

Tabela 1: Descrição das variáveis

Variável	Descrições
Matrícula em algum curso de graduação	0 = Não   1 = Sim
Sexo	0 = Masculino   1 = Feminino
Etnia	0 = Preta, Parda ou Indígena   1 = Branca ou Amarela
Região de Residência	0 = Sudeste   1 = Norte   2 = Nordeste   3 = Sul   4 = Centro-Oeste
Grau de instrução mais elevado alcançado pelo chefe do domicílio	0 = Até Ensino Fundamental   1 = Ensino Médio   2 = Ensino Superior
Rendimento domiciliar <i>per capita</i>	0 = Até 1,5 salário mínimo (s.m.) inclusive   1 = Mais de 1,5 salário mínimo (s.m.)

Fonte: Microdados da PNADc 2019.

foram consideradas as variáveis qualitativas listadas na Tabela 1, cujas categorias de referência estão simbolizadas pelo valor “0”.

A variável etnia foi classificada dessa forma tendo em vista as categorias de cor/raça as quais podem ser destinadas cotas de acesso ao ensino superior, sendo que as vagas de estudantes cotistas podem ser de pessoas da etnia preta, parda ou indígena que cursaram integralmente o ensino médio em escolas públicas, mas não da etnia branca ou amarela (BRASIL, 2012).

Já para a classificação da variável “Rendimento domiciliar *per capita*”, considerou-se o Decreto n° 9.661/2019, que reajustou o salário mínimo brasileiro para o valor de R\$ 998,00 (RF, 2019). A faixa de 1,5 salário mínimo (equivalente a R\$ 1.497,00 em 2019), categoria de referência para esta variável, foi escolhida porque é critério para se fazer valer os benefícios das ações afirmativas nas instituições de ensino superior públicas estaduais e federais e ainda pelo programa Universidade para Todos (ProUni) em faculdades privadas (CARVALHO; WALTENBERG, 2015).

## 2.2 Estimação pontual

Pelas definições de Bolfarine e Bussab (2005), define-se que **população** é um conjunto de elementos com uma ou mais características de interesse. Já **amostra** é um subconjunto da população com elementos selecionados a fim de estimar os parâmetros de interesse.

A respeito de dados populacionais, considerando  $Y$  característica de interesse da população  $U$  a ser pesquisada, definem-se o total e a média populacional, respectivamente,

como:

$$\tau = \sum_{i=1}^N Y_i, \quad (2.1)$$

e

$$\mu = \bar{Y} = \frac{\tau}{N} = \frac{1}{N} \sum_{i=1}^N Y_i, \quad (2.2)$$

onde  $\{Y_1, \dots, Y_N\}$  é o vetor populacional, cujos componentes denotam o valor da variável  $Y$  para o  $i$ -ésimo elemento populacional,  $i = 1, \dots, N$ , e  $N$  é a quantidade de elementos que compõem a população.

Já o total amostral (considerando pesos amostrais), estimador não viciado do total populacional  $\tau$ , é dado pela seguinte fórmula:

$$\hat{\tau} = \sum_{i=1}^n W_i Y_i = \sum_{i=1}^n \frac{Y_i}{\pi_i} = \sum_{i=1}^n \frac{N}{n} Y_i = N \frac{1}{n} \sum_{i=1}^n Y_i, \quad (2.3)$$

onde  $\{Y_1, \dots, Y_n\}$  é o vetor amostral, cujos componentes representam o valor da variável  $Y$  para o  $i$ -ésimo elemento amostral,  $i = 1, \dots, n$ ,  $n$  é a quantidade de elementos que compõem a amostra  $s \subset U$  ( $n < N$ ), e  $W_i$  e  $\pi_i = \frac{1}{W_i} = \frac{n}{N}$  são, respectivamente, o peso e a probabilidade de inclusão da  $i$ -ésima unidade amostral.

A expressão usual da média amostral está expressa na equação (2.4).

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n Y_i. \quad (2.4)$$

A partir das equações (2.3) e (2.4), é possível definir uma outra expressão do estimador não viciado para a média populacional, a média amostral, considerando os pesos amostrais, com desenvolvimento mostrado a seguir.

$$\hat{\tau} = N \frac{1}{n} \sum_{i=1}^n Y_i = N \hat{\mu} \Rightarrow \hat{\mu} = \frac{\hat{\tau}}{N} = \frac{1}{N} \sum_{i=1}^n W_i Y_i. \quad (2.5)$$

Para o caso deste estudo, a média a ser estimada é o valor esperado de uma distribuição Bernoulli, onde a média é igual a  $p$  (proporção ou percentual).

## 2.3 Modelos Lineares Generalizados

Em certas ocasiões, como no presente trabalho, o interesse está em estudar a relação entre um conjunto de variáveis explicativas e uma variável resposta binária, isto é, que pode apresentar como resultado os eventos “sucesso” ou “fracasso”. Neste caso, não é viável considerar que sua distribuição seja do tipo normal e, sim, que  $Y_i \sim \text{Bernoulli}(\pi_i)$ , onde  $\pi_i$  representa a probabilidade de sucesso (ou de  $Y_i = 1$ ). (DOBSON; BARNETT, 2018). Em consequência disso, de acordo com Paula (2004), para tratar de casos em que a distribuição normal não é apropriada para ser modelada, foi proposto, na década de 1970, o modelo linear generalizado, descrito abaixo.

Esta seção utilizará definições do livro de Dobson e Barnett (2018) e apresentará a família exponencial de distribuições e as características dos modelos lineares generalizados e, em especial, o modelo de regressão logística.

### 2.3.1 Família Exponencial de distribuições

Nesta subseção, serão definidos a família exponencial de distribuições e os modelos lineares generalizados.

**Definição 2.1** *Uma variável aleatória  $Y$  com função de densidade (caso contínuo) ou de probabilidade (caso discreto)  $f(y; \theta)$  que depende somente de um parâmetro  $\theta$  pertencerá à família exponencial de distribuições se:*

$$f(y|\theta) = \exp \{a(y)b(\theta) + c(\theta) + d(y)\} , \quad (2.6)$$

em que as funções  $a$ ,  $b$ ,  $c$  e  $d$  são reais e conhecidas.

Pode-se mostrar que algumas distribuições como a de Poisson, Normal, Binomial e Binomial Negativa podem ter suas funções de densidade ou de probabilidade descritas através de (2.6).

Em particular, se  $a(y) = y$ , pode-se afirmar que a equação (2.6) está na sua forma canônica e  $b(\theta)$  pode ser denominado parâmetro natural da distribuição.

Nesse sentido, os Modelos Lineares Generalizados podem ser definidos como:

**Definição 2.2** *Seja  $Y_1, \dots, Y_N$  um conjunto de  $N$  variáveis aleatórias com média  $E(Y_i) = \mu_i$  ( $i = 1, \dots, N$ ), mesma distribuição pertencente à família exponencial e que dependem*

de somente um parâmetro, não necessariamente igual. Seja  $\mathbf{X}_i^T = (1, X_{1,i}, \dots, X_{p,i})$  matriz com dimensão  $(p+1) \times (p+1)$ , onde a primeira coluna é completamente formada pela constante 1 e nas demais estão presentes os valores observados de todas as covariáveis;  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$  um conjunto de parâmetros e  $g$  uma função monótona e diferenciável. Os Modelos Lineares Generalizados (MLG) serão aqueles que associam a cada variável  $Y_i, i = 1, \dots, N$ , (componente aleatório) o seu correspondente preditor linear  $\eta_i = \mathbf{X}_i^T \boldsymbol{\beta}$  (componente sistemático), através da função de ligação  $g$ , de forma que:

$$g(\mu_i) = \eta_i = \mathbf{X}_i^T \boldsymbol{\beta}, \quad (2.7)$$

onde  $\mu_i = E(Y_i)$ , para  $i = 1, \dots, N$ .

### 2.3.2 Regressão Logística para dados binários

Serão apresentados nesta subseção os detalhes acerca do modelo de regressão logística, com a caracterização dos eventos binários, a função de ligação utilizada, a definição de razão de chances e informações sobre a seleção do modelo quanto à sua adequabilidade.

Geralmente, quando o interesse é modelar uma variável dicotômica, as duas categorias que essa variável pode assumir são designadas de “sucesso” (valor 1) e “fracasso” (valor 0). No contexto do presente trabalho, a variável mais relevante pode ser definida da seguinte forma:

$$Y_i = \begin{cases} 1, & \text{se o indivíduo } i \text{ está matriculado em algum curso de graduação} \\ 0, & \text{se o indivíduo } i \text{ não está matriculado em algum curso de graduação} \end{cases}. \quad (2.8)$$

Sendo  $\pi_i$  a probabilidade da pessoa estar matriculada em um curso de graduação, as respectivas probabilidades de sucesso e de fracasso são dadas por  $P(Y_i = 1) = \pi_i$  e  $P(Y_i = 0) = 1 - \pi_i$ . Deste modo, pode-se dizer que  $Y_i$  tem distribuição Bernoulli, com o parâmetro  $\pi_i$  e sua função de distribuição de probabilidade é dada por:

$$\begin{aligned} P(Y_i = y_i | \pi_i) &= \pi_i^{y_i} (1 - \pi_i)^{(1-y_i)} \\ &= \exp [y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)] \\ &= \exp \left[ y_i \ln \left( \frac{\pi_i}{1 - \pi_i} \right) + \ln(1 - \pi_i) \right], \end{aligned} \quad (2.9)$$

para  $y_i = 0, 1$ .

Ao realizar uma comparação entre as equações (2.6) e (2.9), observa-se que a distribuição de Bernoulli pertence à família exponencial na sua forma canônica, pois  $a(y_i) = y_i$ .

Considerando  $N$  variáveis aleatórias  $Y_1, \dots, Y_N$  independentes, cada uma com probabilidade de sucesso, respectivamente, igual a  $\pi_i, i = 1, \dots, N$ , utilizando a equação (2.9), sua função de verossimilhança será dada por:

$$\begin{aligned} \prod_{i=1}^N P(Y_i = y_i | \pi_i) &= \prod_{i=1}^N \exp \left[ y_i \ln \left( \frac{\pi_i}{1 - \pi_i} \right) + \ln(1 - \pi_i) \right] \\ &= \exp \left[ \sum_{i=1}^N y_i \ln \left( \frac{\pi_i}{1 - \pi_i} \right) + \sum_{i=1}^N \ln(1 - \pi_i) \right], \end{aligned} \quad (2.10)$$

que, por sua vez, também pertence à família exponencial.

Levando em conta o cenário de uma amostra aleatória de  $N$  variáveis  $Y_i \sim \text{Bernoulli}(\pi_i)$ , usando a equação (2.10), pode-se mostrar que o logaritmo da função de verossimilhança será:

$$l(y_1, \dots, y_N | \pi_1, \dots, \pi_N) = \sum_{i=1}^N \left[ y_i \ln \left( \frac{\pi_i}{1 - \pi_i} \right) + \ln(1 - \pi_i) \right]. \quad (2.11)$$

Como  $Y_i \sim \text{Bernoulli}(\pi_i)$ , então  $E(Y_i) = \pi_i$ . Se for do interesse modelar a chance da ocorrência do fenômeno, pode-se desenvolver MLG que relacionem o valor esperado de  $Y_i$  às covariáveis  $\mathbf{X}_i^T$ , através da equação (2.7), fazendo

$$g(\pi_i) = \eta_i = \mathbf{X}_i^T \boldsymbol{\beta}, \quad (2.12)$$

onde  $\eta_i$  é denominado preditor linear associado ao  $i$ -ésimo indivíduo e  $g$  é a função de ligação escolhida.

A função de ligação deve ser selecionada de maneira que os valores ajustados das probabilidades ( $\hat{\pi}_i$ ) estejam contidos no intervalo  $[0,1]$ . Usualmente, o modelo faz uso de uma função de distribuição acumulada, que é dada por:

$$\pi_i = g^{-1}(\mathbf{X}_i^T \boldsymbol{\beta}) = \int_{-\infty}^s f(t) dt, \quad (2.13)$$

onde  $f(t)$  é a função de densidade de probabilidade denominada distribuição de tolerância, em que  $f(t) \geq 0$  e  $\int_{-\infty}^{\infty} f(t) dt = 1$ .

No modelo de regressão logística, admite-se que a distribuição de tolerância  $f(t)$ , a função de distribuição acumulada  $F(\eta)$  e a função de ligação logit são dadas, respectivamente, por:

$$f(t) = \frac{\exp(t)}{[1 + \exp(t)]^2}, \quad (2.14)$$

$$F(\eta) = \pi = \int_{-\infty}^{\eta} f(t)dt = \frac{\exp(\eta)}{1 + \exp(\eta)} \quad (2.15)$$

e

$$g(\pi_i) = F^{-1}(\pi_i) = \eta_i = \mathbf{X}_i^T \boldsymbol{\beta} = \text{logit}(\pi_i) = \ln \left( \frac{\pi_i}{1 - \pi_i} \right) . \quad (2.16)$$

As equações (2.7) e (2.16) podem ser relacionadas de modo que:

$$g(\pi_i) = \ln \left( \frac{\pi_i}{1 - \pi_i} \right) = \mathbf{X}_i^T \boldsymbol{\beta} . \quad (2.17)$$

Assim, a probabilidade  $\pi_i$  do  $i$ -ésimo indivíduo frequentar um curso de graduação será dada pela equação:

$$\pi_i = \frac{\exp(\mathbf{X}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i^T \boldsymbol{\beta})} . \quad (2.18)$$

### 2.3.2.1 Estimação dos parâmetros

O método usualmente utilizado para estimar o vetor de parâmetros  $\boldsymbol{\beta}$  é o Método do Escore, que maximiza o logaritmo da função de verossimilhança, anteriormente descrita na equação (2.11), porém, agora definida em função do vetor de parâmetros  $\boldsymbol{\beta}$ :

$$l(y_1, \dots, y_N | \boldsymbol{\beta}) = \sum_{i=1}^N \left[ y_i (\mathbf{X}_i^T \boldsymbol{\beta}) + \ln \left( 1 - \frac{\exp(\mathbf{X}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i^T \boldsymbol{\beta})} \right) \right] . \quad (2.19)$$

O procedimento de máxima verossimilhança consiste em maximizar a função descrita na equação (2.19) em relação ao conjunto  $\boldsymbol{\beta}$ . Para maximização do logaritmo da verossimilhança, iguala-se a zero a derivada dessa função em relação ao vetor de parâmetros  $\boldsymbol{\beta}$ , do seguinte modo:

$$\frac{\partial l(y_1, \dots, y_N | \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^N \mathbf{u}_i(\boldsymbol{\beta}) = 0 , \quad (2.20)$$

onde  $\mathbf{u}_i(\boldsymbol{\beta})$  é a função escore de  $\boldsymbol{\beta}$  e é dada por:

$$\mathbf{u}_i(\boldsymbol{\beta}) = \frac{\partial \ln[P(Y_i = y_i | \boldsymbol{\beta})]}{\partial \boldsymbol{\beta}} . \quad (2.21)$$

A equação (2.20) não tem forma analítica fechada, sendo necessário recorrer a métodos iterativos para poder estimar o vetor de parâmetros  $\boldsymbol{\beta}$ . Neste caso, serão utilizados os métodos de Newton-Raphson para chegar nas estimativas desses parâmetros. Mais

detalhes podem ser vistos em Dobson e Barnett (2018).

A partir da equação (2.18), as probabilidades ajustadas agora podem ser determinadas por:

$$\hat{\pi}_i = \frac{\exp(\mathbf{X}_i^T \hat{\boldsymbol{\beta}})}{1 + \exp(\mathbf{X}_i^T \hat{\boldsymbol{\beta}})} = \frac{\exp(\hat{\eta}_i)}{1 + \exp(\hat{\eta}_i)}, \quad (2.22)$$

sendo que

$$\hat{\eta}_i = \mathbf{X}_i^T \hat{\boldsymbol{\beta}}. \quad (2.23)$$

### 2.3.2.2 Inclusão do plano amostral na estimação dos parâmetros

A PNADc é uma pesquisa domiciliar de maior abrangência geográfica e faz uso de uma amostragem complexa. Para captar as informações do plano amostral de forma adequada, deve-se considerar o Método da Máxima Pseudo-Verossimilhança (MPV) para gerar estimativas pontuais dos parâmetros e de seus respectivos erros padrões de forma consistente (PESSOA; SILVA, 1998; SOUZA, 2019).

Dando início à aplicação do MPV, define-se  $T(\boldsymbol{\beta})$  como a soma dos escores na população  $U$ , um vetor de totais populacionais; tendo equação

$$T(\boldsymbol{\beta}) = \sum_{i \in U} \mathbf{u}_i(\boldsymbol{\beta}), \quad (2.24)$$

com solução dada por  $\boldsymbol{\beta}_U$ , estimador de Máxima Verossimilhança para  $\boldsymbol{\beta}$ .

Um estimador linear ponderado para  $T(\boldsymbol{\beta})$  é definido como:

$$\hat{T}(\boldsymbol{\beta}) = \sum_{i \in s} W_i \mathbf{u}_i(\boldsymbol{\beta}), \quad (2.25)$$

onde  $s$  é a amostra pertencente à população  $U$  ( $s \subset U$ ) e  $W_i$  é o peso da  $i$ -ésima unidade amostral.

Binder (1983) define que o estimador de máxima pseudo-verossimilhança para  $T(\boldsymbol{\beta})$  pode ser encontrado ao igualar a equação (2.25) a 0 em que:

$$\hat{T}(\boldsymbol{\beta}_{MPV}) = \sum_{i \in s} W_i \mathbf{u}_i(\boldsymbol{\beta}) = 0, \quad (2.26)$$

sendo que, como o tamanho da população em pesquisas do tipo por amostragem é geralmente grande,  $\hat{\boldsymbol{\beta}}_{MPV}$  funciona como estimador adequado tanto para  $\boldsymbol{\beta}_U$  como para

$\beta$ .

Por meio da técnica da Linearização de Taylor, obtém-se as fórmulas correspondentes à variância do estimador  $\hat{\beta}_{MPV}$ ,  $V(\hat{\beta}_{MPV})$ , e seu respectivo estimador,  $\hat{V}(\hat{\beta}_{MPV})$ :

$$V(\hat{\beta}_{MPV}) \simeq [J(\beta_U)]^{-1} V \left[ \sum_{i \in s} W_i \mathbf{u}_i(\beta_U) \right] [J(\beta_U)]^{-1}, \quad (2.27)$$

onde

$$J(\beta_U) = \left. \frac{\partial T(\beta)}{\partial \beta} \right|_{\beta = \beta_U} = \sum_{i \in U} \left. \frac{\partial \mathbf{u}_i(\beta)}{\partial \beta} \right|_{\beta = \beta_U}$$

e

$$\hat{V}(\hat{\beta}_{MPV}) = [\hat{J}(\hat{\beta}_{MPV})]^{-1} \hat{V} \left[ \sum_{i \in s} W_i \mathbf{u}_i(\hat{\beta}_{MPV}) \right] [\hat{J}(\hat{\beta}_{MPV})]^{-1}, \quad (2.28)$$

onde

$$\hat{J}(\hat{\beta}_{MPV}) = \left. \frac{\partial \hat{T}(\beta)}{\partial \beta} \right|_{\beta = \hat{\beta}_{MPV}} = \sum_{i \in s} W_i \left. \frac{\partial \mathbf{u}_i(\beta)}{\partial \beta} \right|_{\beta = \hat{\beta}_{MPV}}.$$

Neste caso,  $V \left[ \sum_{i \in s} W_i \mathbf{u}_i(\beta_U) \right]$  é a matriz de variâncias e covariâncias do desenho do estimador do total populacional dos escores  $\hat{T}(\beta)$  e  $\hat{V} \left[ \sum_{i \in s} W_i \mathbf{u}_i(\hat{\beta}_{MPV}) \right]$  é estimador consistente desta matriz.

Para amostras grandes, é possível realizar inferências a respeito de  $\beta$  a partir da estatística  $\hat{\beta}_{MPV}$ , que possui distribuição assintótica Normal Multivariada (NM) (BINDER, 1983; PESSOA; SILVA, 1998). Assim sendo, é possível mostrar que:

$$[\hat{V}(\hat{\beta}_{MPV})]^{-1/2} (\hat{\beta}_{MPV} - \beta_U) \sim \mathbf{NM}(\mathbf{0}; \mathbf{I}). \quad (2.29)$$

A fim de averiguar a significância do coeficiente  $\beta_j$ , que é relativo ao efeito da  $j$ -ésima covariável ( $j = 1, \dots, p$ ) sobre a variável resposta, pode ser realizado o seguinte teste de hipóteses:

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases}.$$

Este teste de significância pode ser feito utilizando a estatística de Wald ou também levando em conta que o estimador  $\hat{\beta}_{MPV}$  tem distribuição assintótica Normal Multivariada, como evidenciado em (2.9). Outra maneira de execução, que foi a considerada neste

trabalho, seria avaliar diretamente o intervalo de confiança da razão de chances (conceito definido na subsubseção 2.3.2.3), verificando se contém o valor 1 ou não.

### 2.3.2.3 Razão de chances

A razão de chances permite a interpretação dos coeficientes do modelo de regressão logística como medidas de associação entre a ocorrência de um evento de interesse com uma dada variável explicativa.

**Definição 2.3** *A razão das probabilidades de ocorrência e de não ocorrência de um evento  $A$  caracteriza a chance (odds) do evento em questão acontecer.*

$$odds(A) = \frac{P(A)}{1 - P(A)}. \quad (2.30)$$

**Definição 2.4** *Ao serem confrontados dois grupos distintos, a razão de chances ou odds ratio (OR) de ocorrência de um dado evento é a razão entre a chance de ocorrência do evento para o 1º grupo ( $odds_1$ ) e a chance de ocorrência para o 2º grupo ( $odds_2$ ).*

$$OR = \frac{odds_1}{odds_2}. \quad (2.31)$$

Sendo o modelo de regressão logística com  $p$  variáveis preditoras e função de ligação logit definida tal como a equação (2.17),  $x_j$  um valor observado de cada uma das variáveis igual para todos os indivíduos, condicionalmente à presença das demais no modelo,  $j = 1, \dots, p$ , e  $\pi(x_j)$  probabilidade de sucesso, descrita pela equação (2.18), a chance de ocorrência do evento “sucesso”  $\{Y = 1\}$  será igual a:

$$odds(x_j) = \frac{\pi(x_j)}{1 - \pi(x_j)}. \quad (2.32)$$

Supondo que as  $p$  variáveis explicativas são quantitativas, considera-se uma variável  $X_j$  ( $j = 1, \dots, p$ ) e são mantidos constantes os valores observados das demais ( $p - 1$ ) variáveis preditivas. Os pontos  $(x_j + 1)$  e  $x_j$  serão observações da variável  $X_j$  para o qual  $(x_j + 1)$  significa o aumento de 1 unidade no valor de  $x_j$ . No caso qualitativo, em que as  $p$  variáveis são categóricas, se  $X_j$  for uma variável do tipo binária, os níveis  $(x_j + 1)$  e  $x_j$  indicarão a presença e a ausência do fator de investigação, nessa ordem. Para ambos os casos, a razão de chances (OR) para ocorrência do evento “sucesso” poderá ser analisada ao se considerar o preditor linear somente com a variável  $X_j$  (MONTGOMERY; PECK; VINING, 2021). Portanto, sendo o preditor linear ajustado  $\hat{\eta}(x_j)$  uma função associada a

um valor fixado da  $j$ -ésima variável explicativa ( $x_j$ ) para todos os indivíduos, caso especial da equação (2.23), incluindo apenas o intercepto e a covariável em questão, os preditores estimados para os pontos  $x_j$  e  $(x_j + 1)$  serão, respectivamente:

$$\hat{\eta}(x_j) = \hat{\beta}_0 + \hat{\beta}_j x_j \quad (2.33)$$

e

$$\hat{\eta}(x_{j+1}) = \hat{\beta}_0 + \hat{\beta}_j (x_j + 1) . \quad (2.34)$$

Relacionando as equações (2.33) e (2.34):

$$\hat{\eta}(x_{j+1}) - \hat{\eta}(x_j) = \hat{\beta}_j . \quad (2.35)$$

Dessa forma, pela equação (2.16), obtém-se:

$$\begin{aligned} \hat{\beta}_j &= \ln \left( \frac{\hat{\pi}(x_j + 1)}{1 - \hat{\pi}(x_j + 1)} \right) - \ln \left( \frac{\hat{\pi}(x_j)}{1 - \hat{\pi}(x_j)} \right) \\ &= \ln(\hat{odds}(x_j + 1)) - \ln(\hat{odds}(x_j)) \\ &= \ln \left( \frac{\hat{odds}(x_j + 1)}{\hat{odds}(x_j)} \right) \\ &= \ln(\hat{OR}_j) . \end{aligned} \quad (2.36)$$

Aplicando a função exponencial, chega-se a:

$$\hat{OR}_j = \exp(\hat{\beta}_j) ; \quad (2.37)$$

assim, a razão de chances  $OR_j$  entre os níveis  $(x_j + 1)$  e  $x_j$  poderá ser estimada por  $\exp(\hat{\beta}_j)$ .

A fim de encontrar um intervalo de confiança (IC) para a razão de chances, parte-se da estatística  $\hat{\beta}_{MPV}$  e sua variância estimada,  $\hat{V}(\hat{\beta}_{MPV})$ .

Pode-se demonstrar que o intervalo de  $100(1 - \alpha)\%$  de confiança para  $\beta_{MPV}$  será dado por:

$$\left[ \hat{\beta}_{MPV} - z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\beta}_{MPV})}; \hat{\beta}_{MPV} + z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\beta}_{MPV})} \right] = [L_{Inf}; L_{Sup}] , \quad (2.38)$$

onde  $z_{1-\alpha/2}$  é o percentil obtido da tabela da distribuição normal padrão, o qual garante que  $P[-z_{1-\alpha/2} < Z < z_{1-\alpha/2}] = 1 - \alpha$  e  $\sqrt{\hat{V}(\hat{\beta}_{MPV})}$  é o erro padrão (EP) estimado de  $\hat{\beta}_{MPV}$ .

Aplicando a função exponencial nos limites inferior ( $L_{Inf}$ ) e superior ( $L_{Sup}$ ) do intervalo de confiança dado pela equação (2.38), é possível encontrar a expressão do IC de  $100(1 - \alpha)\%$  para a razão de chances ( $OR$ ):

$$\left[ \exp\left(\hat{\beta}_{MPV} - z_{1-\alpha/2}\sqrt{\hat{V}(\hat{\beta}_{MPV})}\right); \exp\left(\hat{\beta}_{MPV} + z_{1-\alpha/2}\sqrt{\hat{V}(\hat{\beta}_{MPV})}\right) \right]. \quad (2.39)$$

Na equação (2.39), intervalos que incluem o valor 1 indicam que as chances de ocorrência do evento estudado para a categoria de referência e uma outra investigada não são significativamente diferentes. Nesta pesquisa, vai ser utilizado o nível de confiança de 95,0%.

Para exemplificação, supondo um modelo de regressão logística com função de ligação definida tal como a equação (2.17), porém que considera apenas duas variáveis como preditivas, sendo elas idade, medida em anos, e sexo.

Ao se analisar o modelo depois de ajustado, supondo que  $X_j$  ( $j = 1, \dots, p$ ) define a idade (variável quantitativa) e mantendo a variável sexo num valor fixado, a razão de chances  $\exp(\hat{\beta}_j)$  apontará, em termos de porcentagem, o aumento ou a redução nas chances de ingresso num curso de graduação ao se acrescentar em 1 ano o valor da variável idade. Por exemplo, se  $\hat{\beta}_j = 0,15$ ,  $\exp(\hat{\beta}_j) = 1,16$  e haverá um acréscimo de 16,0% na chance de ingressar no ensino superior ao se aumentar o idade em 1 ano.

Analogamente, para o caso de  $X_j$  representar a variável sexo (qualitativa), ajustando o modelo e mantendo a idade em um valor constante, se o sexo masculino for a categoria de referência,  $\exp(\hat{\beta}_j)$  representará o valor referente à razão entre a chance das mulheres ingressarem num curso de graduação e a chance de ingresso dos homens. Novamente, se  $\hat{\beta}_j = -0,05$ ,  $\exp(\hat{\beta}_j) = 0,95$ , podemos dizer que as chances de ingresso para as mulheres são 5,0% menores que as dos homens.

#### 2.3.2.4 Avaliação da qualidade de ajuste do modelo

Em relação à qualidade do ajuste do modelo de regressão logística, uma das medidas que podem ser utilizadas é o Pseudo- $R^2$ , para o qual, quanto maior seu valor, melhor é a adequabilidade do modelo (GUO; FRASER, 2014).

O estudo de Lumley (2017) propôs a estatística do Pseudo- $R^2$  de Cox-Snell adaptado para analisar o ajuste em MLG que consideram um processo de estimação por amostras complexas, tendo como interpretação o ganho proporcional de informações para o logaritmo da função de pseudo-verossimilhança decorrente dos termos do modelo com-

pleto (com todas as variáveis explicativas) em relação ao modelo nulo (com apenas o intercepto).

O cálculo do Pseudo- $R^2$  de Cox-Snell leva em consideração o logaritmo da função de pseudo-verossimilhança ponderada pelos pesos amostrais tal como a equação a seguir.

$$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^n w_i \ln P(Y_i = y_i | \pi_i) , \quad (2.40)$$

onde  $n$  é o tamanho amostral,  $P(Y_i = y_i | \pi_i)$  é a função de probabilidade, definida pela equação (2.9) e  $w_i$  é o peso amostral do  $i$ -ésimo indivíduo.

Assim, fica permitido encontrar a expressão do Pseudo- $R^2$  de Cox-Snell resumido por Lumley (2017) como:

$$R_{CS}^2 = 1 - \exp \left\{ 2 \left( \frac{\ln L(\hat{\boldsymbol{\beta}}_0) - \ln L(\hat{\boldsymbol{\beta}})}{\sum_{i=1}^n w_i} \right) \right\} , \quad (2.41)$$

onde  $w_i$  é o peso amostral do  $i$ -ésimo indivíduo e, conforme a equação (2.40),  $\ln L(\hat{\boldsymbol{\beta}}_0)$  e  $\ln L(\hat{\boldsymbol{\beta}})$  são, respectivamente, os logaritmos das funções de pseudo-verossimilhança ponderadas para o modelo nulo e para o modelo de interesse.

A expressão de  $R_{CS}^2$  não assume valor 1, ou seja, não considera o caso em que a adequação do modelo é perfeita. Para solucionar esse problema, foi proposta a estatística do Pseudo- $R^2$  de Nagelkerke, que Lumley (2017) adaptou para o cenário da modelos que contemplam o plano de amostragem complexo (MARÔCO, 2010). O Pseudo- $R^2$  de Nagelkerke é apresentado logo abaixo.

$$R_N^2 = \frac{R_{CS}^2}{1 - \exp \left( \frac{2 \ln L(\hat{\boldsymbol{\beta}}_0)}{\sum_{i=1}^n w_i} \right)} , \quad (2.42)$$

onde  $R_{CS}^2$  é o Pseudo- $R^2$  de Cox-Snell,  $w_i$  é o peso amostral do  $i$ -ésimo indivíduo e  $\ln L(\hat{\boldsymbol{\beta}}_0)$  é o logaritmo da função de pseudo-verossimilhança ponderada do modelo nulo.

A contagem de  $Y$ 's preditos corretamente é outra medida que permite investigar a adequabilidade do modelo, com enfoque sobre a capacidade discriminatória. Como o modelo de regressão é do tipo binário, a variável resposta somente pode assumir valores 0 ou 1 (KUTNER et al., 2004). Através da probabilidade estimada de cada indivíduo frequentar um curso de graduação ( $\pi_i$ ), fica possível encontrar um  $y_i$  ajustado para cada um deles, determinando-se um ponto de corte - que pode ser, por exemplo, igual a 0,5,

em que:

$$\begin{cases} \hat{y}_i = 0, \text{ se } \pi_i \leq 0,5 \\ \hat{y}_i = 1, \text{ se } \pi_i > 0,5 \end{cases} \quad (2.43)$$

Posteriormente, constrói-se uma tabela comparando os desfechos das predições com os  $y_i$ 's verdadeiros, a fim de calcular um percentual de acertos, que ocorre quando para um mesmo indivíduo tanto o  $y_i$  observado como o previsto apresentam o mesmo desfecho binário: ambos sendo “sucessos” ou ambos sendo “fracassos”. Quanto maior a porcentagem de  $Y$ 's estimados corretamente, melhor a capacidade preditiva do modelo.

Em seguida, será apresentado o Índice de Oportunidades Humanas (IOH), objeto deste estudo, e como seus resultados podem ser interpretados.

## 2.4 Índice de Oportunidades Humanas

Existem muitos indicadores, na literatura, capazes de medir desigualdades sociodemográficas. Um exemplo é o índice de Gini, usado comumente para mensurar a desigualdade de renda, mas que ainda pode ser aplicado para avaliar outras distribuições, como a concentração de riqueza ou de terra (IPECE, 2015). Outro indicador que caracteriza a desigualdade socioeconômica, através das disparidades em relação ao acesso a serviços essenciais, é o Índice de Oportunidades Humanas (IOH).

Muitas vezes o não desenvolvimento de habilidades de um indivíduo está associado à desigualdade de acesso a um conjunto de oportunidades durante a infância, envolvendo tanto suas escolhas sociais como circunstâncias que estejam fora do seu controle (CATALÃO, 2009).

A fim de alavancar políticas públicas de longo prazo, foi concebido, no trabalho de Barros et al. (2009), o Índice de Oportunidades Humanas (IOH). Pode-se dizer que, o IOH é formado por uma medida de cobertura de acesso a determinado bem ou serviço básico e está, também, baseado em um índice de dissimilaridade ( $D$ ), que tem por objetivo mensurar a desigualdade de oportunidade de acordo com características da população de interesse.

Seguindo a metodologia utilizada por Carvalho e Waltenberg (2015) com enfoque no acesso ao ensino superior, para o cálculo do IOH, primeiramente, devem ser estimadas para cada indivíduo, denotado pelo índice  $i$  ( $i = 1, \dots, n$ ), as probabilidades condicionais

de “sucesso” (por exemplo, o acesso à graduação em instituição de ensino superior). Esse cálculo se baseia na equação abaixo, que é igual à equação (2.22), referente ao modelo de regressão logística, trabalhado na subseção 2.3.2.

$$\hat{\pi}_i = \frac{\exp(\mathbf{X}_i^T \hat{\boldsymbol{\beta}})}{1 + \exp(\mathbf{X}_i^T \hat{\boldsymbol{\beta}})} . \quad (2.44)$$

Em seguida, será possível encontrar  $\bar{\pi}$ , a taxa de cobertura média de um bem ou serviço na sociedade, que é medida ao se ponderar as probabilidades de sucesso  $\hat{\pi}_i$  para os  $n$  indivíduos investigados, descritas pela equação (2.44), pelos seus respectivos pesos amostrais  $w_i$ . Desta forma:

$$\bar{\pi} = \sum_{i=1}^n w_i \hat{\pi}_i . \quad (2.45)$$

No estudo, esta taxa se refere à proporção média de indivíduos pertencentes à demanda potencial ao ensino superior que tiveram acesso a um curso de graduação no Brasil e estavam matriculados em instituição pública ou privada, no ano de 2019.

Ao possuir conhecimento sobre o valor de  $\bar{\pi}$ , pode-se calcular o índice de dissimilaridade D, que busca mensurar o quanto as circunstâncias que se encontram fora do alcance dos indivíduos, como etnia, gênero, renda familiar e região da residência, podem influenciar a taxa de cobertura sobre os bens e serviços.

Para isso, antes, é feita uma separação dos indivíduos em  $G$  grupos ( $G < n$ ), sendo que essa quantidade depende do número de combinações de categorias de variáveis que serão incorporadas na modelagem. Supondo que os grupos sejam representados pelo índice  $g$  ( $g = 1, \dots, G$ ), cada indivíduo pertencente ao  $g$ -ésimo grupo terá a mesma probabilidade estimada  $\hat{\pi}_g$  de sucesso, ou seja,  $\hat{\pi}_g$  representa a correspondente probabilidade prevista para o  $g$ -ésimo grupo. No presente trabalho utilizou-se um total de 120 grupos de indivíduos, agregados de acordo com as variáveis descritas pela Tabela 1.

Nessa perspectiva, serão considerados “não-vulneráveis” os  $K$  grupos ( $K < G$ ) cuja probabilidade  $\hat{\pi}_k$  ( $k = 1, \dots, K$ ) de acesso ao bem ou serviço for maior que a taxa de cobertura média para essa oportunidade ( $\hat{\pi}_k > \bar{\pi}$ ). Para o caso contrário, os grupos são considerados “vulneráveis” no tocante ao acesso à oportunidade estudada ( $\hat{\pi}_k < \bar{\pi}$ ).

A partir dessas informações, o índice de dissimilaridade é construído com base nos

grupos de indivíduos com acesso aos bens ou serviços (não-vulneráveis), da seguinte forma:

$$D = \frac{1}{2} \sum_{k=1}^K w_k \left( \frac{\hat{\pi}_k - \bar{\pi}}{\bar{\pi}} \right), \quad (2.46)$$

onde:

- $w_k$  indica a participação (proporção) do  $k$ -ésimo grupo não-vulnerável em relação ao total da população,  $k = 1, \dots, K$ ;
- $\hat{\pi}_k$  é igual a probabilidade estimada de acesso ao bem ou serviço para o  $k$ -ésimo grupo não-vulnerável,  $k = 1, \dots, K$  e
- $\bar{\pi}$  é a taxa de cobertura média do bem ou serviço na população.

A demonstração da construção da equação (2.46) se encontra disponível nos trabalhos de Jahn, Schmid e Schrag (1947) e Botassio (2017).

Esse indicador assume valores entre 0 e 1 e pode ser interpretado como o percentual de pessoas/recursos que deveria ser deslocado dos grupos não-vulneráveis para os grupos vulneráveis para se chegar em um cenário de igualdade em relação ao acesso a oportunidade em questão. O cenário ideal ocorre quando  $D = 0$ , em que o acesso ao bem ou serviço independe das circunstâncias fora das capacidades de alteração por parte dos indivíduos, não havendo desigualdade no acesso. Já se  $D$  for próximo a 1, tem-se a maior dissimilaridade possível em relação ao acesso à oportunidade analisada.

O Índice de Oportunidades Humanas, por sua vez, relaciona os componentes  $D$  e  $\hat{\pi}$  por meio da seguinte equação:

$$IOH = \bar{\pi} (1 - D). \quad (2.47)$$

Através dele, busca-se captar o quão perto ou distante a sociedade se encontra de um panorama de acesso amplo e igualitário ao bem ou serviço, de modo a indicar como as características sociodemográficas dos indivíduos influenciam na probabilidade de acesso a oportunidade. O valor do IOH varia entre 0 e 1, sendo que este último valor representa a igualdade de oportunidade e completa cobertura de acesso ao serviço avaliado pela população (CARVALHO; WALTENBERG, 2015).

A partir das equações (2.46) e (2.47), pode-se mostrar que, se o acesso dos indivíduos aos bens ou serviços independe das circunstâncias ( $\hat{\pi}_i = \bar{\pi}$ , para  $i = 1, \dots, n$ ), a taxa de cobertura para todos os grupos será igual à taxa de cobertura média da população

( $\hat{\pi}_g = \bar{\pi}$ ), assim o índice de dissimilaridade será nulo ( $D = 0$ ) e o IOH será igual a  $\bar{\pi}$ . Ainda nesse cenário, em que não há desigualdade no acesso ao bem ou serviço ( $D = 0$ ), se o nível de cobertura médio for universal ( $\bar{\pi} = 1$ ), o IOH será igual a 1.

## 3 Análise dos Resultados

Este capítulo introduz os resultados do estudo, dividindo-os em 3 seções: a primeira descreve o perfil dos indivíduos integrantes da demanda potencial ao ingresso no ensino superior após investigação feita sobre os microdados da PNADc 2019; a segunda apresenta os resultados da estimação dos modelos de regressão logística, e a última explicita o valor estimado para o Índice de Oportunidades Humanas.

### 3.1 Análise Exploratória

Nesta seção foram identificadas características capazes de traçar o perfil sociodemográfico dos indivíduos pertencentes ao grupo da demanda potencial ao ensino superior descrito na subseção 2.1.1 e das pessoas desse grupo que frequentavam alguma instituição de nível superior. As análises exibidas a seguir tomaram em conta os microdados da PNADc (Pesquisa Nacional por Amostras de Domicílios Contínua) relativa ao ano de 2019 e se referem aos dados provenientes da amostra ponderados pelos expansores da população, sendo que a inclusão do plano amostral na estimação dos parâmetros foi feita por meio do Método da Máxima Pseudo-Verossimilhança, como apresentado na subsubseção 2.3.2.2.

#### 3.1.1 Demanda e Ingresso no Ensino Superior

A fim de possibilitar o cálculo do IOH para dados educacionais a partir da PNADc, torna-se interessante entender a distribuição das variáveis que traçam um perfil das pessoas quanto ao seu gênero, características dos pais, situação do domicílio em relação à posição geográfica e medidas relacionadas às condições socioeconômicas (HOYOS; NARAYAN, 2011).

Sob essa perspectiva, a base extraída dos microdados da 5<sup>a</sup> visita da PNADc de 2019 foi ponderada considerando os pesos amostrais dos indivíduos e totalizou informações

de 12.157.074 pessoas, tornando possível, por meio da Tabela 2, traçar um perfil dos indivíduos geradores da “demanda potencial ao ensino superior” para o ano de 2019. Percebeu-se que 36,0% do grupo de interesse estavam matriculados em uma instituição de ensino superior. As maiores taxas de ingresso à universidade foram de pessoas que apresentaram como características predominantes a cor/raça branca ou amarela (41,3%), o sexo feminino (40,5%), a localização de seu domicílio na Região Centro-Oeste (43,9%), o grau de instrução mais alto alcançado pelo responsável pelo domicílio como sendo o ensino superior (61,6%) e a renda domiciliar *per capita* de mais de 1,5 salário mínimo (50,2%).

Tabela 2: Perfil da demanda potencial ao ensino superior (em mil pessoas) — Brasil (2019)

Variáveis	Categorias	Total	Matriculados em IES	
			<i>n</i>	%
Total		12.157	4.380	36,0
Sexo	Masculino	6.307	2.010	31,9
	Feminino	5.850	2.370	40,5
Etnia	Preta, Parda ou Indígena	6.963	2.237	32,1
	Branca ou Amarela	5.194	2.143	41,3
Região de Residência	Sudeste	6.019	1.958	32,5
	Norte	1.006	384	38,2
	Nordeste	2.740	1.021	37,3
	Sul	1.520	635	41,8
Escolaridade do chefe	Centro-Oeste	872	383	43,9
	Até fundamental	5.768	1.461	25,3
	Médio	4.447	1.723	38,7
Renda domiciliar <i>per capita</i>	Superior	1.942	1.196	61,6
	Até 1,5 s.m. (inclusive)	9.252	2.921	31,6
	Mais de 1,5 s.m.	2.905	1.459	50,2

Fonte: Elaboração própria baseada em Carvalho e Waltenberg (2015). Microdados da PNADc 2019.

A Tabela 3 permite observar que, em 2019, o percentual de pessoas do grupo da demanda potencial que acessaram o ensino superior era de 18,2% para o ambiente público e de 17,8% para o privado. Citando os valores referentes às IES públicas e particulares, respectivamente, as maiores taxas de ingresso em ambos os tipos de IES ocorreram para indivíduos do sexo feminino (19,5% e 21,0%), da etnia branca ou amarela (18,5% e 22,8%) e cujo chefe do domicílio tinha como grau de escolaridade mais alto o ensino superior (23,5% e 38,0%). O padrão de maior índice de matrículas apenas diferiu para a região geográfica de residência e a renda familiar, sendo que, as maiores porcentagens de ingressantes às IES públicas foram de indivíduos residentes da Região Centro-Oeste (22,4%) e com renda domiciliar *per capita* de até 1,5 salário mínimo inclusive (19,0%),

e para às IES privadas, daqueles que residiam na Região Sul (22,2%) e com rendimento domiciliar *per capita* ultrapassando 1,5 salário mínimo (34,6%).

Tabela 3: Perfil dos ingressantes e não ingressantes do ensino superior entre a demanda potencial (em mil pessoas) — Brasil (2019)

Variáveis	Categorias	IES Públicas				IES Privadas			
		Não Ingr. <sup>1</sup>		Ingr. <sup>2</sup>		Não Ingr.		Ingr.	
		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Total		9.946	81,8	2.211	18,2	9.989	82,2	2.168	17,8
Sexo	Masculino	5.237	83,0	1.070	17,0	5.367	85,1	940	14,9
	Feminino	4.708	80,5	1.142	19,5	4.622	79,0	1.228	21,0
Etnia	Preta, Parda ou Indígena	5.712	82,0	1.251	18,0	5.977	85,8	986	14,2
	Branca ou Amarela	4.234	81,5	960	18,5	4.012	77,2	1.182	22,8
Região de Residência	Sudeste	5.124	85,1	895	14,9	4.956	82,3	1.063	17,7
	Norte	785	78,0	221	22,0	843	83,8	163	16,2
	Nordeste	2.137	78,0	603	22,0	2.322	84,7	418	15,3
	Sul	1.222	80,4	298	19,6	1.183	77,8	337	22,2
	Centro-Oeste	677	77,6	195	22,4	684	78,4	188	21,6
Escolaridade do chefe	Até fundamental	4.859	84,2	909	15,8	5.217	90,4	551	9,6
	Médio	3.602	81,0	845	19,0	3.569	80,3	878	19,7
	Superior	1.485	76,5	457	23,5	1.204	62,0	738	38,0
Renda dom. <sup>3</sup> <i>per capita</i>	Até 1,5 s.m. (inclusive)	7.496	81,0	1.756	19,0	8.087	87,4	1.165	12,6
	Mais de 1,5 s.m.	2.450	84,3	455	15,7	1.901	65,4	1.004	34,6

Fonte: Elaboração própria baseada em Carvalho e Waltenberg (2015). Microdados da PNADc 2019.

Ainda sobre os microdados da PNADc 2019, é importante retomar os critérios racial e de renda, levados em conta para a separação de 50,0% das vagas para acesso às IES públicas, conjuntamente com o tipo de ensino médio cursado, conforme a Lei n° 12.711/2012 (BRASIL, 2012), para entender seus impactos sobre o acesso ao ensino superior. Neste sentido, tomando como base as informações da Tabela 3, a taxa de ingresso dos indivíduos integrantes do grupo da demanda potencial em instituições dessa rede se manteve próxima entre os dois tipos de etnia, sendo que foi um pouco superior para brancos ou amarelos (18,5%) em relação a pretos, pardos ou indígenas (18,0%). Analisando-se o rendimento familiar, as porcentagens de ingresso nas IES públicas para pessoas que apresentavam renda domiciliar *per capita* inferior ou igual a 1,5 salário mínimo (19,0%) eram maiores do que para aquelas cujo domicílio possuía renda *per capita* acima desse limite (15,7%).

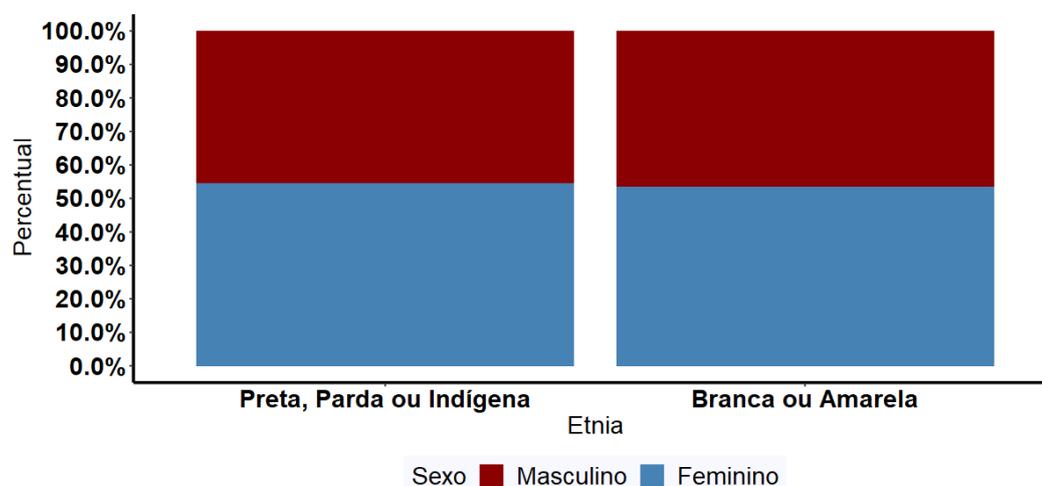
<sup>1</sup>Não Ingr.= Não Ingressantes

<sup>2</sup>Ingr.= Ingressantes

<sup>3</sup>dom.= domiciliar

### 3.1.2 Outros descritores da demanda potencial ao ensino superior matriculada em alguma IES

Ao avaliar a distribuição dos integrantes da demanda potencial ao ensino superior matriculados em alguma instituição de ensino por sexo e etnia (Figura 1), notou-se que ela se dá de forma uniforme entre as categorias de etnia. Para a etnia branca ou amarela, a porcentagem de mulheres foi 7,0% maior do que a de homens. Já para a etnia preta, parda ou indígena, essa diferença foi um pouco mais expressiva (9,4%).



Fonte: Microdados da PNADc 2019 - IBGE.

Figura 1: Perfil dos indivíduos da demanda potencial ao ensino superior que frequentavam alguma IES de acordo com o sexo e a etnia ( $n = 4.379.783$ )

Pela Figura 2, pôde-se concluir que as regiões Sul (47,4%), Sudeste (37,9%) e Centro-Oeste (37,9%), nessa ordem, detinham os maiores percentuais de pessoas cujo domicílio possuía renda *per capita* acima de 1,5 salário mínimo. Por outro lado, os residentes das regiões Nordeste e Norte com rendimento domiciliar *per capita* de até 1,5 s.m. (inclusive) ultrapassavam a faixa dos 70,0%, com porcentagens, respectivamente, iguais a 81,5% e 78,9%.

A Figura 3 permitiu constatar que as mulheres têm maior participação em ambos os tipos de IES. A diferença entre as frequências percentuais por sexo na rede pública foi de apenas 3,2%, porém, na rede privada, chegou a 13,2%.

Através da Figura 4, averiguando a distribuição das categorias de etnia por tipo de IES, notou-se que, na rede pública, havia maior presença de pessoas da etnia preta, parda ou indígena (56,6%), o que não ocorreu na rede privada, onde esse padrão se alterou, já que a maioria (54,5%) era de brancos ou amarelos.

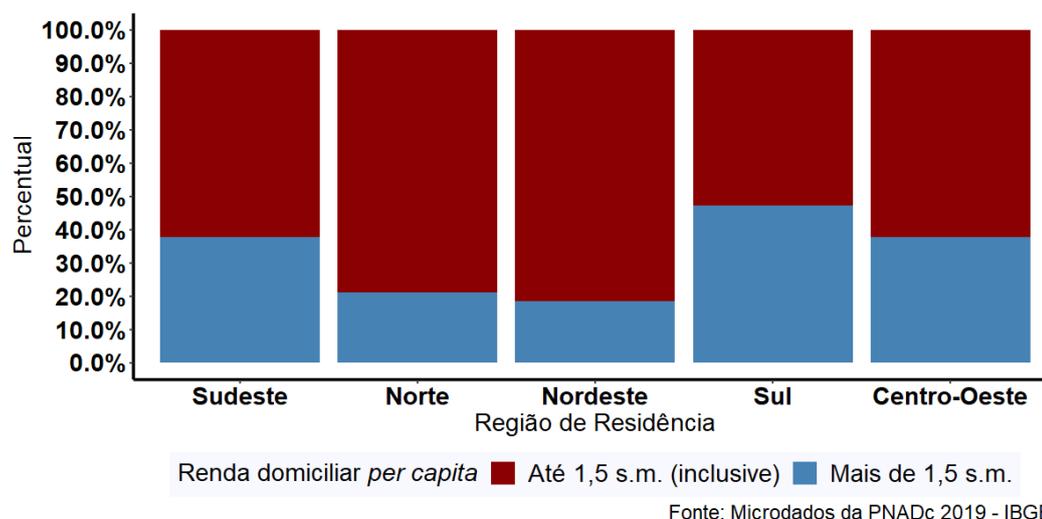


Figura 2: Perfil dos indivíduos da demanda potencial ao ensino superior que frequentavam alguma IES de acordo com a região de residência e a renda domiciliar *per capita* ( $n = 4.379.783$ )

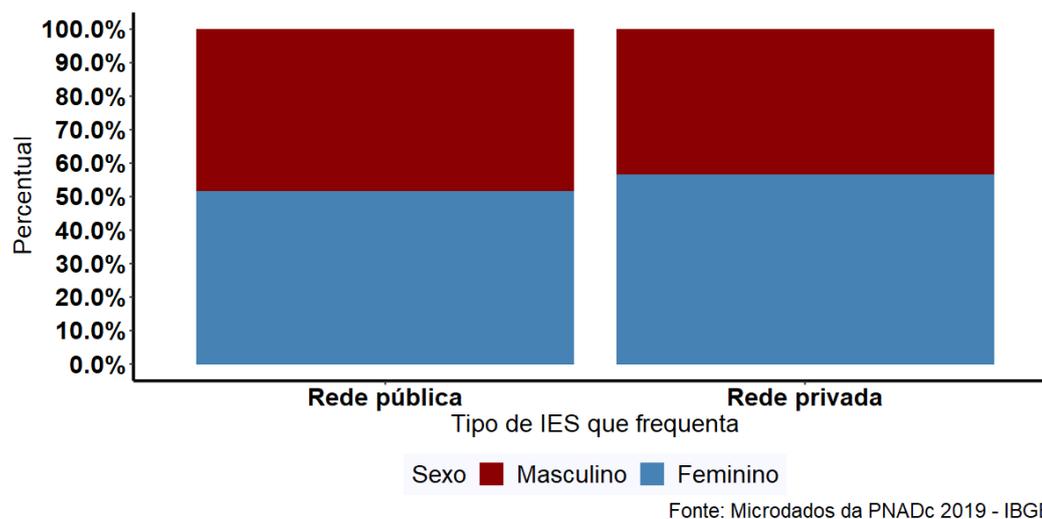


Figura 3: Perfil dos indivíduos da demanda potencial ao ensino superior de acordo com o tipo de IES que frequentava e o sexo ( $n = 4.379.783$ )

Cruzando informações de escolaridade do chefe do domicílio e a rede da instituição que frequentava para os integrantes da demanda potencial, tornou-se perceptível uma diferença no padrão observado por tipo de IES (Figura 5). O grupo de indivíduos cujo chefe domiciliar possuía como grau de escolaridade mais elevado o nível fundamental apareciam com maior percentual entre os que estudavam em IES pública (41,1%). Em contrapartida, nas universidades particulares, esse grupo apresentava a menor porcentagem observada (25,4%) e a maior delas estava associada ao caso em que o chefe do domicílio possuía o nível médio como o mais alto atingido (40,5%).

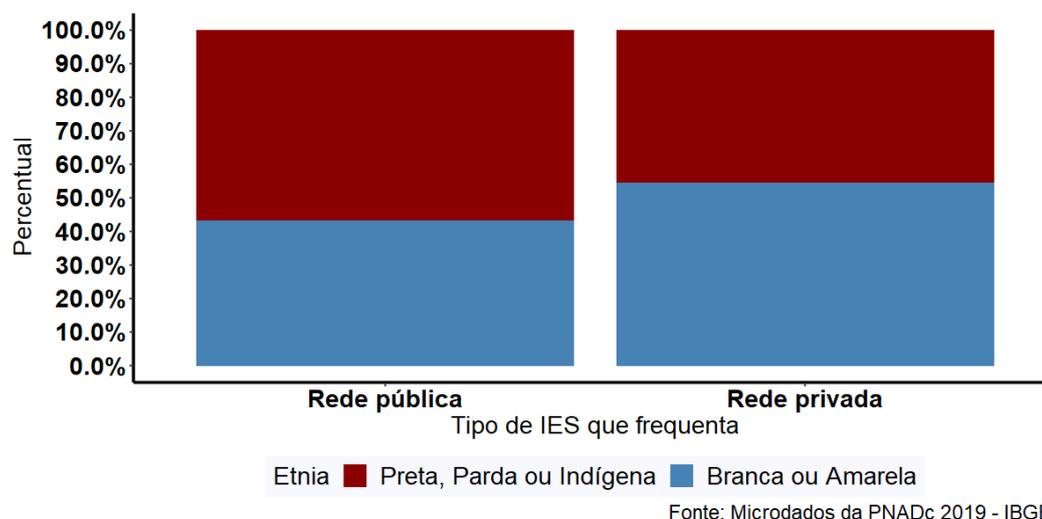


Figura 4: Perfil dos indivíduos da demanda potencial ao ensino superior de acordo com o tipo de IES que frequentava e a etnia ( $n = 4.379.783$ )

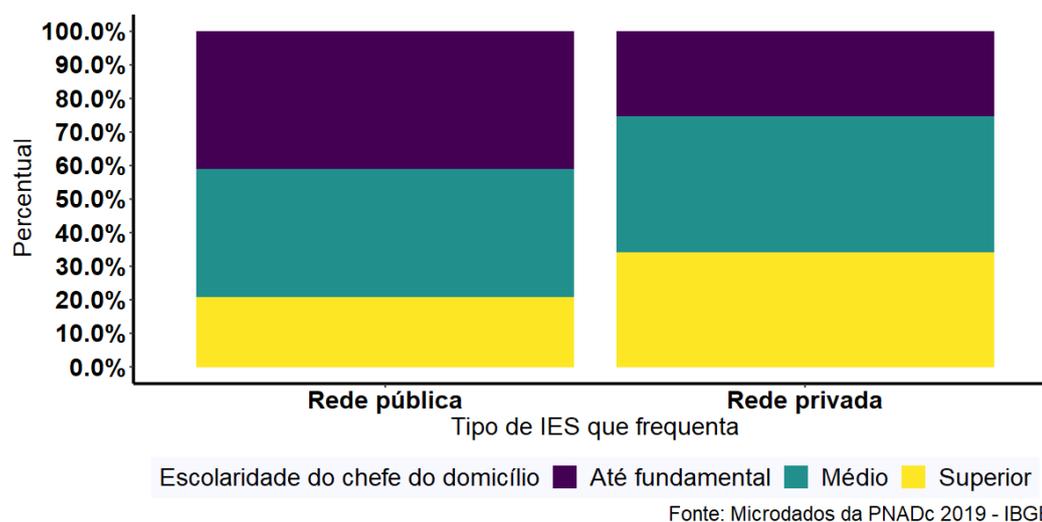


Figura 5: Perfil dos indivíduos da demanda potencial ao ensino superior de acordo com o tipo de IES que frequentava e a escolaridade do chefe do domicílio ( $n = 4.379.783$ )

Posteriormente, investigando sobre a distribuição das faixas de rendimento domiciliar *per capita* por tipo de IES que o indivíduo frequentava (Figura 6), percebeu-se que, em 2019, havia maior concentração de pessoas cujo domicílio possuía renda baixa (até 1,5 salário mínimo inclusive) tanto para a rede particular como para a pública, sendo que para a última, esse valor era mais expressivo (79,4%).

Por fim, ao se avaliar a região de residência e o tipo de universidade que frequentava (Figura 7), observou-se que as regiões Sudeste (54,3%) e Sul (53,0%) apresentavam os maiores percentuais de participação em IES privadas. Por outro lado, as regiões Nordeste (59,0%), Norte (57,6%) e Centro-Oeste (50,9%) apresentavam mais da metade dos

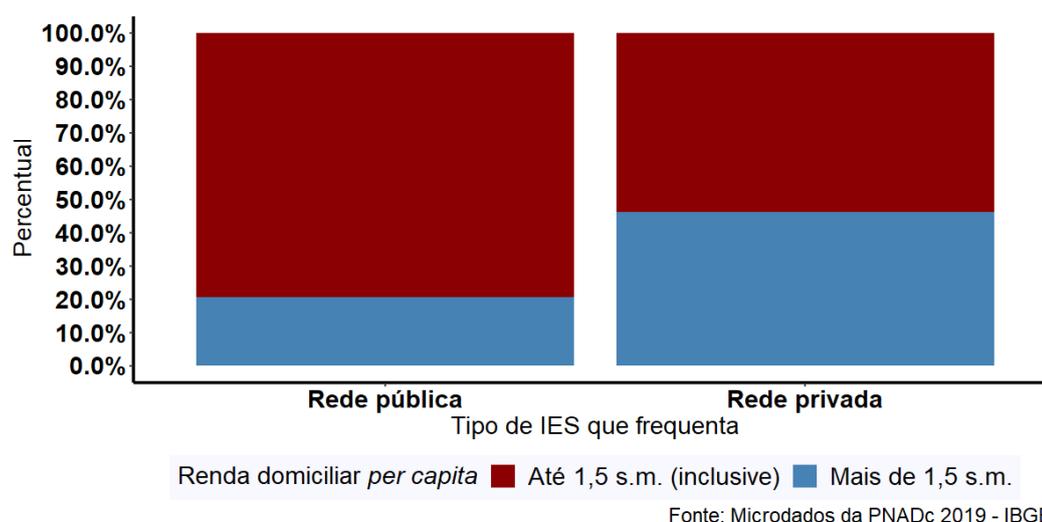


Figura 6: Perfil dos indivíduos da demanda potencial ao ensino superior de acordo com o tipo de IES que frequentava e a renda domiciliar *per capita* ( $n = 4.379.783$ )

indivíduos frequentando instituições de ensino públicas.

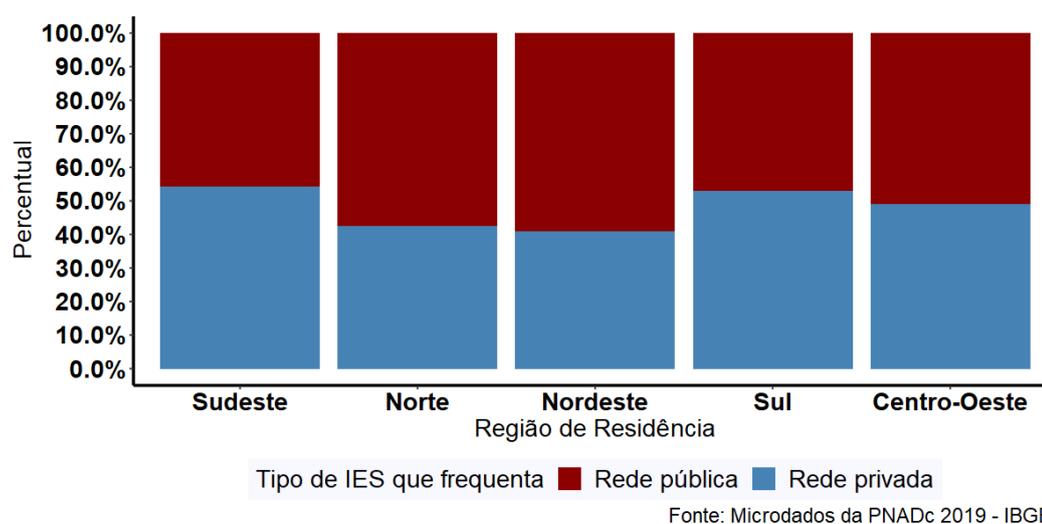


Figura 7: Perfil dos indivíduos da demanda potencial ao ensino superior de acordo com a região de residência e o tipo de IES que frequentava ( $n = 4.379.783$ )

As seções a seguir introduzem os resultados da estimação dos modelos de regressão logística e, logo em seguida, o procedimento realizado para a obtenção do índice de dissimilaridade e do IOH para o ano de 2019.

## 3.2 Estimação e Ajuste do Modelo

Seguindo a metodologia usada por Carvalho e Waltenberg (2015), buscou-se definir 3 modelos de regressão logística, da mesma forma como o apresentado na seção 2.3, com

o objetivo de avaliar a desigualdade no acesso às IES, especificamente para as públicas e para as privadas e, ainda, desconsiderando o tipo de instituição de ensino (modelo geral).

Para o modelo geral, que não levou em conta o tipo de IES que cada indivíduo frequentava, considerou-se:

$$Y_i \sim \text{Bernoulli}(\pi_i), i = 1, \dots, 12.157.074 \quad (3.1)$$

e

$$g(\pi_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{X}_i^T \boldsymbol{\beta}, \quad (3.2)$$

onde

- $Y_i$  é a variável binária que indica se o indivíduo  $i$  está matriculado em algum curso de graduação (com 1 identificando a ocorrência de matrícula e 0 a não ocorrência);
- $\beta_0$  denota o intercepto;
- $\beta_k$  corresponde ao coeficiente associado a variável  $k$ ,  $k = 1, \dots, 5$ .

A matriz  $\mathbf{X}_i^T$  é formada pelas variáveis:

- A primeira coluna de 1's, representando o nível global;
- $X_{1,i}$  está associada ao sexo do indivíduo  $i$ , onde o sexo masculino é a categoria base;
- $X_{2,i}$  representa a etnia do indivíduo  $i$ , em que a categoria de referência é “preta, parda ou indígena”;
- $X_{3,i}$  simboliza a região de residência do indivíduo  $i$ , em que a região Sudeste é a categoria de referência;
- $X_{4,i}$  está relacionada ao grau de instrução mais elevado alcançado pelo chefe do domicílio em que o indivíduo  $i$  residia, onde a categoria base é o nível de ensino até fundamental;
- $X_{5,i}$  remete ao rendimento *per capita* do domicílio em que o indivíduo  $i$  residia, em que a categoria “até 1,5 salário mínimo (inclusive)” é a de referência.

Já nos modelos estimados apenas com base nos dados dos indivíduos que ingressaram em IES pública ou privada, separadamente, a única mudança em relação ao modelo anterior foi o componente aleatório  $Y_i$ , que correspondia à indicação da existência (ou não) de matrícula no correspondente tipo de IES.

As Tabelas 4, 5 e 6 exibem os resultados para a estimação do modelo geral, do modelo que considerou apenas as IES públicas e o que tomou em consideração somente as instituições particulares, respectivamente. Para cada categoria de variável, foram apresentados os valores dos coeficientes estimados, com os respectivos erros padrão, razão de chances ( $OR$ ) e seu intervalo de 95,0% de confiança. Ao nível de significância de 5,0%, percebeu-se que apenas para o modelo relativo ao acesso às IES privadas, os coeficientes correspondentes às categorias de região Norte, Nordeste e Sul não foram significativos.

O modelo logit permitiu avaliar quais as categorias de variáveis cujos coeficientes apresentaram maior valor (positivo) de razão de chances, ou seja, mais contribuíram para o ingresso em IES quando comparadas às suas respectivas categorias de referência. A variável cujo coeficiente apresentou a mais alta razão de chances para o modelo geral foi o grau de instrução mais elevado do chefe do domicílio onde, morar junto com um responsável que possui nível superior leva a uma chance 3,873 vezes maior de ingressar em uma IES do que para aqueles indivíduos que residem em domicílios onde o chefe de família alcançou no máximo o nível fundamental (Tabela 4).

Investigando separadamente as IES públicas (Tabela 5), a categoria de variável com maior razão de chance observada para o ingresso numa graduação também foi a escolaridade do chefe do domicílio sendo o ensino superior, em que a chance de ingresso numa instituição dessa rede de ensino é 1,984 vezes maior quando comparado a um indivíduo que possui um responsável pelo domicílio com até o ensino fundamental completo.

Para as universidades privadas (Tabela 6), as categorias de variáveis que mais interferiram positivamente para o acesso aos cursos de graduação foram o grau de instrução superior do chefe do domicílio e a renda domiciliar *per capita* acima da faixa de 1,5 s.m., sendo que as razões de chances considerando essas categorias, quando comparadas às categorias de referência nível de escolaridade até fundamental do responsável domiciliar e rendimento *per capita* do domicílio menor ou igual a 1,5 s.m., eram de 3,688 e 2,535, respectivamente. Dessa forma, em todos os modelos estimados, a maior razão de chances observada foi referente à categoria de variável “grau de instrução do chefe do domicílio como sendo o ensino superior”, repetindo o padrão do alto *background* familiar e reduzindo a mobilidade social.

Apenas para o modelo que avaliava o ingresso no ensino superior público, a categoria de renda domiciliar *per capita* acima de 1,5 s.m. foi associada a um coeficiente estimado de sinal negativo (-0,420), o que indicou que essa categoria era um fator diminuidor para o ingresso a IES públicas, uma vez que a chance de acesso associada a essa categoria é 0,343 vezes menor do que para a categoria de referência rendimento do domicílio de até 1,5 s.m. (inclusive). Um dos motivos que poderiam existir para que a renda *per capita* do domicílio acima do limite de 1,5 s.m. não contribua mais do que o rendimento abaixo desse ponto de corte para a entrada nas IES públicas é que um dos critérios considerados para destinar vagas referentes às cotas para entrada nesse tipo de IES é possuir domicílio que não ultrapasse a faixa de 1,5 s.m. *per capita* (BRASIL, 2012). Nos demais modelos, todas as categorias de variáveis apresentaram sinais dos coeficientes estimados positivos, ou seja, potencializavam o evento “ingresso em uma IES”, quando comparadas às categorias de referência.

Tabela 4: Resultados da estimação do Modelo Logit geral - Brasil (2019)

Variáveis/Categorias	IES Geral			
	$\hat{\beta}$	EP( $\hat{\beta}$ )	OR	IC[OR; 95,0%]
Intercepto	-1,686	0,053	0,185	[0,167 ; 0,206]
Sexo Feminino	0,463	0,040	1,589	[1,468 ; 1,719]
Etnia Branca ou Amarela	0,253	0,043	1,288	[1,183 ; 1,401]
Região Norte	0,387	0,062	1,473	[1,304 ; 1,664]
Região Nordeste	0,416	0,052	1,516	[1,368 ; 1,679]
Região Sul	0,295	0,058	1,343	[1,200 ; 1,505]
Região Centro-Oeste	0,527	0,067	1,694	[1,486 ; 1,931]
Escolaridade do chefe Ens. Médio	0,588	0,045	1,800	[1,648 ; 1,967]
Escolaridade do chefe Ens. Superior	1,354	0,061	3,873	[3,435 ; 4,370]
Renda dom. <i>per capita</i> >1,5 s.m.	0,427	0,054	1,533	[1,378 ; 1,705]

Fonte: Elaboração própria baseada em Carvalho e Waltenberg (2015). Microdados da PNADc 2019.

Para testar a validação dos modelos logit estimados, foram calculadas as medidas de qualidade do ajuste porcentagem de  $Y$ 's estimados corretamente e Pseudo- $R^2$  de Nagelkerke (Tabela 7). O valor da contagem de  $Y$ 's corretamente preditos foi próxima ou superou a faixa de 70,0% para os três modelos estimados, indicando boa capacidade discriminatória. Já o Pseudo- $R^2$  de Nagelkerke apontou que, para o modelo geral e para aqueles relativos às IES públicas e às privadas, em separado, o ganho proporcional de informações com a inclusão das variáveis explicativas foi de 13,0%, 3,0% e 16,0%, respectivamente. Para chegar nesse resultado, levou-se em consideração que as variáveis sociodemográficas e as suas categorias relacionadas foram incluídas no modelo de modo a respeitar a metodologia utilizada no trabalho de Carvalho e Waltenberg (2015) para construção do IOH. Em termos de comparação, originalmente, no estudo de Barros et al. (2009), trabalhou-se com

Tabela 5: Resultados da estimação do Modelo Logit, para a categoria administrativa de IES pública - Brasil (2019)

Variáveis/Categorias	IES Públicas			
	$\hat{\beta}$	EP( $\hat{\beta}$ )	OR	IC[OR; 95,0%]
Intercepto	-2,000	0,061	0,135	[0,120 ; 0,153]
Sexo Feminino	0,161	0,046	1,175	[1,072 ; 1,286]
Etnia Branca ou Amarela	0,104	0,050	1,110	[1,006 ; 1,225]
Região Norte	0,447	0,072	1,564	[1,358 ; 1,801]
Região Nordeste	0,472	0,063	1,603	[1,417 ; 1,815]
Região Sul	0,344	0,072	1,411	[1,223 ; 1,625]
Região Centro-Oeste	0,509	0,080	1,664	[1,423 ; 1,945]
Escolaridade do chefe Ens. Médio	0,286	0,055	1,331	[1,195 ; 1,483]
Escolaridade do chefe Ens. Superior	0,685	0,070	1,984	[1,729 ; 2,277]
Renda dom. <i>per capita</i> >1,5 s.m.	-0,420	0,067	0,657	[0,576 ; 0,749]

Fonte: Elaboração própria baseada em Carvalho e Waltenberg (2015). Microdados da PNADc 2019.

Tabela 6: Resultados da estimação do Modelo Logit, para a categoria administrativa de IES privada - Brasil (2019)

Variáveis/Categorias	IES Privadas			
	$\hat{\beta}$	EP( $\hat{\beta}$ )	OR	IC[OR; 95,0%]
Intercepto	-2,889	0,074	0,056	[0,048 ; 0,064]
Sexo Feminino	0,559	0,053	1,749	[1,575 ; 1,942]
Etnia Branca ou Amarela	0,288	0,057	1,334	[1,192 ; 1,493]
Região Norte	0,122	0,080	1,130	[0,966 ; 1,321]
Região Nordeste	0,130	0,067	1,139	[1,000 ; 1,298]
Região Sul	0,107	0,074	1,113	[0,963 ; 1,287]
Região Centro-Oeste	0,271	0,077	1,311	[1,127 ; 1,525]
Escolaridade do chefe Ens. Médio	0,725	0,061	2,065	[1,833 ; 2,327]
Escolaridade do chefe Ens. Superior	1,305	0,075	3,688	[3,182 ; 4,273]
Renda dom. <i>per capita</i> >1,5 s.m.	0,930	0,062	2,535	[2,247 ; 2,859]

Fonte: Elaboração própria baseada em Carvalho e Waltenberg (2015). Microdados da PNADc 2019.

as seguintes variáveis: gênero, cor/raça, localidade de nascimento, graus de escolaridade do pai e da mãe e ocupação principal do pai.

Sabendo que, neste trabalho, a quantidade de características pessoais que pudessem ser empregadas como variáveis nos modelos era limitada, a qualidade do ajuste não foi satisfatória, ou seja, somente com essas variáveis, ainda que apresentassem categorias cujos coeficientes correspondentes eram significativos na maior parte dos modelos, esse fato não foi suficiente para que o comportamento da variável dependente pudesse ser bem explicado.

Uma sugestão para trabalhos que busquem construir o IOH para investigar o acesso ao ensino superior é verificar a oferta de informações sobre as variáveis sociodemográficas

que possam ser utilizadas pelos modelos de regressão logística, de forma semelhante ao estabelecido na pesquisa de Barros et al. (2009), em busca de alcançar resultados melhores quanto à qualidade da previsão das estimativas das probabilidades de ingresso de cada indivíduo e dos indicadores D e IOH, com base em modelos de melhores ajustes.

Tabela 7: Medidas de qualidade do ajuste para os Modelos Logit - Brasil (2019)

Medidas de qualidade do ajuste	Modelos Logit		
	IES Geral	IES Públicas	IES Privadas
Pseudo-R <sup>2</sup> de Nagelkerke	0,130	0,030	0,160
Contagem de Y's estimados corretamente (%)	68,298	81,809	82,452

Fonte: Microdados da PNADc 2019.

Na seção a seguir serão expostos e interpretados os valores calculados para os Índices de dissimilaridade e de Oportunidades Humanas.

### 3.3 Estimação do IOH

A partir do resultado da estimação dos coeficientes dos modelos de regressão logística na seção anterior, foram calculadas as probabilidades condicionais para o ingresso em IES e a taxa média de acesso ao ensino superior, assim como descritas nas equações (2.44) e (2.45), respectivamente.

O Apêndice A apresenta as Figuras 8, 9 e 10 em que, para os 3 modelos estimados, as probabilidades de acesso foram ordenadas de forma crescente para cada um dos 120 grupos de indivíduos integrantes da demanda potencial ao ensino superior, que consideraram todas as combinações entre as categorias de circunstância (2 categorias para a variável sexo, 2 para a variável etnia, 5 categorias para a região de residência, 3 para o grau de instrução do chefe do domicílio e 2 para a renda *per capita* do domicílio).

Percebeu-se, pelo modelo geral, que desconsidera o tipo de instituição de ensino superior que o indivíduo está matriculado, que o grupo com menor probabilidade de acesso à educação superior (mais vulnerável) era composto por homens, da etnia preta, parda ou indígena, moradores da região Sudeste, com chefe do domicílio com o ensino fundamental como sendo o grau de instrução mais alto alcançado e a renda domiciliar *per capita* abaixo ou igual a 1,5 salário mínimo, tendo probabilidade de ingresso igual a 0,156. Por outro lado, o grupo menos vulnerável era formado por mulheres brancas ou amarelas, residentes na região Centro-Oeste, cujo responsável domiciliar possuía ensino superior e rendimento *per capita* do domicílio superando a faixa de 1,5 s.m., sendo que a probabilidade de acesso

ao ensino superior era de 0,792 (Figura 8).

Com relação ao ingresso nas IES públicas (Figura 9), o grupo social mais vulnerável era formado por homens não brancos, residentes na região Sudeste, cujo chefe domiciliar tinha como grau de educação mais elevado o ensino fundamental e renda *per capita* do domicílio ultrapassando o limite de 1,5 s.m., com probabilidade de acesso igual a 0,082. Por outro lado, o grupo que tinha a maior probabilidade de ingresso em um curso de graduação dessa rede (0,368) era composto por indivíduos do sexo feminino, da etnia branca ou amarela, que moravam na região Centro-Oeste, cujo responsável pelo domicílio possuía educação superior e rendimento domiciliar *per capita* de até 1,5 s.m. (inclusive).

Avaliando as instituições de ensino superior particulares, através da Figura 10, nota-se que o grupo com menor probabilidade de ingresso a graduações dessa rede (0,053) era constituído de pessoas do sexo masculino, da etnia preta, parda ou indígena, que residiam na região Sudeste, cujo chefe do domicílio possuía grau de instrução até o ensino fundamental e rendimento domiciliar *per capita* menor ou igual a 1,5 s.m. Em contrapartida, o grupo social menos vulnerável para o ingresso às IES privadas era formado por mulheres, brancas ou amarelas, que moravam no Centro-Oeste, cujo responsável domiciliar tinha grau de escolaridade mais elevado como sendo a educação superior e com rendimento *per capita* do domicílio acima do limite de 1,5 s.m., com probabilidade de acesso de 0,614.

A Tabela 8 apresenta, para cada um dos três modelos investigados, nessa ordem, os valores das probabilidades de ingresso no ensino superior para os grupos mais vulneráveis (probabilidade mínima encontrada) e para os menos vulneráveis (probabilidade máxima), exibindo também a taxa média de cobertura relativa ao acesso a essa oportunidade ( $\bar{\pi}$ ) e os índices de dissimilaridade (D) e de Oportunidades Humanas (IOH) calculados para o Brasil a partir dos dados da PNADc de 2019, como descrito nas equações (2.46) e (2.47).

Tabela 8: Taxas de cobertura e índices D e IOH para o acesso ao ensino superior - Brasil (2019)

Indicadores	IES Geral	IES Públicas	IES Privadas
Taxa mínima	0,156	0,082	0,053
Taxa máxima	0,792	0,368	0,614
Taxa média	0,360	0,182	0,178
D	0,080	0,057	0,135
IOH	0,331	0,172	0,154

Fonte: Elaboração própria baseada em Carvalho e Waltenberg (2015). Microdados da PNADc 2019.

Tendo em vista o modelo geral, foi encontrado um valor do índice de dissimilaridade (D) igual a 0,080, indicando que 8,0% das vagas de pessoas da demanda integrantes dos

grupos não vulneráveis deveriam ser realocadas para aquelas que pertencem aos grupos vulneráveis, a fim do reestabelecimento da igualdade de oportunidades no ingresso a uma graduação.

A taxa de cobertura média para o ingresso na educação superior pública para a sociedade em 2019 foi maior do que aquela para o ingresso nas IES privadas: valores de 0,182 e 0,178, respectivamente, relativamente próximos. Além disso, a desigualdade no acesso ao ensino superior, medida pelo índice D, para as instituições particulares (0,135) superou aquela encontrada para as públicas (0,057). Também pode-se enfatizar que o índice de dissimilaridade está próximo de zero para o ingresso nas instituições públicas, situação em que o acesso seria independente das características pessoais fora da capacidade de alteração por parte dos indivíduos pertencentes ao grupo da demanda potencial ao ensino superior, sinalizando que o IOH estava mais próximo de 1 para o acesso a esse tipo de IES, ou seja, havia mais igualdade de oportunidades com relação ao acesso à rede pública de ensino superior em 2019.

Em relação ao IOH, os valores desse indicador para o acesso ao ensino superior, independentemente do tipo de IES, se encontraram distantes de um cenário ideal, em que assumiriam o valor 1. O IOH para o acesso às instituições de ensino superior privadas (0,154) foi menor do que para às IES públicas (0,172), indicando que as maiores oportunidades de ingresso ao ambiente universitário se encontravam no setor público. A desigualdade mais presente para o ingresso ao ambiente universitário privado pode estar relacionada aos fenômenos da crise econômica enfrentada pelo país e a falta de fomento do governo federal a incentivos à entrada nesse tipo de ensino, como financiamentos estudantis (PEDUZZI, 2017; JUSTINO, 2018).

Para traçar um paralelo com os resultados obtidos, foram verificadas as conclusões apresentadas por um dos trabalhos mais recentes feitos para analisar a população brasileira utilizando esse mesmo indicador, sendo esse o de Carvalho e Waltenberg (2015), que se baseava nos dados das PNADs (Pesquisas Nacionais por Amostras de Domicílios) realizadas nos anos de 2003 e 2013. É importante ressaltar que, para essa pesquisa, não houve indicação se foi considerado algum método de estimação na análise dos resultados e que, embora a intenção de retratar o perfil sociodemográfico da população seja a mesma, a PNAD e a PNADc (usada neste presente estudo) seguem metodologias diferentes de trabalho.

Levando em consideração que durante o período investigado pelo trabalho de Carvalho e Waltenberg (2015) a PNADc não era realizada (em 2003) e que atualmente essa é a

principal pesquisa para representar a população entre censos, foram registrados valores de IOH iguais a 0,240; 0,091 e 0,195 para o ano de 2003 e a 0,281; 0,102 e 0,230 para o ano de 2013, ao se considerar, respectivamente, o modelo geral, aquele referente ao acesso às IES públicas e o relacionado ao ingresso nas universidades privadas.

## 4 Conclusões

Neste trabalho, partindo dos microdados da PNADc de 2019, último ano antes da disseminação de Covid-19 pelo mundo, buscou-se mensurar as desigualdades existentes para o ingresso em graduações, ao se tomar como referência o grupo da demanda potencial ao ensino superior.

Pelas análises exploratórias realizadas sobre o perfil de indivíduos integrantes da demanda potencial ao ensino superior em relação aos critérios que podem ser consideradas pela “Lei de Cotas”, percebeu-se que o maior número de ingressos às IES dessa rede são daqueles que possuem etnia preta, parda ou indígena e rendimento domiciliar *per capita* de até 1,5 s.m. (inclusive). Essas informações possibilitam realçar os efeitos da lei em questão com respeito às características étnicas e relativas à renda familiar dos estudantes universitários ingressantes, após sua implementação.

Através dos resultados dos modelos de regressão logística, foi possível calcular os índices de dissimilaridade e de Oportunidades Humanas para o Brasil em 2019.

No que se refere ao objetivo do trabalho, o IOH referente ao ingresso em instituições de ensino superior particulares em 2019 foi menor que o indicador relativo às IES públicas, retratando que as oportunidades de acesso foram maiores para esse último setor, o que se deve, possivelmente, a questões econômicas enfrentadas pelo país no período em pauta. Vale destacar, ainda, que, para os três casos estudados (analisando o acesso às universidades públicas, privadas ou de forma independente do tipo de rede da instituição), os valores de IOH encontrados ficaram distantes do valor ideal 1, sinalizando que o Brasil ainda está longe de apresentar um cenário de igualdade em relação ao acesso à oportunidade “ingresso no ensino superior”.

Diante das dificuldades encontradas, sugere-se, para projetos futuros utilizando o mesmo tema, verificar a disponibilidade de informações na base de dados empregada, a fim de escolher outras variáveis que reflitam as circunstâncias que possam efetivamente, de forma conjunta, colaborar para uma melhor estimativa das probabilidades de cada

pessoa ter acesso a um curso de graduação e, por consequência, dos indicadores D e IOH, a partir de modelos melhores em termos de ajuste. Outra opção seria a construção de um novo indicador, com abordagem distinta de modelagem, para medir a desigualdade de oportunidades com relação ao acesso ao ensino superior.

## Referências

- ALMEIDA, M. E. et al. Comparando bases de dados: o caso do cadastro geral de empregados e desempregados (CAGED) e da pesquisa nacional por amostra de domicílios contínua (PNAD Contínua). Instituto de Pesquisa Econômica Aplicada (Ipea), 2018.
- ANDRADE, C. Y. de. Acesso ao ensino superior no Brasil: equidade e desigualdade social. *Revista Ensino Superior Unicamp*, v. 6, p. 18–27, 2012.
- BARROS, R. P. D. et al. *Measuring inequality of opportunities in Latin America and the Caribbean*. [S.l.]: World Bank Publications, 2009.
- BARROS, R. P. d. O.; FOGUEL, M. N. O.; ULYSSEA, G. O. Desigualdade de renda no Brasil: uma análise da queda recente. Instituto de Pesquisa Econômica Aplicada (IPEA), 2006.
- BINDER, D. A. On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review/Revue Internationale de Statistique*, JSTOR, p. 279–292, 1983.
- BOLFARINE, H.; BUSSAB, W. de O. *Elementos de amostragem*. [S.l.]: Editora Blucher, 2005.
- BOTASSIO, D. *Segregação e desigualdade: Analogia na mensuração e análise da segregação por gênero em setores de atividade no Brasil [Dissertação de mestrado, Universidade de São Paulo]*. Biblioteca Digital de Teses e Dissertações da USP. 2017.
- BRASIL, C. Lei nº 10.172, de 9 de janeiro de 2001. aprova o plano nacional de educação e dá outras providências. *Diário Oficial da União*, p. 1–1, 2001.
- BRASIL, C. 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. *Diário Oficial da União*, 2012.
- BRASIL, C. Lei nº 13.005, de 25 de junho de 2014. aprova o plano nacional de educação-pne e dá outras providências. *Diário Oficial da União*, Poder Executivo Brasília, DF, v. 26, p. 1–7, 2014.
- BRITO, J. R. S. de et al. Oferta de serviços públicos e desenvolvimento regional: Uma análise a partir do índice de oportunidade humana (IOH). *Revista INTERFACE-UFRN/CCSA ISSN Eletrônico 2237-7506*, v. 14, n. 1, p. 27–49, 2017.
- CARVALHO, M.; WALTENBERG, F. d. Desigualdade de oportunidades no acesso ao ensino superior no Brasil: Uma comparação entre 2003 e 2013. *Economia Aplicada*, SciELO Brasil, v. 19, p. 369–396, 2015.

- CATALÃO, I. Entre a institucionalização e a vida cotidiana: elementos para repensar o espaço metropolitano de Brasília. *Cadernos Metr pole.*, v. 11, n. 22, 2009.
- DILL, H. C.; GONÇALVES, F. d. O. Índice de oportunidade humana: estimação e decomposição através do valor de shapley para o Brasil de 1999-2009. *Anais do 39<sup>o</sup> Encontro Nacional de Economia–ANPEC*, 2011.
- DOBSON, A. J.; BARNETT, A. G. *An introduction to generalized linear models*. [S.l.]: CRC press, 2018.
- FACHINETTI, T. A.; CANDIDO, E. A. P.; CARNEIRO, R. R. U. C. A democratização do acesso à educação superior brasileira: realidade e desafios. *DOXA: Revista Brasileira de Psicologia e Educação*, p. 83–95, 2020.
- FEM. *The Global Gender Gap Report 2016*. 2016. <[https://www3.weforum.org/docs/GGGR16/WEF\\_Global\\_Gender\\_Gap\\_Report\\_2016.pdf](https://www3.weforum.org/docs/GGGR16/WEF_Global_Gender_Gap_Report_2016.pdf)>. [Online; accessed 17-December-2021].
- FERREIRA, P. C.; VELOSO, F. A. O desenvolvimento econômico brasileiro no pós-guerra. FGV Crescimento e Desenvolvimento, 2015.
- FERREIRA, U. C. S. et al. Um estudo sobre a desigualdade no acesso as instituições de ensino superior públicas e privadas em 2013 no Brasil. Universidade Federal Fluminense, 2015.
- FIGUEIREDO, E. A. d.; SILVA, C. R. d. F. Desigualdade de oportunidades no Brasil: uma decomposição quantílica contrafactual. Instituto de Pesquisa Econômica Aplicada (Ipea), 2012.
- GONÇALVES, F. de O. et al. Índice de oportunidade humana (IOH) no Distrito Federal. Companhia de Planejamento do Distrito Federal, 2015.
- GROUP, W. B. *Buscando Oportunidades para Todos*. 2016. <<https://documents1.worldbank.org/curated/pt/496211476446558735/pdf/109075-BRI-HumanOpportunityIndex-Brief-oct2016FINAL-PUBLIC-PORTUGUESE.pdf>>. [Online; accessed 30-November-2021].
- GUO, S.; FRASER, M. W. *Propensity score analysis: Statistical methods and applications*. [S.l.]: SAGE publications, 2014. v. 11.
- HOYOS, A.; NARAYAN, A. Inequality of opportunities among children. World Bank, Washington, DC, 2011.
- IBGE. *PNAD Cont nua Informa es sobre os arquivos de Microdados*. 2020. <[https://ftp.ibge.gov.br/Trabalho\\_e\\_Rendimento/Pesquisa\\_Nacional\\_por\\_Amostra\\_de\\_Domicilios\\_continua/Anual/Microdados/Visita/LEIA-ME.pdf](https://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_continua/Anual/Microdados/Visita/LEIA-ME.pdf)>. [Online; accessed 19-January-2022].
- IBGE. *Pesquisa nacional por amostra de domic lios cont nua : notas t cnicas : vers o 1.9*. 2021. <[https://biblioteca.ibge.gov.br/visualizacao/livros/liv101868\\_notas\\_tecnicas.pdf](https://biblioteca.ibge.gov.br/visualizacao/livros/liv101868_notas_tecnicas.pdf)>. [Online; accessed 28-January-2022].

- IBGE. *Pesquisa nacional por amostra de domicílios contínua : O que é*. 2021. <<https://www.ibge.gov.br/estatisticas/sociais/educacao/9173-pesquisa-nacional-por-amostra-de-domicilios-continua-trimestral.html?=&t=o-que-e>>. [Online; accessed 28-December-2021].
- IPECE. *Entendendo o ÍNDICE DE GINI*. 2015. <[https://www.ipece.ce.gov.br/wp-content/uploads/sites/45/2015/02/Entendendo\\_Indice\\_GINI.pdf](https://www.ipece.ce.gov.br/wp-content/uploads/sites/45/2015/02/Entendendo_Indice_GINI.pdf)>. [Online; accessed 17-December-2021].
- JAHN, J.; SCHMID, C. F.; SCHRAG, C. The measurement of ecological segregation. *American Sociological Review*, JSTOR, v. 12, n. 3, p. 293–303, 1947.
- JUSTINO, G. *Com menos matrículas e ingresso de alunos, universidades privadas buscam reinvenção*. 2018. <<https://shre.ink/m5Gs>>. [Online; accessed 23-June-2022].
- KONE, H. R. *Desenvolvimento e gênero na África subsaariana: Ods e políticas de cooperação internacional*. Universidade Federal de Uberlândia, 2019.
- KUTNER, M. H. et al. *Applied linear regression models*. [S.l.]: McGraw-Hill/Irwin New York, 2004. v. 4.
- LUMLEY, T. Pseudo- $R^2$  statistics under complex sampling. *Australian & New Zealand Journal of Statistics*, Wiley Online Library, v. 59, n. 2, p. 187–194, 2017.
- MARÔCO, J. *Análise de equações estruturais: Fundamentos teóricos, software e aplicações. Pêro Pinheiro: Report Number*. [S.l.]: Retirado de [www.reportnumber.pt/ae](http://www.reportnumber.pt/ae), 2010.
- MEC. *Perguntas frequentes*. 2012. <<http://portal.mec.gov.br/cotas/perguntas-frequentes.html>>. [Online; accessed 01-December-2021].
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to linear regression analysis*. [S.l.]: John Wiley & Sons, 2021.
- NAÇÕES UNIDAS, O. d. *Saiba mais sobre Cooperação Sul-Sul e Triangular*. [S.l.]: Organização Internacional do Trabalho, 2019. <<https://www.ilo.org/brasilia/temas/south-south/WCMS.660547/lang--pt/index.htm>>. [Online; accessed 13-January-2022].
- NASCIMENTO, V. M. B. *Evolução do índice de oportunidade humana-IOH-após os grandes programas de investimentos governamentais*. 2016.
- NERI, M. C. *Mapa da inclusão digital*. [S.l.], 2012.
- NOGUEIRA, L. V. *Efeito das instituições públicas de ensino superior sobre o crescimento econômico local*. 2018.
- OLIVEIRA, L. V. N. d. et al. *A desigualdade de oportunidades no acesso ao ensino superior no estado de Mato Grosso do Sul*. Universidade Federal da Grande Dourados, 2018.
- PAULA, G. A. *Modelos de regressão: com apoio computacional*. [S.l.]: IME-USP São Paulo, 2004.

- PEDUZZI, P. *Ritmo de crescimento no número de matrículas no ensino superior diminui em 2016*. 2017. <<https://agenciabrasil.ebc.com.br/educacao/noticia/2017-08/ritmo-de-crescimento-no-numero-de-matriculas-no-ensino-superior-diminui-em>>. [Online; accessed 23-June-2022].
- PESSOA, D. G. C.; SILVA, P. L. N. Análise de dados amostrais complexos. *São Paulo: Associação Brasileira de Estatística*, v. 112, 1998.
- REIS, M. C.; RAMOS, L. Escolaridade dos pais, desempenho no mercado de trabalho e desigualdade de rendimentos. *Revista brasileira de economia*, SciELO Brasil, v. 65, p. 177–205, 2011.
- RF, B. Decreto nº 9.661, de 1º de janeiro de 2019. *Diário Oficial da União*, 2019.
- ROEMER, J. Equality of opportunity, harvard u. *Press, Cambridge*, 1998.
- SOUZA, M. J. d. Modelagem da chance de diabetes em mulheres brasileiras: uma avaliação do efeito do plano amostral. 2019.
- SULIANO, D.; MIRO, V. Avaliando as condições de oportunidade a partir do IOH ao longo da década de 2000: Evidências para o Ceará. Instituto de Pesquisa e Estratégia Econômica do Ceará, 2014.
- SULIANO, D. C.; FILHO, J. J. Um estudo empírico utilizando o índice de oportunidade humana para o Brasil. *Análise Econômica*, v. 35, n. 68, 2017.

# APÊNDICE A

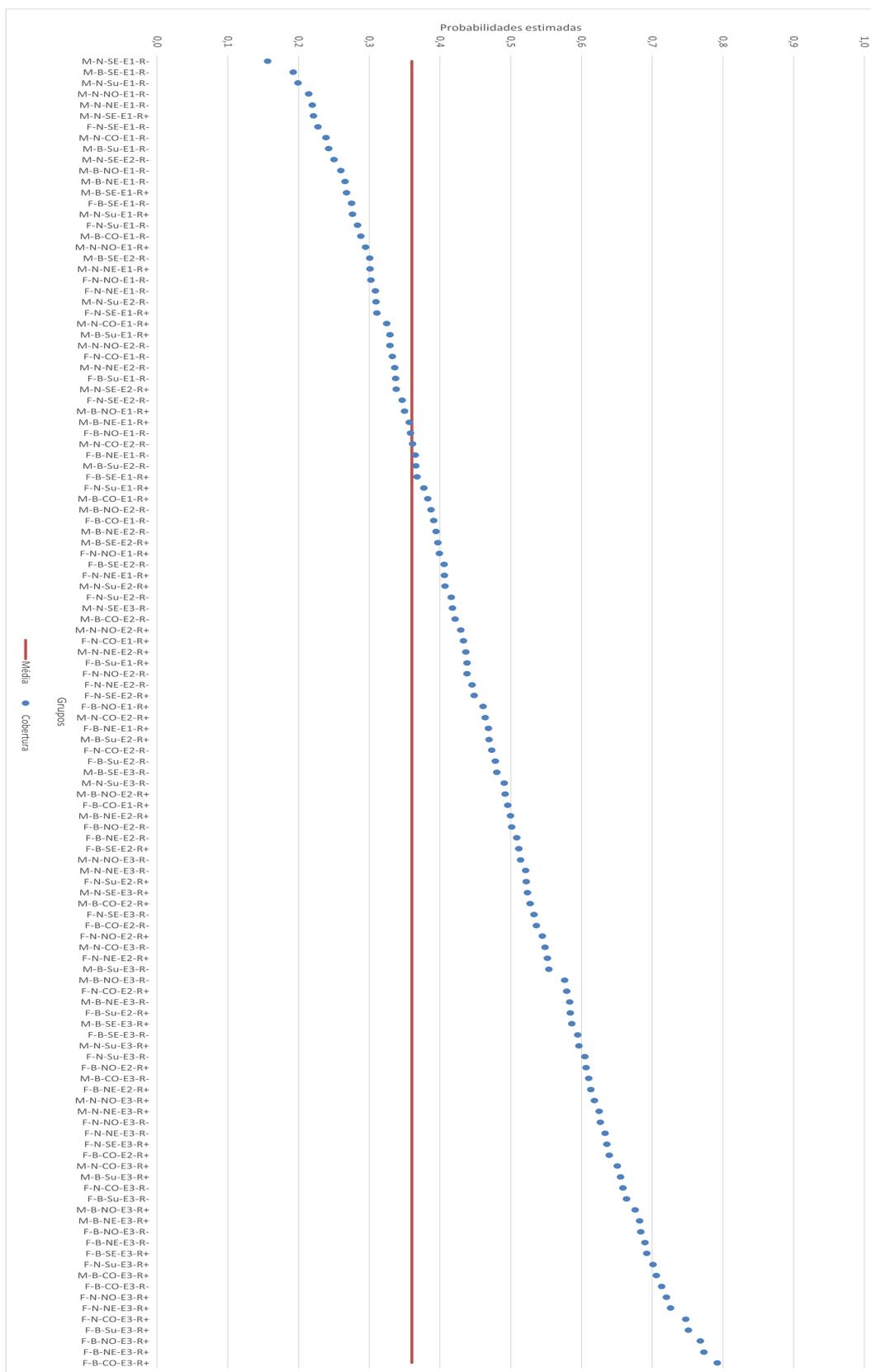


Figura 8: Probabilidades estimadas de ingresso em instituições de ensino superior segundo os grupos

Fonte: Elaboração própria baseada em Carvalho e Waltenberg (2015). Microdados da PNADc 2019. Legenda: M=Sexo Masculino, F=Sexo Feminino, B=Branco ou Amarelo, N=Preto, Pardo ou Indígena, SE= Sudeste, NO= Norte, NE= Nordeste, Su= Sul, CO= Centro-Oeste, E1= Até Ensino Fundamental, E2= Ensino Médio, E3= Ensino Superior, R-=Renda domiciliar *per capita* de até 1,5 s.m. (inclusive), R+=Renda domiciliar *per capita* superior a 1,5 s.m.

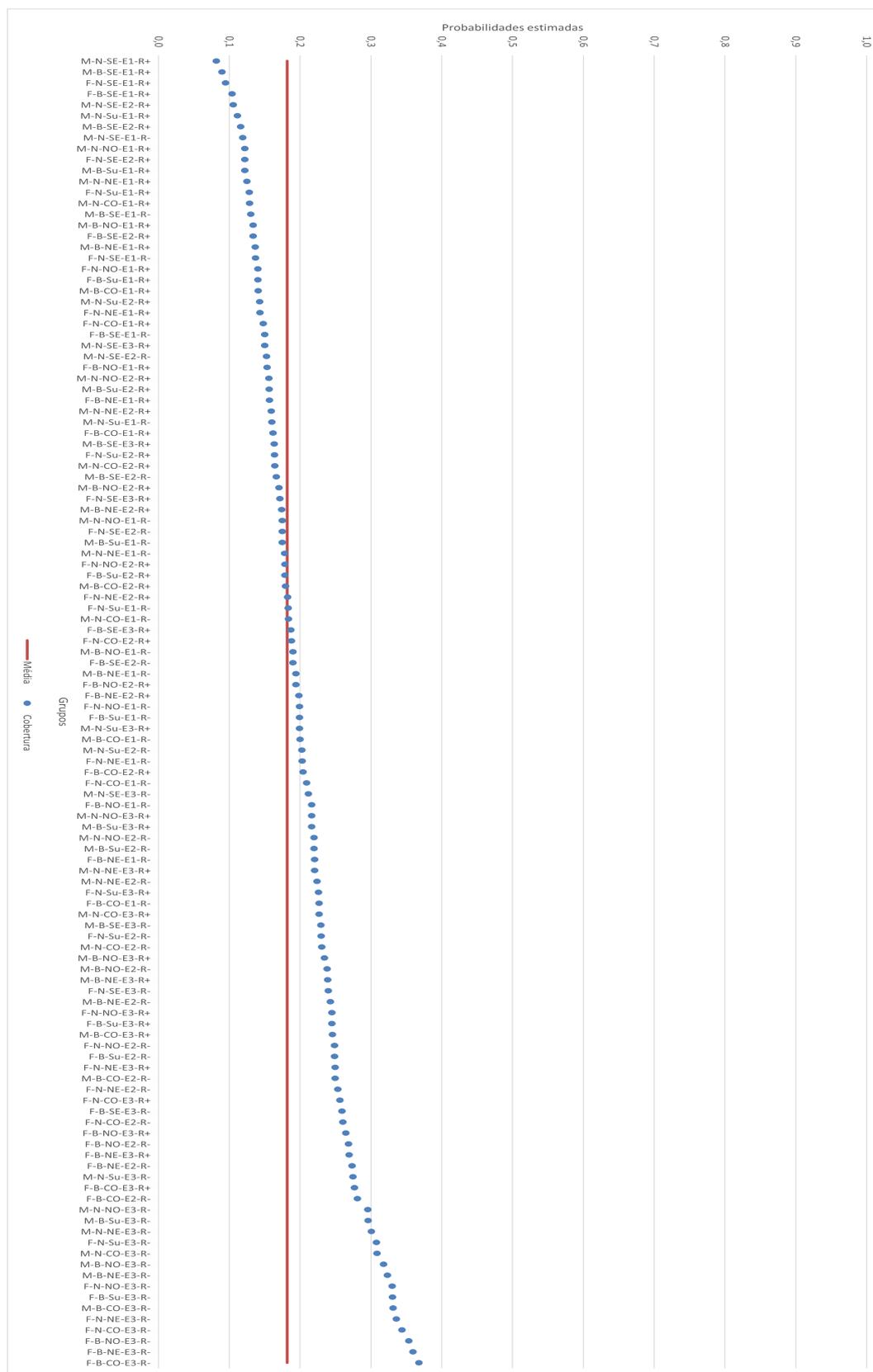


Figura 9: Probabilidades estimadas de ingresso em instituições de ensino superior públicas segundo os grupos

Fonte: Elaboração própria baseada em Carvalho e Waltenberg (2015). Microdados da PNADc 2019. Legenda: M=Sexo Masculino, F=Sexo Feminino, B=Branco ou Amarelo, N=Preto, Pardo ou Indígena, SE= Sudeste, NO= Norte, NE= Nordeste, Su= Sul, CO= Centro-Oeste, E1= Até Ensino Fundamental, E2= Ensino Médio, E3= Ensino Superior, R-=Renda domiciliar *per capita* de até 1,5 s.m. (inclusive), R+=Renda domiciliar *per capita* superior a 1,5 s.m.

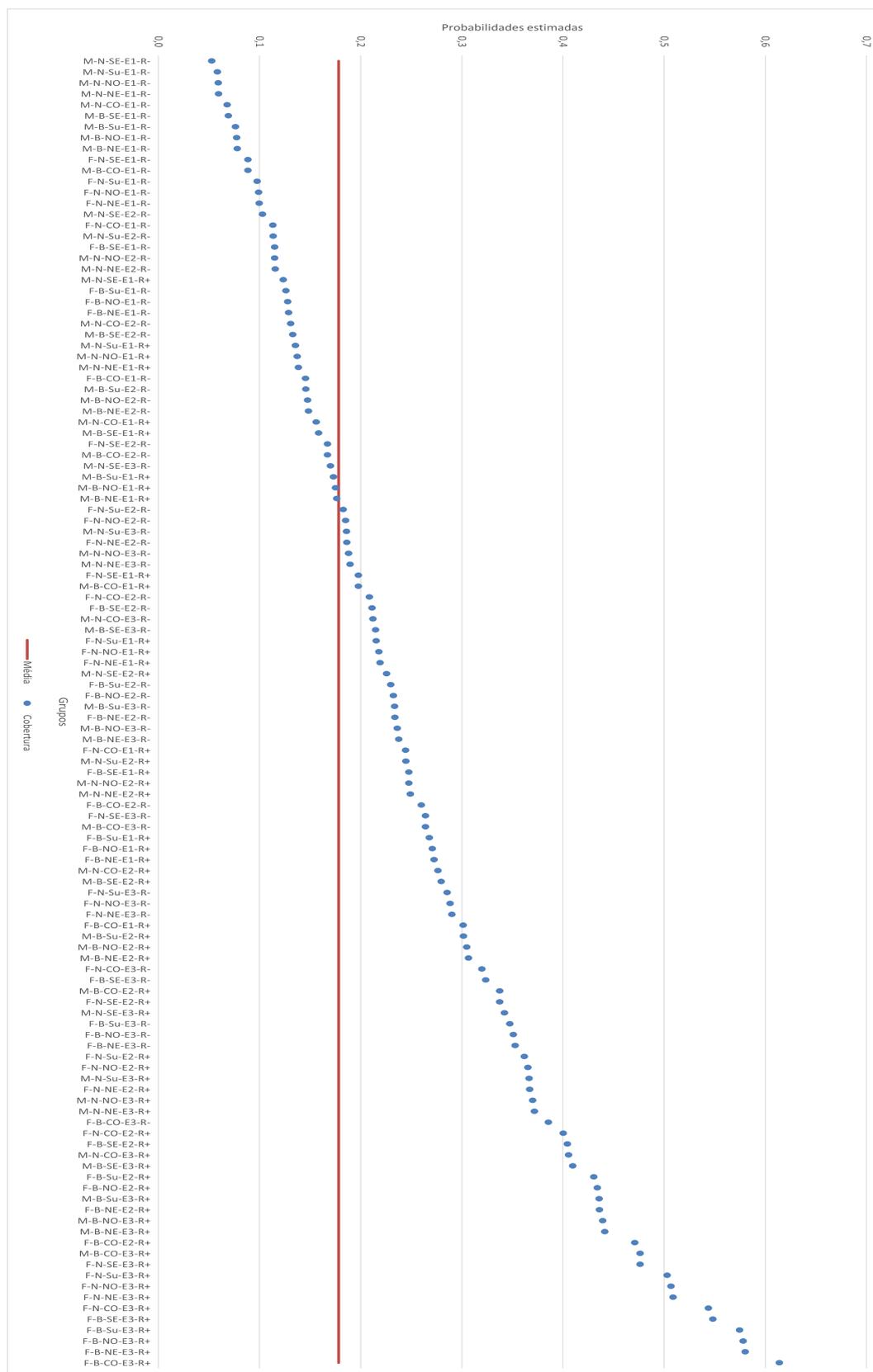


Figura 10: Probabilidades estimadas de ingresso em instituições de ensino superior privadas segundo os grupos

Fonte: Elaboração própria baseada em Carvalho e Waltenberg (2015). Microdados da PNADc 2019. Legenda: M=Sexo Masculino, F=Sexo Feminino, B=Branco ou Amarelo, N=Preto, Pardo ou Indígena, SE= Sudeste, NO= Norte, NE= Nordeste, Su= Sul, CO= Centro-Oeste, E1= Até Ensino Fundamental, E2= Ensino Médio, E3= Ensino Superior, R-=Renda domiciliar *per capita* de até 1,5 s.m. (inclusive), R+=Renda domiciliar *per capita* superior a 1,5 s.m.