



Universidade de Brasília  
Faculdade de Economia, Administração e Contabilidade  
Departamento de Economia

**Acesso ao Ensino Superior e Disponibilidade de  
Internet Domiciliar: Uma Análise de Impacto  
Probabilístico**

Vinícius de Oliveira Alves  
Brasília - DF

## Resumo

Este trabalho busca verificar, para o caso brasileiro, qual o impacto do acesso domiciliar à internet na probabilidade de uma pessoa estar cursando ensino superior. Para tanto, controlando para um conjunto de características da pessoa e do local em que vive, estima-se um modelo Logit. Com a restrição de indivíduos com entre 15 e 30 anos de idade, realizam-se as estimações a partir de dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) 2015. Como resultado, encontra-se impacto positivo e estatisticamente significativo do acesso a tal tecnologia na probabilidade de se cursar graduação. De fato, sem acesso à internet, essa probabilidade cai para menos da metade, mesmo considerando indivíduos cujas características os tornam mais propensos a realizar ensino superior.

**Palavras - chave:** ensino superior, internet, probabilidade

**Classificação JEL:** C35, I20

## Abstract

This study aims to evaluate, with respect to Brazil, the impact of household internet access on people's probability of going to college. In order to do so, controlling for several people's characteristics and where they live, a Logit model is estimated. With the restriction of individuals aged between 15 - 30, the estimations are done using data from National Household Sample Survey (PNAD) 2015. As a result, we find positive and statistically meaning impact of household internet access on the probabability of becoming an undergraduate student. Indeed, without internet acess, this probability falls to less than a half, even considering people whose characteristics turn then more likely to go to college

**Keywords:** higher education, internet, probability

**JEL classification:** C35, I20

# Sumário

<b>Introdução</b>	<b>4</b>
<b>Capítulo 1: Arcabouço Teórico - Teoria do Capital Humano</b>	<b>5</b>
<b>Capítulo 2: Revisão de Literatura</b>	<b>8</b>
<b>Capítulo 3: Exercício Econométrico</b>	<b>12</b>
3.1 Dados . . . . .	12
3.1.1 Variáveis utilizadas e derivadas . . . . .	12
3.1.2 Visualização dos dados . . . . .	12
3.2 Modelos de regressão . . . . .	20
3.3 Resultados . . . . .	24
<b>Conclusão</b>	<b>28</b>
<b>Referências</b>	<b>30</b>
<b>Anexos</b>	<b>32</b>

# Introdução

A educação, sobretudo a partir da segunda metade do século XX, tem despertado o interesse dos economistas de forma sistemática, de modo que a inclusão dessa variável nos modelos econômicos tornou-se recorrente. Nos anos 1960, essa tendência se inicia com o desenvolvimento da Teoria do Capital Humano, a partir da qual se consolida, por exemplo, a associação de maiores níveis de escolaridade a salários mais elevados. Dado esse impulso inicial, essa tendência adquiriu, também, outros contornos, de modo a abarcar temas associados, por exemplo, a distribuição pessoal de renda e ao crescimento econômico<sup>1</sup>.

Dessa forma, seja em razão de incentivos microeconômicos, seja em razão de desdobramentos macroeconômicos, é de se esperar que exista uma demanda por qualificação. De fato, no Brasil, em se tratando de ensino superior, o número de matriculados em cursos de graduação presencial cresceu, entre 2001 e 2015, cerca de 120% . Mais ainda, do lado da oferta, também no Brasil, o número de Instituições de Ensino Superior (IES) cresceu, no mesmo período, cerca de 70%<sup>2</sup>.

Paralelo a isso, o advento da internet tem representado, no período recente, uma expansão dos meios de comunicação<sup>3</sup>. Isso se reflete na educação, inclusive no ensino superior<sup>4</sup>. Tanto é que, no Brasil, enquanto, em 2002, havia menos de 50 cursos de graduação a distância, esse número totalizou, em 2015, mais de 1400. Nesse mesmo período, também no Brasil, o número de matriculados nessa modalidade de ensino cresceu mais de 3000%<sup>5</sup>. Apesar disso, o acesso à internet domiciliar não é universal neste país. Com efeito, em 2015, cerca de 42% dos brasileiros não dispunham de acesso a essa tecnologia em seu domicílio<sup>6</sup>.

Nesse contexto, objetiva-se, neste trabalho, responder a pergunta que se segue: considerando o Brasil e, controlando para um conjunto de características (do indivíduo e do local em que vive), pretende-se verificar qual o impacto do acesso domiciliar à internet na probabilidade de uma pessoa estar cursando ensino superior. Para tanto, a ideia é estimar modelos econométricos de resposta binária, de modo a ser possível averiguar se esse impacto é estatisticamente significativo.

Nesse sentido, além desta introdução e de uma conclusão, este trabalho estrutura-se em 3 capítulos. No primeiro, buscar-se-á uma base teórica para respaldar a análise subsequente da pergunta de pesquisa escolhida. Para isso, a ideia é adotar a Teoria do Capital Humano. Por sua vez, o segundo capítulo realizará uma revisão de literatura, explorando as abordagens mais usadas empiricamente na análise de problemas associados a pergunta em pauta. Já o terceiro capítulo consistirá na apresentação da base de dados e das variáveis utilizadas, bem como na exposição e discussão dos resultados da estimação dos modelos econométricos.

---

<sup>1</sup>Essa exposição sobre a relevância da educação enquanto objeto de pesquisa baseia-se em Ramos (2012)

<sup>2</sup>Esses percentuais foram calculados com base em sinopses estatísticas (sobre ensino superior) do INEP de 2001 e de 2015. Disponíveis em <http://inep.gov.br/sinopses-estatisticas-da-educacao-superior>

<sup>3</sup>Para mais detalhes do papel da internet em se tratando de meios de comunicação, veja Comparato (2000)

<sup>4</sup>Em capítulo posterior, discutiremos com mais detalhes o papel da internet no ensino superior

<sup>5</sup>Esse percentual, assim os totais de cursos de graduação a distância, foram calculados com base em sinopses estatísticas do INEP de 2002 e de 2015. Disponíveis em <http://inep.gov.br/sinopses-estatisticas-da-educacao-superior>

<sup>6</sup>Esse percentual foi calculado com base em dados da PNAD 2015. Nesse cálculo, desconsideramos indivíduos cuja renda domiciliar era, na amostra da PNAD, igual a 0

## Capítulo 1: Arcabouço Teórico

Segundo o modelo neoclássico, sabemos que os salários são determinados pela produtividade marginal do fator de produção trabalho. Por sua vez, essa produtividade, segundo a Teoria do Capital Humano, deriva de duas fontes. A primeira, em verdade, fora do escopo da teoria econômica, diz respeito a características genéticas que podem interferir na inteligência e na saúde do indivíduo. Já a segunda se refere a habilidades adquiridas, cuja obtenção ocorre mediante a qualificação do agente. Ou seja, este buscaria educação com intuito de tornar-se mais produtivo e, assim, lograr maiores rendimentos.<sup>7</sup>

A exposição acima esclarece o benefício (maiores salários) associado a maiores níveis de educação sobre a ótica do indivíduo. Contudo, a decisão de se elevar o grau de instrução implica alocação de recursos e, destarte, envolve custos. Trata-se não só de dispêndios atrelados ao financiamento do ensino, como também de custos de oportunidade. Afinal, a decisão de se qualificar pode exigir adiamento da inserção da pessoa no mercado de trabalho e, portanto, abre-se mão de salários no presente. Além disso, há características e condições do indivíduo as quais podem constituir uma barreira a entrada na obtenção de instrução.<sup>8</sup>

Com efeito, o processo de qualificação pode ser visto como um investimento cuja viabilidade demanda benefícios superiores aos custos associados.<sup>9</sup> Nesse sentido, nosso estudo circunscreve a identificação de fatores que, em virtude de influenciarem esses custos, afetam a decisão da pessoa quanto a obtenção de educação. Começamos, então, pelo fator renda. Evidentemente, em geral, recursos financeiros são necessários para o custeio, em algum grau, dos estudos. Nesse contexto, em se tratando de ensino superior, a renda familiar é especialmente importante para a inserção do público jovem nesse grau de ensino.

Conforme esclarece Becker (1993), na verdade, a família tem influência, na obtenção de capital humano do indivíduo, que vai além de disponibilidade financeira. De fato, atributos como conhecimentos, habilidades, hábitos e valores são influenciados pela estrutura familiar da pessoa. Contudo, nossa análise se restringe a enfatizar a relevância da questão monetária: pais de famílias mais pobres simplesmente podem não dispor de condições para custear os estudos de seus filhos.<sup>10</sup> Logo, a renda domiciliar do jovem apresenta, em tese, relação direta com a propensão deste a cursar graduação.

A participação no mercado de trabalho também pode ser entendida como um aspecto relevante na obtenção de instrução. Por um lado, argumenta-se que a aquisição de educação superior é dispendiosa em termos de tempo alocado e, portanto, requer dedicação exclusiva. Por outro lado, indivíduos empregados elevam a renda do domicílio em que vivem e, destarte, podem contornar, no caso de serem jovens, a questão previamente explanada da dependência familiar no financiamento dos estudos. Além disso, esse suposto *trade off* entre trabalho e estudo (em razão de restrições de tempo disponível)

---

<sup>7</sup>Essa breve introdução a Teoria do Capital Humano baseia-se em Becker (1993) e em Ramos (2012)

<sup>8</sup>Essas características e condições serão explicitadas ao longo desta exposição. Além disso, cabe ressaltar que, neste trabalho, estamos interessados numa forma particular de capital humano, qual seja: ensino superior

<sup>9</sup>Essa analogia foi extraída de Becker (1993)

<sup>10</sup>Em princípio, essa insuficiência de recursos poderia ser solucionada via mercado de crédito. Isto é, a família ou o próprio estudante poderia tomar, nesse mercado, recursos emprestados, com os quais seria possível financiar os estudos. Contudo, existe uma assimetria de informação entre tomadores de empréstimo e credores. Estes, desprovidos de informação suficiente (sobretudo sobre o estudante), ofertam fundos diante de seleção adversa. Isso se reflete no custo de obtenção do crédito, o qual se torna, nesse sentido, inviável ao estudante como forma financiamento dos estudos. Para mais detalhes sobre informação imperfeita e seleção adversa, veja Akerlof (1970).

pode ter sido amenizado, na contemporaneidade, com advento da *internet*.<sup>11</sup>

O destaque até aqui conferido à parcela mais jovem da população não foi ao acaso.<sup>12</sup> Afinal, é nessa faixa etária que se espera maior propensão da pessoa a obter qualificação, inclusive ensino superior. A fim de facilitar o entendimento desse ponto, recordemos a analogia da aquisição de capital humano com a realização de um investimento. A efetivação deste depende, também, do horizonte temporal no qual é possível colher os rendimentos dessa operação. Nesse sentido, pessoas mais jovens, em comparação com aquelas em idade mais avançada, podem usufruir dos rendimentos (maiores salários) do investimento em capital humano por mais tempo.<sup>13</sup>

Ademais, a influência da idade também se manifesta na variável esforço. Pois, se se admite que, conforme se transcorrem os anos, o esforço para se lograr aprendizado é crescente, resulta que, em idades mais avançadas, o investimento em capital humano torna-se demasiado custoso e, por conseguinte, inviável. Esse comportamento crescente do esforço em relação à idade decorre, por exemplo, da ocorrência de redução, ao longo do tempo, da capacidade ou flexibilidade de se adquirir novos conhecimentos.<sup>14</sup>

A despeito de todos os fatores abordados acima, constante é o fato de que o incentivo ao investimento em capital humano depende da expectativa do retorno na forma de maiores salários. Não obstante, se esse retorno difere entre grupos, ainda que igualmente produtivos, ou seja, se há discriminação, os incentivos à obtenção de instrução se alteram. Dessa forma, a ocorrência dessa distinção em razão de, por exemplo, sexo e/ou cor reduz os incentivos, à obtenção de educação, dos grupos afetados nesse sentido.<sup>15</sup>

As mulheres, que são um dos grupos afetados pela existência de discriminação, possuem um condicionante adicional: o fato de possuírem, ou não, filhos. Isso se revela não só em razão dos custos financeiros envolvidos na maternidade, visto que o exercício desta implica também alocação considerável de tempo. Ou seja, tal qual no argumento utilizado previamente na suposta dicotomia entre trabalho e estudo, haveria, também, um possível *trade off* entre se dedicar aos cuidados de um filho e aos estudos.<sup>16</sup>

O penúltimo condicionante (à obtenção de educação) que será abordado neste estudo trata-se de desigualdades regionais. Isso se relaciona ao fato de os incentivos que o agente econômico possui para se educar dependem também do tamanho do mercado de trabalho. Com efeito, mercados de trabalho maiores, encontrados, por exemplo, em cidades metropolitanas, são mais atrativos, do ponto de vista da aquisição de capital humano, que mercados mais limitados, encontrados, por exemplo, em cidades periféricas, do interior.

Isso decorre do fato de mercados maiores serem, também, mais diversificados, e, portanto, requerem mais pessoas qualificadas e eficientes. Por conseguinte, o incentivo a se obter educação guarda relação direta com o tamanho do mercado. Por exemplo, no Brasil, essa comparação de mercados poderia ser feita entre algum município situado na região Sudeste e outro localizado na região Norte<sup>17</sup>

---

<sup>11</sup> Aprofundaremos, posteriormente, o argumento sobre o papel da *internet* nesse *trade off*

<sup>12</sup> Utilizamos, recorrentemente, o termo "jovem" de forma genérica. Contudo, na parte empírica deste trabalho, adotaremos uma definição mais precisa em termos de idade.

<sup>13</sup> O capítulo 4 de Becker (1993) apresenta mais detalhes sobre o papel da idade nesse sentido

<sup>14</sup> A referência para o argumento da variável esforço é o capítulo 5 de Ramos (2012)

<sup>15</sup> O capítulo 5 de Becker (1993) apresenta com mais cuidado os efeitos da discriminação nesse contexto

<sup>16</sup> Assume-se implicitamente que o custo de criação de um filho, seja em termos financeiros, seja em termos de tempo alocado, recai, em média, mais sobre mulheres que sobre homens. Na verdade, essa hipótese não é tão distante da realidade brasileira. Pois, segundo o Censo Demográfico de 2010, mais de um quarto das famílias com filho constituem-se de mulheres sem cônjuge

<sup>17</sup> A referência para a influência do tamanho do mercado é o capítulo 4 de Becker (1993)

Apresentamos, anteriormente, uma série de fatores consolidados na literatura. Feito isso, propomos, agora, um novo condicionante, qual seja, o acesso domiciliar à *internet*. De fato, no período recente, sabe-se que o uso da tecnologia de informação nos sistemas de ensino é uma realidade. Nesse sentido, os assim chamados cursos de graduação a distância se aproveitam dessa tecnologia utilizando a *internet* como insumo essencial.<sup>18</sup> Além disso, até mesmo em cursos presenciais, a bibliografia para estudo pode ser comumente adquirida digitalmente por meio da *internet*.<sup>19</sup> O uso desta, desse modo, propicia economia de tempo alocado nos estudos e, portanto, pode tornar mais factível a realização simultânea de trabalho e curso superior.

Na verdade, o acesso à *internet* se configura como um instrumento de economia de tempo e, concomitantemente, como uma barreira a entrada ao ensino superior. Afinal, no cenário atual, a realização de curso de graduação a distância é inviável ao indivíduo desprovido de acesso à *internet*. Mais ainda, esse indivíduo fica impossibilitado até mesmo de participar de processos seletivos para ingresso em curso superior. Por exemplo, a inscrição para o Exame Nacional do Ensino Médio (ENEM) de 2018 ocorreu exclusivamente via *internet*.<sup>20</sup>

Em suma, dissertamos sobre diversos fatores que, do ponto de vista teórico, influenciam a decisão do agente quanto a obtenção de educação. A princípio, introduzimos alguns dos fundamentos da Teoria do Capital Humano. Num segundo momento, apresentamos condicionantes, tais como renda e idade, consolidados na literatura dessa teoria. Em seguida, propusemos que o acesso à *internet* seria, também, um desses fatores. Nessa argumentação, concluímos que esse acesso facilita os estudos e, ao mesmo tempo, atua como barreira a entrada ao ensino superior.

---

<sup>18</sup>Uma discussão aprofundada sobre o uso da tecnologia de informação no ensino superior a distância encontra-se em Maia (2003)

<sup>19</sup>Um exemplo é a plataforma JSTOR (Journal Storage) que oferece acesso à artigos acadêmicos a universidades e bibliotecas de todo o mundo

<sup>20</sup>Conforme disposto no item 1.5 do edital do ENEM 2018

## Capítulo 2: Revisão de Literatura

O impacto positivo da renda na demanda por educação é, de fato, amplamente documentado na literatura. No Brasil, por exemplo, Costa, Becker e Pavão (2013) encontram, a partir de um modelo *Probit* Binomial, que a frequência escolar, em 2009, considerando jovens de 15 a 24 anos, seria função crescente da renda domiciliar *per capita*. Considerando essa mesma faixa etária juvenil, Cabanas, Komatsu e Menezes-Filho (2015) reforçam essa ideia. Para isso, concluem, por meio de um modelo *Logit* Multinomial, que o aumento da renda média dos adultos, observado ao longo deste século até 2012, eleva a probabilidade de os jovens se manterem estudando.

No resto do mundo, essa relação também foi verificada em diversos países. Por exemplo, Nguyen e Taylor (2003) encontram, por meio de um modelo *Logit* Multinomial, para os Estados Unidos, que a probabilidade de um recém graduado do ensino médio ingressar num curso superior de 4 anos é função crescente da renda familiar. Por sua vez, Jimenes e Velasco (2000), inferem relação similar para a Espanha<sup>21</sup>. Além disso, Nakata e Mosk (1987), numa análise do período pós segunda grande guerra, também encontram essa relação positiva entre renda familiar e demanda por educação superior para o Japão<sup>22</sup>

A respeito da influencia do exercício de uma profissão na realização dos estudos, restringimos nossa análise ao caso brasileiro. Em geral, a literatura reconhece, sobretudo em se tratando de jovens, a existência de todas as combinações possíveis entre essas duas atividades. Ou seja, só trabalhar, só estudar, trabalhar e estudar ou não trabalhar e não estudar. Dessa forma, uma conclusão recorrente é que a probabilidade de o jovem trabalhar é função decrescente da renda familiar, cujo aumento, por outro lado, torna o jovem mais propenso se manter estudando<sup>23</sup>. Com base nisso, inferimos que, conforme a renda familiar aumenta, a dedicação exclusiva aos estudos torna-se possível e preferível à realização simultânea dessas duas atividades.

No tocante a idade, os estudos aplicados sugerem, para o Brasil, que o avanço dela reduz a probabilidade de se frequentar ensino superior. Em particular para os jovens, isso ocorre em contraste com um aumento da probabilidade de inserção no mercado de trabalho conforme se avança a idade. Um exemplo concreto disso está no Ceará, onde Lima et al (2015) estimam o seguinte: em 2009, um ano adicional de idade aumenta em 9,5% a probabilidade de um jovem não estudar e trabalhar. Por sua vez, Costa, Becker e Pavão (2013) e Tillmann e Comim (2016) encontram resultado nessa direção para o Brasil como um todo.

Para o resto do mundo, as conclusões sobre a influência da idade variam de acordo com o país e com o período de tempo analisado. Por exemplo, Psacharopoulos (1982) encontra, para Portugal, que, em 1979, cada ano adicional de idade reduz a probabilidade de um jovem (recém saído do ensino médio) cursar ensino superior em mais de 10%. Por outro lado, Sá, Florax e Rietveld (2006) chegam a esta conclusão: na Holanda, entre 1998 e

---

<sup>21</sup>A direção da conclusão, relação positiva entre renda e escolha por se manter estudando, é a mesma. A diferença, no caso espanhol, é que os autores não restringem a amostra a recém saídos do ensino médio. Além disso, o modelo, embora também seja *Logit*, é binomial

<sup>22</sup>Esses autores definem a demanda por educação de forma distinta dos cientistas anteriormente citados. Ao invés de usarem uma variável binária, usam a assim chamada *application rate*. Trata-se de uma variável computada anualmente, em que se divide o fluxo de alunos que ingressaram no ensino superior pelo número de estudantes que concluíram ensino médio. Além disso, estimam regressões lineares, e não logísticas

<sup>23</sup>Exemplos disso são encontrados em Silva e Kassouf (2002), Costa, Becker e Pavão (2013) e em Komatsu e Menezes-Filho (2015)

1999, a idade do jovem recém graduado do ensino médio não tem impacto estatisticamente significativo na escolha por prestar ensino superior<sup>24</sup>

Um caso interessante, nos Estados Unidos, foi estudado por Cruce e Hillman (2012). Esses autores estudam a demanda por educação pós secundária de pessoas com entre 55 e 79 anos de idade.<sup>25</sup> Nesse sentido, concluem, por meio de um modelo *Logit*, que, em 2005, apesar de a idade reduzir a probabilidade de a pessoa exercer cursos relacionados a sua profissão, isso não acontece quando o curso escolhido relaciona-se ao gosto pessoal do indivíduo. Isso reforça esta hipótese: embora a busca por investimento em capital humano decaia com o avanço da idade, isso não necessariamente ocorre se o ensino for demandado como consumo<sup>26</sup>

Assim como do ponto de vista teórico, a literatura aplicada tem documentado a relevância da inclusão de variáveis as quais indicam existência de discriminação. No Brasil, em se tratando de cor, os estudos convergem para a seguinte conclusão: em geral, jovens negros têm menor probabilidade (se comparados a não negros) de continuarem adquirindo capital humano após o ensino médio. Além disso, sabe-se que o rendimento médio da população negra é inferior ao observado para a branca. Dessa forma, dada a influência das condições financeiras da família na aquisição do capital humano do jovem, percebe-se um ciclo vicioso na formação desse segmento populacional<sup>27</sup>

Por outro lado, nos Estados Unidos, a situação é diferente. Pois, de acordo com Nguyen e Taylor (2003), em geral, para esse país, os estudos mostram que jovens negros possuem maior probabilidade de continuar adquirindo capital humano após a conclusão do ensino médio. Isso não significa que não haja discriminação nos EUA. Na verdade, segundo esses autores, o investimento em capital humano, nesse caso, é também motivado pela possibilidade de se compensar a discriminação a qual esse grupo étnico espera enfrentar no mercado de trabalho.

Apesar disso, Wetzel, O'Toole e Peterson (1998), identificam, também nos Estados Unidos, uma questão financeira a qual afeta comparativamente mais esse grupo étnico. Para facilitar o entendimento disso, apresentamos a seguinte definição: o *yield* trata-se do percentual de estudantes efetivamente matriculados do total de alunos aceitos por uma instituição de ensino. Nesse sentido, o *yield* de jovens negros é mais sensível a variações no custo financeiro líquido de se exercer curso superior.<sup>28</sup> Ou seja, se, por exemplo, esse custo cair em 1000 dólares, espera-se que, em geral, o *yield* aumente. Contudo, esse aumento é percentualmente maior para negros comparativamente a brancos

Com relação a variável sexo, Costa, Becker e Pavão (2013) inferem o seguinte: jovens brasileiras têm maior probabilidade de se encontrarem estudando, se comparadas a indivíduos do sexo masculino da mesma faixa etária (entre 15 e 24 anos de idade). Essa

---

<sup>24</sup>Esses dois estudos derivaram suas conclusões com base em modelos Logit estimados. No caso do estudo para a Holanda, a idade só influencia, na verdade, a modalidade de educação superior a ser escolhida

<sup>25</sup>Ensino Pós Secundário significa, nesse caso, qualquer forma de educação formal cujo requisito de entrada seja a conclusão do ensino médio. Não se trata necessariamente de curso superior com diploma. De fato, segundo esses autores, tais cursos não são os mais procurados nessa faixa etária

<sup>26</sup>Essa hipótese da demanda por educação como consumo foi extraída de Ramos (2012)

<sup>27</sup>Essa última inferência (do ciclo vicioso) é explicada com mais profundidade em Silva e Kassouf (2002). Já a conclusão sobre a menor probabilidade do jovem negro foi encontrada, por exemplo, em Silva e Kassouf (2002), Komatsu e Menezes-Filho (2015), e em Tillmann e Comim (2016). Além disso, a comparação entre renda média (de acordo com a cor da pessoa) é ratificada, por exemplo, pelos dados da PNAD (2015)

<sup>28</sup>O termo custo líquido é justificado por esses autores pelo fato de se subtrair auxílios financeiros dos custos

conclusão também é obtida por Cabanas, Komatsu e Menezes-Filho (2015), os quais, a despeito disso, verificam que mulheres nessa faixa etária possuem maior probabilidade de estarem em situação na qual não trabalham e não estudam.<sup>29</sup>

Não obstante, esse último resultado pode apresentar um viés de definição da variável ocupação. Pois, Silva e Kassouf (2002), ao invés de considerarem somente as 4 combinações entre trabalho e estudo, admitem a existência de mais uma atividade, um trabalho diferenciado: "cuidar de afazeres domésticos". Dessa forma, trabalham com 6 combinações possíveis de ocupação.<sup>30</sup> Assim sendo, é possível que a direção, ou, pelo menos, a magnitude do resultado inferido por Cabanas, Komatsu e Menezes-Filho (2015) fosse diferente se essa possibilidade adicional de atividade fosse incluída na análise.

Ainda a respeito do sexo feminino, a literatura empírica aponta outra variável relevante para análise: o fato de se possuir, ou não, filho(s). Por exemplo, Tillmann e Comim (2016) encontram, para o Brasil, evidência favorável a hipótese segundo a qual a maternidade adolescente está associada a abandono da escola e a baixa escolaridade. Ou seja, jovens nessa condição tornam-se, de fato, menos propensas a buscar investimentos em capital humano.

Em outros países, especificamente no tocante a ensino superior, essa diferença na probabilidade entre homens e mulheres nem sempre é verificada. Por exemplo, na Alemanha, Spiess e Wrohlich (2010) não encontram, a partir de um modelo *Logit* Multinomial, diferenças estatisticamente significativa entre sexos na propensão de uma pessoa a ingressar numa Universidade. Similarmente, esse resultado também é encontrado, na Holanda, por Sá, Florax e Rietveld (2006).<sup>31</sup> Por outro lado, Albert (2000) conclui, por meio de um modelo *Logit* Binomial, que, na Espanha, mulheres têm probabilidade maior de cursarem graduação que homens. A explicação para isso, segundo essa autora, é análoga àquela apresentada anteriormente sobre a discriminação de negros nos EUA.

A literatura aplicada também identifica desigualdades regionais como relevantes na propensão de a pessoa obter capital humano. No Brasil, isso é verificado por meio de uma comparação entre as regiões brasileiras, de modo a se concluir o seguinte: em algumas regiões, o jovem está mais propenso a estar estudando comparativamente a outras.<sup>32</sup> Além disso, comparações são feitas entre pessoas as quais residem em área urbana e aquelas que moram em área rural. Por exemplo, Costa, Becker e Pavão (2013) concluem que jovens do sexo masculino os quais residem em área rural têm menor probabilidade de estar estudando, se comparados aqueles cuja moradia situa-se em área urbana.

Em outros países, estudos também verificam o impacto de desigualdades regionais. Os resultados, todavia, não necessariamente apontam na mesma direção do caso brasileiro. Por exemplo, Spiess e Wrohlich (2010) inferem que, nos estados do Sul da Alemanha, estudantes recém saídos do ensino médio têm maior probabilidade de cursarem ensino superior, se comparados com jovens do Centro - Oeste alemão. Contudo, esses autores não encontram diferença estatisticamente significativa entre indivíduos de áreas urbanas e rurais na propensão a ingressar em cursos de graduação<sup>33</sup>

---

<sup>29</sup>Recorde que, nessa literatura, é comum a análise das 4 combinações possíveis entre trabalho e estudo

<sup>30</sup>Ou seja, somente estudar, somente trabalhar, estudar e cuidar de afazeres domésticas, estudar e trabalhar, somente cuidar dos afazeres domésticos, e não trabalhar nem estudar

<sup>31</sup>A diferença, no caso da Holanda, é que há 2 modalidades de ensino superior (*University* e *Professional College*), Nesse sentido, só há diferença, entre homens e mulheres, na probabilidade de escolha entre uma dessas modalidades

<sup>32</sup>Exemplos disso são encontrados em Silva e Kassouf (2002) e em Costa, Becker e Pavão (2013)

<sup>33</sup>O Sul da Alemanha compreende os estados de Bavaiera e Baden-Wuerttemberg. Por sua vez, o Centro - Oeste, segundo esses autores, da Alemanha constitui-se dos estados Renânia do Norte Vestfália, Hesse

Por outro lado, na Holanda, Sá, Florax e Rietveld (2006) inferem o seguinte: recém saídos do ensino médio localizados em províncias rurais possuem maior probabilidade de ingressarem em curso superior que jovens residentes de províncias mais urbanizadas. Por sua vez, Nguyen e Taylor (2003) identificam uma diferenciação regional, porém no sentido de, em alguns estados, haver maior propensão de jovens cursarem cursos superior de 2 anos, em comparação com cursos de 4 anos.

Em síntese, discorremos sobre diversos trabalhos empíricos os quais apontam um conjunto de condicionantes para a obtenção de capital humano. Com exceção da variável trabalho, buscamos apresentar a evidência nacional e internacional, de modo a compará-las. Nesse processo, percebemos que nem sempre os resultados encontrados no Brasil são os mesmos verificados em outros países. Além disso, notamos que, na maioria dos estudos apresentados, utilizam-se modelos econométricos de resposta binária. Em particular, o modelo *Logit* é recorrentemente adotado.

# Capítulo 3: Exercício Econométrico

## 3.1 Dados

### 3.1.1 Variáveis utilizadas e derivadas

Os dados que norteiam este exercício empírico são oriundos da Pesquisa Nacional por Amostra de Domicílios (PNAD) 2015. Uma das modalidades de informação dessa base de dados se dá ao nível da pessoa, havendo, a princípio, 356904 observações. Como pretendemos investigar a propensão do indivíduo a estar cursando ensino superior, desconsideramos as pessoas que já cursaram pelo menos graduação. Além disso, restringimos nossa amostra a pessoas com idade entre 15 e 30 anos<sup>34</sup>. Feito isso, restaram 80.760 observações. Isso significa, após a aplicação dos pesos amostrais da PNAD 2015, um total de 45.436.767 de observações, a partir das quais coletamos e construímos algumas variáveis, que serão descritas a seguir.<sup>35</sup>

A partir de informações sobre escolaridade, construímos uma variável dummy para indivíduos que estão cursando ensino superior. Ademais, construímos dummy's para sexo, cor, trabalho, região brasileira e para mulheres com algum filho vivo<sup>36</sup>. Uma última variável binária (que foi construída a partir dos dados sobre domicílios dessa base) indica se a pessoa possui acesso domiciliar à internet. Ademais, foram coletadas variáveis que informam a idade e a renda domiciliar per capita da pessoa.

### 3.1.2 Visualização dos dados

A princípio, descreveremos como todos os estudantes de graduação estão distribuídos em termos de idade. Em seguida, em relação aos indivíduos com entre 15 e 30 anos, iremos comparar, quanto a um conjunto de características, aqueles que são estudantes de ensino superior com a população total, ou seja, com a totalidade da amostra ponderada descrita na subseção 2.1.

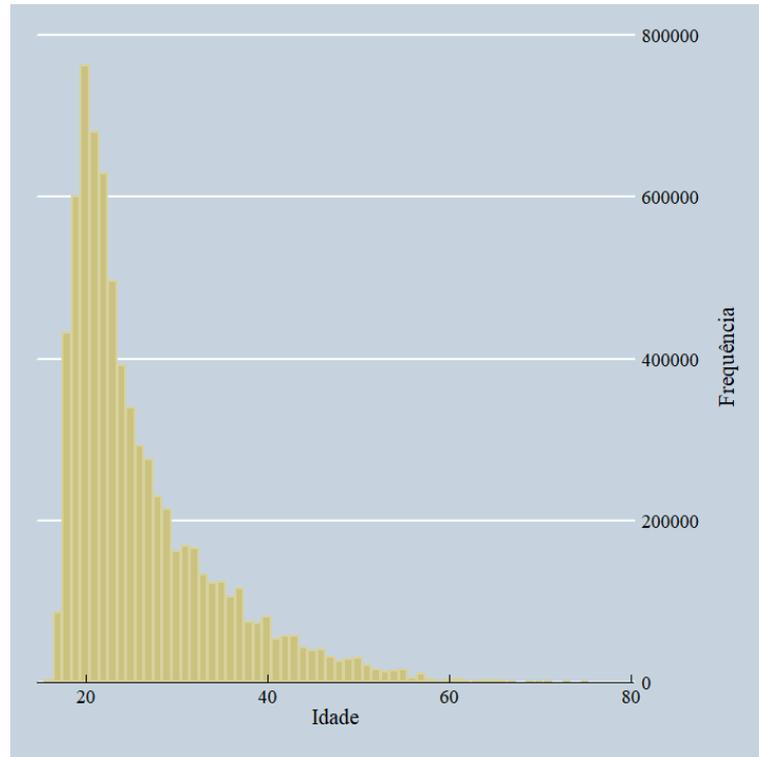
---

<sup>34</sup>O argumento, apresentado em capítulos anteriores, de que a idade reduz a probabilidade de se adquirir capital humano, em tese, dá respaldo a essa restrição. Além disso, a próxima seção deste capítulo deixará ainda mais clara o motivo dessa restrição

<sup>35</sup>Essas variáveis são justamente aquelas justificadas teoricamente e empiricamente, nos capítulos anteriores. Além disso, cabe notar que, nessa amostra, excluímos observações com renda domiciliar per capita igual a zero

<sup>36</sup>A dummy de cor indica se o indivíduo é branco. Por sua vez, a dummy de ocupação assume valor 1 se o indivíduo trabalhou ou se estava afastado do serviço temporariamente na data em que a pesquisa foi realizada.

Figura 1: Idade dos estudantes de ensino superior



Fonte: Elaboração própria com base em dados da PNAD (2015)

Em média, estudantes de graduação possuem 26 anos de idade, com um desvio padrão da ordem de 8,2.<sup>37</sup> Além disso, o histograma acima evidencia que existe uma faixa etária na qual a população de estudantes de graduação se concentra. De fato, cerca de 76,4% do total de estudantes tem até 30 anos. Isso sugere que, apesar de restringimos nossa análise para pessoas com entre 15 e 30 anos, nossa amostra ainda é representativa.

Tabela 1: População (15 - 30 anos) de acordo com o sexo

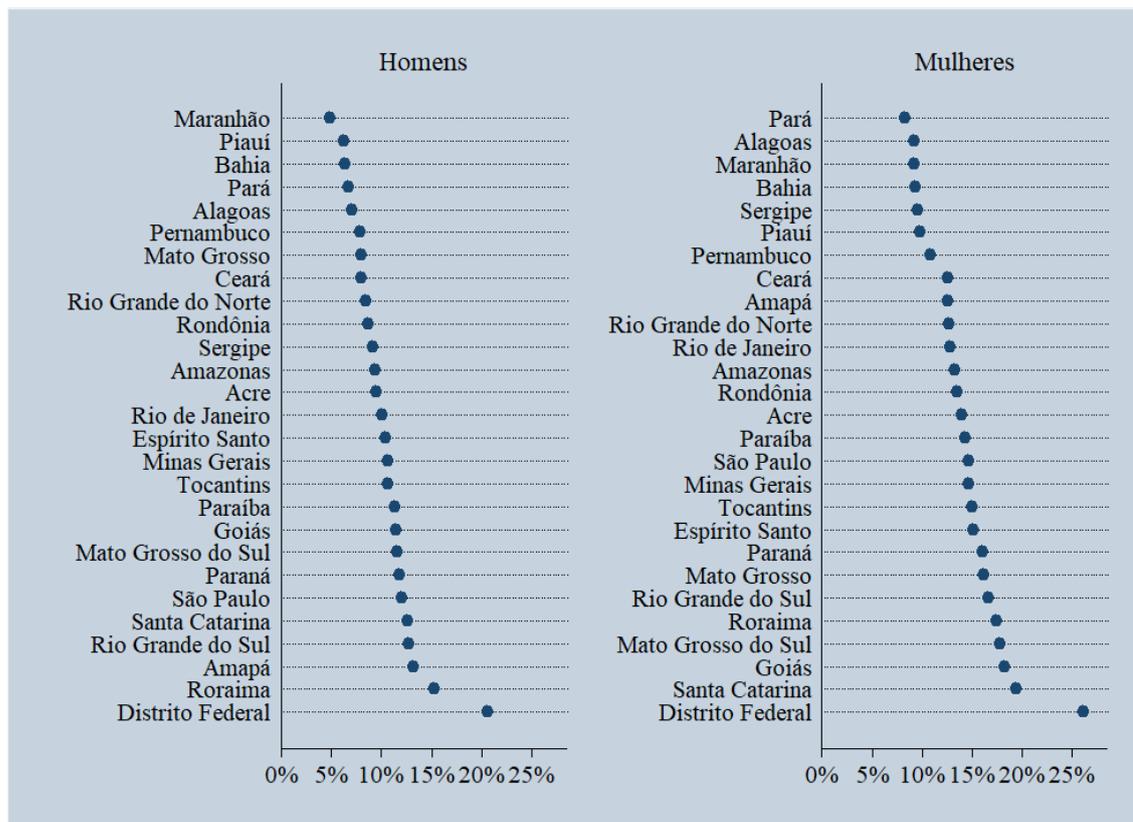
(a) Estudantes de ensino superior			(b) População total		
Indivíduo	Freq	Percent	Indivíduo	Freq	Percent
Homem	2.334.710	43,70%	Homem	23.268.092	51,21%
Mulher	3.008.101	56,30%	Mulher	22.168.675	48,79%
Total	5.342.811	100,00%	Total	45.436.767	100,00%

Fonte: Elaboração própria com base em dados da PNAD (2015)  
 Exclusive quem já concluiu graduação

Com relação a sexo, conforme a Tabela 1, percebe-se que, na população total, a proporção de homens é maior comparativamente a de mulheres. Por outro lado, o contrário ocorre na população de estudantes de graduação. Mais ainda, essa tabela nos permite deduzir que o percentual de mulheres (da população total) as quais cursam graduação (cerca de 14%) é superior aquele observado para homens (aproximadamente 10%). Essa comparação também pode ser feita a nível estadual, conforme realizado na figura a seguir.

<sup>37</sup>Essas estatísticas foram calculadas com base na amostra total de estudantes de graduação da PNAD (2015)

Figura 2: Percentual de pessoas (15 - 30 anos) que frequentam ensino superior (de acordo com o sexo)



Fonte: Elaboração própria com base em dados da PNAD (2015)  
 Exclusive quem já concluiu graduação

A Figura 2 esclarece, num primeiro momento, a existência, entre os estados brasileiros, de uma heterogeneidade no percentual de pessoas as quais realizam ensino superior. Em geral, para ambos sexos, os estados do Centro-Sul possuem maior proporção de indivíduos cursando ensino superior, se comparados aos estados da região Nordeste e Amazônia<sup>38</sup>. Mais especificamente, notamos que o percentual de homens (da população total) os quais cursam graduação varia, aproximadamente, de 5% a 21%, enquanto o percentual de mulheres oscila entre 8% e 26%. Ademais, percebe-se o seguinte: tal qual a nível nacional, com exceção do Amapá, em todos os outros estados brasileiros, o percentual de homens os quais realizam graduação é inferior aquele observado para mulheres.

<sup>38</sup>Estamos utilizando, nesse caso, a divisão regional geoeconômica proposta pelo Geógrafo Pedro Pinchas Geiger. Para mais detalhes sobre essa divisão regional, veja Boscariol, Renan Amabile et al (2017)

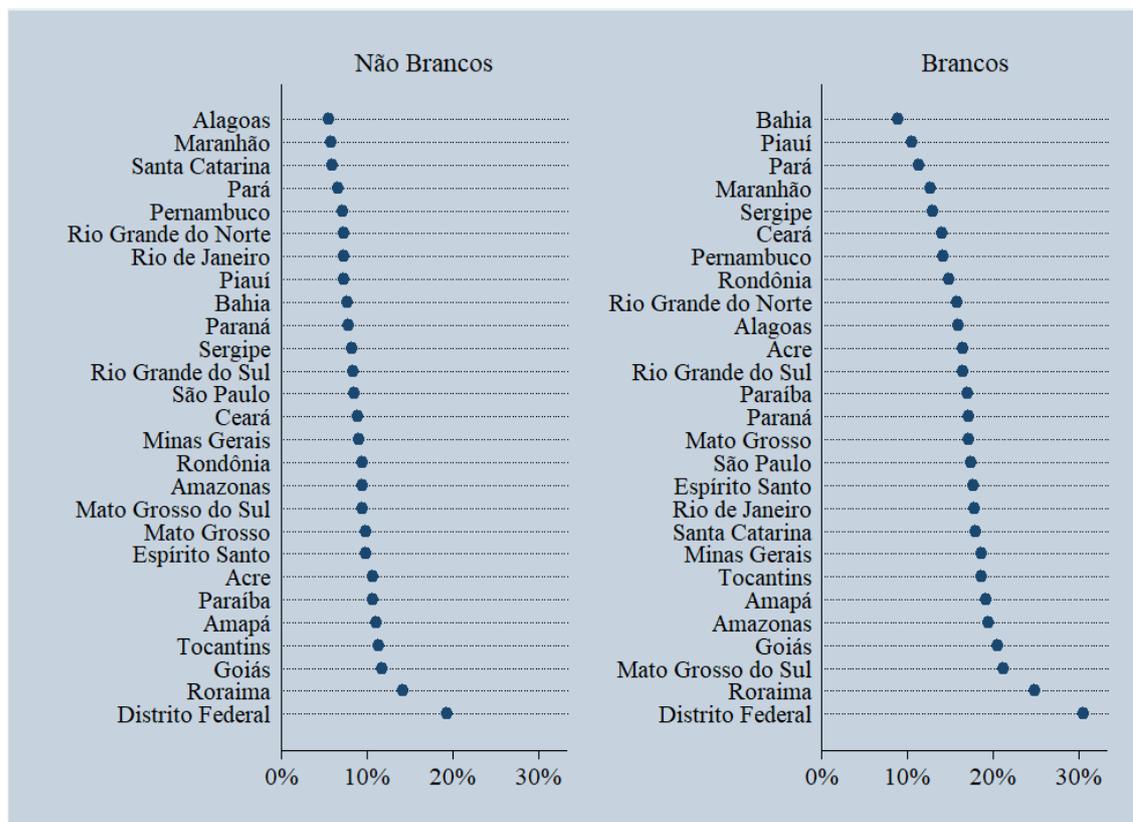
Tabela 2: População (15 - 30 anos) de acordo com a cor

(a) Estudantes de ensino superior			(b) População total		
Indivíduo	Freq	Percent	Indivíduo	Freq	Percent
Branco	3.021.239	56,55%	Branco	17.889.499	39,37%
Não Branco	2.321.572	43,45%	Não Branco	27.547.268	60,63%
Total	5.342.811	100,00%	Total	45.436.767	100,00%

Fonte: Elaboração própria com base em dados da PNAD (2015)  
 Exclusive quem já concluiu graduação

No que diz respeito a cor, segundo a Tabela 2, os não brancos são majoritários na população total. Por outro lado, a situação inverte-se na população de graduandos. Nesta, afinal, os não brancos são minoria. Além disso, essa tabela nos permite obter a proporção de indivíduos, dentro de cada grupo racial, que realizam graduação. De fato, percebemos que o percentual de brancos os quais cursam ensino superior (cerca de 17%) é superior aquele observado para não brancos (aproximadamente 8%). Essa última comparação pode ser estendida a nível estadual, consoante a figura abaixo:

Figura 3: Percentual de pessoas (15 - 30 anos) que frequentam ensino superior (de acordo com a cor)



Fonte: Elaboração própria com base em dados da PNAD (2015)  
 Exclusive quem já concluiu graduação

A Figura 3, assim como a Figura 2, revela expressiva variabilidade, entre os estados deste país, no percentual de pessoas as quais frequentam ensino superior. Essa variação,

además, oscila, dentro de cada grupo racial, em intervalos distintos. Por um lado, para não brancos, somente em 7 estados, o percentual em análise é superior a 10%. Por outro lado, em relação a brancos, somente na Bahia o percentual em pauta é inferior a 10%. Com efeito, em todos os estados brasileiros, essa proporção é maior para brancos, se comparados a não brancos.

Tabela 3: População (15 - 30 anos) de acordo com a ocupação

(a) Estudantes de ensino superior			(b) População total		
<b>Indivíduo</b>	<b>Freq</b>	<b>Percent</b>	<b>Indivíduo</b>	<b>Freq</b>	<b>Percent</b>
Não Trabalha	2.459.215	46,03%	Não Trabalha	22.535.180	49,60%
Trabalha	2.883.596	53,97%	Trabalha	22.901.587	50,40%
Total	5.342.811	100,00%	Total	45.436.767	100,00%

Fonte: Elaboração própria com base em dados da PNAD (2015)  
 Exclusive quem já concluiu graduação

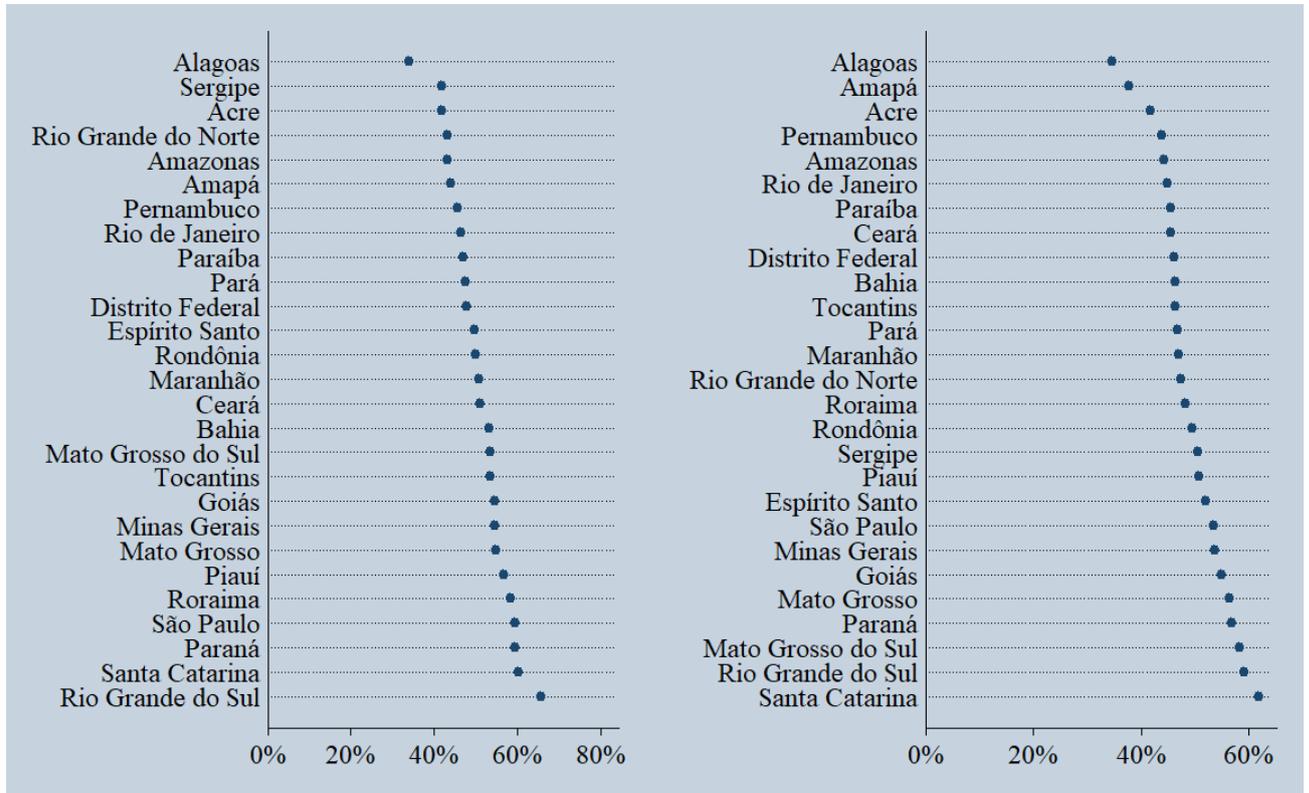
Por sua vez, em se tratando de Ocupação, a Tabela 3 sugere que a proporção de indivíduos que trabalham é maior no total de graduandos, se comparados a população total<sup>39</sup>. Na figura a seguir, desagregamos essa comparação a nível estadual

<sup>39</sup>Para fazer essa comparação, consideramos como pertencentes a População em Idade Ativa indivíduos a partir de 15 anos. Como nossa amostra é restrita a pessoas com idade entre 15 e 30 anos, esse critério é satisfeito

Figura 4: Percentual de pessoas (15 - 30 anos) que trabalham

(a) Estudantes de ensino superior

(b) População total



Fonte: Elaboração própria com base em dados da PNAD (2015)  
 Exclusive quem já concluiu graduação

Em geral, a Figura 4 evidencia que, nos estados brasileiros, o percentual de pessoas as quais trabalham varia, nos dois grupos analisados, numa faixa similar, embora esse intervalo de oscilação seja maior para os estudantes de graduação, se comparados a população total<sup>40</sup>. Contudo, a nível estadual, a conclusão a respeito de qual grupo possui percentual maior de indivíduos que trabalham nem sempre é a mesma do resultado obtido a nível nacional. Por exemplo, no Mato Grosso do Sul, o percentual de indivíduos que trabalham é maior na população total (cerca de 58%) comparativamente aos estudantes de graduação (cerca de 53%).

Tabela 4: População (15 - 30 anos) com acesso domiciliar à internet

(a) Estudantes de ensino superior

(b) População total

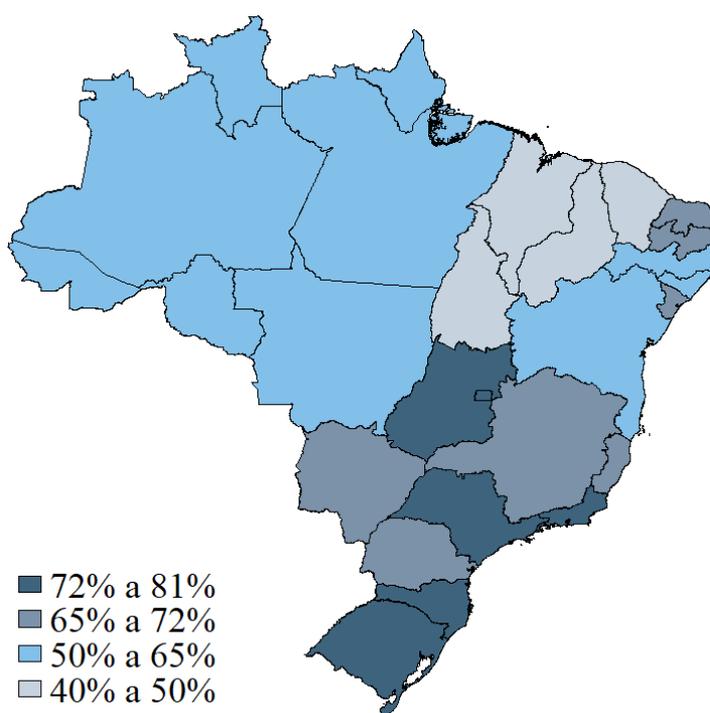
Indivíduo	Freq	Percent	Indivíduo	Freq	Percent
Sem internet	462.234	8,65%	Sem Internet	15.489.656	34,09%
Com internet	4.880.577	91,35%	Com Internet	29.947.111	65,91%
Total	5.342.811	100,00%	Total	45.436.767	100,00%

Fonte: Elaboração própria com base em dados da PNAD (2015)  
 Exclusive quem já concluiu graduação

<sup>40</sup>Especificamente, o percentual de estudantes de ensino superior varia de 33,98% a 65,68%, enquanto o percentual da população total oscila entre 34,65 e 61,74%

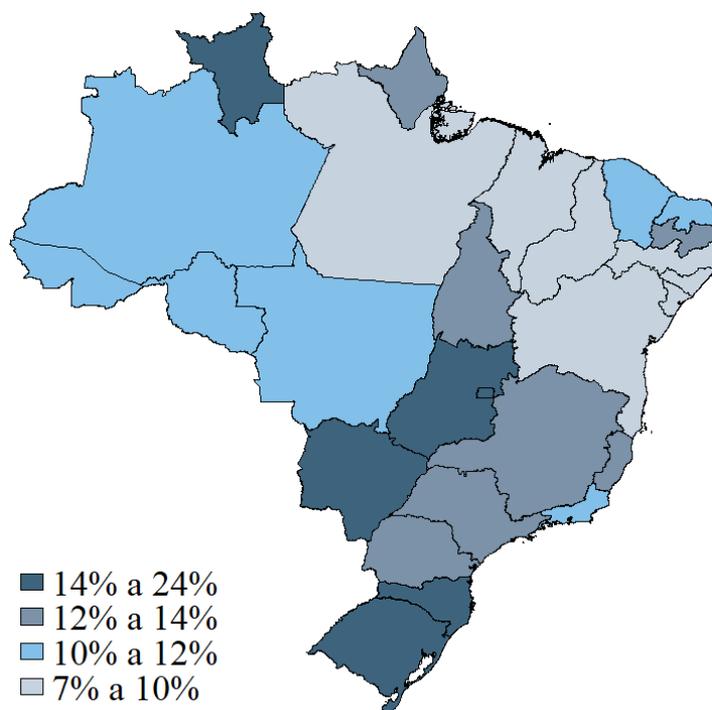
Finalmente, no tocante a acesso à internet, consoante a Tabela 4, nota-se uma diferença expressiva entre esses 2 grupos. Pois, embora, nas 2 populações, a maioria das pessoas possuam acesso à internet domiciliar, a proporção de indivíduos com acesso a essa tecnologia é substancialmente maior para estudantes de ensino superior do que para a população total. Trata-se, nesse sentido, de um indício da existência de alguma relação entre cursar graduação e possuir acesso domiciliar à internet. Estudemos, então, isso mais a fundo. Para tanto, faremos, por meio das figuras abaixo, uma comparação a nível estadual.

Figura 5: Percentual de pessoas que cursam ensino superior (idade: 15 - 30 anos; por UF)



Fonte: Elaboração própria com base em dados da PNAD (2015)  
Exclui quem já concluiu graduação

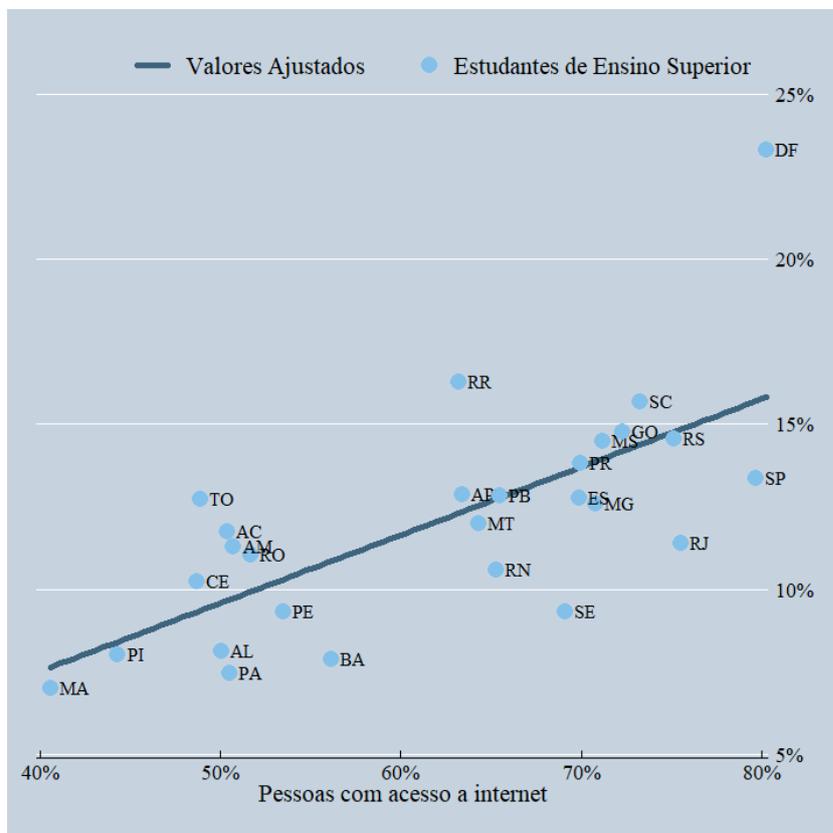
Figura 6: Percentual de pessoas com acesso à internet (idade: 15 - 30 anos; por UF)



Fonte: Elaboração própria com base em dados da PNAD (2015)  
Exclusive quem já concluiu graduação

Comparando as Figuras 5 e 6, inferimos, inicialmente, o seguinte: há, nos estados deste país, uma discrepância entre os percentuais de pessoas que cursam graduação e aquelas com acesso domiciliar à internet. Afinal, enquanto, no primeiro grupo, esse percentual varia entre aproximadamente 7% e 24%, no segundo grupo essa proporção oscila entre cerca de 40% e 81%. Apesar disso, notamos que, em geral, os estados nos quais o percentual de pessoas com acesso à internet é maior (cor mais escura nos mapas) são também aqueles com maior percentual de estudantes de graduação. Similarmente, os estados com menor percentual (cor mais clara nos mapas) de pessoas com acesso essa tecnologia são também aqueles com menor percentual de estudantes de ensino superior. Com efeito, essa correlação pode ser visualizada de forma mais explícita, consoante a figura a seguir.

Figura 7: Percentual de pessoas que cursam ensino superior e aquelas com acesso à internet (idade: 15 - 30 anos; por UF)



Fonte: Elaboração própria com base em dados da PNAD (2015)  
 Exclusive quem já concluiu graduação

A Figura 7 apresenta um gráfico de dispersão o qual sugere uma relação positiva entre o percentual de pessoas que cursam graduação e o percentual de indivíduos com acesso à internet. Adicionalmente, os valores ajustados formam uma reta de regressão (dessas duas variáveis), a qual enfatiza essa correlação. Esta, especificamente, é da ordem de 0,6989<sup>41</sup>. Evidentemente, toda essa análise descritiva, bem como a estatística de correlação servem como motivação para as próximas seções, nas quais avaliamos se o impacto do acesso à internet é, de fato, estatisticamente significativo na propensão a se exercer curso de graduação.

### 3.2 Modelos de regressão

Como pretendemos investigar a probabilidade de a pessoa estar cursando ensino superior, a variável dependente  $Y_i$  do modelo seria binária. Ou seja,  $Y_i$  assumiria valor 1 se o indivíduo  $i$  estiver frequentando curso de graduação, e 0 caso contrário. Por sua vez, a variável explicativa de interesse, que também é binária, assume valor 1 caso a pessoa possua acesso à internet, e 0 caso contrário.

Com o intuito de mitigar questões ligadas a viés de variável omitida, adicionamos, ao modelo, variáveis de controle, que foram descritas na subseção 2.1. Nesse sentido,

<sup>41</sup>Essa correlação é significativa a 1%

uma primeira tentativa de modelar esse problema será feita por meio de um Modelo de Probabilidade Linear (MPL), de acordo com a seguinte especificação:

$$Y_i = \beta_0 + \beta_1 internet_i + \beta_2 mulher_i + \beta_3 mulher * filho_i + \beta_4 idade_i + \beta_5 idade2 + \beta_6 trabalha_i + \beta_7 rendapercapita_i + \beta_8 branco_i + \beta_9 SU + \dots + \beta_{12} NO + u_i$$

$$i = 1, \dots, n$$

Tabela 5: Variáveis

Variável	Descrição
$Y_i$	dummy para pessoa que está cursando ensino superior
<i>internet</i>	dummy para pessoa com acesso domiciliar à internet
<i>mulher</i>	dummy para mulher
<i>mulher * filho</i>	dummy de interação para mulher com filho(s) vivo(s)
<i>idade2</i>	idade ao quadrado
<i>trabalha</i>	dummy para pessoa que trabalha
<i>rendapercapita</i>	renda per capita domiciliar
<i>branco</i>	dummy para pessoa branca
<i>SU..NO</i>	regiões brasileiras, exceto Centro Oeste (Categoria Base)

Se se admite que o valor esperado do erro do modelo  $u_i$ , condicional nas variáveis explicativas, é nulo, podemos, com efeito, estimar o modelo acima por meio do tradicional estimador Mínimos Quadrados Ordinários (MQO)<sup>42</sup>. Como o MPL apresenta heterocedasticidade por construção, a estimação (na seção seguinte) desse modelo ocorrerá com erros padrão robustos a heterocedasticidade.

Desse modo, o propósito da estimação desse modelo seria verificar se há uma diferença estatisticamente significativa (na probabilidade de estar cursando ensino superior) entre indivíduos com e sem acesso domiciliar à internet.

Contudo, sabe-se que o MPL possui algumas limitações, tais como a hipótese de a probabilidade de sucesso da variável dependente ser função linear das variáveis explicativas. Então, com o intuito de lidar com essas limitações, recorreremos a um modelo de regressão não linear, conhecido na literatura como *Logit*. De fato, conforme exposto no capítulo "Revisão de Literatura", esse modelo é recorrentemente adotado em situações nas quais a variável dependente é binária. Desse modo, trabalharemos com a seguinte especificação:

$$\ln\left(\frac{P_i}{1-P_i}\right) = Z_i + u_i \quad (1)$$

$$i = 1, \dots, n$$

Em que:

$$P_i = \frac{1}{1+e^{-Z}}$$

$$Z_i = \beta_0 + \beta_1 internet_i + \beta_2 mulher_i + \beta_3 mulher * filho_i + \beta_4 idade_i + \beta_5 idade2 + \beta_6 trabalha_i + \beta_7 rendapercapita_i + \beta_8 branco_i + \beta_9 SU + \dots + \beta_{12} NO$$

$\left(\frac{P_i}{1-P_i}\right)$  = razão de chances em favor da ocorrência da variável dependente  $Y_i$ , isto é, da variável binária para pessoa que cursa ensino superior

<sup>42</sup>Consideramos implicitamente como satisfeitas mais um conjunto de condições estatísticas (tais como ausência de multicolinearidade perfeita) padrão para estimação duma Regressão Linear

De modo análogo ao caso do MPL, consideramos como implicitamente satisfeitas um conjunto de hipóteses estatísticas, as quais tornam adequada a estimação do modelo Logit por Máxima Verossimilhança. Na verdade, como a interpretação da magnitude dos coeficientes de (1) não é muito intuitiva, iremos estimar, também, outra equação na qual simplesmente aplicamos a função exponencial em ambos lados de (1), de modo a obter:

$$\frac{P_i}{1-P_i} = e^{Z_i+u_i} \quad (2)$$

A partir de (2), temos o impacto marginal de uma variável explicativa qualquer  $x_k$  (dado por  $e^{\beta_k \Delta x_k}$ ) sobre a Razão de Chances ( $\frac{P_i}{1-P_i}$ ). Em particular, se  $x_k$  for binária, esse impacto medirá uma relação entre  $P_i$  e  $(1 - P_i)$ , de modo a comparar o grupo para qual  $x_k = 1$  com o grupo para o qual  $x_k = 0$ . Por exemplo, para a variável *internet*, se  $e^{\beta_1} = 3$ , interpretamos que pessoas com acesso à internet domiciliar têm 3 vezes mais chance de estar cursando ensino superior do que pessoas sem acesso a tal tecnologia (tudo o mais mantido constante)<sup>43</sup>.

Antes de partir para as estimações dos modelos, resta uma última análise, qual seja: os sinais esperados dos coeficientes das regressões. Na verdade, a expectativa desse sinal está implicitamente justificada pelos capítulos "Arcabouço Teórico" e "Revisão de Literatura", nos quais se baseia toda a argumentação que se segue. Embora distintos, os modelos MPL e Logit devem, em tese, apresentar coeficientes com o mesmo sinal. Por isso, a análise a seguir não é dividida entre essas 2 especificações econométricas.

Em se tratando de renda *per capita*, espera-se um coeficiente positivo, em função dos custos financeiros associados a realização de curso superior. Por sua vez, quanto a variável *trabalha*, o sinal esperado não é claro. Isso ocorre em virtude de o argumento da necessidade de dedicação exclusiva aos estudos sugerir sinal negativo, ao passo que ideia do trabalho prover fonte adicional de renda sugere sinal positivo<sup>44</sup>.

Já no tocante as variáveis *idade* e *idade2*, espera-se, respectivamente sinais positivo e negativo. Ou seja, supõem-se uma relação côncava, de modo que, para idades iniciais, próximas de 15 anos, o avanço dos anos aumente a chance de a pessoa cursar graduação. Afinal, em geral, nessa faixa etária o indivíduo está na fase de conclusão do ensino médio, o que o tornaria, *ceteris paribus*, mais propenso a seguir para a faculdade. Contudo, em razão do argumento de o custo de se adquirir capital humano ser função crescente da idade, argumenta-se que, a partir de uma certa idade, o avanço desta passe a diminuir a probabilidade de se cursar graduação.

Quanto as variáveis indicadoras de discriminação, esperamos sinal positivo para a variável *branco*. Isso significa que a variável dummy complementar, ou seja, aquela que indicaria pessoa não branca, assumiria sinal negativo. Isso é compatível tanto com os achados da literatura empírica para o Brasil, como também com o argumento sobre o desincentivo a obtenção de capital humano ocasionado pelo expectativa de discriminação no mercado de trabalho.<sup>45</sup>

Em tese, esse mesmo argumento teórico valeria para a variável *mulher*, de modo que seria de se esperar um coeficiente com sinal negativo. Contudo, a literatura aplicada tem encontrado, para o Brasil, sinal positivo para essa variável. Uma possível explicação

<sup>43</sup>Para mais detalhes sobre o modelo Logit, veja Gujarati e Porter (2011).

<sup>44</sup>Veja o capítulo "Arcabouço Teórico" para mais detalhes sobre essa discussão

<sup>45</sup>Apresentamos, na verdade, no capítulo "Revisão de Literatura", trabalhos aplicados que encontram sinal negativo para a dummy indicadora de pessoa negra. Contudo, os negros, usualmente definidos, nesses trabalhos, como aqueles que se declaram, nas pesquisas do IBGE, de cor preta ou parda, constituem a maioria dos indivíduos não brancos (segundo, por exemplo, a PNAD 2015). Por isso, as conclusões para negros e não brancos caminham na mesma direção.

para isso é a ideia segundo a qual o investimento em capital humano pode, na verdade, compensar a expectativa de discriminação no mercado de trabalho.

A respeito de desigualdades regionais, esperamos o seguinte: regiões com mercados de trabalho mais desenvolvidos devem apresentar sinal positivo em comparação com regiões dotadas de mercados de trabalho menos desenvolvidos. Finalmente, com relação a variável *internet*, conjectura-se sinal positivo, seja em razão de essa tecnologia propiciar economia de tempo nos estudos, seja devido ela funcionar como barreira a entrada, no cenário atual, a realização de cursos de graduação.

### 3.3 Resultados

Tabela 6: Regressões estimadas

	<i>Variável Dependente</i>		
	(MPL)	Ensino superior (Logit)	(RC)
internet	0.0818*** (0.0001)	1.4942*** (0.0017)	4.4557*** (0.0074)
mulher	0.0827*** (0.0001)	0.8121*** (0.0011)	2.2526*** (0.0025)
branco	0.0450*** (0.0001)	0.5062*** (0.0011)	1.6589*** (0.0018)
trab	-0.0382*** (0.0001)	-0.3882*** (0.0011)	0.6783*** (0.0008)
mulher_filho	-0.1338*** (0.0001)	-1.5103*** (0.0019)	0.2208*** (0.0004)
sudeste	-0.0354*** (0.0002)	-0.3827*** (0.0019)	0.6820*** (0.0013)
sul	-0.0314*** (0.0002)	-0.3267*** (0.0022)	0.7213*** (0.0016)
nordeste	-0.0103*** (0.0002)	-0.1722*** (0.0020)	0.8418*** (0.0017)
norte	0.0001 (0.0002)	-0.0041 (0.0025)	0.9960 (0.0025)
idade	0.1575*** (0.0001)	2.3539*** (0.0017)	10.5263*** (0.0182)
idade2	-0.0033*** (0.0000)	-0.0498*** (0.0000)	0.9514*** (0.0000)
rend_pcapita	0.0001*** (0.0000)	0.0006*** (0.0000)	1.0006*** (0.0000)
Constant	-1.7716*** (0.0010)	-30.6769*** (0.0197)	0.0000*** (0.0000)
$R^2$	0.17		
Pseudo $R^2$		0.24	0.24
N	45436767.00	45436767.00	45436767.00

Fonte: Elaboração própria

$p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Erros padrão entre parêntesis

RC = Razão de Chances =  $\frac{P_i}{1-P_i}$

A tabela acima resume o resultado da estimação dos modelos propostos, nos quais a maioria dos coeficientes estimados são estatisticamente significativos a 1%. Nesse sentido, procedemos nossa exposição da seguinte forma: para cada variável explicativa, interpretamos os coeficientes do MPL, bem como os do Logit. Neste, na verdade, interpretamos

a Razão de Chance (RC). Em particular, em relação a variável de interesse, realizamos, também, um exercício de previsão.

### **Renda domiciliar *per capita***

Embora positivo e estatisticamente significativo, a magnitude do coeficiente da variável *renda per capita* é, no MPL estimado, pequena do ponto de vista econômico, em função da argumentação teórica e da evidência empírica apresentada nesse sentido. Similarmente, esse coeficiente também apresenta magnitude pequena no caso modelo Logit estimado.

### **Trabalho**

O MPL estimado sugere que, *ceteris paribus*, em média, indivíduos os quais trabalham possuem probabilidade de cursar ensino superior cerca de 3,8% menor, se comparados aqueles que não trabalham. Já o modelo Logit estimado prediz que, tudo o mais mantido constante, em média, pessoas as quais exercem alguma profissão possuem chance (de cursar graduação) cerca de 1,5 vezes menor em comparação com aquelas as quais não trabalham <sup>46</sup>.

### **Idade**

No MPL, os sinais dos coeficientes *idade* e *idade2* sugerem uma relação côncava entre idade e a probabilidade em pauta. Similarmente, o modelo Logit estimado também sugere uma relação côncava, porém entre idade e a razão de chance em favor de se cursar ensino superior. Em ambos modelos, tudo o mais mantido constante, a idade ótima estimada para se cursar graduação é de aproximadamente 24 anos <sup>47</sup>. Ou seja, após essa idade, no MPL estimado, o impacto marginal dessa variável passa a reduzir a probabilidade de se cursar graduação. De forma análoga, no modelo Logit estimado, após 24 anos, um ano adicional de idade passa a reduzir a razão de chance em pauta.

### ***Dummy* para mulher e *dummy* de interação entre mulher e possuir filho**

A partir do MPL estimado, inferimos o seguinte: em média, tudo o mais mantido constante, mulheres têm probabilidade de cursar ensino superior cerca de 8,3% maior em relação a homens. Por sua vez, o modelo Logit estimado sugere que, *ceteris paribus*, em média, mulheres possuem cerca de 2,3 vezes mais chances de cursarem graduação, se comparadas a homens.

Por outro lado, mulheres com filho, em comparação com aquelas sem filho, possuem, *ceteris paribus*, em média, segundo o MPL estimado, probabilidade (de cursar ensino superior) cerca de 13,4% menor. Já o modelo Logit estimado sugere que mulheres sem

<sup>46</sup> O coeficiente da *dummy trabalha* é menor que 1 (0,6783). Dessa forma, para facilitar a interpretação do resultado, apresentamos o coeficiente da variável *dummy* inversa, ou seja, da variável "não trabalha", que assumiria valor 1 se o indivíduo não trabalha, e 0 caso contrário. Para obtê-lo, basta fazer o inverso do coeficiente da variável *trabalha*. De fato,  $0,6783^{(-1)} = 1,4743$

<sup>47</sup> Para os 2 modelos estimados, essa idade (*id*) foi encontrada a partir do instrumental de cálculo diferencial. No MPL, note que  $\frac{\partial Y}{\partial id} = 0 = \beta_{id} + 2\beta_{id2}idade \Rightarrow idade^* = \frac{\beta_{id}}{-2\beta_{id2}}$  (3) Substituindo os coeficientes do MPL estimado em (3), temos:  $idade^* = \frac{0,1575}{2x0,0033} = 23,86$

Similarmente, no modelo Logit, temos:  $\frac{\partial L}{\partial id} = 0 = \beta_{id} + 2\beta_{id2}idade \Rightarrow idade^* = \frac{\beta_{id}}{-2\beta_{id2}}$  (4) Substituindo os coeficientes do modelo Logit estimado em (4), temos:  $idade^* = \frac{2,3539}{2x0,0498} = 23,63$

filho possuem, tudo o mais mantido constante, em média, chance (de cursar graduação) cerca de 4,5 vezes maior, se comparadas a mulheres com filho <sup>48</sup>

### ***Dummy para cor***

O MPL estimado sugere que, *ceteris paribus*, pessoas brancas têm, em média, probabilidade de frequentar ensino superior cerca de 4,5% maior comparativamente a não brancos. Por sua vez, a partir do modelo Logit estimado, inferimos o seguinte: indivíduos brancos possuem chance (de cursar graduação) cerca de 1,7 vezes maior em relação a não brancos.

### **Regiões Brasileiras**

Se desconsiderarmos a região Norte (única cujo coeficiente não é estatisticamente significativo a 1%), podemos realizar comparações. De fato, no MPL, como os outros coeficientes das dummies de região são todos negativos, concluímos que, *ceteris paribus*, o Centro - Oeste (categoria base) é a região na qual, em média, uma pessoa tem a maior probabilidade de estar cursando ensino superior. Similarmente, o modelo Logit sugere que, tudo o mais mantido constante, o Centro - Oeste trata-se da região na qual há, em média, a maior chance estimada de se cursar graduação.

### **Acesso domiciliar à *internet***

No MPL, o coeficiente estimado da variável de interesse indica que pessoas com acesso domiciliar à internet possuem, em média, uma probabilidade de estar cursando graduação cerca de 8,2% maior do que a estimada para pessoas sem acesso a tal tecnologia (tudo o mais mantido constante). Por sua vez, no modelo Logit, inferimos o seguinte: pessoas com acesso domiciliar à internet possuem chance estimada de estar cursando graduação cerca de 4,5 vezes maior do que indivíduos sem acesso a tal tecnologia (tudo o mais mantido constante).

Dado esse conjunto de resultados, iremos, por fim, executar um exercício de previsão. Em essência, iremos comparar, quanto a probabilidade prevista (pelo modelo Logit) de se cursar graduação, indivíduos (entre as regiões brasileiras) com e sem acesso a internet. Nesse sentido, como essa probabilidade depende da magnitude das variáveis explicativas, estas precisam assumir um determinado valor, de modo que seja possível realizar essa previsão. Para tanto, adotamos o seguinte critério: escolhemos valores, para os regressores, tais que a razão de chance em pauta  $\left(\frac{P_i}{1-P_i}\right)$  seja a maior possível. Essa ideia será esclarecida a seguir.

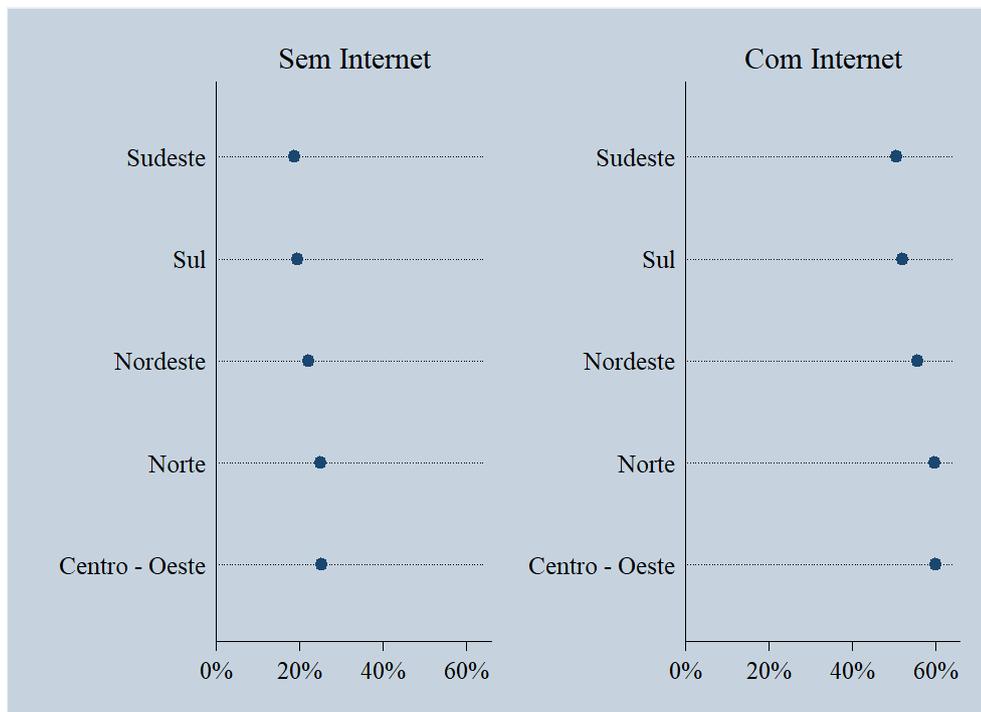
No caso das variáveis binárias, isso significa escolher a categoria que apresenta, em média, maior razão de chance estimada. Por isso, trabalhamos com mulheres brancas (sem filho) que não trabalham. Já no caso da variável idade, nosso critério implica escolher a idade após a qual o impacto marginal dessa variável sobre a razão de chance torna-se negativo. Isto é, trabalhamos com a idade ótima estimada, 24 anos. Excepcionalmente, para a renda *per capita*, adotamos um critério distinto. Pois, como essa variável depende linearmente do logaritmo da razão de chance, não faz sentido pensar num valor para

---

<sup>48</sup>De forma análoga ao caso da variável *trabalha*, calculamos o coeficiente da variável "mulher\*sem filho", que seria o inverso (com relação a possuir filho) da dummy de interação *mulher \* filho*. Ou seja, fizemos  $0,2208^{(-1)} = 4,5290$

renda cuja razão de chance associada seja a maior possível.<sup>49</sup> Por isso, simplesmente trabalhamos com a renda *per capita* média de nossa amostra.

Figura 8: Probabilidades previstas (por região) de se cursar ensino superior para mulheres (brancas, sem filho, com 24 anos, com renda domiciliar per capita média e que não trabalham):



Fonte: Elaboração própria com base em dados da PNAD (2015)

A figura acima traduz nossa proposta de exercício preditivo. Por exemplo, no Centro - Oeste, uma mulher branca (com 24 anos de idade e sem filho) a qual não trabalha e com renda domiciliar *per capita* de aproximadamente 789 reais (a valores correntes de 2015) pode ser vista de duas óticas<sup>50</sup>. Caso ela possua acesso domiciliar à *internet*, sua probabilidade estimada de cursar graduação é igual a cerca de 60%. Por outro lado, se ela não possuir acesso a essa tecnologia, sua probabilidade prevista cai para aproximadamente 25,1%.

Esse padrão comparativo, de fato, mantém-se para as demais regiões brasileiras<sup>51</sup>. Em geral, entre as regiões deste país, a probabilidade prevista de mulheres (com as características acima descritas) com acesso domiciliar à *internet* varia de aproximadamente 50,6% (Sudeste) a 60% (Centro - Oeste). Por sua vez, a probabilidade estimada de mulheres sem acesso a essa tecnologia oscila entre cerca de 18,6% (Sudeste) e 25,1% (Centro - Oeste)<sup>52</sup>.

<sup>49</sup>Recorde que, no modelo Logit proposto, na equação (1), especificamos a renda *per capita* como função linear do logaritmo da razão de chance

<sup>50</sup>789,2976 reais é justamente a renda domiciliar *per capita* média de nossa amostra

<sup>51</sup>Exceto a região Norte. Pois, a rigor, a probabilidade associada a essa região deve ser desconsiderada, pois o coeficiente estimado para tal região, no modelo Logit, não é estatisticamente significativo nem mesmo a 10%

<sup>52</sup>O Anexo A deste trabalho apresenta cada uma dessas probabilidades estimadas com seus respectivos intervalos de confiança a 95%

## Conclusão

O presente trabalho buscou verificar qual o impacto do acesso domiciliar à internet na probabilidade de uma pessoa estar cursando ensino superior. Para isso, o primeiro passo foi reconhecer que a decisão de se qualificar, inclusive em se tratando de ensino superior, depende, na verdade, de uma série de fatores, tais como renda e idade, consolidados na literatura teórica e na empírica. Por isso, com vistas a entender o papel de cada um desses condicionantes, apresentamos e discutimos uma série de análises teóricas e trabalhos aplicados nesse sentido.

Em seguida, para responder especificamente a pergunta de pesquisa escolhida, utilizamos modelos econométricos de resposta binária. Em particular, partimos de uma abordagem mais simples, um Modelo de Probabilidade Linear (MPL), para uma abordagem mais sofisticada e recorrentemente adotada na literatura, o modelo Logit. Dessa forma, além das variáveis dependente (cursar, ou não, ensino superior) e da de interesse (acesso domiciliar a internet), incluímos um conjunto de variáveis de controle, justificadas teoricamente e empiricamente. Todas essas variáveis foram obtidas e derivadas a partir de dados da PNAD 2015.

Os resultados das estimações dos modelos, em geral, para as variáveis de controle, foram consistentes com os achados da literatura aplicada. Inclusive, cabe notar, o MPL, a despeito de suas limitações, apresentou coeficientes com o mesmo sinal do modelo Logit, embora esses modelos não sejam diretamente comparáveis em termos da magnitude dos coeficientes. De toda forma, isso sugere que, pelo menos, os resultados desses 2 modelos apontaram na mesma direção.

Quanto a variável de interesse, nossas estimações sugerem, nos 2 modelos, impacto positivo e estatisticamente significativo do acesso à internet na probabilidade de a pessoa cursar ensino superior. De forma a reforçar essa inferência, executamos, a partir do modelo Logit, um exercício preditivo. Por meio desse exercício, concluímos que, mesmo considerando pessoas com as características que as tornam mais propensas a cursar ensino superior, essa probabilidade cai para menos da metade caso esses indivíduos não possuam acesso a tecnologia em pauta.

Esses resultados são favoráveis ao nosso argumento teórico, segundo o qual o acesso a essa tecnologia facilita os estudos e, ao mesmo tempo, atua como barreira a entrada a realização de ensino superior. Dado essa evidência inicial, trabalhos futuros podem aprofundar a análise da pergunta de pesquisa em pauta a partir de dois focos de atuação. O primeiro refere-se ao tratamento de problemas econométricos atrelados a endogeneidade. Esta se manifesta devido a este trabalho estar suscetível a questões de viés de variável omitida, particularmente a escolaridade dos pais.<sup>53</sup>

Em verdade, a não inclusão dessa variável justifica-se pela indisponibilidade dessa informação na PNAD 2015. Nesse sentido, trabalhos futuros poderiam buscar alternativas para a inclusão dessa variável. Outra possível fonte de endogeneidade neste estudo trata-se de viés de simultaneidade. Isso decorre da existência de causalidade reversa entre capital humano e renda.<sup>54</sup> Dessa forma, trabalhos posteriores poderiam lidar com esse problema

---

<sup>53</sup>O argumento para inclusão dessa variável relaciona-se a influência da família, além da renda, na decisão do agente quanto a obtenção de educação. Para mais detalhes sobre essa influência, veja o capítulo "Arcabouço Teórico" deste trabalho"

<sup>54</sup>Mais especificamente, essa causalidade reversa se daria, neste estudo, entre realizar ensino superior e renda domiciliar per capita. Uma análise mais aprofundada dessa forma de causalidade entre renda e capital humano encontra-se em Bonelli (2002), conforme citado por Cangussu, Salvato e Nakabashi (2010)

a partir da utilização de variáveis instrumentais.

Por sua vez, o segundo foco de atuação consiste na investigação mais a fundo do papel da internet nos incentivos a realização de curso superior. Por exemplo, seria possível verificar se esse impacto depende dos meios, tais como celular, tablet e computador pessoal, por meio dos quais se acessa essa tecnologia. Outra possibilidade de estudo seria averiguar se esse impacto é afetado pela qualidade, em termos de velocidade e estabilidade de conexão, da internet utilizada.

## Referências

- AKERLOF, George A. The market for ‘Lemons’: Quality uncertainty and the market mechanism. **Quarterly Journal of Economics**, v. 84, n. 3, p. 488-500, 1970.
- ALBERT, Cecilia. Higher education demand in Spain: The influence of labour market signals and family background. **Higher Education**, v. 40, n. 2, p. 147-162, 2000.
- BECKER, Gary S. **Human capital: A theoretical and empirical analysis, with special reference to education**. University of Chicago press, 1993.
- BOSCARIOL, Renan Amabile et al. **Região e regionalização no Brasil: uma análise segundo os resultados do Índice de Desenvolvimento Humano Municipal (IDHM)**. Territórios em números, p. 185-208, 2017. Disponível em: <<http://tiny.cc/t3my7y>>. Acesso em: 09 maio. 2019.
- CABANAS, Pedro; KOMATSU, Bruno; MENEZES FILHO, Naercio. **O crescimento da renda dos adultos e as escolhas dos jovens entre estudo e trabalho**. São Paulo: Insper: Centro de Políticas Públicas, 2015. Disponível em: <<http://tiny.cc/hcny7y>>. Acesso em: 06 maio. 2019.
- CANGUSSU, Ricardo Corrêa; SALVATO, Márcio Antônio; NAKABASHI, Luciano. Uma análise do capital humano sobre o nível de renda dos estados brasileiros: MRW versus Mincer. **Estudos Econômicos (São Paulo)**, v. 40, n. 1, p. 153-183, 2010.
- COMPARATO, Fábio Konder. A democratização dos meios de comunicação de massa. **Revista USP**, n. 48, p. 6-17, 2000.
- CRUCE, Ty M.; HILLMAN, Nicholas W. Preparing for the silver tsunami: The demand for higher education among older adults. **Research in Higher Education**, v. 53, n. 6, p. 593-613, 2012.
- DA COSTA, Jaqueline Severino; BECKER, Kalinca Léia; PAVÃO, Andressa Rodrigues. Influência da renda domiciliar per capita na alocação do tempo dos jovens no Brasil. **Revista de Economia**, v. 39, n. 1, p. 7-24, 2013.
- DE DIOS JIMÉNEZ, Juan; SALAS-VELASCO, Manuel. Modeling educational choices. A binomial logit model applied to the demand for higher education. **Higher Education**, v. 40, n. 3, p. 293-311, 2000.
- DE LIMA, Vinícius Misael Alves et al. Determinantes da inserção ocupacional dos jovens no Ceará. **Revista de Economia e Agronegócio**, v. 13, n. 1, 2, 3, 2015.
- GUJARATI, Damodar N.; PORTER, Dawn C. **Econometria Básica**. Amgh Editora, 2011.
- MAIA, Marta de Campos. **O uso da tecnologia de informação para a educação a distância no ensino superior**. Tese de Doutorado. FGV. 2003. Disponível em: <<http://tiny.cc/afny7y>>. Acesso em: 09 maio. 2019.
- NAKATA, Yoshi-fumi; MOSK, Carl. The demand for college education in postwar Japan. **Journal of Human Resources**, v.22, n.2, p. 377-404, 1987.
- NGUYEN, Anh Ngoc; TAYLOR, Jim. Post-high school choices: New evidence from a multinomial logit model. **Journal of Population Economics**, v. 16, n. 2, p. 287-306, 2003.
- PSACHAROPOULOS, George. An analysis of the determinants of the demand for upper secondary education in Portugal. **Economics of Education Review**, v. 2, n. 3, p. 233-251, 1982.
- RAMOS, Carlos Alberto. **Economia do Trabalho: modelos teóricos eo debate no Brasil**. Editora CRV, 2012.
- SÁ, Carla; FLORAX, Raymond JGM; RIETVELD, Piet. Does accessibility to higher

education matter? Choice behaviour of high school graduates in the Netherlands. **Spatial Economic Analysis**, v. 1, n. 2, p. 155-174, 2006.

SILVA, Nancy de Deus Vieira; KASSOUF, Ana Lúcia. **O trabalho e a escolaridade dos brasileiros jovens**. Anais do XIII Encontro da ABEP. Ouro Preto: ABEP, 2002. Disponível em:

<<https://bit.ly/2MzCU4L>>. Acesso em: 06 maio. 2019.

SPIESS, C. Katharina; WROHLICH, Katharina. Does distance determine who attends a university in Germany?. **Economics of Education Review**, v. 29, n. 3, p. 470-479, 2010.

TILLMANN, Eduardo; COMIM, Flavio. **Os determinantes da decisão entre estudo e trabalho dos jovens no Brasil e a geração Nem-Nem**. Pesquisa e Planejamento Econômico, 46(2), 2016. Disponível em:

<<http://tiny.cc/0mny7y>>. Acesso em: 06 maio. 2019.

WETZEL, James; O'TOOLE, Dennis; PETERSON, Steven. An analysis of student enrollment demand. **Economics of Education Review**, v. 17, n. 1, p. 47-54, 1998.

## Anexos

### Anexo A - Probabilidades previstas para se cursar graduação

Tabela 7: Mulheres com internet

Região	PE	IC a 95%
CO	60,06%	[59,96%; 60,17%]
SE	50,64%	[50,56%; 50,71%]
SU	52,03%	[51,94%; 52,12%]
NE	55,87%	[55,79%; 55,95%]
NO	59,97%	[59,86%; 60,14%]

Fonte: Elaboração própria com base em dados da PNAD (2015)

PE: Probabilidade Estimada

IC: intervalo de Confiança

Tabela 8: Mulheres sem internet

Região	PE	IC a 95%
CO	25,14%	[25,05%; 25,24%]
SE	18,64%	[18,58%; 18,70%]
SU	19,50%	[19,43%; 19,57%]
NE	22,04%	[21,97%; 22,11%]
NO	25,07%	[24,97%; 25,16%]

Fonte: Elaboração própria com base em dados da PNAD (2015)

PE: Probabilidade Estimada

IC: intervalo de Confiança