Pontifícia Universidade Católica do Rio De Janeiro
Departamento de Economia

Monografia de Final de Curso

O Impacto do Nível de Escolaridade dos Pais, Gênero e Outros Fatores Observáveis Sobre a Probabilidade de Conclusão de Curso Superior

Ana Beatriz Dantas Machado Trindade
Matrícula: 1511653

Professor Orientador: Mauricio Cortez Reis
Coordenador de Monografia: Márcio Garcia

Junho 2019

Pontifícia Universidade Católica do Rio De Janeiro
Departamento de Economia

Monografia de Final de Curso

O Impacto do Nível de Escolaridade dos Pais, Gênero e Outros Fatores Observáveis Sobre a Probabilidade de Conclusão de Curso Superior

Professor Orientador: Mauricio Cortez Reis
Coordenador de Monografia: Márcio Garcia

Junho 2019

"Declaro que o presente trabalho é de minha autoria e que não recorri para realizá-lo, a nenhuma forma de ajuda externa, exceto quando autorizado pelo professor tutor".

Ana Beatriz Dantas Machado Trindade (1511653)

“As opiniões expressas neste trabalho são de responsabilidade única e exclusiva do autor.”

AGRADECIMENTOS

À todos meus amigos e família que me apoiaram ao longo desses anos, com seus altos e baixos. Meu muito obrigada à cada pessoa que foi especial para mim. Seria injusto deixar de fora as tecnologias, da Wikipedia ao YouTube e Stack Overflow. Sem cada um de vocês o caminho seria muito mais difícil.

ÍNDICE DE FIGURAS

ÍNDICE DE TABELAS

ÍNDICE DE GRÁFICOS

SUMÁRIO

1 INTRODUÇÃO

Além de aumentar a eficiência da economia, quem possui um diploma de nível superior no Brasil ganha 140% a mais em média em relação a um profissional que estudou apenas até o ensino médio, segundo um estudo realizado pela OCDE (Organização para a Cooperação e Desenvolvimento Econômico). O Brasil lidera com a maior diferença entre 40 países analisados.

Ao mesmo tempo, a porcentagem de jovens que concluem o nível superior no país é extremamente pequena. Com base na PNAD Contínua de 2016, 84,7% dos brasileiros não possuía ensino superior completo, ou seja, cerca de 17% da população ganha, em média, 140% a mais que os trabalhadores com ensino médio, sendo esse diferencial ainda mais elevado em relação ao restante, o que acentua o problema centenário de concentração de renda no país.

Este trabalho visa entender o impacto do ambiente no qual o jovem está inserido sobre a probabilidade de entrar em um curso de graduação e concluí-lo. Utilizaremos fatores observáveis como sexo, idade, cor ou raça, ter ou não registro de nascimento, local de nascimento, escolaridade dos pais, entre outros, para isso.

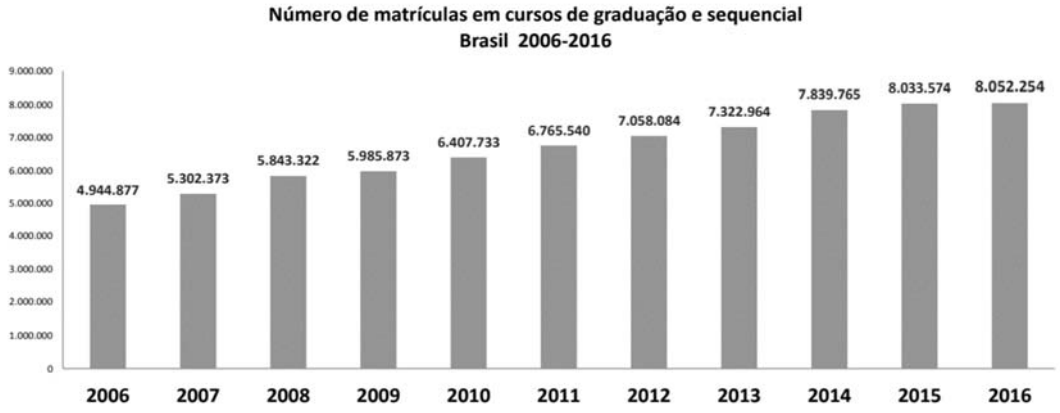
A análise será baseada nos dados do suplemento da PNAD 2014, fornecidos pelo IBGE, de onde conseguiremos tanto nossa variável de interesse quanto as explicativas. Usaremos um modelo ‘logit’ e depois expressaremos o efeito marginal de cada variável sobre a probabilidade de concluir o ensino superior.

Os fatores observáveis que se mostrem mais relevantes são um sinal de atenção para onde focar esforços com o objetivo de reduzir a desigualdade no país de maneira mais eficaz possível, balanceando injustiças sociais.

2 REFERENCIAL TEÓRICO

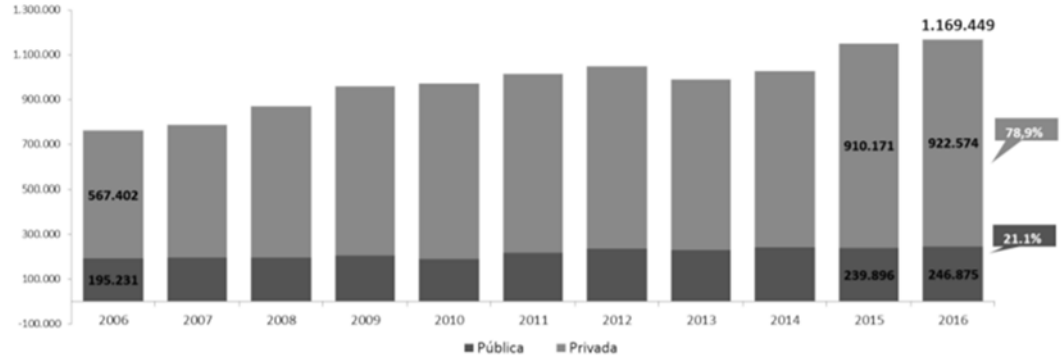
O número de pessoas que se matriculam em uma graduação no Brasil é bastante baixo, porém não tão alarmante quanto às que de fato se formam. Supondo que os 1.169 milhões que se formaram em 2016 entraram na faculdade em 2011 (6.765 milhões), aproximadamente, temos uma taxa de apenas 17,3% de concluintes. Conforme os dados abaixo e as figuras 2.1 e 2.2 vemos gravidade da situação.

Figura 1: Número de matrículas em cursos de graduação e sequencial – Brasil – 2006-2016



Fonte: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira Legislação e Documentos (INEP)

Figura 2: Número de concluintes em cursos de graduação, por categoria administrativa – Brasil – 2006-2016



Fonte: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira Legislação e Documentos (INEP) – 2016 (Acesso em: 06/10/2018)

Apesar do aumento constante em matrículas e conclusões, exceto pelas quedas em 2013 e 2014, o percentual da população com formação superior é ainda muito pequeno. Com base na PNAD Contínua de 2016, 84,7% dos brasileiros não possuía ensino superior completo.

2.1 Educação no brasil

O Brasil foi o último país a abolir a escravidão africana no Ocidente. Ainda era fortemente escravocrata enquanto países desenvolvidos ampliavam a cidadania e educação básica. Não há como entender a história brasileira sem este fator e suas consequências. (GOLDEMBERG, 1993)

Sobre as implicações no sistema educacional, Goldemberg (1993, p.66) destaca 2 fatores:

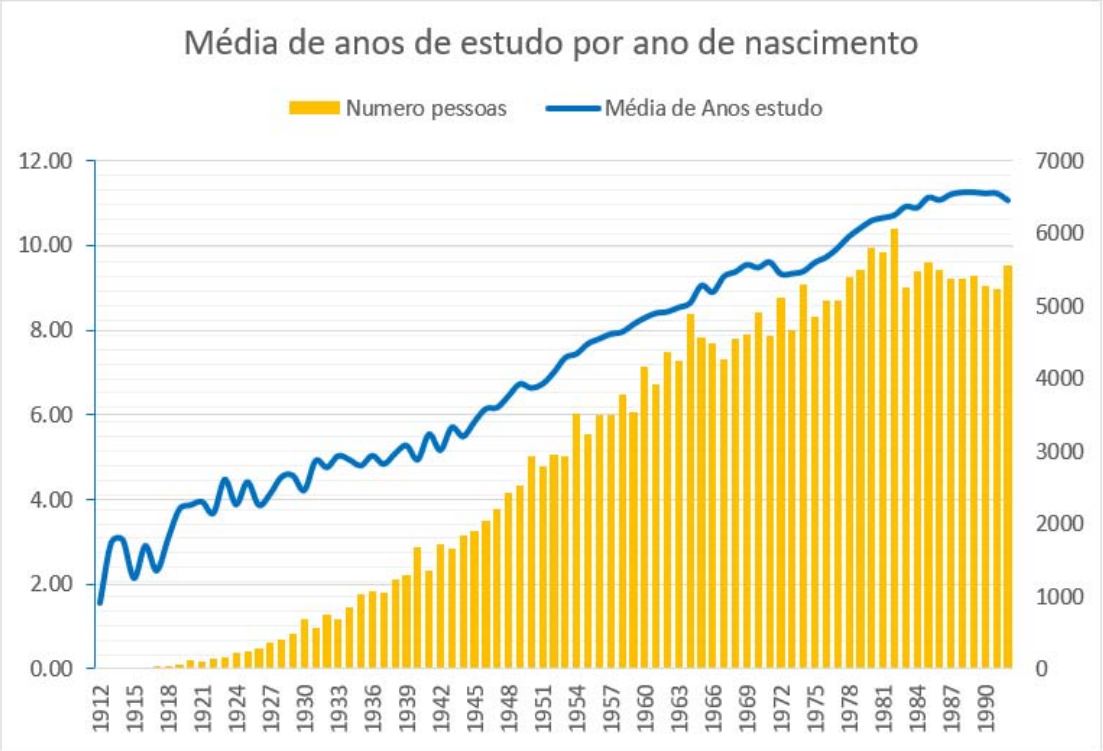
> "De um lado, pelas mudanças de tradições, valores e hábitos exigidas de uma população para a qual a escola não faz parte da perspectiva normal de vida nem integra sua tradição cultural. De outro lado, pela resistência das elites tradicionais em estenderem a cidadania a escravos e ex-escravos e, portanto, pela dificuldade em aceitarem e promoverem o ideal da escolarização universal como fundamento das políticas públicas."

Ainda em Goldemberg (1993) vemos que a expansão da educação básica se dá a partir de 1950 e se torna um direito público apenas na Constituição de 1988. É também duplicado o tempo obrigatório de escolarização. Ainda ressalta que existe relação entre o péssimo sistema educacional e a desigualdade na sociedade brasileira.

Com dados da PNAD 2014, no gráfico abaixo, podemos ver a média de anos de estudo ao longo do tempo, baseado no ano de nascimento. O intuitivo é que quanto mais idade, mais anos de estudo uma pessoa teria até chegar a um nível estabilizado. Porém podemos notar que quanto mais jovem, maior o nível de educação.

Nos últimos anos tem se estabilizado no 11, ou seja, 10 anos de estudo aproximadamente, ou quase o fim do ensino médio, o que ainda é baixo. Vale lembrar que mesmo assim, não podemos falar de qualidade com esses dados somente.

Gráfico 1: Média de anos de estudo por ano de nascimento



Fonte: Dados PNAD Contínua, 2014; elaboração própria

Estes e outros fatores contribuíram para o cenário atual onde o desempenho dos alunos brasileiros se encontra inferior à média dos outros países da OCDE. As notas se mantiveram praticamente estáveis desde o início dos anos 2000 exceto para matemática que cresceu significativamente entre 2003 e 2015. Ressaltando que o PIB per capita brasileiro (USD 15 893) não chega à metade da média dos outros países do grupo (USD 39 333) e que o gasto com alunos de 6 a 15 anos é de apenas 42% também relativo ao grupo (PISA, 2016).

2.2 Fatores sociais e desempenho escolar

Guerreiro-Casanova, et. al. (2011) pesquisaram, entre outras informações, o impacto do nível educacional dos pais e mães sobre a percepção de Autoeficácia Acadêmica das crianças. Este quesito engloba 3 dimensões sobre percepções individuais muito ligadas à confiança percebida na 1. Autoeficácia para aprender, 2. para atuar na vida escolar e 3. para decisão de carreira (GUERREIRO-CASANOVA, et al.,2011).

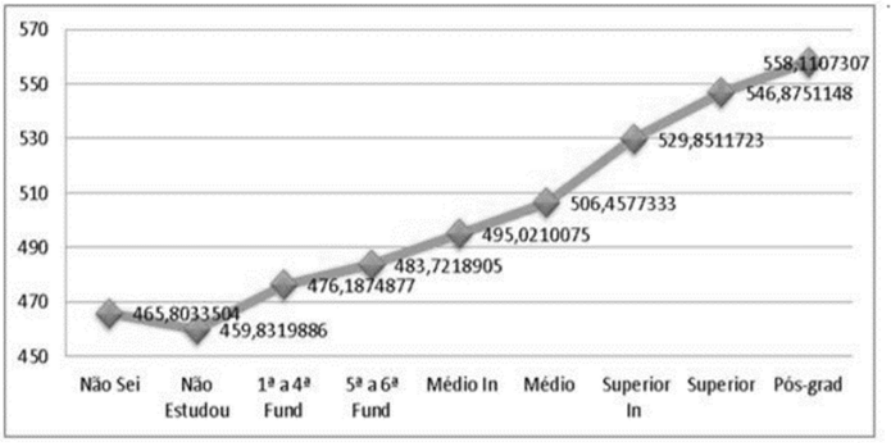
No geral o grupo com crianças cujas mães possuem Ensino Superior Completo teve as notas mais altas nos 3 quesitos, enquanto os que não sabiam o nível de estudo, as notas mais baixas. O mesmo ocorre quando considerada a educação dos pais, concluindo que pode haver sim uma interferência entre Autoeficácia Acadêmica e a escolaridade dos pais (GUERREIRO-CASANOVA, et al., 2011).

Lima e Silva et. al. (2017) demonstram correlações interessantes entre escolaridade dos pais e renda com a nota do estudante no Exame Nacional do Ensino Médio (ENEM) usando dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) no ano de 2013 (LIMA E SILVA et. al., 2017).

Dois pontos importantes foram verificados. Primeiro que a maioria populacional apresenta baixa escolaridade e está nos níveis mais baixos de renda familiar. Segundo que tanto o nível de escolaridade da mãe e do pai quanto a renda familiar têm impacto positivo na nota do estudante. Quanto maior a renda e escolaridade, maior a nota, como demonstrado nas figuras 3.2.1, 3.2.2 e 3.2.3 abaixo (LIMA E SILVA et. al., 2017).

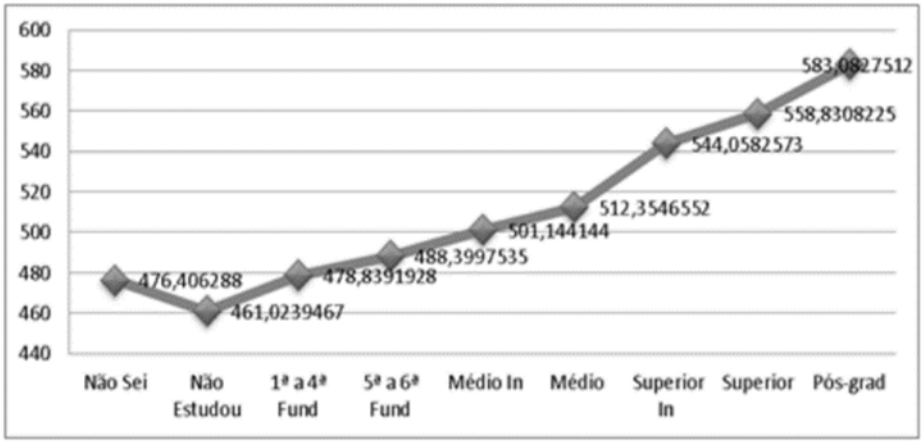
Somados esses dois fatores, temos uma situação alarmante cíclica onde famílias com menos renda tem desempenho cada vez pior. Relembro que este trabalho visa entender qual impacto gerado por cada um destes fatores na probabilidade de se chegar ao final de uma graduação e assim saber o modo mais eficiente de atuar.

Figura 3: Média das notas em relação à escolaridade da mãe



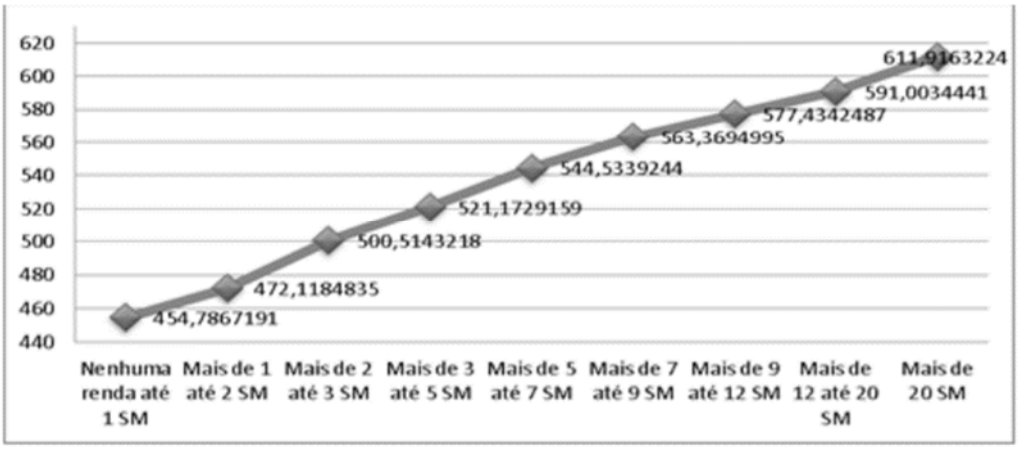
Fonte: (LIMA E SILVA, A. C; MOTA, R. O; LIMA, J. C. F; QUEIROZ, F. C. B. P; NORONHA, S. L. 2017. "Média das notas em relação à escolaridade da mãe". *A influência da escolaridade dos pais e da renda familiar no desempenho dos candidatos do ENEM,* 11. Santa Catarina: XXXVII Encontro Nacional De Engenharia De Produção.)

Figura 4: Média das notas em relação à escolaridade do pai



Fonte: (LIMA E SILVA, A. C; MOTA, R. O; LIMA, J. C. F; QUEIROZ, F. C. B. P; NORONHA, S. L. 2017. "Média das notas em relação à escolaridade do pai". *A influência da escolaridade dos pais e da renda familiar no desempenho dos candidatos do ENEM*, 11. Santa Catarina: XXXVII Encontro Nacional De Engenharia De Produção.)

Figura 5: Média das notas em relação à renda familiar



Fonte: (LIMA E SILVA, A. C; MOTA, R. O; LIMA, J. C. F; QUEIROZ, F. C. B. P; NORONHA, S. L. 2017. "Média das notas em relação à renda familiar". *A influência da escolaridade dos pais e da renda familiar no desempenho dos candidatos do ENEM*, 11. Santa Catarina: XXXVII Encontro Nacional De Engenharia De Produção.)

2.3 Transição para o ensino superior

A partir daqui queremos entender fatores que ajudem a explicar a decisão de se cursar um Ensino Superior, e permanecer nele até o final, principalmente nos baseando

nos quesitos de interesse, sendo renda familiar, escolaridade dos pais e inclusive sexo, quando possível.

Em um estudo realizado com dados da Alemanha, se conclui que jovens cujos pais são ‘altamente educados’ aumentam a probabilidade de iniciar uma faculdade em aproximadamente dez pontos percentuais em relação àqueles com pais com educação básica. A renda está também positivamente relacionada com esta probabilidade (RIPHAHN, R. T; SCHIEFERDECKER, F., 2010).

Contudo, R.T. Riphahn, F. Schieferdecker (2010) ao controlarem por seleção veem que a correlação perde significância estatística, o motivo pode ser o tamanho pequeno da amostra. Mesmo assim o resultado mais importante é que mesmo após a correção, a renda familiar é um fator positivo e significativo na decisão de transição para o ensino superior. Os 25% mais ricos tem dez pontos percentuais de vantagem relativamente aos 25% mais pobres.

2.4 Evasão escolar

Para o cenário brasileiro um fator muito relevante é a evasão escolar, como vimos anteriormente os números são alarmantes.

Silva Filho et. al. (2007) especifica a evasão em 2 aspectos:

> “1. A evasão anual média mede qual a percentagem de alunos matriculados em um sistema de ensino, em uma IES, ou em um curso que, não tendo se formado, também não se matriculou no ano seguinte. (...)”
>
> “2. A evasão total mede o número de alunos que, tendo entrado num determinado curso, IES ou sistema de ensino, não obteve o diploma ao final de um certo número de anos. É o complemento do que se chama índice de titulação. (...)”

Entre 2001 e 2005 a *evasão anual média* foi de 22% no ensino superior brasileiro, sendo maior nas Instituições de Ensino Privadas (IEP). Um ponto interessante mostrado é que quanto mais “seletiva” a instituição, menores as taxas de evasão. Sendo que, a taxa de evasão do ensino superior é entre duas a três vezes maior no primeiro ano em comparação com os seguintes (SILVA FILHO et. al., 2007).

3 FONTE DE DADOS

Os principais dados utilizados serão os disponibilizados pelo Governo, mais especificamente aqueles disponibilizados no portal do IBGE (Instituto Brasileiro de Economia e Estatística) e no do Inep (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira Legislação e Documentos). Isto é, o suplemento da PNAD de 2014, que inclui dados sobre Educação e Qualificação Profissional, e o Questionário do estudante do ENADE dos anos 2015, 2016 e 2017.

Cada curso/habilitação é avaliado de 3 em 3 anos pelo ENADE, assim, utilizaremos dados dos 3 anos agrupados para termos todos os cursos incluídos. Destes dados podemos retirar distribuição de idade, sexo, renda total familiar, anos de estudos da mãe e do pai para reafirmar e validar os dados retirados da PNAD.

Da base de dados da PNAD de 2014, utilizaremos o arquivo de pessoas que conta com 362.627 observações sobre diversas variáveis de interesse. Dentre elas foram selecionadas 19 para a elaboração deste trabalho, descritas abaixo:

Tabela 1: Variáveis Identificadoras de Indivíduo

Variáveis Identificadoras de Indivíduo	
UF	Unidade da Federação
V0302	Sexo
V8005	Idade do morador na data de referência
V0401	Condição na unidade domiciliar
V0404	Cor ou raça
V0408	Tem registro de nascimento
V5030	Lugar de nascimento
V0602	Frequenta escola ou creche
V6003	Curso que frequenta
V6007	Curso mais elevado que frequentou anteriormente
V0611	Concluiu este curso que frequentou anteriormente
V4803	Anos de estudo (todas as pessoas)
V4742	Rendimento mensal domiciliar *per capita*
V32012	Quando tinha quinze anos de idade, o curso de nível mais elevado que seu pai (ou o homem responsável pela sua criação) com quem morava, frequentou era:
V32013	Quando tinha quinze anos de idade, seu pai (ou o homem responsável pela sua criação) havia concluído esse curso
V32026	Quando tinha quinze anos de idade, o curso de nível mais elevado que sua mãe (ou a mulher responsável pela sua criação) com quem morava,frequentou era:
V32027	Quando tinha quinze anos de idade, sua mãe (ou a mulher responsável pela sua criação) havia concluído esse curso
V32039	Peso do morador selecionado para o Suplemento de Mobilidade Sócio-Ocupacional COM ajuste pela projeção de população - usado no cálculo de indicadores de morador selecionado
V9993	Data de geração do arquivo de microdados

Fonte: Dados PNAD Contínua, 2014; elaboração própria

3.1 Ajuste de dados

Para que o estudo tenha mais relevância serão excluídas as observações onde a idade do morador é menor que 22 ou maior que 80. Isto para que jovens que ainda não tenham tido tempo de se formar gerem viés na amostra e, também, para que não tenham dados de pessoas nascidas antes de 1934, onde o cenário de ensino superior no país era completamente diferente do atual. Aqui nossa amostra reduz de 362.627 para 233.738.

Na variável da "Condição na unidade domiciliar" serão utilizados dados de "Pessoa de referência", "Cônjuge", "Filho", "Agregado" ou "Outro parente", excluindo-se "Pensionista", "Empregado doméstico" e "Parente do empregado doméstico", fazendo nossa amostra diminuir para 233.248 observações.

Duas variáveis são utilizadas para determinar região de nascimento. A variável 'Lugar de nascimento' (V5030) é preenchida quando o morador não reside na UF que nasceu. Portanto quando esta variável estiver vazia, faremos V5030 = UF.

Ao retirarmos da amostra as observações onde 'Anos de estudo' = 17 (Não determinados) ou quem não tem declaração na variável 'Cor ou raça'. Assim, nos sobram 232.682 observações.

Perdemos o maior número de amostras ao filtramos pelos que não possuem a variável 'peso do morador', já que estes representam 179.940 amostras. E para que nosso estudo seja válido, são retirados os que não tem nenhuma informação de educação de pelo menos um dos pais.

Por fim, conseguimos nossa base de dados final, com 27.636 amostras válidas. Essa redução é acentuada pois, por definição, apenas um morador é selecionado para responder tais perguntas, mas é necessária pela para nossa pesquisa.

3.2 Variáveis de interesse

Como variável dependente (binária) utilizaremos a V4803 (anos completos de estudo). Se V4803=16 (15 anos de estudo ou mais), Y=1 e se V4803 está entre 1 e 15, Y=0, ou seja, uma dummy igual a 1 para quem possui ensino superior.

Como variáveis explicativas utilizaremos Sexo (V0302), Idade (V8005), Cor ou raça (V0404), Lugar de nascimento (V5030) e principalmente a Escolaridade do Pai

(V32012 e V32013) e a Escolaridade da Mãe (V32026 e V32027), conforme descritas abaixo.

A variável ‘Sexo’ é binária e utilizaremos ‘Feminino’=1 e ‘Masculino’=0.

Já a variável ‘Cor ou raça’ será dividida em 2 grupos para gerar uma Dummy. Nela a resposta ‘Branca’ será igual a 0 e as respostas ‘Preta’, ‘Amarela’, ‘Parda’, ‘Indígena’, terão valor 1.

Em relação ao ‘Lugar de Nascimento’ dividiríamos em 6 categorias. A primeira sendo ‘País estrangeiro’ e as outras 5 baseadas nas regiões brasileiras Centro-Oeste, Nordeste, Norte, Sul e Sudeste. Para facilitar, o Distrito Federal seria avaliado como Centro-Oeste.

Tabela 2: Estados brasileiros por Macrorregião

Região de Nascimento	
Rondônia	Norte
Acre	Norte
Amazonas	Norte
Roraima	Norte
Pará	Norte
Amapá	Norte
Tocantins	Nordeste
Maranhão	Nordeste
Piauí	Nordeste
Ceará	Nordeste
Rio Grande do Norte	Nordeste
Paraíba	Nordeste
Pernambuco	Nordeste
Alagoas	Nordeste
Sergipe	Nordeste
Bahia	Nordeste
Minas Gerais	Sudeste
Espírito Santo	Sudeste
Rio de Janeiro	Sudeste
São Paulo	Sudeste
Paraná	Sul
Santa Catarina	Sul
Rio Grande do Sul	Sul
Mato Grosso do Sul	Centro-Oeste
Mato Grosso	Centro-Oeste
Goiás	Centro-Oeste
Distrito Federal	Centro-Oeste
País estrangeiro	País estrangeiro

Fonte: Dados PNAD Contínua, 2014; elaboração própria

Para segmentar a ‘Escolaridade do Pai’ faremos 5 grupos:

Tabela 3: Segmentação de escolaridade

Escolaridade	
Primário incompleto	V32012 < 3
	V32012 = 4 & V32013 = 2
Fundamental incompleto	V32012 = 4 & V32013 = 1
	V32012 = 5 & V32013 = 2
	V32012 = 7 & V32013 = 2
Médio incompleto	V32012 = 5 & V32013 = 1
	V32012 = 6 & V32013 = 2
	V32012 = 7 & V32013 = 1
	V32012 = 8 & V32013 = 2
Médio completo	V32012 = 6 & V32013 = 1
	V32012 = 8 & V32013 = 1
	V32012 = 9 & V32013 = 2
Superior completo	V32012 = 9 & V32013 = 1
	V32012 = 10

Fonte: Dados PNAD Contínua, 2014; elaboração própria

A ‘Escolaridade da Mãe’ segue exatamente o mesmo modelo substituindo-se V32012 por V32026 e V32013 por V32027.

Usaremos a variável V32039 (Peso do morador selecionado para o Suplemento de Mobilidade Sócio-Ocupacional COM ajuste pela projeção de população - usado no cálculo de indicadores de morador selecionado) para identificar o peso da pessoa nesta amostra e ter um ajuste pela projeção de população.

4 METODOLOGIA

A metodologia é inspirada por R.T. Riphahn, F. Schieferdecker (2010) ao estimarem a importância de fatores como o que estimaremos neste trabalho sobre a decisão de cursar um ensino superior na Alemanha.

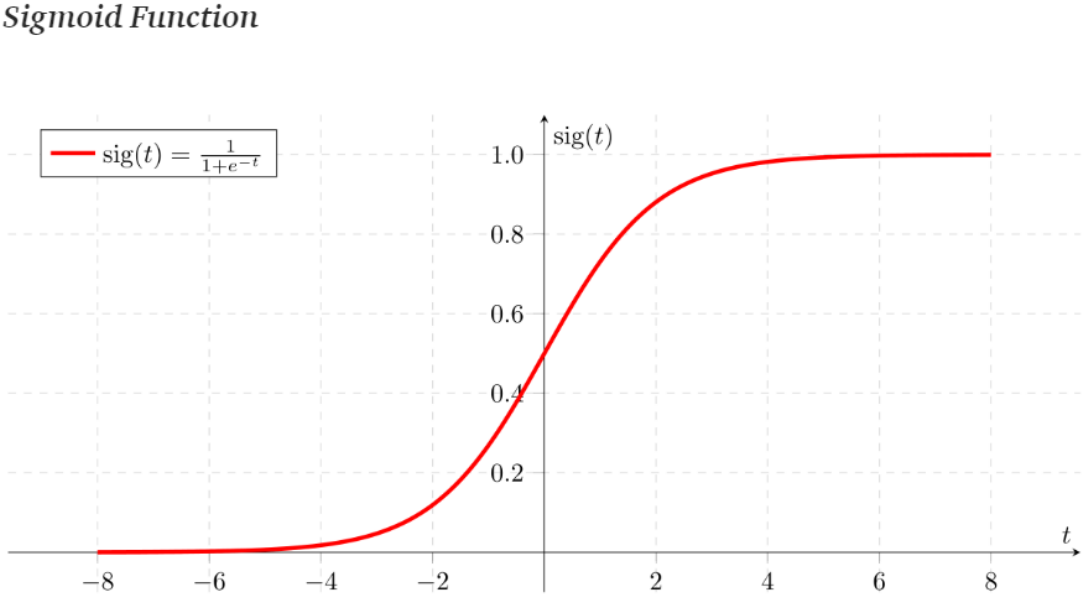
Será implementado um modelo LOGIT, onde possuir (=1) ou não (=0) ensino superior completo é a variável dependente e as variáveis explicativas são sexo, idade, cor ou raça, região de nascimento e uma variável que define a maior escolaridade entre os pais.

Retirando-se variáveis irrelevantes, o modelo final definido pode ser expresso como:

$$GRADUADO \sim Sexo + Idade + Cor.ou.raça + Região + Maior.educação$$

O esperado neste modelo são coeficientes relevantes que se traduzam em probabilidades, algo parecido com uma função sigmoide, demostrada no gráfico abaixo.

Gráfico 2: Função Sigmoide



Fonte: Towards Data Science. Disponível em: <https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc>. Acesso em 22 de junho de 2019.

Para clarear os resultados, podemos expressar o efeito marginal de cada variável independente sobre a probabilidade de conclusão do ensino superior. Assim, conseguimos descriminar o impacto específico de cada uma sobre nossa variável de interesse, o que nos ajuda a chegar a uma conclusão mais relevante de forma geral.

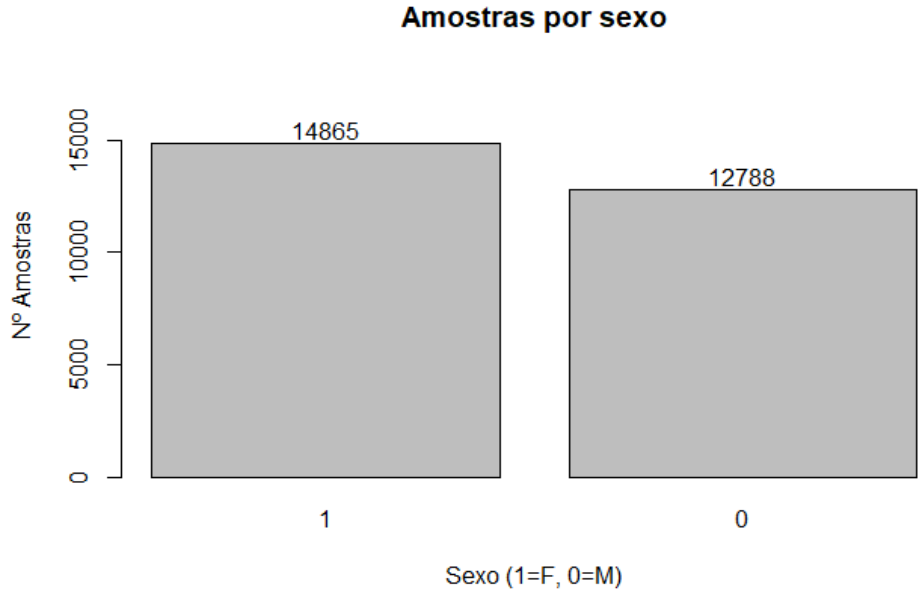
Seria interessante rodar uma segunda análise com as mesmas variáveis explicativas, porém alterando a dependente. Ao invés de investigar as diferenças entre pessoas com graduação ou não em geral, utilizar dados de quem já cursou uma graduação e ver as diferenças de quem a concluiu (=1) ou não (=0). Contudo, o número de amostras neste caso fica muito pequeno e nossa regressão, inconclusiva.

5 RESULTADOS

Inicialmente, é interessante observar a distribuição da base para nossos fatores observáveis de modo geral.

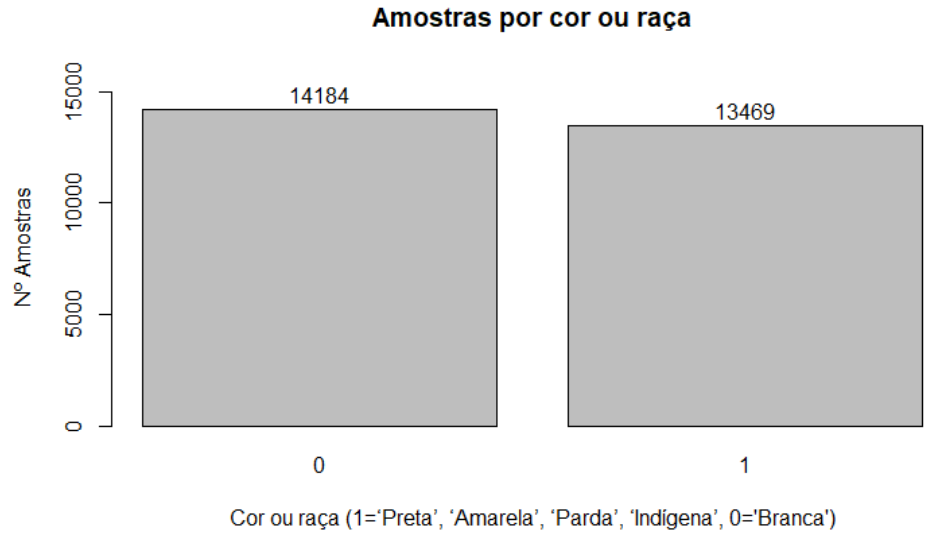
Em questão de sexo e ‘cor ou raça’ a nossa amostra parece relativamente bem distribuída como demonstram as tabelas 4 e 5 abaixo.

Tabela 4: Amostras por sexo



Fonte: Dados PNAD Contínua, 2014; elaboração própria

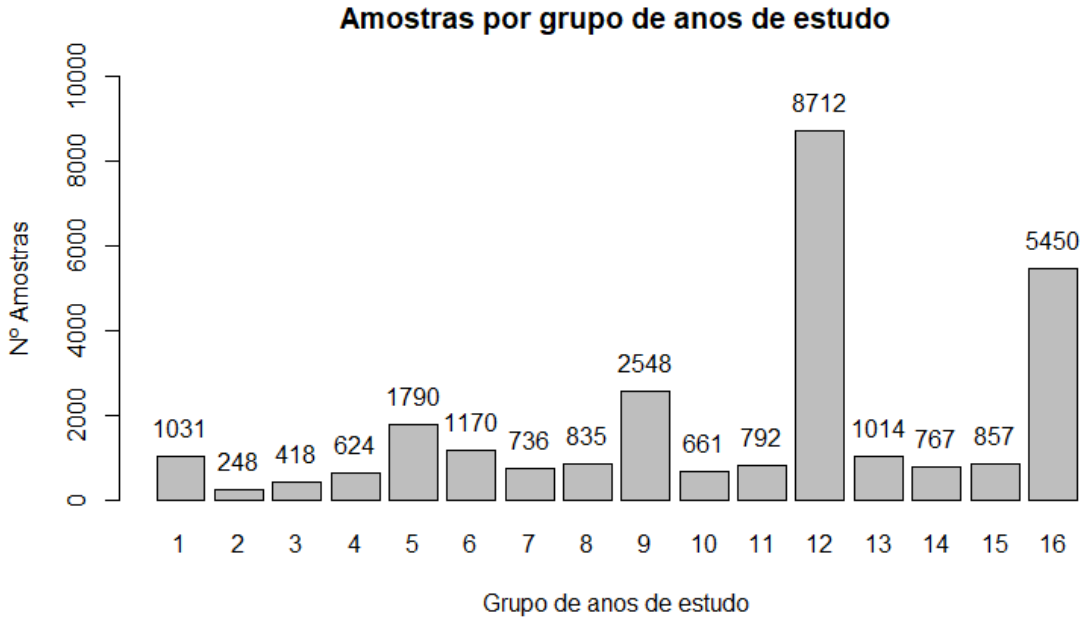
Tabela 5: Amostras por ‘cor ou raça’



Fonte: Dados PNAD Contínua, 2014; elaboração própria

É interessante observar as grandes concentrações que temos na distribuição por grupos de anos de estudo. Principalmente em 12 que equivale a 11 anos de estudo, ou aproximadamente, até o fim do ensino médio. Em 9 que equivale a 8 anos de estudo, ou seja, 3 anos a menos que o anterior e significa que provavelmente cursou o ensino fundamental. Por último o grupo 16, 15 anos ou mais de estudo, que agrupa muito provavelmente todos os que possuem um ensino superior completo ou mais.

Tabela 6: Amostras por grupo de anos de estudo



Fonte: Dados PNAD Contínua, 2014; elaboração própria

É válido observaremos uma tabela descritiva com os principais fatores de interesse. Com esta tabela inicial é notória a diferença entre os que possuem graduação principalmente nos fatores ‘sexo’, ‘cor ou raça’ e rendimento mensal domiciliar per capita.

Já as variáveis ‘idade’, ‘ter registro de nascimento’ e inclusive ‘frequenta escola ou creche’ não são muito diferentes e podem não influenciar muito na nossa regressão. As variáveis de anos de estudo e graduação servem para verificação de dados. Os 16 anos e 100%, respectivamente, demostram a quebra correta da amostra.

Tabela 7: Tabela descritiva

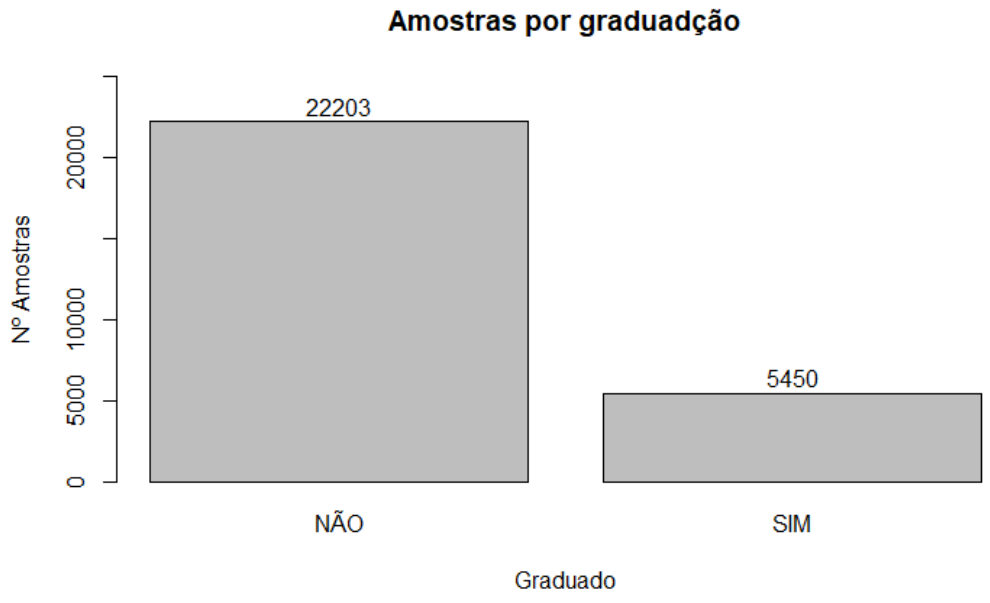
	Não Graduado - Média	Não Graduado - Desv.P.	Graduado - Média	Graduado - Desv.P
Sexo	52%	50%	59%	49%
Idade.do.morador.na.data.de.referência	41	14	42	13
Cor.ou.raça	52%	50%	33%	47%
Tem.registro.de.nascimento	99%	10%	99%	9%
Frequenta.escola.ou.creche	8%	26%	5%	22%
Anos.de.estudo..todas.as.pessoas.	9.58	3.63	16.00	0.00
Rendimento.mensal.domiciliar.per.capita	2761	1639	5468	2274
GRAD	0%	0%	100%	0%
Tamanho amostra: 27.653				

Fonte: Dados PNAD Contínua, 2014; elaboração própria

Podemos notar claramente que os que possuem graduação apresentam maiores participações de mulheres, de brancos e possuem, em média, maior renda domiciliar. Este já é um primeiro ponto de atenção.

Por fim, como podemos notar na tabela 8, o número de amostras onde o entrevistado possui uma graduação em ensino superior é extremamente baixa e isso atrapalhará na precisão do nosso modelo como veremos logo em seguida.

Tabela 8: Amostras por critério de graduação



Fonte: Dados PNAD Contínua, 2014; elaboração própria

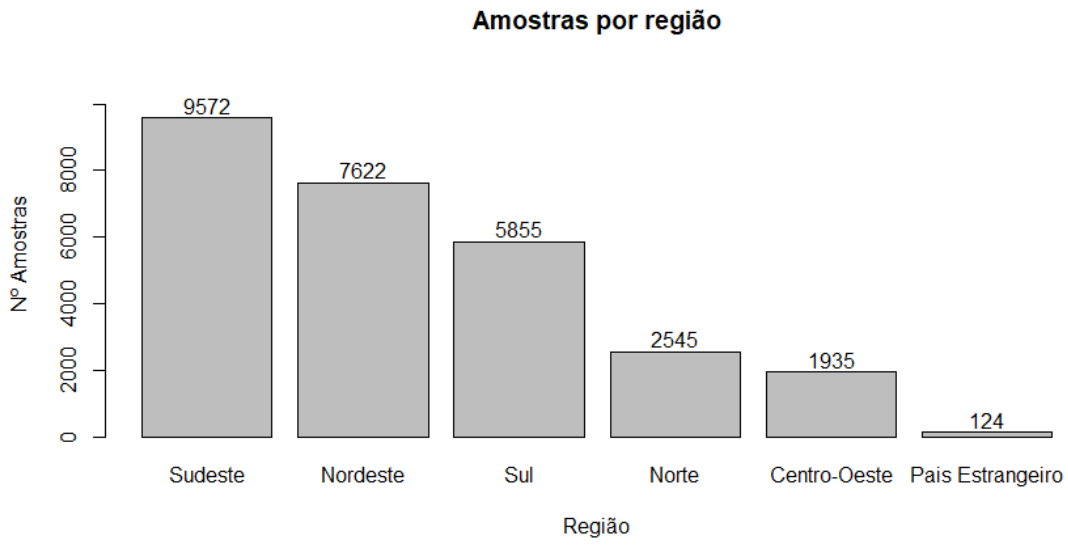
5.1 Resultados regressão

Aqui analisaremos o impacto das nossas variáveis explicativas sobre a probabilidade de possuir ensino superior. Como dito anteriormente, nossa base amostral é pequena e um número muito grande de variáveis pode gerar resultados equivocados.

Para reduzir o número de variáveis explicativas retiraremos a 'tem ou não registro de nascimento' que não se mostrou relevante durante as modelagens. Tampouco usaremos a renda domiciliar per capita pois é muito correlacionada com a nossa variável de interesse, por mais que seja interessante ver a média de renda, não é boa para a regressão. Isso ocorre porque a renda já é influenciada pela decisão de passar fazer um curso superior. Ou seja, a conclusão afeta a renda domiciliar per capita.

Por fim, alguns ajustes foram feitos na variável 'região'. As amostras cujo nascimento foi em um país estrangeiro foram retiradas por serem muito poucas, como pode ser visto na tabela 9, abaixo. As regiões foram agrupadas em 2 grandes grupos, Norte/Nordeste e Sul/Sudeste/Centro-Oeste. Essa divisão foi baseada por fatores macroeconômicos semelhantes entre elas no ano de 2014, como taxa de desocupação, rendimento médio real de todos os trabalhos e o PIB per capita, fornecidos pelo IPEA em seu boletim regional.

Tabela 9: Amostras por região de nascimento



Fonte: Dados PNAD Contínua, 2014; elaboração própria

Portanto a regressão final é descrita como:

$$GRADUADO \sim Sexo + Idade + Cor.ou.raça + Região + Maior.educação$$

$onde,$

$Sexo$:

$Homem = 0$;

$Mulher = 1$

$Idade$:

$valor\ discreto\ entre\ 22\ e\ 80$

$Cor\ ou\ raça$:

$Branca = 0$;

$Preta/Amarela/Parda/Indígena\ = 1$

$Região$:

$Subdividida\ em$:

1. Norte/Nordeste
2. Sul/Sudeste/CentroOeste

$Maior.educação$:

$Subdividida\ em$:

1. $Primário.Incompleto$
2. $Fundamental.Incompleto$
3. $Médio.Incompleto$
4. $Médio.Completo$
5. $Superior.Completo$

Lembrando que as últimas duas variáreis omitirão a sua primeira opção e os nossos betas estimados são em relação as opções omitidas.

Rodando esta regressão logística obtemos os seguintes coeficientes estimados na tabela 10, abaixo. Como podemos notar de início, 'cor ou raça' é a única variável com beta negativo, isso indica que simplesmente não ser branco, diminui a probabilidade de ter um ensino superior. Lembrando que tanto para educação dos pais quanto para região de nascimento, precisamos usar comparações, isto é, nascer no nosso segundo grupo, Sul/Sudeste/CO, aumenta a probabilidade de se graduar relativamente a nascer no Norte/Nordeste. O mesmo se aplica para educação dos pais.

Tabela 10: Coeficientes regressão logística

```
Call:
glm(formula = GRAD ~ Sexo + Idade.do.morador.na.data.de.referência +
    Cor.ou.raça + regiao.final + maior.educacao, family = binomial(link = "logit"),
    data = pnad_pes_2014_final)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8722  -0.6140  -0.4473  -0.3001   2.6948

Coefficients:
                                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                             -3.526345   0.086742 -40.653  < 2e-16 ***
Sexo1                                    0.362096   0.034039  10.638  < 2e-16 ***
Idade.do.morador.na.data.de.referência   0.016975   0.001232  13.783  < 2e-16 ***
Cor.ou.raça1                            -0.468080   0.036604 -12.788  < 2e-16 ***
regiao.final2.Sul_Sudeste_Centro-Oeste   0.267536   0.038407   6.966 3.27e-12 ***
maior.educacao2.Fundamental.Incompleto   0.635354   0.056030  11.340  < 2e-16 ***
maior.educacao3.Médio.Incompleto         1.212116   0.065067  18.629  < 2e-16 ***
maior.educacao4.Médio.Completo           1.997102   0.057767  34.572  < 2e-16 ***
maior.educacao5.Superior.Completo        3.101003   0.064822  47.839  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 27236  on 27528  degrees of freedom
Residual deviance: 22876  on 27520  degrees of freedom
AIC: 22894
```

Fonte: Dados PNAD Contínua, 2014; elaboração própria

Quebramos nossa base em duas partes, uma de treinamento e uma de teste do modelo. Assim, conseguimos estimar o percentual de acerto dele, com base na quebra demonstrada na tabela 11, que no nosso caso foi de 82,52%. Não é um valor excelente, mas dado o tamanho da amostra é razoável.

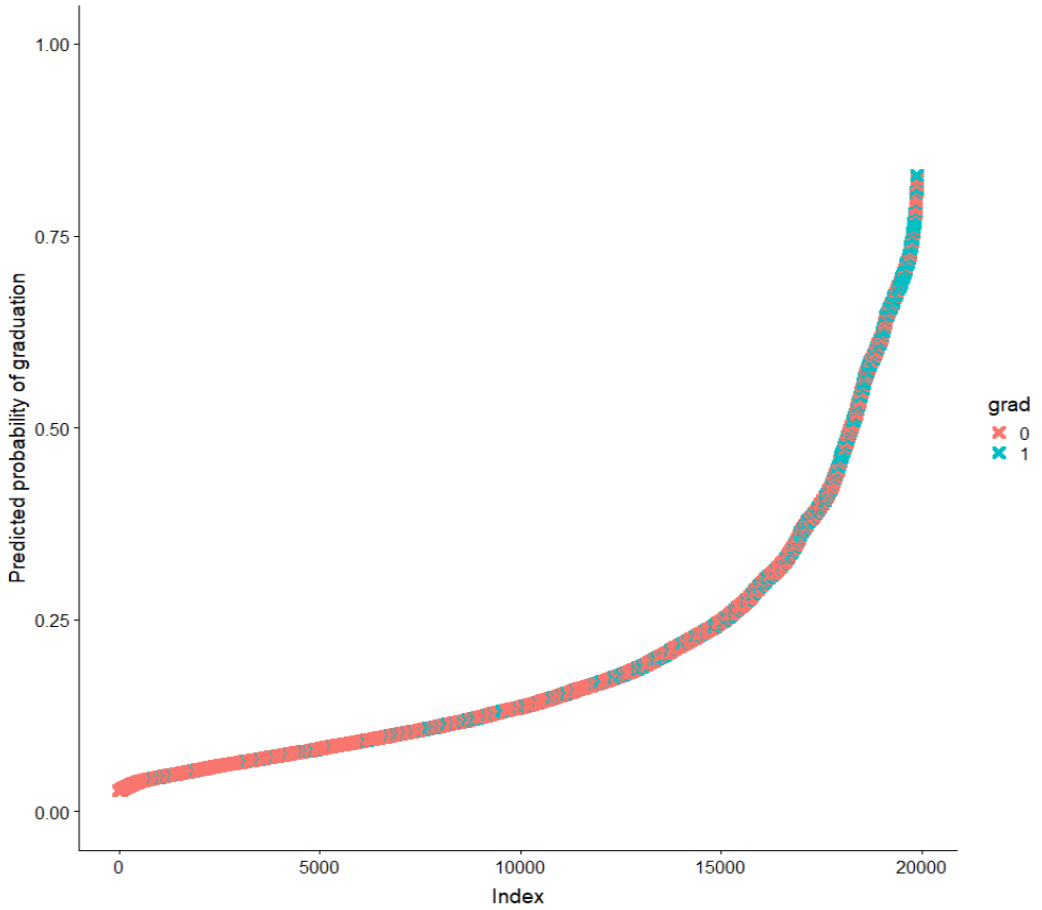
Tabela 11: “ConfusionMatrix” - Matriz de acerto do modelo

		Predicted_value	
		FALSE	TRUE
Actual_ value	0	16580	650
	1	3093	1089

Fonte: Dados PNAD Contínua, 2014; elaboração própria

Podemos também, plotar a probabilidade estimada do modelo dado o que de fato aconteceu, como no gráfico 3. O esperado é que quanto maior a probabilidade estimada, de fato tenham mais pessoas com ensino superior completo. Como podemos notar abaixo, isso ocorre, mas com uma margem de erro. O desejado é que quanto mais para direita mais turquesa deveria ficar e para a esquerda, abaixo dos 50% mais rosado.

Gráfico 3: Probabilidade estimada pelo modelo dado o que de fato ocorreu



Fonte: Dados PNAD Contínua, 2014; elaboração própria

Por último então, expressaremos o efeito marginal de cada variável independente sobre a probabilidade de conclusão do ensino superior dado o este modelo para a magnitude de cada variável ficar mais clara.

Isso é feito de maneira simples, queremos saber como a probabilidade muda com o aumento de uma unidade na variável explicativa mantendo as outras constantes. Ou seja, sendo 'p' a probabilidade de um evento ocorrer e 1-p a probabilidade de ele não ocorrer,

$$\left(\frac{p}{1-p}\right) \text{ é } a\ chance\ ou\ 'odds'$$

$$Logo, o\ logit\ da\ propabilidade \text{ é } o\ logaritimo\ dos\ 'odds':$$

$$logit(p) = \log\left(\frac{p}{1-p}\right) =$$

$$\log(p) - \log(1-p) = -log\left(\frac{1}{p} - 1\right)$$

Tabela 12: Efeitos marginais - Logit

Ter ensino superior completo	Efeitos marginais Logit
(Intercept)	-0.47
Sexo	0.05
Idade.do.morador.na.data.de.referência	0.002
Cor.ou.raça	-0.06
regiao.final_2.Sul_Sudeste_Centro-Oeste	0.04
maior.educacao_2.Fundamental.Incompleto	0.08
maior.educacao_3.Médio.Incompleto	0.16
maior.educacao_4.Médio.Completo	0.26
maior.educacao_5.Superior.Completo	0.40

Fonte: Dados PNAD Contínua, 2014; elaboração própria

Com a tabela 12, podemos ver claramente os efeitos de cara variável. Fica claro que a educação dos pais tem a maior influência na educação dos filhos dados esses quesitos. Quando pelo menos um dos pais possui ensino superior completo, as chances do filho também possuir são 40 pontos percentuais maiores comparados aos que os pais que possuem primário incompleto.

Esse percentual diminui conforme a educação dos pais diminui. Para ensino médio completo as chances são 26 pontos percentuais maiores comparados aos que os pais que possuem primário incompleto, para ensino médio incompleto as chances são 16p.p. maiores e por fim, para ensino fundamental incompleto as chances são 8p.p. maiores.

A cor da pele tem a segunda maior influência, pessoas não brancas tem 6 pontos percentuais a menos de chances que as brancas. Seguido do sexo, ser mulher aumenta em 5p.p. e região de nascimento, nascer nas regiões mais ‘ricas’ do país aumentam em 4p.p. as chances em comparação com o Norte e Nordeste.

6 CONCLUSÃO

Conseguimos concluir que certos fatores observáveis exercem grande influência sobre a probabilidade de conclusão de um curso superior no brasil. Sem dúvida, a escolaridade dos pais tem uma influência tremenda o que pode ser uma boa ou má notícia.

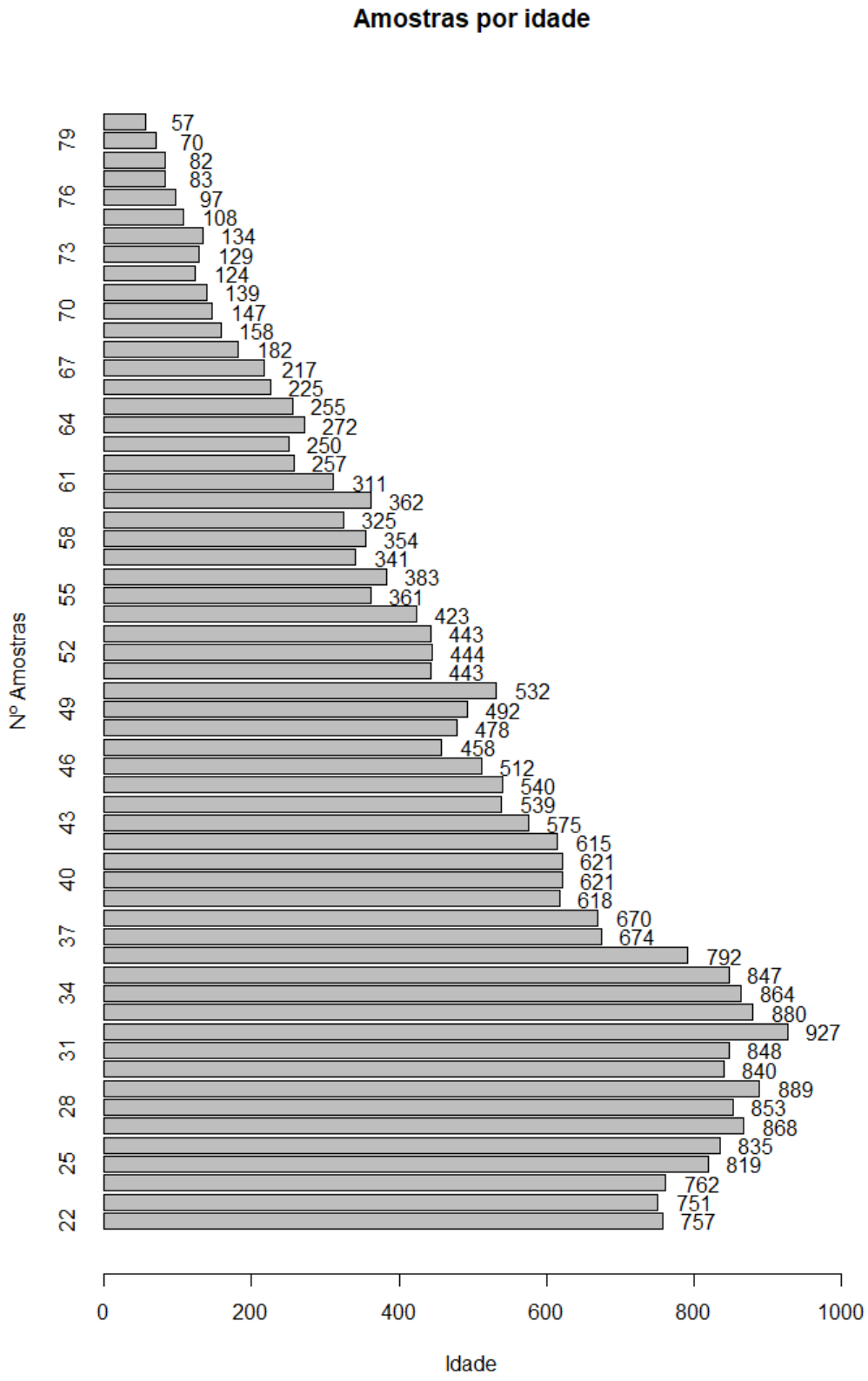
Conforme mais se educa a população, provavelmente mais será passado adiante de forma ‘inercial’. Porém, pode ser um fator perigoso também caso se crie um ciclo vicioso. Pessoas mais educadas ficando cada vez mais educadas e o mesmo no sentido oposto. Como educação se traduz claramente em salário, em renda familiar, é eminente um aumento da, já grande, concentração de renda no Brasil.

É necessário, portanto, garantir um acesso à educação de qualidade principalmente para aqueles que não possuem o exemplo em suas casas, normalmente pessoas mais carentes. É uma das formas mais efetivas de proporcionar outra realidade e outro futuro. Quando mais educação no país, mais desenvolvimento e aumento de qualidade de vida.

Por fim, lembrando que o ideal era uma amostra maior de concluintes para uma análise mais aprofundada neste assunto, muito mais ainda pode ser feito. Esperamos que esses números aumentes nos próximos anos.

ÍNDICE

Tabela 13: Quebra da amostra por idade



Fonte: Dados PNAD Contínua, 2014; elaboração própria

REFERÊNCIAS BIBLIOGRÁFICAS

GOLDEMBERG, J. (1993). O repensar da educação no Brasil. **Revista estudos avançados**. v. 8, n. 18. São Paulo. 1993.

GUERREIRO-CASANOVA, C.; DANTAS, A.; AZZI, G. (2011). Autoeficácia de alunos do ensino médio e nível de escolaridade dos pais. **Estudos Interdisciplinares em Psicologia.**

IBGE (2010). Distribuição da população por sexo, segundo os grupos de idade. Disponível em: <*https://censo2010.ibge.gov.br/sinopse/index.php?dados=12*>. Acesso em 06 de outubro de 2018.

INEP (2016). Censo Da Educação Superior 2016. Disponível em: <*http://download.inep.gov.br/educacao_superior/censo_superior/apresentacao/2016/apresentacao_censo_educacao_superior.pdf*>. Acesso em 22 de junho de 2019.

IPEA (2017). Boletim regional, urbano e ambiental. Disponível em: <http://www.ipea.gov.br/portal/images/stories/PDFs/boletim_regional/170531_bru_16_indicadores01.pdf>. Acesso em 22 de junho de 2019.

LIMA E SILVA, A. C; MOTA, R. O; LIMA, J. C. F; QUEIROZ, F. C. B. P; NORONHA, S. L. (2017). A influência da escolaridade dos pais e da renda familiar no desempenho dos candidatos do ENEM. **XXXVII Encontro Nacional De Engenharia De Produção**.

PARRODE, A. (2017). "Apenas 15% dos brasileiros têm ensino superior completo, mostra IBGE". Disponível em: <*https://www.jornalopcao.com.br/ultimas-noticias/apenas-15-dos-brasileiros-tem-ensino-superior-completo-mostra-ibge-13091*>. Acesso em 30 de maio de 2019.

PISA (2016). Programme for International Student Assessment. Results from PISA 2015. **Country Note**. OECD.

RIPHAHN, R. T; SCHIEFERDECKER, F. (2010). The transition to tertiary education and parental background over time. **Journal of Population Economics.**

SWAMINATHAN, S. (2019). Logistic Regression—Detailed Overview. Disponível em: <*https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc*>. Acesso em 22 de junho de 2019.

SILVA FILHO, R. L. L; MOTEJUNAS, P. R; HIPÓLITO, O; LOBO; M. B. C. M. (2007). A evasão no ensino superior brasileiro. **Cadernos de Pesquisa, v. 37, n. 132**. Instituto Lobo para o Desenvolvimento da Educação, da Ciência e da Tecnologia

WIKIPEDIA (2019). Logit. Disponível em: <*https://en.wikipedia.org/wiki/Logit*>. Acesso em 22 de junho de 2019.