

**FUNDAÇÃO GETULIO VARGAS
ESCOLA DE POLÍTICAS PÚBLICAS E GOVERNO**

FREDERICO PEREIRA VASCONCELOS

ESTUDO SOBRE OS DETERMINANTES DO ABISMO DIGITAL NO BRASIL

BRASÍLIA

2021

FREDERICO PEREIRA VASCONCELOS

ESTUDO SOBRE OS DETERMINANTES DO ABISMO DIGITAL NO BRASIL

Dissertação apresentada à Escola de Políticas Públicas e Governo da Fundação Getulio Vargas, como requisito para a obtenção do título de Mestre em Políticas Públicas e Governo.

Área de concentração: Políticas Públicas e Governo

Orientador: Edson Kenji Kondo, PhD

BRASÍLIA

2021

Vasconcelos, Frederico Pereira.

Estudo sobre os determinantes do abismo digital no Brasil / Frederico Pereira Vasconcelos. - 2021.

121 f.

Orientador: Edson Kenji Kondo.

Dissertação (mestrado MPPG) – Fundação Getulio Vargas, Escola de Políticas Públicas e Governo.

1. Inclusão digital - Política governamental - Brasil. 2. Tecnologia da informação - Aspectos sociais. 3. Letramento digital. 4. Políticas públicas - Brasil. I. Kondo, Edson Kenji. II. Dissertação (mestrado MPPG) – Escola de Políticas Públicas e Governo. III. Fundação Getulio Vargas. IV. Título.

CDU 35::007

FREDERICO PEREIRA VASCONCELOS

ESTUDO SOBRE OS DETERMINANTES DO ABISMO DIGITAL NO BRASIL

Dissertação apresentada à Escola de Políticas Públicas e Governo da Fundação Getulio Vargas, como requisito para a obtenção do título de Mestre em Políticas Públicas e Governo

Área de concentração: Políticas Públicas e Governo

Data da aprovação: 15 de junho de 2021

Banca examinadora:

Edson Kenji Kondo, PhD (Orientador)
Escola de Políticas Públicas e Governo (EPPG)

Kaizô Iwakami Beltrão, PhD
Escola de Políticas Públicas e Governo (EPPG)

Claudio Chauke Nehme, PhD
Diretor do Gartner

A meus pais e avós,
que me incentivaram e me deram a oportunidade e as condições
necessárias para que eu pudesse seguir os meus sonhos.

AGRADECIMENTOS

Ao professor Edson Kenji Kondo, que, prontamente, aceitou o convite de ser meu orientador e que, ao longo de todo este trabalho científico, sempre esteve disponível para compartilhar de seu conhecimento e me aconselhar com relação à melhor direção a ser tomada.

Ao professor Benjamin Miranda Tabak, pelo seu comprometimento incessante com a qualidade deste programa de mestrado, pela parceria sólida com os alunos no decorrer de seus estudos e suas pesquisas científicas e pela disponibilidade, transparência e cordialidade costumeiras, que pautaram nossas aulas e nossas conversas.

Ao professor Kaizô Iwakami Beltrão, pelas orientações e apoio nas atividades no campo da Estatística, que enriqueceram o modelo proposto e os achados deste estudo.

Ao professor Claudio Chauke Nehme, que se dispôs a participar da defesa desta dissertação e cujos apontamentos, sugestões e críticas são fundamentais para a melhoria da qualidade e do resultado final deste trabalho.

À Tânia Maria de Almeida, cujo apoio, ao longo de todo este trabalho científico, foi fundamental para a organização, planejamento e desenvolvimento de suas atividades.

A todo o corpo docente e administrativo da Escola de Políticas Públicas e Governo (EPPG) da Fundação Getúlio Vargas (FGV) – particularmente à, Camila Ferreira de Meneses, Débora Sena, Fernanda Miranda de Souza e Júlia Noronha Cruz Rios – que trabalham, diariamente, com afinco e profissionalismo para garantir a melhor experiência aos alunos e que me abasteceram de informações, orientações e conhecimentos valiosos.

Ao Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), pela disponibilidade e presteza em esclarecer as dúvidas e responder os questionamentos a respeito da pesquisa TIC Domicílios e seus respectivos microdados.

Aos colegas de mestrado, que me ajudaram ao longo dos cursos e das atividades acadêmicas e que me incentivaram durante o desenvolvimento desta dissertação de mestrado.

Aos amigos que me ajudaram com discussões acerca de temas a serem abordados nesta dissertação.

À minha família, que me apoiou desde o início e ao longo de toda esta jornada.

“Man’s greatest asset is the unsettled mind.”

Isaac Asimov

RESUMO

A busca pela compreensão acerca do abismo digital, da sua definição, das suas causas e dos seus efeitos e, finalmente, das ações que podem ser tomadas sob a ótica de políticas públicas em prol do acesso inclusivo à Internet é um esforço que já dura mais de duas décadas. Ainda que a inclusão digital não esteja associada, exclusivamente, à Internet, por conta da sua relevância na vida dos indivíduos, organizações, Estado e sociedade, o seu uso tornou-se um indicador comumente utilizado para o estudo dos determinantes desta desigualdade social. No Brasil, ainda que o número de usuários cresça a cada ano, segundo o Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), aproximadamente, um em cada quatro brasileiros, com 10 anos ou mais de idade, não se conectou à rede no ano de 2019. No intuito de compreender melhor os fatores que influenciam no estabelecimento do abismo digital no país e contribuir com informações atualizadas para a discussão de ações que o combatam, esta pesquisa buscou, por meio da análise dos microdados da pesquisa TIC Domicílio, conduzida anualmente pelo Cetic.br, e uso de regressão logística, identificar e mensurar a relevância dos efeitos dos indicadores sociais, econômicos, demográficos e culturais que determinaram a desigualdade digital no período de 2015-2019. Os achados desse estudo sugerem que, no período analisado, as variáveis idade, grau de instrução, classe econômica, população economicamente ativa, religião e área de residência apresentaram, com relevância estatística, características de determinantes do abismo digital no Brasil.

Palavras-chaves: Abismo digital. Analfabetismo digital. Inclusão digital. Transformação digital. Brasil. Políticas Públicas.

ABSTRACT

The search for the understanding of the digital divide, its definition, its causes and effects and, finally, the actions that can be taken from a public policies standpoint in favor of inclusive access to the Internet is an effort that has already lasted longer of two decades. Although digital inclusion is not exclusively associated with the Internet, due to its relevance in the lives of individuals, organizations, the State and society, its use has become a commonly used indicator for the study of the determinants of this social inequality. In Brazil, although the number of users grows every year, according to the Regional Center for Studies for the Development of the Information Society (Cetic.br), approximately one in four Brazilians, aged 10 or over, did not have access to the network in the year 2019. In order to better understand the factors that influence the establishment of the digital divide in the country and contribute with updated information to the discussion of actions to tackle it, this research sought, through the analysis of microdata from the TIC Domicílio survey, conducted annually by Cetic.br, and use of logistic regression, to identify and measure the relevance of the effects of social, economic, demographic and cultural indicators that determined digital inequality in the period 2015-2019. The research findings suggest that, in the period analyzed, the variables age, level of education, economic class, economically active population, religion and area of residence presented, with statistical significance, characteristics of determinants of the digital divide in Brazil.

Keywords: Digital divide. Digital literacy. Digital inclusion. Digital transformation. Brazil. Public Policies.

LISTA DE ILUSTRAÇÕES

FIGURA 1 – LACUNAS DE CONHECIMENTO DOS MEIOS DE COMUNICAÇÃO.....	29
FIGURA 2 – CATEGORIAS DE DETERMINANTES DO ABISMO DIGITAL.....	34
FIGURA 3 – EVOLUÇÃO DA IDADE COMO DETERMINANTE DIGITAL.....	52
FIGURA 4 – EVOLUÇÃO DO GRAU DE INSTRUÇÃO COMO DETERMINANTE DIGITAL.....	54
FIGURA 5 – EVOLUÇÃO DA CLASSE ECONÔMICA COMO DETERMINANTE DIGITAL – CLASSE A	55
FIGURA 6 – EVOLUÇÃO DA CLASSE ECONÔMICA COMO DETERMINANTE DIGITAL – CLASSE B	56
FIGURA 7 – EVOLUÇÃO DAS CLASSES ECONÔMICA COMO DETERMINANTE DIGITAL – CLASSE C.....	57
FIGURA 8 – EVOLUÇÃO DA POPULAÇÃO ECONOMICAMENTE ATIVA COMO DETERMINANTE DIGITAL ...	58
FIGURA 9 – EVOLUÇÃO DA RELIGIÃO COMO DETERMINANTE DIGITAL.....	60
FIGURA 10 – EVOLUÇÃO DA ÁREA RESIDENCIAL COMO DETERMINANTE DIGITAL.....	61
FIGURA 11 – CURVA CARACTERÍSTICA DE OPERAÇÃO DE RECEPTOR (ROC)	64

LISTA DE TABELAS

TABELA 1 – EVOLUÇÃO DO NÚMERO DE USUÁRIOS DA INTERNET NO BRASIL (2015-2019).....	15
TABELA 2 – INFORMAÇÕES A RESPEITO DA VARIÁVEL DEPENDENTE.....	43
TABELA 3 – INFORMAÇÕES A RESPEITO DAS VARIÁVEIS INDEPENDENTES.....	43
TABELA 4 – RESULTADOS DE PROCESSO DE TRATAMENTO DE TABELAS EXPERIMENTAIS.....	45
TABELA 5 – REGRESSÃO LOGÍSTICA PARA DETERMINANTES DO ABISMO DIGITAL NO BRASIL.....	49
TABELA 6 – MÉTRICAS DE DESEMPENHO DE PREDIÇÃO.....	64

LISTA DE SIGLAS

CEPAL	Comissão Econômica para a América Latina e o Caribe
Cetic.br	Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação
CGI.br	Comitê Gestor da Internet no Brasil
EUA	Estados Unidos da América
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FUST	Fundo de Universalização dos Serviços de Telecomunicações
ITU	International Telecommunication Union
NIC.br	Núcleo de Informação e Coordenação do Ponto BR
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
PC	Personal Computer
PNBL	Programa Nacional de Banda Larga
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ROC	Característica de Operação do Receptor
TELEBRAS	Telecomunicações Brasileiras S.A.
TIC	Tecnologia da Informação e Comunicação
WEF	World Economic Forum
WWW	World Wide Web
VPN	Virtual Private Network

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Apresentação	13
1.2	Justificativa do estudo	21
1.3	Delimitação do estudo	22
2	REVISÃO DE LITERATURA.....	24
2.1	Introdução ao abismo digital	24
2.2	Evolução e diferenciação de abordagens.....	27
2.3	Os determinantes do abismo digital.....	33
2.4	Bibliometria.....	37
3	METODOLOGIA.....	39
3.1	Delimitação de conceitos.....	39
3.2	Fontes de dados secundários.....	41
3.3	Modelo teórico.....	41
3.4	Tratamento e apresentação de dados	45
4	APRESENTAÇÃO DE RESULTADOS	47
4.1	Discussão dos resultados	50
4.2	Análise de robustez e desempenho do modelo	62
5	CONCLUSÃO.....	66
	REFERÊNCIA BIBLIOGRÁFICA.....	71
	APÊNDICE A - INFORMAÇÕES BIBLIOMÉTRICAS	81
	APÊNDICE B - INFORMAÇÕES ESTATÍSTICAS.....	94
	APÊNDICE C - TRATAMENTO DE TABELAS EXPERIMENTAIS	105
	APÊNDICE D - PROGRAMAÇÃO EM PYTHON.....	107

1 INTRODUÇÃO

1.1 Apresentação

A Internet vem conquistando, inexoravelmente, novos adeptos ano após ano e o seu uso vem impactando, de forma inovadora e irrevogável, o dia a dia das pessoas, das corporações e organizações, dos diversos segmentos da economia, da academia e do meio científico, da política e dos governos – enfim, dos mercados, das nações e da sociedade globais.

Muito mais do que uma infraestrutura robusta de cabos e ativos eletrônicos interconectados, que suporta um total estimado de 4,66 bilhões de usuários (59,5% da população mundial) (JOHNSON, 2021) e produz, aproximadamente, um volume diário de dados na casa de 2,5 Exabytes (quintilhão) (MARR, 2018), a Internet se converteu em um ícone de comunicação da sociedade pós-moderna e um habilitador da transformação digital (STOLTERMAN; FORS, 2004) na iniciativa privada e no setor público.

A rede contribuiu, nas últimas décadas, com importantes mudanças em diversos segmentos de mercado (e.g. cinematográfico, musical, literário, transporte, atacadista, financeiro, etc.), no estabelecimento de novos modelos econômicos (e.g. economia colaborativa (HAMARI; SJOKLINT; UKKONEN, 2016)), na promoção do ativismo político e no surgimento de movimentos populares contra regimes autoritários (e.g. Primavera Árabe (KHONDKER, 2011) e Revolução Laranja (LYSENKO; DESOUZA, 2014)), na possibilidade de prestação remota de serviços públicos (e.g. telemedicina (BRASIL, 2020a) e videoconferência para atos processuais no Judiciário (BRASIL, 2020b)) e em profundas alterações nas relações interpessoais, entre clientes e empresas e cidadãos e governos.

Infelizmente, ela também está associada a alguns comportamentos condenáveis. Por exemplo, o assédio e a discriminação virtual (TOKUNAGA, 2010)), ao submundo dos crimes no ciberespaço¹, à disseminação em larga escala de narrativas alternativas (“*fake news*”) de fatos verídicos e científicos (HOPF et al., 2019; IYENGAR; MASSEY, 2019) e aos distúrbios psicológicos/psiquiátricos resultantes do seu uso em excesso (e.g. o cibervício (DAVIS, 2001; KUSS; GRIFFITHS, 2011) e a nomofobia (KING et al., 2013; YILDIRIM; CORREIA, 2015).

Em suma, a Internet se faz presente no cotidiano das pessoas, das organizações, dos governos e das nações e se tornou uma ferramenta importante de socialização, aprendizagem, comunicação, emprego e atividades econômicas.

¹Outra forma de referir-se à Internet. Fonte: National Institute of Standards and Technology (NIST).

Estudar os benefícios e os efeitos negativos associados ao seu uso é primordial para compreender os desdobramentos, todavia, parcialmente desvendados, do uso dos recursos dessa rede nos diversos setores da sociedade e nas distintas áreas do conhecimento.

E, particularmente, com relação a políticas públicas, faz-se necessário propor ações que viabilizem o acesso igualitário e universal à rede, reduzam a burocracia e aumentem a efetividade da Administração Pública e regulem aspectos importantes do uso da rede (e.g. Marco Civil da Internet (Lei nº 12.965) e Lei Geral de Proteção de Dados (Lei nº 13.709)).

No intuito de obter uma compreensão mais aprofundada a respeito desta temática e poder municiar, com informações concretas e apropriadas, iniciativas de políticas públicas, locais ou nacionais, com foco na digitalização inclusiva da sociedade em geral, a comunidade científica busca entender as razões e os desdobramentos das desigualdades de acesso, uso, capacidades necessárias, motivações e benefícios da adoção da Internet.

O estudo do “*digital divide*”² tem abordado perspectivas importantes relativas às desigualdades digitais que podem ser identificadas entre países ou grupos de países (ONO; ZAVODNY, 2007), dentro de uma única nação (LOO; NGAN, 2012; VAN DEURSEN; VAN DIJK, 2011), em grupos específicos de indivíduos (HARGITTAI; HINNANT, 2008) ou setores estratégicos de um dado país (e.g. educação (WARSCHAUER, 2003) e saúde (CHOI; DINITTO, 2013)).

E, depois de quase 30 anos de estudos, o tema continua relevante e atual, sobretudo, em um cenário de crise sanitária global, em que as evidências e consequências do abismo digital se apresentam expostas a todos.

1.1.1 O acesso desigual à Internet na sociedade brasileira

Ano após ano, o número de usuários da Internet cresce no país. Segundo o último levantamento do Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), entidade ligada ao Comitê Gestor da Internet no Brasil (CGI.Br), o Brasil possuía, em 2019, aproximadamente 133 milhões de usuários, o que representava uma cobertura de 74% na população brasileira com, no mínimo 10 anos de idade naquele ano (CETIC.BR, 2020, p. 66).

Adicionalmente, de acordo com o Cetic.br, “desde 2009, a proporção de domicílios conectados no Brasil cresce a uma média de 4,3 pontos percentuais ao ano, bem acima da média mundial (2,7), da dos países em desenvolvimento (2,9) e da dos países desenvolvidos (2,6)

² Referenciado, em Português, como abismo digital, lacuna digital e desigualdade digital.

(CETIC.BR, 2020, p. 61). A penetração da Internet na população brasileira é superior “às estimativas de 2019 para o mundo (47%) e para os países em desenvolvimento (54%)”, mas ainda abaixo para a dos países desenvolvidos (87%) (CETIC.BR, 2020, p. 68).

Se analisado em um contexto regional, América Latina e Caribe, o Brasil possui o maior número absoluto de usuários. No entanto, segundo a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), o país possuía, em 2017, uma posição bem abaixo da média dos países membros da organização e um pouco superior à média dos países que compõem a região com relação à penetração do acesso à Internet em suas respectivas populações – ficando atrás de países como Bahamas, Chile, Argentina, Venezuela, Uruguai, República Dominicana e outros (OECD, 2020).

Tabela 1 – Evolução do número de usuários da Internet no Brasil (2015-2019)

Indicador (% população)	2015	2016	2017	2018	2019
População brasileira	58	61	67	70	74
Área de Residência					
Urbana	63	65	71	74	77
Rural	34	39	44	49	53
Faixa Etária					
De 10 a 15 anos	76	76	84	83	84
De 16 a 24 anos	86	86	88	90	92
De 25 a 34 anos	74	80	85	86	90
De 35 a 44 anos	64	66	76	80	83
De 45 a 59 anos	40	46	54	61	68
De 60 anos ou mais	16	19	25	28	34
Grau de Instrução					
Superior	92	95	95	95	97
Médio	81	82	87	88	89
Fundamental	43	46	54	57	60
Analfabeto / Educação Infantil	4	6	9	14	16
Classe Econômica					
A	95	95	96	92	95
B	82	86	89	91	93
C	57	66	74	76	78
D e E	28	35	42	48	57
Regionalidade					
Sudeste	64	69	74	75	75
Nordeste	49	50	58	64	71
Sul	61	60	69	70	75
Norte	51	58	58	70	74
Centro-Oeste	59	63	76	70	76

Fonte: TIC Domicílios (2015-2019)

O uso da rede no país ocorre de maneira particular e em tempos distintos nos diferentes segmentos da sociedade brasileira. Ela começou, primeiramente, nos domicílios de classe A e B, localizados em centros urbanos e formados por pessoas de um alto nível de escolaridade.

Em 2015, mais de 90% dos indivíduos pertencentes à classe A fazia uso da rede, enquanto o número para os indivíduos da classe D e E era de apenas 28% (Tabela 1).

Esse quadro evidencia desigualdades tecnológicas que dificultam a universalização do serviço e a obtenção de forma igualitária dos benefícios associados ao uso frequente da Internet.

A Tabela 1 informa, com base na pesquisa TIC Domicílios, como se deu a evolução da adoção da Internet no Brasil, considerando a área de residência, faixa etária, grau de instrução, classe econômica e regionalidade, no período de 2015-2019 (CETIC.BR, 2016, 2017, 2018, 2019, 2020). Ainda que a taxa de penetração deste serviço venha aumentando para cada indicador presente nela, algumas desigualdades chamam muito a atenção.

Por exemplo, ainda que, em 2019, no meio rural, a maioria das pessoas tenha feito o uso da rede, há uma defasagem, de pelo menos cinco anos, com relação aos centros urbanos. A situação das classes D e E em 2019 era a mesma que a dos indivíduos da classe C em 2015. E a da classe C, em 2019, igual à da classe B anterior a 2015.

Enquanto 84% das pessoas dentro da faixa etária de 10 a 15 anos fez uso da Internet em 2019, nesse mesmo ano, o número para pessoas com, no mínimo, 60 anos foi de 34%. E, por fim, se comparada a adoção daqueles com ensino superior e ensino fundamental, os números são, respectivamente, 97% e 60%.

Segundo Cetic.br, 1 em cada 4 brasileiros, aproximadamente, não possui acesso à rede. Isso representa um universo de marginalizados digitais de, aproximadamente, 20 milhões de domicílios, e 47 milhões de brasileiros com idade igual ou superior a 10 anos, que não têm a oportunidade de gozar dos benefícios esperados do uso desta tecnologia (CETIC.BR, 2020).

Além do acesso propriamente dito, outros aspectos trazem preocupações com relação à desigualdade digital existente no país.

Com o advento da banda larga móvel, a disponibilidade de *smartphones*, a popularização dos aplicativos e o preço mais acessível dos serviços de conexão com a Internet, o celular tornou-se o principal dispositivo de acesso ao ciberespaço para os brasileiros. Em 2015, o telefone celular foi utilizado por 89% dos usuários, enquanto 65% utilizaram computadores de mesa, notebooks ou *tablets* (CETIC.BR, 2016, p. 28).

Essa diferença continuou aumentando e foi acompanhada pela redução do percentual de domicílios que possui computadores. Em 2015, 50% dos domicílios possuía computadores de mesa, notebooks ou *tablets*. Em 2019, o número caiu para 39%.

A presença de computadores nos domicílios brasileiros é evidenciada, principalmente, quando analisado o aspecto econômico. Em 2019, 95% dos domicílios de classe A possuíam

computadores, enquanto, esses aparelhos estavam presentes em somente 12% dos domicílios das classes D e E (CETIC.BR, 2020, p. 63-64).

Se comparados o uso do celular e do computador para o acesso à Internet dentro das classes econômicas, verifica-se que, no caso do celular, independente da classe, 99% dos usuários o utilizam. No caso do computador, os números são, respectivamente 89%, para classe A, 73%, para a classe B, 39% para a classe C e 15% para a classe D.

Esta situação traz uma preocupação, entre os pesquisadores, com relação a uma diferenciação de uso e uma possível desigualdade com relação aos benefícios e às oportunidades associadas ao uso da rede.

Segundo Ono e Zavodny (2007, p. 1135), as diferenças de acesso e uso de recursos de Tecnologia da Informação e Comunicação (TIC), quando considerados os contextos demográficos e socioeconômicos, demonstram os reflexos das desigualdades da própria nação estudada, visto que esses elementos estão inter-relacionados.

Os autores sugerem, ainda, que o fato de grupos vulneráveis terem sido beneficiados por programas nacionais de democratização do acesso à Internet, as desigualdades demográficas e socioeconômicas acabam por contribuir para o estabelecimento de disparidades no uso dos recursos da rede (ONO; ZAVODNY, 2007, p. 1137). Esta visão é consubstanciada pela categorização do abismo digital em níveis de exclusão, sendo que o primeiro está associado ao acesso e o segundo, ao uso e aos conhecimentos necessários para o uso.

1.1.2 As políticas públicas de inclusão e governo digitais no Brasil

O Brasil possui inúmeras políticas públicas que buscam contribuir com a democratização da adoção de TIC junto às comunidades locais, promovendo, particularmente, o acesso à Internet e estimulando o seu uso.

Iniciativas de Wi-Fi público, como o Wi-Fi Brasil (BRASIL, 2021a), buscam disponibilizar livremente o sinal de Internet em áreas públicas e disponibilizar serviços públicos de interesse comum a comunidades locais e seus visitantes. Subsídios de planos de dados de banda larga móvel estimulam a adoção da Internet para usos específicos (e.g. educação a distância (SAO PAULO, 2021)), sem que os usuários tenham que arcar com os custos adicionais deste tráfego.

O Programa de Inovação Educação Conectada (PIEC) do Ministério da Educação busca “apoiar a universalização do acesso à Internet de alta velocidade, por via terrestre e satelital, e fomentar o uso de tecnologia digital na Educação Básica” (BRASIL, 2017a). A reativação da

empresa Telecomunicações Brasileiras S.A. (TELEBRAS) (BRASIL, 2010) e o lançamento do Satélite Geoestacionário de Defesa e Comunicações Estratégicas contribuem com a interiorização da conectividade e o fortalecimento do ecossistema de provedores regionais de Internet. O estabelecimento de parcerias com a participação de organizações públicas (e.g. Nordeste Conectado (BRASIL, 2017b)) buscam trazer agilidade na ampliação dos pontos de acesso, reduzir custos e interligar instituições de ensino e pesquisa.

Esses e vários outros são exemplos de ações de governos municipais, estaduais e federais de promoverem a inclusão digital da população em geral e de ampliarem a adoção dos serviços de governo no formato digital.

Muitas dessas ações são suportadas por marcos legais e regulatórios federais que estabeleceram, como uma prioridade da administração pública, a universalização da conectividade à Internet e da inclusão digital.

Com o aumento da adoção dos serviços digitais por parte da população brasileira e com a necessidade, conseqüentemente, de estabelecer “princípios, garantias, direitos e deveres para o uso da Internet no Brasil” (BRASIL, 2014), o Congresso Nacional decretou e a Presidente da República Dilma Rousseff sancionou, no dia 23 de abril de 2014, a Lei nº 12.965, popularmente conhecida como “Marco Civil da Internet”.

A lei, que regula o uso da Internet no país, estabelece, em seu art. 7º, que o acesso à Internet possui caráter de essencialidade ao exercício da cidadania e estabelece uma série de direitos e garantias aos seus usuários (e.g. privacidade, acessibilidade, qualidade, não suspensão do serviço, etc.) (BRASIL, 2014).

Outro exemplo mais recentemente, foi a alteração da modalidade de outorga, de concessão para autorização, referente aos serviços de telecomunicações, pela Lei nº 13.879 (BRASIL, 2019a). Essa alteração visa ao aumento de investimentos de infraestrutura de banda larga nacional por meio da liberação de investimentos obrigatórios, pelas empresas do setor, em serviços ultrapassados (e.g. telefones públicos) e o acesso a recursos de fundos de telecomunicações (e.g. Fundo de Universalização dos Serviços de Telecomunicações (FUST)).

Por fim, há uma possibilidade de que a licitação das faixas de frequência para o serviço de 5G no Brasil traga, como contrapartida da empresa arrematante, investimentos tecnológicos em áreas tecnologicamente carentes do país (BRASIL, 2021b).

Além da questão do acesso universal, a administração pública busca, adicionalmente, se adaptar aos novos tempos e promover a digitalização dos seus serviços. A história do uso de tecnologia no setor público brasileiro é vasta e muito rica. Mas foi a partir do ano 2000, que as

ações direcionadas ao que, hoje, é chamado de Governo Digital, começaram a ganhar corpo – particularmente no âmbito federal.

Naquela época, no entanto, o uso de tecnologia estava mais voltado à informatização dos serviços e a democratização da informação, o chamado Governo Eletrônico (FILGUEIRAS; FLÁVIO; PALOTTI, 2019). Hoje, as ações estão mais voltadas para a maior participação da população, por meio de plataformas digitais, no estabelecimento de políticas públicas com foco nas tecnologias digitais e na democracia nacional (BRASIL, 2018).

Seguem, abaixo, algumas dessas ações:

- Estratégia de Governança Digital (BRASIL, 2018)
- Estratégia de Governo Digital (BRASIL, 2020c)
- Plano Nacional de Internet das Coisas (BRASIL, 2019b)
- Estratégia Nacional de Segurança Cibernética (BRASIL, 2020d)

No contexto interno da Administração Pública, a busca pela redução das ineficiências e a necessidade de contenção de gastos vem motivando cada vez mais a digitalização de serviços públicos e gerando uma expectativa de maior eficiência. Desde o início do atual governo, 800 serviços já passaram por esta transformação e dos 3.600 serviços disponíveis no sítio Gov.br, 58% deles já são considerados digitais. Pela Estratégia de Governo Digital, a meta é de que o país chegue a 100% em 2022, proporcionando uma redução das despesas governamentais de R\$ 38 bilhões (nos âmbitos federais, estaduais e municipais) entre 2020 e 2025 (BRASIL, 2020e).

Segundo pesquisa TIC Domicílios, no ano de 2019, 68% dos usuários de Internet com 16 anos ou mais, realizaram alguma atividade de governo eletrônico nos 12 meses anteriores à pesquisa (CETIC.BR, 2020).

1.1.3 A pandemia da COVID-19³

A conjuntura da pandemia da COVID-19 afetou, positivamente, a adoção tecnológica nos setores produtivos, na administração pública federal, estadual e municipal e pela população brasileira. Segundo a Comissão Econômica para a América Latina e o Caribe (CEPAL), as “tecnologias digitais têm sido essenciais para o funcionamento da economia e da sociedade durante a crise da pandemia da doença por coronavírus (COVID-19). [...] Avanços que

³ A COVID-19 é uma doença causada pelo coronavírus, denominado SARS-CoV-2, que apresenta um espectro clínico variando de infecções assintomáticas a quadros graves. Fonte: Ministério da Saúde.

deveriam levar anos para se materializar ocorreram em apenas alguns meses” (CEPAL, 2020, p.1, tradução nossa).

No Brasil, como resposta à crise sanitária e aos seus impactos, as empresas e o setor privado tiveram que se reinventar, alterar planos e estratégias, propor soluções alternativas (e.g. *home office*⁴) e adotar o uso de recursos tecnológicos (e.g. Virtual Private Network (VPN)⁵ e Videoconferência) para se manterem produtivas e darem continuidade aos seus negócios.

A administração pública, no uso de suas prerrogativas, permitiu, em caráter temporário, a alteração do serviço público e a adoção de tecnologias digitais. O trabalho remoto ou teletrabalho foi uma das primeiras medidas adotadas pelas entidades públicas (DISTRITO FEDERAL, 2020; MATO GROSSO DO SUL, 2020). Essa modalidade de trabalho não é novidade no serviço público. Alguns órgãos já a utilizavam (BRASIL, 2012) e um projeto de lei para instituí-la amplamente na administração pública já havia sido apresentado e arquivado na Câmara dos Deputados antes da pandemia (BRASIL, 2015), mas, por conta da enfermidade, o seu uso foi amplamente adotado e novos projetos de lei a respeito do tema estão em discussão (BRASIL, 2020f).

O teletrabalho colaborou com o distanciamento social. E outras medidas suportaram a continuidade dos serviços públicos e deram resiliência à administração pública e aos governos federais, estaduais e municipais (e.g. autorização de ensino a distância no ensino superior (BRASIL, 2020g), ampliação da adoção de telessaúde (BRASIL, 2020a), estabelecimento de plataforma emergencial de videoconferência no âmbito da Justiça (BRASIL, 2020h) etc.).

No entanto, algumas das ações que contemplam o uso de TIC, propostas como resposta à crise sanitária global, não são compatíveis com a realidade e as condições de parcelas vulneráveis da sociedade brasileira e de outros países da América Latina. O uso do teletrabalho, por exemplo, não é aplicável a toda atividade profissional e costuma beneficiar, particularmente, trabalhadores com maior nível de escolaridade e com salários mais altos (CEPAL, 2020). E ainda que medidas de auxílio financeiro emergencial (BRASIL, 2021c) tenham ajudado, os desafios iniciais de suas implementações (e.g. cadastro de beneficiários por meio de aplicativos, registros errados e saques presenciais) levaram cidadãos mais vulneráveis à exposição ao vírus em filas e aglomerações em repartições públicas e instituições financeiras.

E para os casos nos quais o uso de TIC seria possível, a desigualdade digital de acesso, uso e habilidades para o uso, impediram a obtenção dos seus benefícios. Enquanto o ensino à

⁴ Conhecido também como teletrabalho.

⁵ Rede Privada Virtual. Tecnologia para o estabelecimento de uma conexão remota segura com as redes corporativas e seus centros de dados.

distância foi adotado por muitas escolas (principalmente, do ensino privado), o ensino básico público ficou paralisado. As consultas médicas e a busca por informações a respeito do coronavírus pela Internet não estão disponíveis a todos os cidadãos, visto que necessitam de dispositivos e periféricos (e.g. impressoras) e conexão com a rede. E serviços públicos disponíveis *online*, só podem ser acessados com o auxílio de outras pessoas e o compartilhamento da conexão e de seus respectivos dispositivos de acesso.

Em suma, a situação atual deixou “mais evidente a exclusão digital de parcela significativa da população, que persiste em um contexto de desigualdades no acesso, nas habilidades digitais e na capacidade de realizar atividades de maneira remota” (CETIC.BR, 2020, p. 59).

Pesquisas a respeito do abismo digital durante o período atual ratificam esta constatação no mundo todo. Segundo van Deursen, as desigualdades previamente existentes são reforçadas, pelo fato de grupos sociais mais favorecidos possuírem mais oportunidades de ganhos com a informação e a comunicação relativas à COVID-19 oriundas do uso da Internet (VAN DEURSEN, 2020, p. 11, tradução nossa). De acordo com Nguyen, Hargittai e Marler, norte-americanos mais velhos, com menos renda, menos escolaridade e poucas habilidades para utilizar a Internet estão propensos a diminuir o uso da rede nas suas atividades corriqueiras (e.g. conversar com parentes e familiares). Os autores sugerem que “a perda de suporte digital presencial e locais de acesso gratuito à Internet devido a medidas de isolamento social” possa ser uma justificativa para este achado (2021, p. 8, tradução nossa).

Relatos e pesquisas em áreas específicas como, por exemplo, saúde (RAMSETTY; ADAMS, 2020) e educação (RODICIO-GARCÍA et al., 2020), confirmam as dificuldades enfrentadas por parcelas sociais marginalizadas digitalmente, com base em seus contextos socioeconômicos, demográficos e culturais, durante este período de restrições a atendimentos e serviços presenciais em países em desenvolvimento e desenvolvidos.

1.2 Justificativa do estudo

Diante da crescente importância que a digitalização dos serviços vem ganhando na sociedade brasileira, faz-se necessário, do ponto de vista de políticas públicas, buscar compreender o contexto dos usuários de Internet no Brasil e analisar os aspectos sociais, econômicos, demográficos, culturais, pessoais, materiais e motivacionais que contribuam com o estabelecimento do abismo digital (denominados, na literatura científica, de determinantes digitais ou determinantes do abismo digital).

Dessa forma, é possível contribuir com o desenvolvimento deste campo de pesquisa, estimulando discussões a respeito do tema e ampliando o conhecimento necessário às iniciativas de políticas públicas em prol da inclusão digital igualitária da sociedade.

1.3 Delimitação do estudo

1.3.1 Objetivo

Este estudo tem, como objetivo, a pesquisa acerca dos determinantes do abismo digital de nível I – também conhecidos como determinantes de acesso – no Brasil. Com este propósito, ele busca identificar determinantes que contribuem com o estabelecimento da desigualdade no acesso à Internet e mensurar suas respectivas influências na sociedade brasileira nos últimos anos (2015-2019).

1.3.2 Hipótese

Pesquisas anteriores demonstram que, com a difusão do acesso à Internet, os impactos específicos dos determinantes vão se alterando – ganhando ou perdendo influência.

Além disso, novas formas de acesso, mais móveis e compartilhadas, e a adoção de dispositivos pessoais, com preços mais acessíveis do que computadores de mesa (e.g. celulares e “tablets”), alteram o *status quo* desses fatores e aumentam as expectativas com relação ao crescimento de usuários.

No que se refere aos determinantes digitais, pesquisas apontam a influência de fatores econômicos (por exemplo, renda e emprego) como um aspecto relevante na adoção da Internet (NISHIJIMA; IVANAUSKAS; SARTI, 2017). E, ao longo do período estudado, o Brasil apresentou melhorias em importantes indicadores econômicos – o que levaria a uma expectativa de redução dos impactos negativos dos determinantes econômicos no que tange ao abismo digital no país.

Por exemplo, o Índice Nacional de Preços ao Consumidor Amplo (IPCA), que foi de 10,67% em 2015, apresentou queda, chegando a 4,31% em 2019⁶. E o Produto Interno Bruto (PIB) real saiu de variações negativas em 2015 e 2016 (com o menor valor de -5,14% a.a. no primeiro trimestre de 2016) para variações positivas a partir de 2017 até 2019⁷.

No entanto, a taxa de desemprego (desocupação) seguiu tendência contrária, saindo de 8,9% no último trimestre de 2015 (início da coleta da TIC Domicílios de 2015) para 12,2% (fim

⁶ Fonte: Instituto Brasileiro de Geografia e Estatística. Índice Nacional de Preços ao Consumidor Amplo.

⁷ Fonte: Pesquisa Nacional por Amostra de Domicílios Contínua - PNAD Contínua.

da coleta da TIC Domicílios de 2019), com pico de 13,7% (no primeiro trimestre de 2017)⁸ – apresentando, assim, piora ao longo do período.

Diante deste quadro de deterioração do cenário empregatício no país, este estudo propõe-se a testar a hipótese de que não houve melhoria do determinante do abismo digital relacionado ao emprego durante o período de 2015-2019.

⁸ Fonte: Pesquisa Nacional por Amostra de Domicílios Contínua - PNAD Contínua.

2 REVISÃO DE LITERATURA

2.1 Introdução ao abismo digital

A expressão “*digital divide*” (traduzida para Língua Portuguesa, normalmente, como abismo ou fosso ou lacuna ou exclusão digital) é utilizada amplamente, na literatura científica, para se referir às disparidades dos acessos a recursos de TIC, por exemplo, computadores (WILSON; WALLIN; REISER, 2003), infraestrutura de comunicações (LOO; NGAN, 2012), celulares (NISHIJIMA; IVANAUSKAS; SARTI, 2017) e Internet (DIMAGGIO et al., 2001) e pelos fatores que os explicam.

Ela forma parte do campo semântico, juntamente com outras palavras como, por exemplo, “*digital literacy*” (letramento ou alfabetismo digital) e “*digital inequality*” (desigualdade digital), que cobrem questões relativas às desigualdades sociais associadas à marginalização tecnológica (digital) de comunidades ou segmentos demográficos (SCHEERDER; VAN DEURSEN; VAN DIJK, 2017).

A primeira ocorrência de seu uso, encontrada nesta revisão de literatura, foi em um artigo, no ano de 1997, que analisa, com base em dados de uma pesquisa nacional realizada nos Estados Unidos da América (EUA) no mês de outubro de 1995, as motivações e barreiras do uso da Internet. Nele, os autores apontam a existência de um abismo digital que poderia trazer importantes consequências sociais e políticas, caso o mesmo se perpetuasse a longo prazo (KATZ; ASPDEN, 1997).

Com o passar do tempo, o tema foi ganhando cada vez mais relevância e atenção da comunidade científica, de órgãos multilaterais, da imprensa e dos meios de comunicação e de governos e, eventualmente, a expressão ganhou conotação específica a assuntos relacionados às desigualdades do acesso, do uso e dos benefícios relativos à adoção da Internet. Esse aumento de interesse pode estar associado, em parte, à difusão global da Internet, às expectativas de benefícios e ganhos por indivíduos, organizações e nações, advindos da adoção desta tecnologia, à preocupação política com relação às consequências negativas de um processo de digitalização não-inclusivo e, por fim, ao desejo acadêmico em pesquisar e compreender como se dá esse mesmo processo.

Neste contexto específico, a primeira abordagem adotada pelo meio científico está relacionada, particularmente, à questão do comparativo entre usuários e não usuários da Internet – e os seus respectivos indicadores socioeconômicos e demográficos e outras características relativas ao serviço de conectividade (FERRO; HELBIG; GIL-GARCIA, 2011). Ainda assim,

diferentes focos podem ser identificados em artigos mais específicos. Por exemplo, conforme Loo e Ngan (2012, p. 888), o abismo digital não se refere, somente, à disponibilidade da tecnologia, mas a outras questões relativas à sua confiança, ao seu uso, à velocidade, ao acesso amplo e ao custo acessível.

Com a sofisticação das pesquisas e o interesse por outras questões como, por exemplo, os tipos de uso, as habilidades e as motivações dos indivíduos para adentrarem ao ciberespaço, o conceito foi sendo ampliado para contemplar outras perspectivas. Como, por exemplo, aquela definida por DiMaggio et al., apresentada abaixo:

Por abismo digital, nos referimos às desigualdades no acesso à Internet, extensão de uso, conhecimento das estratégias de busca, qualidade das conexões técnicas e suporte social, capacidade de avaliar a qualidade da informação e diversidade de usos. (DIMAGGIO et al., 2001, p. 310, tradução nossa).

Visões mais amplas e multidisciplinares, que contemplam a realidade da desigualdade digital, os seus contextos socioeconômicos particulares e populações específicas também foram sendo desenvolvidos ao longo dos anos. De acordo com Gil-Garcia, Helbig e Ferro (2006, p. 3, tradução nossa), ela é “[...] muitas vezes caracterizada como algum tipo de relação entre as tecnologias de informação e comunicação (TICs) e grupos de indivíduos, que estão situados em um arranjo complexo de questões sociais, ambientais, políticas e econômicas.”

Em outras palavras, as tecnologias se moldam de acordo com as realidades locais e os seus costumes e suas prioridades (DIMAGGIO et al., 2001) e a sua adoção pode vir a ser a resultante de um fenômeno social, em que indivíduos, mesmo com características socioeconômicas, demográficas e culturais divergentes daquelas normalmente esperadas de um usuário comum, acabam sendo convencidos ou levados a se tornarem adeptos tecnológicos por influência de pessoas ou grupos de suas comunidades locais (AGARWAL; ANIMESH; PRASAD, 2009).

Desde os seus primórdios, o abismo digital tem despertado o interesse do meio acadêmico e gerado divergências com relação aos benefícios sociais que poderiam ser adquiridos pela adoção e universalização do uso da Internet.

Há pesquisadores, que impulsionados pelas possibilidades de ganhos (e.g. oportunidade de emprego, educação ampla e acesso a informação, uso de serviços de governo digital e eletrônico, participação política, estabelecimento de redes sociais, etc.) potencializados pela

Revolução da Informação e a digitalização da sociedade, acreditam na redução da desigualdade social por meio do acesso à informação, que será viabilizado pela redução do seu custo, permitindo, possivelmente, a ascensão socioeconômica daqueles menos afortunados (DIMAGGIO et al., 2001).

Há outros, que, por sua vez, acreditam que os grandes beneficiados pelos ganhos esperados com o uso da Internet serão as classes mais abastadas, que estarão sempre à frente do restante da população, particularmente, por conta das desigualdades de acesso e da natureza distinta entre os serviços de informação (e.g. telefone, cabo, etc.), que necessitam de pagamentos recorrentes, e bens de informação (e.g. rádio, televisão, etc.), que estão associados a compras de mercadorias (DIMAGGIO et al., 2001; NEGROPONTE, 1995).

Nesta situação, a digitalização da sociedade contribuiria com o agravamento da desigualdade social, permitindo o estabelecimento da segregação da sociedade digital entre “usuários” ou “ganhadores” e “não-usuários” ou “perdedores” (BONFADELLI, 2002).

Segundo DIMAGGIO et al., os desdobramentos da adoção desta tecnologia são mais limitados do que as expectativas dessas visões divergentes sugerem e o seu impacto depende de como “atores econômicos, regulamentação governamental e usuários organizam coletivamente” (DIMAGGIO et al., 2001, p. 310, tradução nossa) a evolução desta tecnologia.

E conforme Cullen assinala,

A tecnologia, por si só, não resolve discrepâncias sociais ou econômicas dentro de uma sociedade e [...] novas tecnologias nem sempre substituem tecnologias antigas. Elas podem coexistir e, ao fazê-lo, aumentam a faixa de experiência humana sem, necessariamente, depreciar a experiência daqueles que não a utilizam, pois preferem as opções legadas para alcançar os mesmos fins. (CULLEN, 2001, p. 312, tradução nossa).

É importante ressaltar que a falta de uma consistência terminológica e um arcabouço teórico sólido dificultam o estabelecimento de avanços estruturados e a possibilidade de validação metodológica e sistêmica dos achados. Conforme BONFADELLI constata,

Não só faltam dados empíricos sólidos que possam demonstrar, por exemplo, as vantagens do acesso à Internet sobre o uso dos meios de comunicação tradicionais, mas mesmo de uma perspectiva teórica também não é claro se o postulado político de acesso à Internet para todos será o fator necessário para o sucesso no futuro - ou se o acesso à mídia ou às informações da Internet

serão relevantes de alguma forma. (BONFADELLI, 2002, p. 66, tradução nossa).

Um avanço, nesta direção, foram as publicações de revisões sistêmicas de literatura, genéricas ou temáticas, que ajudaram a estruturar e organizar esse arcabouço teórico (SCHEERDER; VAN DEURSEN; VAN DIJK, 2017).

2.2 Evolução e diferenciação de abordagens

A pesquisa acerca do abismo digital vem sendo norteadada pelo estudo e pela busca da compreensão das lacunas de conhecimento, que proíbem que as pessoas se beneficiem ou maximizem os seus ganhos com relação às promessas do uso da Internet. De acordo com Bonfadelli (2002, p. 71), existem, no mínimo, 04 barreiras associadas a estas lacunas:

- Falta de habilidades e conhecimentos em informática e existência de receios e atitudes negativas, por parte, principalmente, de pessoas idosas ou com baixo nível educacional, com relação à Tecnologia da Informação e Comunicação (TIC).
- Mesmo com conhecimentos tecnológicos prévios, existem barreiras de acesso (principalmente econômicas) aos meios e ferramentas necessários ao acesso.
- Falta de recursos que estimulem ou melhorem a experiência do usuário (“*user friendliness*”).
- Lacunas associadas à forma como se dá o uso da Internet.

Cullen (2001, p. 313-315) ratifica, qualquer proposição de iniciativa para endereçar o abismo digital deve contemplar ações que combatam as 04 barreiras apontadas logo abaixo:

- Existência de estruturas robustas de TIC, que permitam o acesso robusto e confiável à Internet, com banda suficiente compatível o uso variado dos recursos do ciberespaço, a um preço acessível e disponível à sociedade em geral.
- Analfabetismo e falta de conhecimentos, habilidades e suporte em informática (analfabetismo digital).
- Barreiras comportamentais e atitudes negativas com relação ao uso tecnológico.
- Desinteresse ou irrelevância de conteúdos.

2.2.1 A Hipótese da Lacuna do Conhecimento

A Hipótese da Lacuna do Conhecimento (“*Knowledge Gap Hypothesis*”), formulada por Phillip Tichenor, George Donohue e Clarice Olien em 1970, estabelece que:

À medida que a difusão de informações de mídia de massa em um sistema social aumenta, segmentos da população com status socioeconômico mais elevado tendem a adquiri-las em uma taxa mais veloz do que os segmentos com status mais baixos, de modo que a lacuna do conhecimento entre esses segmentos tende a aumentar ao invés de diminuir. (TICHENOR; DONOHUE; OLIE, 1970, tradução nossa).

Bonfadelli (2002) ratifica que a hipótese não afirma que parcelas menos abastadas economicamente da sociedade, ou com menor grau de educação, não terão, eventualmente, acesso a tais informações, permanecendo ignorantes, indefinidamente, com relação assunto em questão.

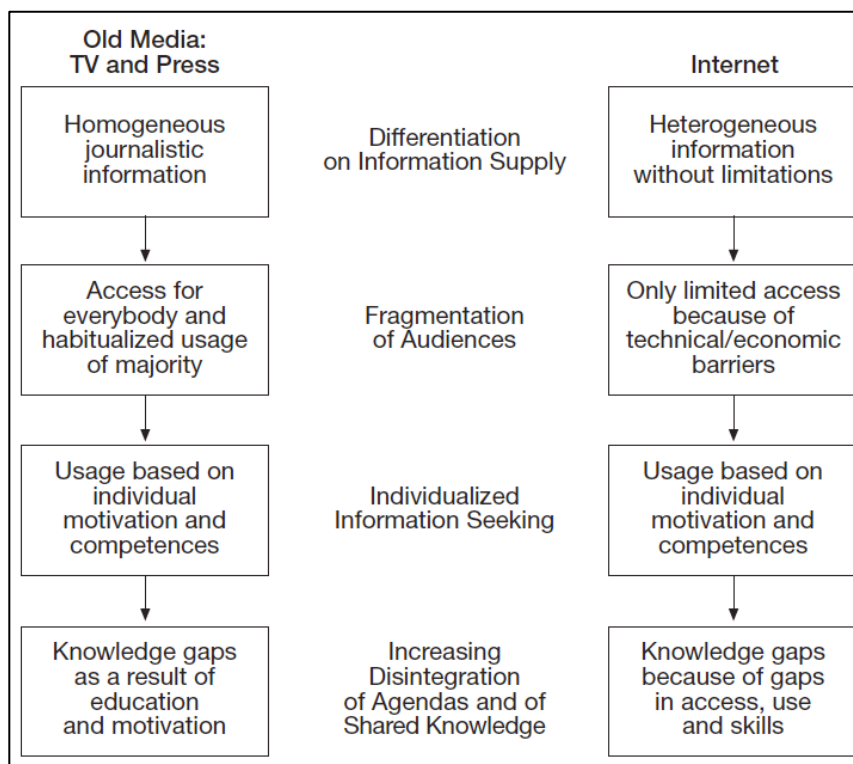
Mas que, na prática, aqueles em uma situação economicamente mais favorável se beneficiarão deste acesso exclusivo, posicionando-se em uma situação privilegiada perante aqueles que não gozam da mesma oportunidade.

A situação caracterizada pela lacuna do conhecimento está associada, à maneira assimétrica e desigual, como se dá o fluxo de informação, que é consequência da estrutura social presente em uma determinada sociedade e pode ser justificada por cinco fatores (habilidade de comunicação, conhecimento prévio, contatos sociais relevantes, uso seletivo, aceitação e armazenamento do conteúdo e, por último, estrutura do sistema de mídia) – desses, quatro, segundo Bonfadelli, são normalmente influenciados pelo grau de educação de um indivíduo qualquer (BONFADELLI, 2002 p. 68).

Com relação à estrutura do sistema de mídia, a Internet não é exatamente igual aos meios de comunicação tradicionais (e.g. jornais, TV, revistas e rádios). E comparações entre eles são inevitáveis. De acordo com Bonfadelli (2002) e ilustrado na Figura 1, no caso das mídias tradicionais, se considerarmos a era anterior ao estabelecimento da Internet como veículo de informação, as lacunas de conhecimento estavam associadas às diferenças de interesse relacionados a temas específicos, visto que essas possuem acesso popular e comunicação em massa, as informações são homogêneas e compartilhada por jornalistas legítimos. Por sua vez, a Internet possui uma enorme heterogeneidade no que tange à natureza das informações, que muitas vezes não são produzidas, necessariamente, por profissionais especializados. Em outras

palavras, “a Internet promove a fragmentação da audiência e a busca individualizada de informações; e isso poderia resultar em uma desintegração crescente de agendas individuais e da quantidade de conhecimento compartilhado.” (BONFADELLI, 2002, p. 73, tradução nossa).

Figura 1 – Lacunas de conhecimento dos meios de comunicação



Fonte: Bonfadelli (2002, p. 73)

Hargittai e Hinnat (2008) apresentam evidências de pesquisas complementares que buscam relacionar a ocorrência da lacuna de conhecimento com questões adicionais além daquelas econômicas – como, por exemplo, motivação, nível educacional e uso de mídias.

Com relação ao conhecimento prévio para o uso da Internet especificamente, estudos apontam que as habilidades do usuário para realizar pesquisas podem interferir no conteúdo que ele consome (BONFADELLI, 2002) e a natureza do uso (e.g. profissional, educacional) que ele faz dos recursos digitais disponibilizados no ciberespaço também são fatores que contribuem com o estabelecimento de lacunas de conhecimentos.

A produção científica a respeito do abismo digital costuma ser classificada de acordo com as temáticas estudadas, as suas abordagens e suas lacunas específicas. Conforme Scheerder, van Deursen e van Dijk (2017), existem:

- Abismo digital de nível I – pesquisas que abordam a lacuna de acesso e as desigualdades da conectividade com a Internet.

- Abismo digital de nível II – pesquisas que abordam as lacunas de uso e de habilidades.
- Abismo digital de nível III – pesquisas que abordam as desigualdades dos benefícios resultantes.

Ferro, Helbig e Gil-Garcia (2011) identificam os mesmos níveis, mas os denominam respectivamente de:

- Abordagem dicotômica
- Abordagem multidisciplinar
- Abordagem multiperspectiva

2.2.2 Abismo digital de nível I

O abismo digital de nível I caracteriza-se pelo foco no acesso à Internet e na propriedade ou disponibilidade das ferramentas ou dos recursos que o provêm. Nesta abordagem, os pesquisadores demonstram as diferenças entre as populações de usuários e não usuários da Internet (abordagem dicotômica) (RICE; KATZ, 2003).

Do ponto de vista das lacunas de conhecimento, essas pesquisas tratam de compreender as causas da inexistência do acesso (“*gap in access*”) e identificar os seus determinantes por meio de análises de informações sócio-demográficas e econômicas.

Alguns trabalhos que adotam esta linha de pesquisa trazem a perspectiva do acesso à Internet como produto ou *commodity* qualquer e da sua adoção como um processo de comercialização (ZHANG, 2013). Neste contexto, as lacunas ou desigualdades existentes referem-se, somente, às consequências temporárias das diferentes taxas do processo de difusão dessa tecnologia, que, com o passar do tempo, deixariam de existir – sem a necessidade de intervenção. Em outras palavras, uma vez conectado, todos teriam as mesmas possibilidades de uso e poderiam se beneficiar, de forma igualitária, das mesmas oportunidades do uso da tecnologia e do acesso a informações (GIL-GARCIA; HELBIG.; FERRO, 2006).

Por considerarem ultrapassada e já resolvida a situação do acesso, alguns pesquisadores se dedicam a outros aspectos multidimensionais do abismo digital.

No entanto, vale destacar que:

- As pesquisas que estabeleceram os alicerces para este tipo de abordagem foram realizadas no início do século XXI e muitas delas utilizam, como paradigma de acesso, o uso de computadores conectados à Internet nas casas e escolas de

usuários (WEI et al., 2011). Atualmente, o acesso tem características móveis e, em alguns casos, compartilhada. No Brasil, o celular, e não o computador, é o dispositivo mais importante para o acesso à Internet – sendo utilizado por 99% dos seus usuários (CETIC.BR, 2020). “Esses resultados, por um lado, refletem mudanças importantes nos padrões de acesso e de uso das TIC.” Por outro lado, podem representar a amplificação de desigualdades entre os que acessam e os que não acessam diferentes dispositivos (CETIC.BR, 2015, p. 132).

- Gil-Garcia, Helbig e Ferro (2006) relembram que recursos de TIC englobam uma série de tecnologias distintas como, por exemplo, computadores, rádios, satélites, telefones celulares e etc. Adicionalmente, os recursos disponíveis na Internet “são acessíveis por meio de uma variedade de plataformas - de computadores a televisão digital a telefones móveis WAP” (SELWYN, 2002, p.7, tradução nossa). No entanto, inúmeras pesquisas tendem a tratar os recursos de TIC de maneira homogênea e indiscriminada, sem levar em consideração as suas particularidades e possíveis impactos na adesão do serviço. Elas não costumam considerar as distinções entre esses recursos, suas evoluções tecnológicas e os seus respectivos impactos com relação ao abismo digital.
- O acesso universal à Internet não é uma realidade na maioria dos países – principalmente, naqueles que não são enquadrados como desenvolvidos. E mesmo em países com números absolutos, é comum encontrar indícios de populações marginalizadas. Em artigo publicado recentemente, van Deursen e Van Dijk sugerem que:

O abismo digital de primeiro nível continua sendo um problema em um dos países mais ricos e tecnologicamente avançados do mundo [Holanda]. Ao estender o acesso físico básico combinado com o acesso a materiais, o estudo conclui que uma diversidade no acesso a dispositivos e periféricos, oportunidades relacionadas a dispositivos e as despesas contínuas necessárias para manter o hardware, software e assinaturas afetam as desigualdades existentes relacionadas às habilidades de Internet, usos e resultados (VAN DEURSEN; VAN DIJK, 2019, p. 354, tradução nossa).

2.2.3 Abismo digital de nível II

O abismo digital de nível II parte da proposição de que não há um único abismo digital (HARGITTAI; HINNANT, 2008) e o acesso, ou a conectividade, ainda que seja requisito básico, não se traduz, por si só, em valor ou benefício ao usuário. “Em vez disso, o seu uso, geralmente, determina quanto valor indivíduos obtêm da Internet” (GOLDFARB; PRINCE, 2008, p. 2, tradução nossa).

Segundo Barzilai-Nahon (2006, p. 2, tradução nossa), “[...] o pensamento tradicional orientado para o acesso se concentrava em questões relacionadas a medidas como propriedade, disponibilidade e acessibilidade da infraestrutura, agora o foco está indo além da tecnologia para os usuários.”

A pesquisa acerca do abismo digital de nível II aborda o uso da Internet e questões como a real motivação, o objetivo final e as dificuldades enfrentadas pelos usuários. A relevância de aspectos como o analfabetismo digital (“*digital illiteracy*”), questões comportamentais e culturais e tipos de conteúdos acessados são tópicos que também podem ser encontrados na literatura (FERRO; HELBIG; GIL-GARCIA, 2011).

Do ponto de vista das lacunas de conhecimento, as pesquisas tratam de compreender as questões multidimensionais relacionadas ao uso e às habilidades (“*gap in use*” e “*gap in skills*”) dos usuários. Os pesquisadores buscam entendê-las por meio de tentativas elaboradas de se mensurar o alfabetismo digital e a fluência digital (HARGITTAI; HINNANT, 2008). Métodos utilizados costumam envolver a criação de tipologias de usuários (e.g. básicos ou avançados, novatos ou veteranos e conectados ou parcialmente conectados) (BONFADELLI, 2002; FERRO; HELBIG; GIL-GARCIA, 2011; RICE; KATZ, 2003), a categorização de habilidades (e.g. relativas ao conteúdo) (SCHEERDER; VAN DEURSEN; VAN DIJK, 2017) e métricas de frequência (BRANDTZAEG; HEIM; KARAHASANOVIC, 2011).

2.2.4 Abismo digital de nível III

Segundo WEI et al. (2011, p.170, tradução nossa), “o abismo digital de resultados (o abismo digital de terceiro nível) é a desigualdade de resultados (e.g. aprendizagem e produtividade) da exploração de TIC decorrente da exclusão digital de segundo nível e outros fatores contextuais”. Ele relaciona-se às lacunas nas capacidades individuais dos usuários de traduzir o acesso e o uso da Internet em resultados favoráveis fora da rede (VAN DEURSEN, A.J.A.M. AND HELSPER, 2015, p. 30, tradução nossa) .

Em outras palavras, o abismo digital de nível III trata das desigualdades do uso de TIC e dos seus efeitos na possibilidade de obtenção de capital econômico (e.g. rede de contatos profissionais e acesso a bens e serviços), capital cultural (e.g. conhecimento) e capital social (e.g. relacionamentos) (GÓMEZ, 2020 p. 4-5).

Essa visão leva em consideração o fato de indivíduos e comunidades empregarem tecnologias para fins específicos, relacionados frequentemente às suas histórias e localidades (Hines, Nelson, & Tu, 2001, p. 5).

Do ponto de vista das lacunas de conhecimento, Gómez refere-se àquela associada ao abismo digital de nível III como lacuna de utilidade (“*utility gap*”) (2020, p.70).

2.3 Os determinantes do abismo digital

O termo determinante é utilizado, na literatura científica, para se referenciar aos aspectos sociais, econômicos, demográficos, culturais, pessoais, materiais e motivacionais que afetam a existência do abismo digital.

Gil-Garcia, Helbig e Ferro argumentam que a desigualdade digital está relacionada a “[...] outras características que definem as oportunidades sociais de um indivíduo como o sexo, a educação, a empregabilidade, as habilidades em informática, a capacidade de falar uma língua estrangeira – entre outras.” (GIL-GARCIA; HELBIG.; FERRO, 2006, p. 2, tradução nossa).

Dentre essas, algumas são particulares a estudos relativos a um determinado nível ou abordagem do tema tratado (e.g. motivação, habilidades). Outras, como, por exemplo, renda e educação, são amplamente utilizadas em toda produção científica desse campo de pesquisa.

A unificação dos conceitos e a criação de uma terminologia universal que trata desses indicadores é algo difícil de ser feito, no entanto, tornaria a literatura mais clara e mais fácil de ser administrada.

No intuito de avançar nesta direção, a revisão sistêmica de literatura que trata dos níveis II e III do abismo digital (SCHEERDER; VAN DEURSEN; VAN DIJK, 2017) estabelece 7 categorias de classificação dos determinantes (Figura 2).

Figura 2 – Categorias de determinantes do abismo digital

Sociodemographics	Age, gender, marital status, residency, living area, living environment, urban/rural dimension, life space.
Economic	Income, household income, household wealth, household poverty, family income, SES, household SES, individual SES, owning goods, financial situation, work situation, employment status, employment type, employment status parents, occupational status, social class, life stage, educational level, Years of schooling, educational resources, parental education, educational level parents, school sector, academic orientation, doing homework online, working hours, type of activity - job seeking.
Social	Household composition, family size, family composition, family living arrangement, parental status, having children, parental mediation, number of children, informal networks, connected family members, connected friends, amount of Facebook friends, network size online, socializing, social activity, social support, assessing digital support networks, Facebook friends' instrumental support, formal volunteering, degree of social isolation, social orientation, loneliness, type of activity - e-mailing, type of activity - Facebook, type of activity - social media, type of activity - social network, type of activity - Instant Messaging, type of activity - Social Networking Sites, Facebook interactions, express political content (Facebook), connections w/ political actors (Facebook), political networks, type of activity - social media for political purposes, cyber political participation, political orientation.
Cultural	Cultural, cultural capital, cultural status, cultural possessions, religion, ethnicity, Internet use language.
Personal	Type of activity - information seeking, type of activity - entertainment, type of activity - web support, groups (health related), type of activity - downloading/listening to music, type of activity - gaming, type of activity - podcast use, type of activity - online news, online news use, media use, traditional news media use, online media multiplexity, amount of media, offline news consumption, language integration, traditional literacy, literacy, language skills, English skills, previous achievements, school performance, academic performance, grade level, mastery orientation, shyness, confidence, self-efficacy, cognitive function, health status, mental health, health condition, health interests, physical activity, offline health activities, seeking offline health information, satisfaction with physician, trust in online health information, personality traits - neuroticism, personality traits - extraversion, personality traits - conscientiousness, personality traits - openness, personality traits - Agreeableness, psychological distress.
Material	Internet availability, Internet access, access locations, home access, home ICT access, school access, access type, access quality, number of electronic devices, PC at home, use of other technologies.
Motivational	Attitude towards ICTs, attitude towards computers, Internet attitude, Internet motivation, perceived Internet relevance, Internet use, frequency of Internet use, usage frequency, (amount of) time spent online, intensity of Internet use, Internet experience, years of experience, digital skills, Internet skills, e-skills, computer skills, ICT skills, operational skills, formal skills, information skills, strategic skills, medium-related skills, creative skills, ICT competence, digital competence, media literacy, Internet literacy, digital literacy, Internet efficacy, eHealth literacy, ICT autonomy, technological efficacy, Internet use at work.

Fonte: Scheerder, Van Deursen e Van Dijk (2017, p. 1620)

2.3.1 Situação econômica

A situação econômica (e.g. a renda e o emprego) é um determinante muito importante na adoção da Internet. Inúmeros artigos corroboram o impacto positivo desta variável independente com a probabilidade de um indivíduo ser usuário dos seus serviços (CHAUDHURI; FLAMM; HORRIGAN, 2005; FERRO; HELBIG; GIL-GARCIA, 2011; NISHIJIMA; IVANAUSKAS; SARTI, 2017).

Segundo Bonfadelli (2002), essa constatação poderia ser explicada por conta das barreiras econômicas associadas aos custos de equipamentos (e.g. computador) e serviços de acesso à rede (e.g. telefonia).

No entanto, se superados esses empecilhos econômicos para o acesso, é possível encontrar resultados contraditórios com relação ao uso. Conforme Goldfarb e Prince (2008) relatam, há observações de uma relação negativa desta variável independente com o tempo de uso. Em outras palavras, indivíduos com menor renda tendem a passar mais tempo conectados do que aqueles com uma situação econômica mais favorável. Os autores especulam que a diferença entre adoção e uso efetivo possa ser explicada levando em consideração o custo de oportunidade, utilidade do serviço, tempo de lazer e valor agregado.

No entanto, Haight, Quan-Haase e Corbett (2014) informam de uma expectativa de que usuários com maior renda realizem mais atividades *online*. Todavia, ao analisar o uso de aplicações específicas, como os casos de sítios de redes sociais, não há indícios de que a renda tenha um efeito estatisticamente significativo.

2.3.2 Nível educacional

Outra variável independente que costuma ser identificada como um importante determinante para a adesão da Internet é o nível educacional.

No que tange ao acesso à rede, estudos demonstram uma maior concentração de usuários entre os indivíduos com maior grau de instrução (HAIGHT; QUAN-HAASE; CORBETT, 2014; VAN DEURSEN; VAN DIJK, 2011).

Adicionalmente, a educação está relacionada com as questões de habilidades, usos e resultados, estudados nas pesquisas de abismo de níveis II e III. “A Internet não é, por si só, educação, ela não alfabetiza e ela requer habilidades altamente desenvolvidas para acessar e interpretar as informações encontradas” (CULLEN, 2001, p. 312, tradução nossa).

Segundo (BONFADELLI, 2002, p. 76, tradução nossa), “pessoas com maior nível educacional [...] dispõem de mais recursos cognitivos como pré-requisitos necessários para o uso da Internet”. Segundo o pesquisador, os usuários mais experientes e mais assíduos são aqueles que possuem um maior grau educacional. Eles usam os recursos presentes na rede de uma forma instrumental (e.g. informações e serviços), enquanto usuários com menor grau educacional recorrem ao entretenimento *online* (SHAH; KWAK; HOLBERT, 2001).

Adicionalmente, segundo (VAN DEURSEN; VAN DIJK, 2014) constata, indivíduos em uma situação educacional inferior usam a Internet por um período maior durante o seu tempo livre.

Por fim, além do nível educacional, há outras questões que, frequentemente, são estudadas. O fato de uma pessoa ser um(a) aluno(a), ou seja, estar estudando, por exemplo, é um fator relevante para predizer se ela será assinante da Internet, segundo (CHAUDHURI; FLAMM; HERRIGAN, 2005).

2.3.3 Sexo

Com relação ao sexo, as pesquisas trazem resultados divergentes. Algumas o apontam como um determinante para o acesso (GOLDFARB; PRINCE, 2008), mas outras apresentam resultados contrários (WILSON; WALLIN; REISER, 2003).

Adicionalmente, há estudos que apontam diferenças significativas das atividades *online* entre homens e mulheres, demonstrando que homens são mais ativos (HARGITTAI, 2007) e completam mais atividades do que as mulheres (HAIGHT; QUAN-HAASE; CORBETT, 2014). Eles também consomem mais conteúdos orientados a notícias e lazer (VAN DEURSEN; VAN DIJK, 2014).

Por outro lado, é conferido, ao público feminino, o maior uso de redes sociais (HAIGHT; QUAN-HAASE; CORBETT, 2014) e jogos (VAN DEURSEN; VAN DIJK, 2014). De acordo com Hargittai (2007), sexo é o fator demográfico, dentre aqueles estudados, estatisticamente mais relevante com relação ao uso de redes sociais.

2.3.4 Idade

Pesquisas apontam a idade com um determinante do abismo digital e identificam a sua relação inversamente proporcional à probabilidade de uso da Internet (CHAUDHURI; FLAMM; HARRIGAN, 2005; GOLDFARB; PRINCE, 2008; RICE; KATZ, 2003).

Segundo Gilleard e Higgs (2008) sugerem, as tentativas de se explicar essa constatação, geralmente, estão associadas ao contexto da desigualdade social relativa à população idosa (e.g. baixa renda, baixa educação, saúde debilitada, exclusão laboral), dificuldades de aprendizagem, aversão natural a novas tecnologias e choque de gerações (mundo mecânico versus mundo eletrônico).

2.3.5 Localidade

Existem vários aspectos no que tange à localidade como um determinante do abismo digital. Um deles diz respeito ao binômio urbano-rural e ao fato de moradores de centros urbanos terem uma maior probabilidade de serem usuários da Internet (GOLDFARB; PRINCE, 2008; HAIGHT; QUAN-HAASE; CORBETT, 2014).

Segundo Loo e Ngan (2012) constata, levar a infraestrutura necessária para suportar a instalação do serviço a localidades remotas pode ser trabalhoso e custoso. Assim, a difusão da Internet, normalmente, acontece, primeiramente, nos centros urbanos, legando às áreas rurais uma lacuna temporal de acesso.

Além disso, de acordo com van Deursen e van Dijk (2014), moradores de centros urbanos permanecem *online* por períodos mais longos e participam de um maior volume de interações sociais do que aqueles que moram em áreas urbanas.

Outro aspecto diz respeito ao número de localidades de onde se dá o uso da Internet. Segundo Hargittai e Hinnat (2005) sugerem, indivíduos que têm a possibilidade de se conectarem de diversos locais (e.g. casa, trabalho, etc.) estão propensos ao uso com propósitos benéficos a si mesmos.

2.3.6 Outros determinantes

Além dos determinantes comentados anteriormente, esta revisão de literatura encontrou artigos que tratam, adicionalmente, dos seguintes aspectos:

- Comunidades específicas – Cullen (2001) sugere a existência de barreiras que podem se aplicar a comunidades indígenas, minorias culturais e comunidades rurais de analfabetos.
- Estado civil – pesquisas revelam um efeito positivo e significativo do estado civil no acesso à Internet nos domicílios de usuários (CHAUDHURI; FLAMM; HERRIGAN, 2005; GOLDFARB; PRINCE, 2008).
- Etnia – Chaudhuri, Flamm e Herrigan (2005) argumentam de aspectos étnicos e raciais como determinantes do abismo digital de nível I – independentemente de outros fatores (e.g. renda).
- Religião – De acordo com pesquisa de D’Haenens, Koeman e Saeys (2007, p. 288), a religião é um fator que influencia o uso de TIC entre os jovens de origem Turca na Holanda. E, segundo Lissitsa e Chachashvili-Bolotin (2015, p. 47-48, tradução nossa), com relação ao abismo digital em Israel, “estudos indicam uma taxa crescente de acesso à Internet e variedade de usos digitais à medida que o nível de religiosidade diminui”. Adicionalmente, os resultados da regressão logística do estudo indicam que “no grupo mais jovem, os efeitos da religiosidade e da renda familiar tornaram-se mais fortes com o tempo, enquanto entre os idosos esses efeitos não mudaram” (LISSITSA; CHACHASHVILI-BOLOTIN, 2015, p. 53, tradução nossa).
- Situação empregatícia – Chaudhuri, Flamm e Herrigan (2005) sugerem que o emprego não está, estatisticamente, relacionado a ser um assinante de um serviço de conectividade à Internet.

2.4 Bibliometria

Para o processo de revisão de literatura desta dissertação, buscou-se, primeiramente, identificar o universo de artigos científicos associados ao tema tratado nesta dissertação. A busca foi realizada na base de dados do sítio *Web of Science* utilizando-se a expressão de pesquisa abaixo:

TOPIC: ("digital divide" OR "digital inequality" OR "digital literacy") AND **TOPIC:** ("Internet")

O resultado foi um total de 3.017 artigos direcionados ao estudo do abismo digital (“*digital divide*”) e assuntos correlatos como alfabetismo digital (“*digital literacy*”), desigualdade digital (“*digital inequality*”) e Internet (“*Internet*”).

Posteriormente, a relação de artigos identificados foi exportada para formato BibTeX e carregada na plataforma *bibliometrix* (ARIA; CUCCURULLO, 2017), para análises de produção científica e identificação das fontes, dos artigos e dos autores mais relevantes e de maiores impactos.

As informações referentes a esses achados podem ser encontradas no APÊNDICE A. Nele, é possível encontrar o histórico da produção científica a respeito do abismo digital (

Figura A-1 e

Figura A-2) e informações referentes às suas fontes (

Figura A-3,

Figura A-4,

Figura A-5 e

Figura A-6), aos seus autores (

Figura A-7,

Figura A-8,

Figura A-9 e

Figura A-10) e às instituições acadêmicas (

Figura A-11).

3 METODOLOGIA

Esta dissertação está fundamentada em epistemologia positivista e objetiva responder, por meio de abordagem quantitativa, mediante uso de análises matemática e estatística, a pergunta que trata da evolução dos determinantes digitais no Brasil e dos seus respectivos efeitos para o estabelecimento do abismo digital – representado, neste estudo, pelas desigualdades do acesso à Internet.

3.1 Delimitação de conceitos

No intuito de facilitar a compreensão deste estudo, faz-se necessário, preliminarmente, definir os significados dos termos “acesso à” e “uso da” Internet dentro deste trabalho acadêmico.

De acordo com a revisão de literatura presente neste documento, essas palavras referem-se diretamente às categorias de abismo digital – nível I (acesso) e nível II (uso) – e, conseqüentemente, tratam de aspectos e dimensões distintas do mesmo problema.

Muitos dos artigos consultados referem-se à questão do acesso à Internet como a propriedade ou disponibilidade de recursos tecnológicos necessários para garantir a conectividade básica com a rede. Esses artigos ora tratam de forma genérica o conceito do acesso, ora abordam questões particulares a seu respeito (e.g. espaço geográfico onde ele ocorre (WEI et al., 2011), tecnologia e dispositivos utilizados (GIL-GARCIA; HELBIG.; FERRO, 2006)).

O conceito do uso da Internet, por sua vez, está associado a outros aspectos relativos ao próprio usuário, como, por exemplo, a questão das suas motivações, da frequência do uso e dos conhecimentos e das habilidades necessárias para tirar proveito desse instrumento.

O manual do *International Telecommunication Unit* (ITU)⁹ para medições do acesso e uso de TIC nos domicílios e pelos respectivos moradores estabelece que o “acesso à TIC refere-se à disponibilidade de TIC dentro dos lares. O uso de TIC refere-se à sua utilização por um ou mais indivíduos de um lar, em casa ou em qualquer outro lugar.” (ITU, 2014, p. 3, tradução nossa).

⁹ A União Internacional de Telecomunicação é uma entidade ligada à Organização das Nações Unidas (ONU) que trata de temas variados relacionadas à TIC (e.g. alocação global de canais de frequência, desenvolvimento de protocolos padronizados para a intercomunicação de redes e o desenvolvimento de iniciativas para conectar pessoas, etc.).

O conceito de uso estabelecido pelo ITU é pertinente e atual e representa uma realidade em todo o globo – e não só de um dado país.

No caso brasileiro, o celular é o dispositivo mais importante, sendo utilizado por 99% dos seus usuários. Em contrapartida, o computador de mesa corresponde à 20,16% dos usuários no país (CETIC.BR, 2020).

Além disso, são múltiplas, as formas como o celular é utilizado para o acesso à Internet. Pode-se, por exemplo, utilizá-lo, individualmente, via banda larga móvel ou compartilhá-lo com um grupo de indivíduos via Wi-Fi.

Adicionalmente, o acesso à Internet está, cada vez mais, disponível, gratuitamente ou não, em centros comunitários, cybercafés, livrarias e bibliotecas, espaços públicos e afins – beneficiando, particularmente, usuários de baixa de renda, que, normalmente, possuem planos de dados limitados e que podem usufruir desta facilidade e economizar os seus dados mensais remanescentes para outras ocasiões.

A própria literatura acerca do abismo digital já não trata mais a questão do acesso ou da conectividade como uma divisão entre aqueles que o têm e aqueles que não o têm, mas o encara como um espectro amplo de possibilidades de acesso de qualquer lugar (i.e. não mais especificamente da casa, do trabalho ou da escola), não necessariamente de um computador de mesa (i.e. sendo possível a utilização de outros dispositivos como, por exemplo, notebooks, tablets, celulares, televisões, etc.) e sem a obrigatoriedade de ter a sua propriedade (i.e. permitindo a possibilidade de seu compartilhamento entre múltiplos usuários, por exemplo, via Wi-Fi). “Não existe um tipo de acesso *online*, mas muitos; o significado e o valor do acesso variam em contextos sociais específicos; o acesso existe em um *continuum*, ao invés de uma oposição bipolar[dicotômica].” (WARSCHAUER, 2003, p. 297-298).

Assim, diante desta realidade, os termos “acesso” e “uso” são utilizados, neste trabalho acadêmico, no contexto do abismo digital de nível I e no sentido de estar “*online*”, ou seja, à capacidade de poder conectar-se à rede e consumir os seus serviços – independentemente, de como ele ocorra.

As questões, frequentemente, relacionadas ao abismo digital de nível II como, por exemplo, as motivações, a frequência e as capacidades e habilidades relativas ao conceito de uso da Internet, caso sejam mencionadas ou referenciadas, serão feitas de forma explícita.

3.2 Fontes de dados secundários

Os microdados utilizados neste trabalho são oriundos da pesquisa TIC Domicílios, realizada, anualmente, pelo Cetic.br, e que tem, por objetivo, “medir a posse e o uso das Tecnologias de Informação e Comunicação (TIC) entre a população residente no Brasil com 10 anos de idade ou mais.” (CETIC.BR, 2020, p.31).

Eles representam as estatísticas domiciliares de TIC, que referem-se a informações relativas à presença de recursos de TIC nos domicílios pesquisados e no uso que os seus moradores fazendo dos mesmos (ITU, 2014).

A pesquisa é realizada com o uso de questionários estruturados, com perguntas fechadas e respostas predefinidas (únicas ou múltiplas) e segue os referenciais da iniciativa *Partnership on Measuring ICT for Development*, liderada pelo ITU (CETIC.BR, 2020). Essa iniciativa global busca criar um arcabouço metodológico padronizado que possa ser utilizado para comparações dos respectivos contextos tecnológicos de países distintos – particularmente, daqueles em desenvolvimento (ITU, 2014).

As séries históricas utilizadas referem-se aos anos de 2015-2019 e as informações originais relativas aos indicadores demográficos, socioeconômicos e culturais obtidos pela pesquisa TIC Domicílios estão disponíveis na Tabela B-1 no APÊNDICE B.

3.3 Modelo teórico

O modelo teórico utilizado neste estudo empírico é composto de variáveis independentes, que se referem às características de naturezas demográficas, socioeconômicas e culturais dos participantes da pesquisa TIC Domicílios 2015-2019 e uma variável dependente, que aponta a condição de um indivíduo ser usuário da Internet.

Como a variável dependente possui natureza dicotômica, a regressão logística é o processo mais recomendado de análise (RICE; KATZ, 2003) e amplamente utilizado em vários artigos do tema pesquisado (GIL-GARCIA; HELBIG.; FERRO, 2006; HAIGHT; QUAN-HAASE; CORBETT, 2014; NISHIJIMA; IVANAUSKAS; SARTI, 2017).

O fundamento matemático associado à regressão logística é o logaritmo natural das razões das chances (“*logito*”), que é utilizado para tratar os desafios de descrever, por meio de uma expressão linear, as características da curva sigmoide (PENG; LEE; INGERSOLL, 2002, p. 4) e a não-linearidade entre as variáveis independentes e dependentes.

Um modelo estatístico, baseado neste tipo de regressão, que estabelece p , como sendo a probabilidade de ocorrência da variável dependente, ou seja, $p=P(Y=1)$, com base nas

variáveis independentes (X_1, X_2, \dots, X_n) e seus respectivos coeficientes ($\beta_1, \beta_2, \dots, \beta_n$), é regido pela equação (1), que estabelece a relação de linearidade das variáveis independentes com o *logito* para o evento $Y=1$:

$$\text{logito}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_n X_{n,i} \quad (1)$$

A expressão $\frac{p_i}{1-p_i}$ é denominada chance (“odds”) e refere-se à razão entre a probabilidade de um evento acontecer e a probabilidade do mesmo não acontecer (PENG; LEE; INGERSOLL, 2002) – neste caso, a probabilidade da ocorrência (ou não) da própria variável dependente (Y). Adicionalmente, é possível reescrever a equação (1) da seguinte maneira:

$$\begin{aligned} \frac{p_i}{1-p_i} &= e^{\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_n X_{n,i}} \\ &= e^{\beta_0} \times e^{\beta_1 X_1} \times e^{\beta_2 X_2} \times \dots \times e^{\beta_n X_n} \end{aligned} \quad (2)$$

Isolando-se p, tem-se a função sigmoide para a regressão logística:

$$\begin{aligned} p_i &= (1 - p_i) \times e^{\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_n X_{n,i}} \\ &= e^{\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_n X_{n,i}} - (p_i \times e^{\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_n X_{n,i}}) \\ e^{\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_n X_{n,i}} &= p_i + (p_i \times e^{\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_n X_{n,i}}) \\ e^{\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_n X_{n,i}} &= p_i \times (1 + e^{\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_n X_{n,i}}) \\ p_i &= \frac{e^{\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_n X_{n,i}}}{(1 + e^{\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_n X_{n,i}})} \end{aligned} \quad (3)$$

3.3.1 Variável Dependente

A variável dependente, transformada em dicotômica, sugere que uma determinada pessoa, que tenha feito uso da Internet nos 03 meses antecedentes à pesquisa, seja considerada como usuária da rede. Esse período é estipulado pelo próprio ITU que define, como período de referência, “o período referido na pesquisa ao perguntar sobre o uso individual das TIC [...] e recomenda um período de referência de 3 meses, em vez de 12 meses como na edição

anterior”(ITU, 2014, p. 37). A pesquisa TIC Domicílios seguiu esta definição metodológica nos anos de 2015-2019.

A codificação utilizada para a variável utiliza o valor 0, para identificar não usuários da Internet, e o valor 1, para identificar usuários da Internet, seguindo recomendação de padronização para variáveis *dummies*, que busca evitar que a interpretação da regressão torne-se “estranha e menos direta” (PENG; LEE; INGERSOLL, 2002, p. 11).

Informações adicionais a respeito da variável dependente estão presentes na Tabela 2.

Tabela 2 – Informações a respeito da variável dependente

Nome	Código de Variável ¹⁰	Pergunta da Pesquisa ¹¹	Respostas	Observação
Usuário(a) da Internet	C3	“Quando o(a) senhor(a) usou a Internet pela última vez?”	1 = “Há menos de 3 meses” 2 = “Entre 3 meses e 12 meses” 3 = “Mais de 12 meses atrás” 99 = “Não se aplica”	Criação de variável “dummy” para resposta “Há menos de 3 meses”

Fonte: Elaborado pelo autor com base em dados de pesquisa TIC Domicílios 2015-2019

3.3.2 Variáveis Independentes

Como variáveis independentes, foram utilizadas as informações de natureza demográficas, socioeconômicas e culturais encontradas no questionário de pesquisa que pudessem ser classificadas em alguma das categorias de determinantes do abismo digital no artigo de revisão sistemática de literatura (SCHEERDER; VAN DEURSEN; VAN DIJK, 2017). Informações detalhadas a respeito dessas variáveis estão disponíveis na Tabela 3. **Error! Reference source not found..**

Tabela 3 – Informações a respeito das variáveis independentes

Nome	Categoria	Código de Variável ¹²	Pergunta da Pesquisa ¹³	Respostas
Sexo	Sociodemográfico	sexo (2015-2016) SEXO (2017-2019)	“Sexo respondente?”	1="Masculino" 2="Feminino"
Idade	Sociodemográfico	idade (2015-2016) IDADE (2017-2019)	“Qual sua idade exata?”	(Valor correspondente a anos de vida)
Grau de Instrução	Econômico	grau_instrucao (2015-2016) GRAU_INSTRUCAO (2017-2019)	“O(A) sr. completou?”	1="Analfabeto" 2="Sabe ler/escrever, mas não cursou escola" 3="Até pré-escola incompleto" 4="Pré-escola completo" 5="1a até 3a série/1a até 3a série" 6="4a série/4a até 5a série" 7="5a até 7a série/1a até 3a série"

¹⁰ Código de variável encontrada no dicionário de variáveis da pesquisa TID Domicílios 2015-2019.

¹¹ Refere-se à pergunta dentro do questionário de pesquisa TIC Domicílios 2015-2019.

¹² Código de variável encontrada no dicionário de variáveis da pesquisa TID Domicílios 2015-2019.

¹³ Refere-se à pergunta dentro do questionário de pesquisa TIC Domicílios 2015-2019.

Nome	Categoria	Código de Variável ¹²	Pergunta da Pesquisa ¹³	Respostas
				8="8a série/4a série" 9="1a e 2a série/1a e 2a série" 10="3a série/vestibular" 11="Superior incompleto" 12="Superior completo"
Renda Familiar	Econômico	RENDA_FAMILIAR (2015-2019)	"O(a). Sr(a). poderia me dizer em qual destas faixas está a renda total da sua família no último mês, somando as rendas mensais de todas as pessoas que moram com o(a) sr(a), inclusive a sua?"	Valores referentes aos salários mínimos anuais. Para 2017, por exemplo, tem-se: 1 = "Até R\$ 937,00" 2 = "De R\$ 937,01 até R\$ 1.874,00" 3 = "De R\$ 1.874,01 até R\$ 2.811,00" 4 = "De R\$ 2.811,01 até R\$ 4.685,00" 5 = "De R\$ 4.685,01 até R\$ 9.370,00" 6 = "De R\$ 9.370,01 até R\$ 18.740,00" 7 = "De R\$ 18.740,01 até R\$ 28.110,00" 8 = "Mais de R\$ 28.110,00" 9 = "Não tem renda" 97 = "Não sabe" 98 = "Não respondeu"
Religião	Cultural	RELIGIAO (2015-2019)	"Qual é a sua religião ou culto?"	1 = "Católica Apostólica Romana" 2 = "Católica Ortodoxa" 3 = "Igreja Evangélica Luterana/Presbiteriana/Methodista/Batista" 4 = "Outras Igrejas Evangélicas" 5 = "Outras religiosidades cristãs" 6 = "Islamismo / Muçulmano" 7 = "Espiritismo" 8 = "Umbanda" 9 = "Candomblé" 10 = "Judaísmo" 11 = "Hinduísmo" 12 = "Budismo" 13 = "Sem religião" 14 = "Agnóstico" 15 = "Ateu" 16 = "Outras religiões" 98 = "Não respondeu"
Etnia	Cultural	ETNIA (2015-2019)	-	1="Branca" 2="Preta" 3="Amarela" 4="Parda" 5="Indígena" 98="Não respondeu"
Região	Sociodemográfico	COD_REGIAO_2 (2015-2019)	-	1 = "Sudeste" 2 = "Nordeste" 3 = "Sul" 4 = "Norte" 5 = "Centro-Oeste"
Área	Sociodemográfico	area (2015) AREA (2016-2019)	-	1="Urbana" 2="Rural"
Classe ¹⁴ Econômica	Econômico	CLASSE_2015 (2015-2019)	Perguntas específicas	1="A" 2="B" 3="C" 4="DE"
População Econômica mente Ativa	Econômico	PEA_2 (2015-2019)	"Não trabalha e não procurou trabalho nos últimos 30 dias?"	1 = "PEA" 2 = "Não PEA"

Fonte: Elaborado pelo autor com base em dados de pesquisa TIC Domicílios 2015-2019

¹⁴ Classe econômica pelo Critério Brasil 2015

3.4 Tratamento e apresentação de dados

Os microdados da pesquisa TIC Domicílios 2015-2019 foram coletados do repositório disponível no sítio do Cetic.br e carregados em um script em Python¹⁵.

Para manter a integridade das tabelas originais, foram criadas tabelas experimentais (cópias), contendo apenas as colunas referentes às variáveis independentes e dependente. Este procedimento permitiu que, ao longo desse estudo, validações e comparações com as informações originais dos microdados pudessem ser realizadas.

Concluída esta etapa, foi realizado um processo de verificação de conformidade dos dados presentes nas tabelas experimentais com base nas definições contidas nos dicionários de variáveis para cada ano avaliado. Os dicionários de variáveis explicam a estruturação original das variáveis independentes, como, por exemplo, os seus respectivos números de identificação, as suas descrições, os tipos de informações (e.g. numérica e textual) e os códigos e os seus respectivos significados (“rótulos”).

O objetivo desta verificação foi de estabelecer uma padronização (normalização) dos dados nas tabelas ao longo de todo o período. As discrepâncias identificadas e tratadas durante a realização desta verificação estão discriminadas na Tabela C-1 do APÊNDICE C. A Tabela 4, abaixo, apresenta o consolidado do antes e depois dessa verificação.

Tabela 4 – Resultados de processo de tratamento de tabelas experimentais

Ano	Número de Registros		
	Antes de Tratamento	Depois de Tratamento	Redução
2015	20.397	18.083	11,34%
2016	20.722	17.936	13,44%
2017	20.490	17.803	13,11%
2018	20.544	17.766	13,52%
2019	20.536	17.811	13,27%

Fonte: Elaborado pelo autor.

Padronizados os dados, as variáveis referentes ao grau de escolaridade e renda familiar foram transformadas de variáveis categóricas para variáveis discretas. O grau de escolaridade foi transformado em anos de educação usando como base as informações da Figura B-1 do APÊNDICE B. A variável relativa à renda familiar foi transformada em valores monetários

¹⁵ Tipo de linguagem de programação com múltiplas bibliotecas voltadas para análise matemática e estatística. Para esta pesquisa, foram utilizadas as seguintes bibliotecas: Pandas, NumPy, Scikit-learn, Statsmodels, Seaborn e Matplotlib.

utilizando as médias entre os pisos e tetos de suas categorias, convertida em milhares de reais e corrigida pelo Índice Nacional de Preços ao Consumidor Amplo (IPCA). Para a última categoria, sem teto, foi utilizada a renda média (45,25 salários mínimos) segundo informações do IBGE para o ano 2015. Os valores monetários referentes a cada ano estudado encontram-se Tabela B-2 do APÊNDICE B.

Posteriormente, foram criadas variáveis “*dummies*”, respectivamente, para a variável dependente e para as demais variáveis independentes com características categóricas e estabelecidos os seus referenciais para a realização da regressão logística (ONO; ZAVODNY, 2007). Para religião, a variável “*dummy*” utilizada representa a característica de um indivíduo não ser católico (referência: católico). A Tabela B-2 e Tabela B-3 do APÊNDICE B apresentam, respectivamente, as informações finais de formatação final e distribuição estatística destas variáveis.

4 APRESENTAÇÃO DE RESULTADOS

A pesquisa relativa a este estudo dos determinantes do abismo digital no Brasil busca identificar, dentre aqueles indicadores socioeconômicos, demográficos e culturais presentes na pesquisa TIC Domicílios 2015-2019, aqueles que expliquem, com relevância estatística, o impacto das características que eles representam (variáveis independentes) e o uso da Internet por parte da sociedade brasileira (variável dependente).

A motivação original para este estudo buscava compreender os fatores que explicassem o processo de desigualdade digital no país e contribuir com as discussões a respeito de propostas de políticas públicas em prol da democratização, ou universalização, do uso da Internet e que busquem beneficiar as comunidades ou grupos sociais marginalizados por esse processo.

A pesquisa TIC Domicílios apresenta, para cada ano em que foi realizada, a fotografia do cenário de uso da Internet não só para toda a sociedade brasileira, mas sob a ótica de alguns desses indicadores socioeconômicos, demográficos e culturais. No entanto, ela não apresenta ou identifica os fatores, com significância estatística, que contribuíram ou explicam o estabelecimento dessa realidade.

O fato de haver distinções de uso desta tecnologia associadas a um indicador específico não lhe confere prontamente a característica de determinante digital. Uma análise matemática e estatística mais apurada precisa ser realizada antes de se chegar a esta conclusão.

Os resultados desse estudo, apresentados a seguir, são fruto dos achados referentes à avaliação dos microdados, à estruturação do modelo estatístico e à sustentação teórica advinda das informações e conclusões dos artigos encontrados na revisão de literatura. É esse arcabouço teórico, empírico e estatístico que, quando avaliado, confere a característica de determinante digital a um dado indicador socioeconômico, demográfico e cultural. Tomemos um exemplo com base nos achados desta pesquisa.

Em todos os anos estudados, o número absoluto de usuárias de Internet foi maior do que o número de usuários. Esse fato não é relevante isoladamente para explicar o impacto do sexo na adoção da Internet, visto que o maior número de mulheres na amostra da pesquisa TIC Domicílios explica, parcialmente, essa situação. No entanto, se analisada a diferença da taxa de penetração do serviço de Internet entre os sexos ao longo de 2015-2019, nota-se o crescimento da taxa de adoção do público feminino com relação ao público masculino em todos os anos, tendo 2018 como ponto de inflexão para o estabelecimento das mulheres como o público de maior adoção, com 63,38% das mulheres fazendo uso da Internet.

Ainda que os dados sugiram que ser do sexo feminino tenha tido uma influência para o aumento de usuários no período estudado, ainda, não é possível afirmar categoricamente esta relação. A teoria acerca do abismo digital suporta a possibilidade do sexo como um determinante digital, mas, a análise dessa variável no modelo proposto, juntamente com as demais, não encontrou respaldos estatísticos para corroborá-la. E, por isso, o sexo foi descartado como determinante do abismo digital no Brasil para o período de 2015-2019.

Outra questão importante com relação a este estudo é a análise dos determinantes ao longo de 05 anos. Desta maneira, é possível identificar suas evoluções e apontar tendências. Essa análise de múltiplos anos contempla a utilização, primeiramente, das lacunas (*gaps*) de acesso (BONFADELLI, 2002) (ou, neste documento, também chamadas de lacunas digitais) e, posteriormente, da regressão logística, para mensurar, utilizando-se como referência os indicadores da pesquisa TIC Domicílios, o seguinte:

- a evolução das lacunas de acesso.
- a evolução dos impactos dos determinantes digitais.

A Tabela B-4 do APÊNDICE B apresenta as lacunas digitais referentes a cada determinante socioeconômico, demográfico e cultural. Elas representam as diferenças mais significativas entre os percentuais de adoção da Internet dentro da amostragem das populações respectivas a cada categoria dessas variáveis categóricas. E quando analisados ao longo de um dado período, eles ajudam a caracterizar como se dão as mudanças no abismo digital e os seus influenciadores.

Os resultados consolidados do modelo estatístico, baseado em regressão logística, para o estudo dos determinantes do abismo digital no Brasil entre 2015-2019 são apresentados na Tabela 5¹⁶. Informações adicionais estão disponíveis na Tabela B-5 (Matriz de Correlação), Figura B-2 (Regressão Logística completa) e Tabela B-6 (*Odds Ratios*) do APÊNDICE B.

Originalmente, uma regressão logística foi gerada para cada ano e os respectivos determinantes identificados. No entanto, por conta da variação plurianual do impacto de alguns determinantes permanecer dentro do intervalo de confiança e essa situação dificultar o estudo da evolução dos seus respectivos impactos no abismo digital existente no país, a abordagem foi alterada para uma única regressão logística com dados dos 05 anos. E para mensurar a evolução

¹⁶ Para a regressão logística, foram testados a função de ligação logito, o complemento log-log e a função probito. A opção com a função de ligação logito apresentou-se como melhor opção e, por isso, foi a escolhida.

dos impactos foram utilizadas uma variável auxiliar que identifica o ano do registro e variáveis de interação dos determinantes digitais com a variável auxiliar.

Os indicadores relativos a idade, grau de instrução, classe econômica, população economicamente ativa, religião e área residencial (urbana/rural) permaneceram no modelo final. Os indicadores relativos a sexo, etnia, região não demonstraram significância estatística e, conseqüentemente, foram desconsiderados do modelo proposto. A renda familiar não foi considerada por conta da correlação com a classe econômica.

Tabela 5 – Regressão Logística para determinantes do abismo digital no Brasil

Variáveis Independentes	Regressão Logística para Anos 2015-2019			
	Variável Dependente: Usuários de Internet			
	Coefficiente β	Erro Padrão	z	P > z
Constante	-0,8324	0,107	-7,760	0,000
ANO	0,1053	0,042	2,506	0,012
Idade ²	-0,0001	0,000	-2,358	0,018
Idade	-0,0587	0,005	-11,488	0,000
Grau de Instrução ²	0,0039	0,000	8,025	0,000
Grau de Instrução	0,1539	0,008	19,372	0,000
Classe A (ref. Classes D e E)	2,7706	0,150	18,445	0,000
Classe B (ref. Classes D e E)	2,2719	0,064	35,494	0,000
Classe C (ref. Classes D e E)	1,1662	0,038	30,567	0,000
População Economicamente Ativa (PEA) (ref. Não PEA)	0,2183	0,041	5,291	0,000
Não Catolicismo (ref. Catolicismo)	0,2469	0,035	7,112	0,000
Área Urbana (ref. Área Rural)	0,5549	0,058	9,596	0,000
ANO:Idade ²	-0,0001	0,000	-6,141	0,000
ANO:Idade	0,0172	0,002	8,549	0,000
ANO:Grau de Instrução ²	-0,0008	0,000	-7,760	0,000
ANO:Classe B (ref. Classes D e E)	-0,1037	0,029	-3,620	0,000
ANO:Classes C (ref. Classes D e E)	-0,0536	0,015	-3,495	0,000
ANO:PEA (ref. Não PEA)	-0,0423	0,016	-2,591	0,010
ANO:Não Catolicismo (ref. Catolicismo)	0,0405	0,014	2,852	0,004
ANO: Área Urbana (ref. Área Rural)	-0,0539	0,023	-2,384	0,017
Pseudo-R ² de McFadden	0,4370			
Valor p (LLR):	0,000			

Fonte: Elaborado pelo autor com base nas informações da Figura B-2 do APÊNDICE B.

A teoria acerca de regressões logísticas informa que:

- no caso de variáveis contínuas ou discretas, a alteração de uma unidade da variável (*ceteris paribus*) impacta, segundo o respectivo coeficiente β , o *logito* de acordo com a equação (1) ou a chance, segundo a equação (2).
- no caso de variáveis categóricas (“*dummies*”), os valores do coeficiente β e do *Odds Ratio* estabelecem as bases comparativas do impacto desta variável na variável

dependente em relação à referência (e.g. referência para área residencial é “área rural”).

No entanto, a idade e o grau de instrução apresentaram características de preditores polinomiais, transformando o modelo estatístico em um modelo de regressão logística polinomial. E, adicionalmente, o modelo estatístico faz uso de variáveis de interação.

Assim, o cálculo do Efeito Ajustado (E.A.) de uma determinada variável independente no uso da Internet, se dá, com base nas informações da regressão logística (Tabela 5), por meio da fórmula geral (4), onde ‘ANO’ refere-se a um ano entre 2015-2019, ‘VAR’, a uma variável independente, e ‘i’, a um valor relativo à variável independente (e.g. 10 anos para idade e 1 para variáveis dicotômicas).

$$E.A. = \beta_{Ano}(ANO - 2014) + \beta_{Var2}VAR_i^2 + \beta_{Var}VAR_i + \beta_{AnoVar2}(ANO - 2014)VAR_i^2 + \beta_{AnoVar}(ANO - 2014)VAR_i + Constante \quad (4)$$

A análise dos coeficientes informa a orientação (positiva ou negativa) do impacto de uma variável independente – além de sua magnitude. Analisando-os para as variáveis dos determinantes (sem levar em consideração as interações com a variável auxiliar referente ao ano) da Tabela 5, nota-se que essas orientações são coerentes com as referências encontradas nos artigos identificados na revisão de literatura – corroborando a relevância dos achados.

O aumento da idade afeta negativamente a chance de uso da Internet. Enquanto, o maior grau de instrução, ser membro das classes A, B ou C, fazer parte da população economicamente ativa, não ser católico e morar em área urbana possuem impacto positivo. (No entanto, por conta do modelo plurianual, o impacto de um determinante digital específico é afetado não só pelo seu coeficiente, mas pela variável auxiliar, pelas variáveis de interação e, no caso da idade e grau de instrução, pelas características polinomiais.)

4.1 Discussão dos resultados

4.1.1 Idade

Os resultados referentes à idade, presentes na Tabela 5, confirmam, com significância estatística (e.g. $\beta = -0,0587$, EP = 0,005, valor $p < 0,001$), o seu impacto, inversamente

proporcional, na adoção da Internet no país. Em outras palavras, quanto maior a idade de um cidadão brasileiro, menor a probabilidade de que ele faça uso da rede.

Essa constatação é respaldada por achados e conclusões de artigos científicos acerca do abismo digital no Brasil e no mundo encontrados na revisão de literatura (HAIGHT; QUAN-HAASE; CORBETT, 2014). Segundo Van Deursen e Van Dijk (2014, p. 511, tradução nossa), em estudo sobre a sociedade holandesa, “a idade parece ser uma das mais significativas variáveis que afetam o uso da Internet”.

Especificamente com relação à realidade brasileira, Nishijima, Ivanauskas e Sarti (2017, p. 19, tradução nossa) afirmam que os “idosos também têm menor acesso à Internet”. Embora constatem a redução do impacto do determinante referente à idade nos anos pesquisados (2003, 2005, 2011 e 2013) e argumentem, como explicação para tal, a relação mais forte do analfabetismo digital no Brasil com a falta de acesso à educação formal do que com as dificuldades do uso da rede, indicando, em sua pesquisa, uma “convergência dos efeitos dos coeficientes de idade [em faixa etária] e educação [em anos]” e sugerindo “uma variável interativa no modelo” (NISHIJIMA; IVANAUSKAS; SARTI, 2017, p. 19, tradução nossa).

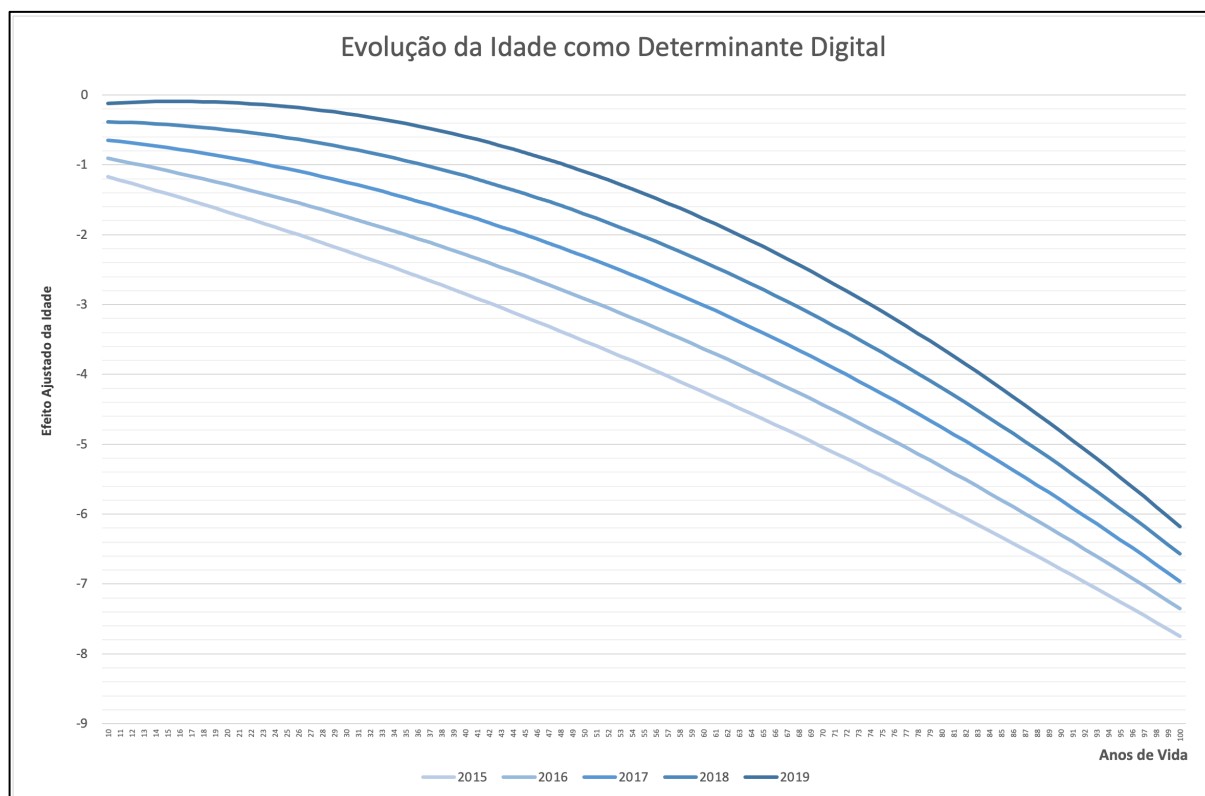
No entanto, artigos identificados na revisão de literatura acerca da realidade de outros países apontam outras possibilidades para esta situação, como, por exemplo, o envelhecimento natural dos jovens (VAN DEURSEN; VAN DIJK, 2014) e a adoção de serviços voltados para o contexto de idosos (e.g. monitoramento de pacientes com doenças crônicas (CHOI; DINITTO, 2013)).

A Figura 3 apresenta a evolução do efeito ajustado da idade nos anos de 2015-2019. Nela, é possível perceber que a idade, enquanto um determinante digital, perde força no período (fato indicado pela aproximação das curvas anuais do eixo das coordenadas). No entanto, os gráficos das parábolas anuais vão se fechando, deixando suas curvas mais proeminentes, o que representa que a probabilidade de uso da Internet cai, com o aumento da idade, para os jovens, mas a queda é intensificada no caso de adultos e idosos.

Ainda assim, houve uma melhoria nas desigualdades entre jovens e idosos. Em 2015, a chance de uma pessoa de 60 anos fazer uso da Internet, com relação a outra de 10 anos, era de 0,0460:1 (*Odds Ratio*¹⁷), o que corresponde a uma probabilidade de 4,40%, calculado segundo (PAMPEL, 2000, p. 25). Em 2019, o número saltou para 0,1925:1, o que corresponde a uma probabilidade de 16,14%.

¹⁷ Primeiramente, foi calculada a chance para cada idade, utilizando a equação 2, presente na página 41. E, posteriormente, calculada a razão das chances para as duas idades.

Figura 3 – Evolução da idade como determinante digital



Fonte: Elaborado pelo autor.

Ressalva-se que, mesmo com a redução da influência da idade, há uma discrepância muito grande do uso da Internet no país entre jovens e idosos. A Tabela B-4, presente no APÊNDICE B, informa que, em 2019, somente 29,60% das pessoas com mais de 60 anos fez uso da Internet – enquanto o número para faixa etária de 16 a 24 anos foi de 91,81%. Se analisada a evolução da desigualdade de acesso, utilizando o conceito de lacuna de acesso entre os anos de 2015-2019, constata-se que, com relação à idade, houve uma redução total de 10,34 pp¹⁸ - com o aumento da adoção de 18,22 pp, para pessoas com mais de 60 anos, e 7,88 pp para faixa etária de 16 a 24 anos (Tabela B-4).

4.1.2 Grau de Instrução

Os resultados presentes na Tabela 5, confirmam, com significância estatística (e.g. $\beta = 0,1539$, EP = 0,008, valor $p < 0,001$), o impacto positivo que o grau de instrução possui na adoção da Internet no Brasil. Em outras palavras, quanto maior a quantidade de anos de estudos que um cidadão brasileiro possui, maior a probabilidade de que ele faça uso da rede.

¹⁸ Pontos percentuais.

A variação do impacto da educação, no entanto, não se dá uniformemente dentro de todos os anos de instrução de acordo com a

Figura 4. O ponto dos gráficos anuais situado entre os 11-12 anos de estudos, praticamente não se altera. A análise gráfica sugere uma tendência de nivelamento ao longo dos anos de 2015-2019, com a melhoria da chance de uso da rede entre aqueles que não estudaram ou que se encontram nos primeiros anos de estudos e a sua redução para aqueles que estão concluindo ou já concluíram o nível superior. Adicionalmente, a redução marginal relativa aos anos superiores é mais rápida, indicando que a relevância do nível superior diminui com relação aos outros níveis de escolaridade e contribuindo para minimizar os efeitos da idade no abismo digital.

Em 2015, por exemplo, uma pessoa que soubesse ler e escrever, mas que não tenha frequentado a escola, e outra com 18 anos de estudos (ensino superior) tinham probabilidades, respectivamente, de 36,12% e 95,44% de usarem a Internet. Em 2019, as probabilidades atualizadas foram, respectivamente, 46,20% e 91,81%. O *Odds Ratio* da pessoa sem escolaridade fazer uso da rede, com relação à outra, em 2015, foi de 0,0270:1 (2,63%) e, em 2019, 0,0766:1 (7,12%).

Em outras palavras, a probabilidade de uso da rede por parte da população com menor escolaridade aumentou e a probabilidade referente àqueles com maior escolaridade caiu, caracterizando, assim, a queda dos efeitos da idade no cenário de desigualdades digitais no país.

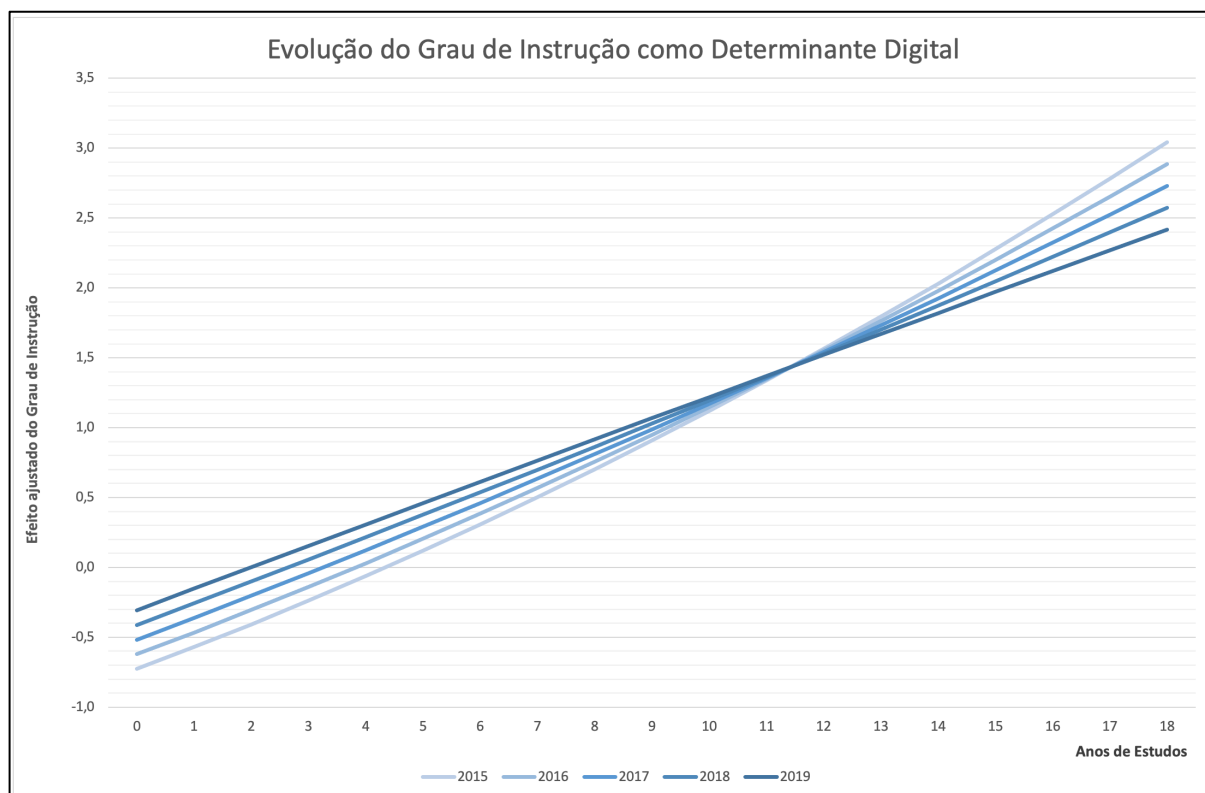
Se considerados os microdados, verifica-se que a lacuna de acesso caiu em todos os graus de instrução ao longo dos cinco anos analisados. No entanto, mesmo com o aumento percentual na casa de dois dígitos para as faixas dos menos escolarizados, a desigualdade existente para esse determinante continua sendo muito alta, chegando, em 2019, a 82,74 pp. Enquanto a adoção da Internet por usuários com curso superior é de 95,88%, para pessoas que sabem ler e escrever, mas não cursaram a escola, ela é 23,15% (Tabela B-4).

Segundo Van Deursen e Van Dijk (2011, p. 897, tradução nossa), “a educação é o preditor global mais consistente do uso das TICs”. E de acordo com Nishijima, Ivanauskas e Sarti (2017, p. 22, tradução nossa), “melhorias na política educacional representam uma estratégia eficaz para reduzir a exclusão digital na população brasileira a longo prazo, reduzindo as barreiras relacionadas ao analfabetismo digital.”

No entanto, Bonfadelli (2002, p. 69, tradução nossa) destaca que, “a pesquisa empírica prova que as desigualdades de conhecimento baseadas em déficits educacionais podem ser alteradas ou revertidas - pelo menos parcialmente - por fatores como relevância ou interesse

peçoal”. Essa perspectiva, abre possibilidades de ganhos mais imediatos de combate ao abismo digital, sem que os alunos tenham que esperar por longos anos dentro do sistema educacional para reverter tal situação.

Figura 4 – Evolução do grau de instrução como determinante digital



Fonte: Elaborado pelo autor.

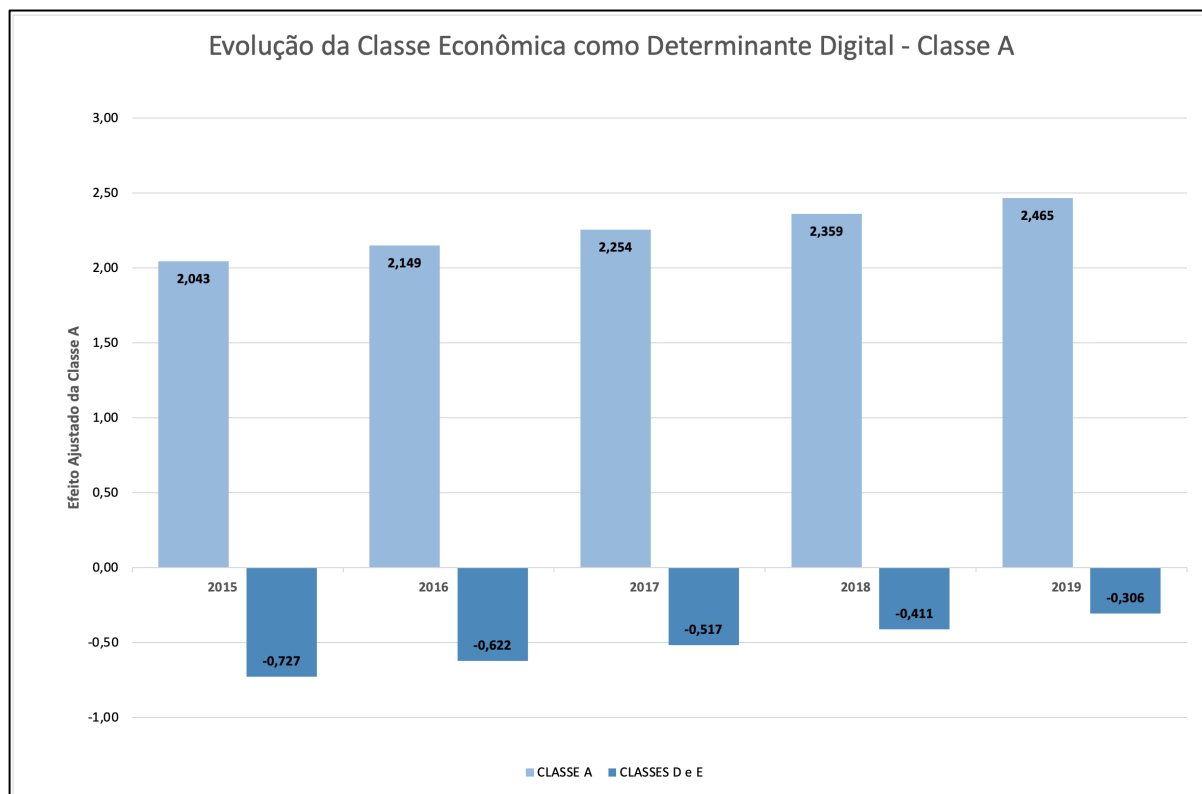
4.1.3 Classe Econômica

A metodologia empregada na pesquisa TIC Domicílios para identificar a classe econômica utiliza-se do Critério Brasil 2015, que é um indicador do poder de compra da população, composto de medidores relativos à posse de bens de consumo (e.g. geladeira e automóveis), inclusive produtos de TIC (e.g. computadores e notebooks), do acesso a serviços públicos (e.g. água encanada e rua pavimentada) e do nível de educação do chefe da família.

Com relação a este indicador econômico como um determinante do abismo digital, os resultados presentes na Tabela 5, confirmam, com significância estatística, o impacto positivo do fato de uma pessoa fazer parte da classe A (e.g. $\beta = 2,7706$, EP = 0,150, valor $p < 0,001$), classe B (e.g. $\beta = 2,2719$, EP = 0,064, valor $p < 0,001$) e classe C (e.g. $\beta = 1,1662$, EP = 0,038, valor $p < 0,001$), tendo as classes D e E sido estabelecidas como classe referencial.

As Figura 5, Figura 6 e Figura 7 apresentam a evolução dos efeitos ajustados para as classes econômicas enquanto determinantes digitais.

Figura 5 – Evolução da classe econômica como determinante digital – Classe A

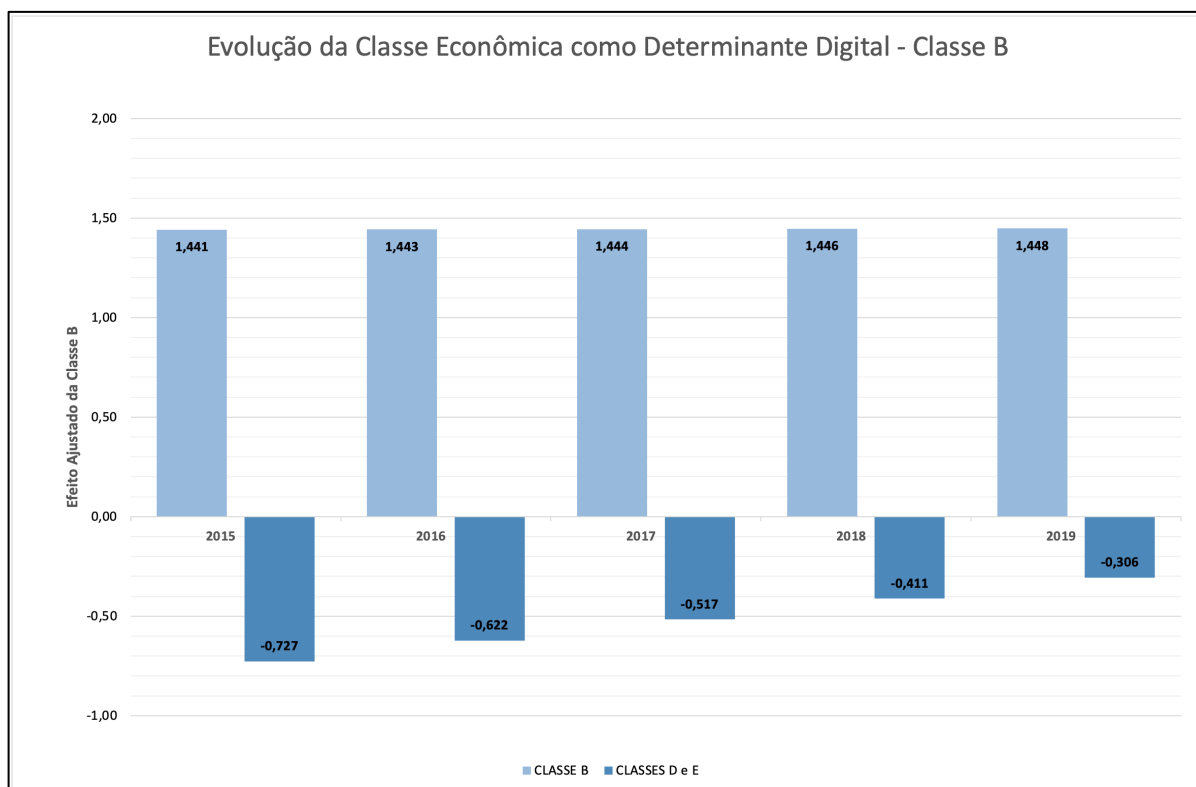


Fonte: Elaborado pelo autor.

Os achados apresentam uma alta relevância, no início do período, de um indivíduo de classe A, em comparação ao das classes D e E, fazer uso da Internet (Figura 6). No entanto, ao longo do período estudado, ela se mantém estagnada. Nos anos de 2015 a 2019, a chance de um indivíduo da classe A fazer uso da rede, com relação a outro das classes D ou E, foi de 15,9682:1 (94,11%).

Os achados apresentam uma relevância, no início do período, de um indivíduo de classe B, em comparação ao das classes D e E, fazer uso da Internet e a sua redução até o final do período (Figura 6). Em 2015, a chance de um indivíduo da classe B fazer uso da rede, com relação à outra das classes D ou E, foi de 8,7426:1 (89,74%). Em 2019, a chance foi de 5,7753 (85,24%).

Figura 6 – Evolução da classe econômica como determinante digital – Classe B



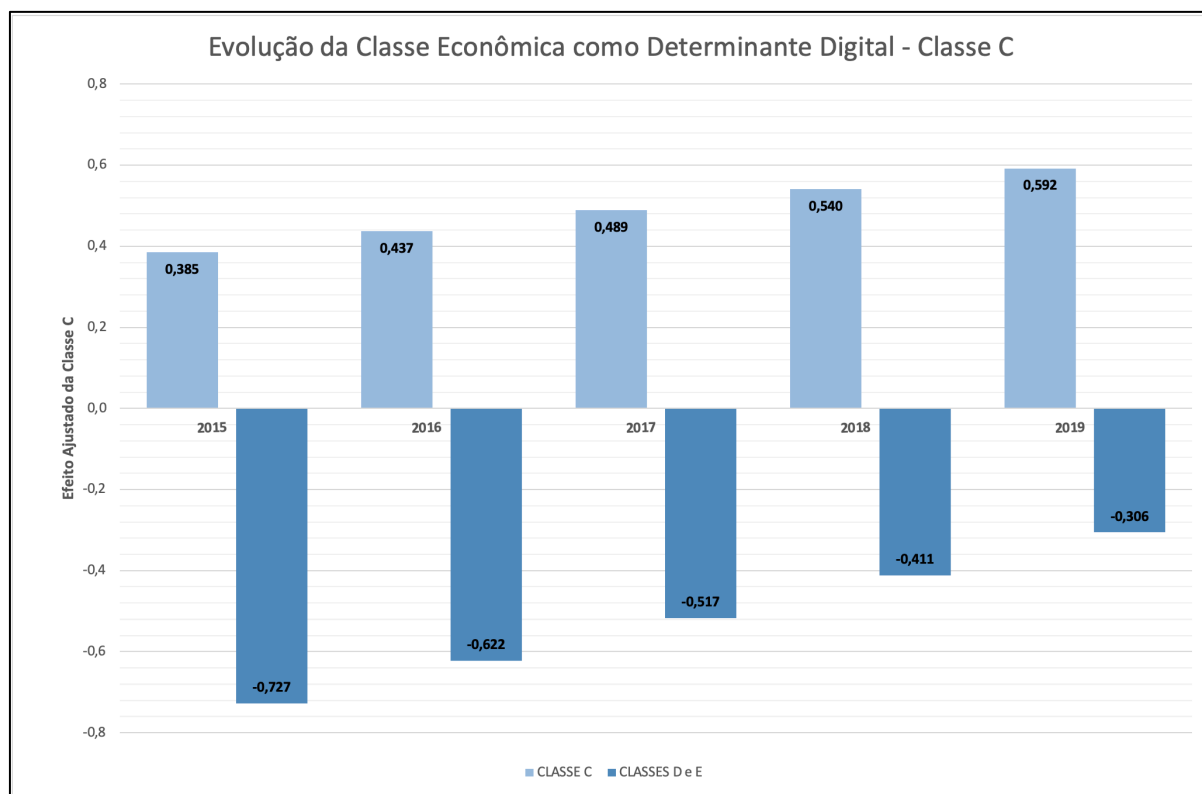
Fonte: Elaborado pelo autor.

Com relação à classe C, a Figura 7 indica a queda constante da influência dessa classe econômica, com relação às classes D e E, e apresenta uma perspectiva de melhoria com relação à desigualdade digital para este determinante entre essas classes.

Em 2015 e 2019, as chances de um membro da classe C com relação aos membros das classes D e E, fazer uso da Internet eram, respectivamente, de 3,0420:1 (75,26%) e 2,4548:1 (71,06%).

No período analisado, o aumento de 23,32 pp da penetração do serviço de Internet nas classes D e E (agrupadas) contribuiu para que a lacuna de acesso associada à classe econômica caísse 22,49 pp (Tabela B-4).

Figura 7 – Evolução das classes econômica como determinante digital – Classe C



Fonte: Elaborado pelo autor.

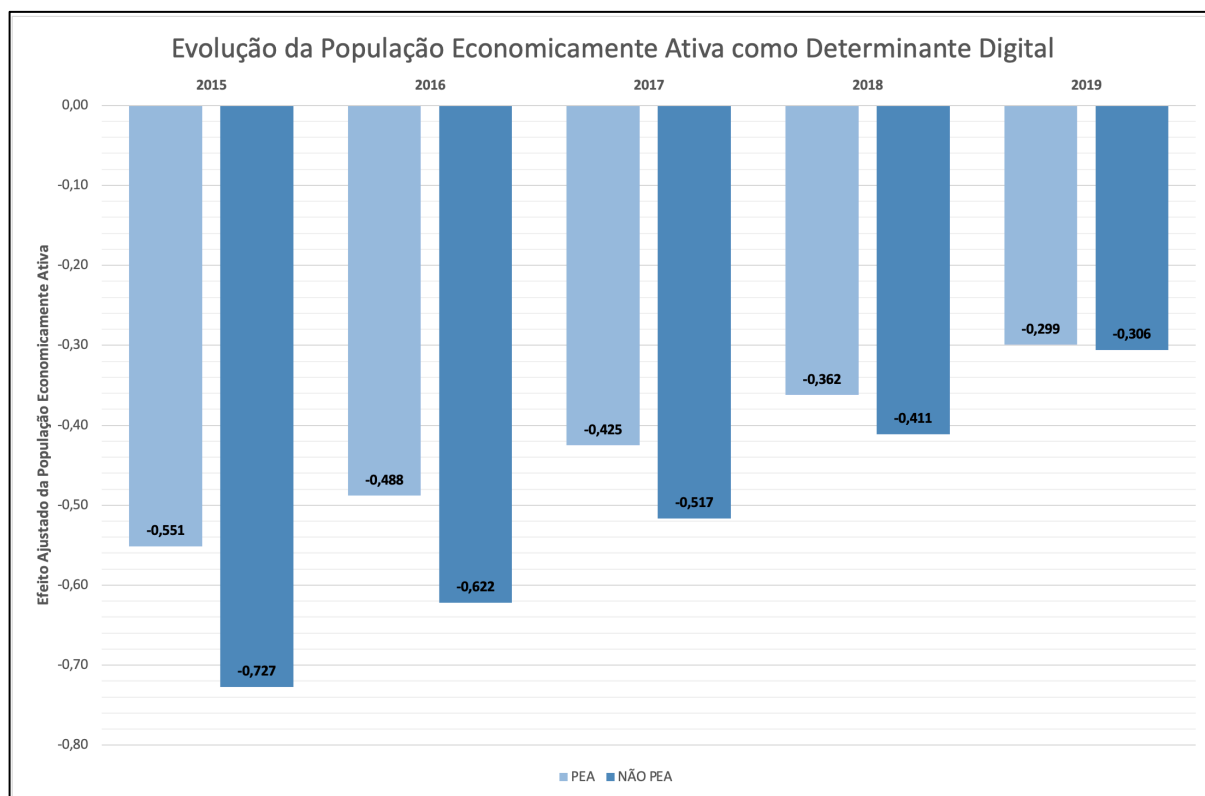
4.1.4 População Economicamente Ativa (PEA)

A maioria dos artigos que abordam a questão do emprego enquanto determinante do abismo digital apontam o seu impacto positivo no uso da Internet (FERRO; HELBIG; GILGARCIA, 2011; VAN DEURSEN; VAN DIJK, 2011).

Para a realidade brasileira, fazer parte da população economicamente ativa afeta, com significância estatística (e.g. $\beta = 0,2183$, EP = 0,041, valor $p < 0,001$), a chance de uma pessoa ser usuária da Internet (Tabela 5).

Nishijima, Ivanauskas e Sarti (2017, p. 22, tradução nossa) concluem, para os anos estudados, que, no Brasil, “renda [baixa], [não possuir] emprego e [alto] número de pessoas em um domicílio são barreiras significantes para o acesso a bens de TIC” e, demonstram uma queda nos seus respectivos impactos no que tange à desigualdade digital.

Figura 8 – Evolução da população economicamente ativa como determinante digital



Fonte: Elaborado pelo autor.

Os achados deste estudo corroboram com o que Nishijima, Ivanauskas e Sarti afirmam com relação à renda, mas apontam uma alteração com relação ao emprego. De acordo com a Figura 8, fazer parte da população economicamente ativa perde relevância, com relação à sua referência, chegando ao final do período, com o efeito similar ao efeito associado a não fazer parte da população economicamente ativa. Em 2015, pessoas com emprego ou procurando emprego nos últimos 30 dias tinham chance de 1,1924:1 (54,39%), em relação àquelas não consideradas como economicamente ativas, de fazer uso da Internet. Em 2019, o número caiu para 1,0067:1 (50,17%).

Ao final do período analisado, de acordo com os microdados, 72,50% daqueles indivíduos que fazem parte da população economicamente ativa eram usuários de Internet. Enquanto, para aqueles que não fazem parte da população economicamente ativa, o número foi de 53,70%. O aumento durante o período foi de 14,89 pp e 17,38 pp, respectivamente, para aqueles economicamente ativos e inativos e redução da lacuna em 2,49 pp (Tabela B-4).

Dentre as categorias da população economicamente ativa, um grupo que se destaca é aquele formado por indivíduos que não estão trabalhando, mas que tomaram alguma providência para conseguir trabalho nos últimos 30 dias. Esse grupo teve adesão superior às demais categorias (81,05%) segundo Tabela B-4 – o que sugere o papel da Internet enquanto

meio para busca de vagas, realocação profissional, estabelecimento de relacionamentos e capacitação.

4.1.5 Religião

Os resultados presentes na Tabela 5, confirmam, com significância estatística (e.g. $\beta = 0,2469$, EP = 0,035, valor $p < 0,001$), o impacto positivo da religião (caracterizado pelo binômio católico e não católico neste estudo) na adoção da Internet no Brasil no período analisado.

De acordo com a Figura 9, a análise gráfica da evolução do efeito ajustado deste determinante informa que ele aumentou durante o período. Não ser católico teve uma importância crescente na adoção da Internet no país entre 2015-2019. Em 2015, a chance de um indivíduo não católico fazer uso da rede, em relação a um adepto desta religião, era de 1,3330:1 (57,14%). Em 2019, o valor subiu para 1,5675:1 (61,05%).

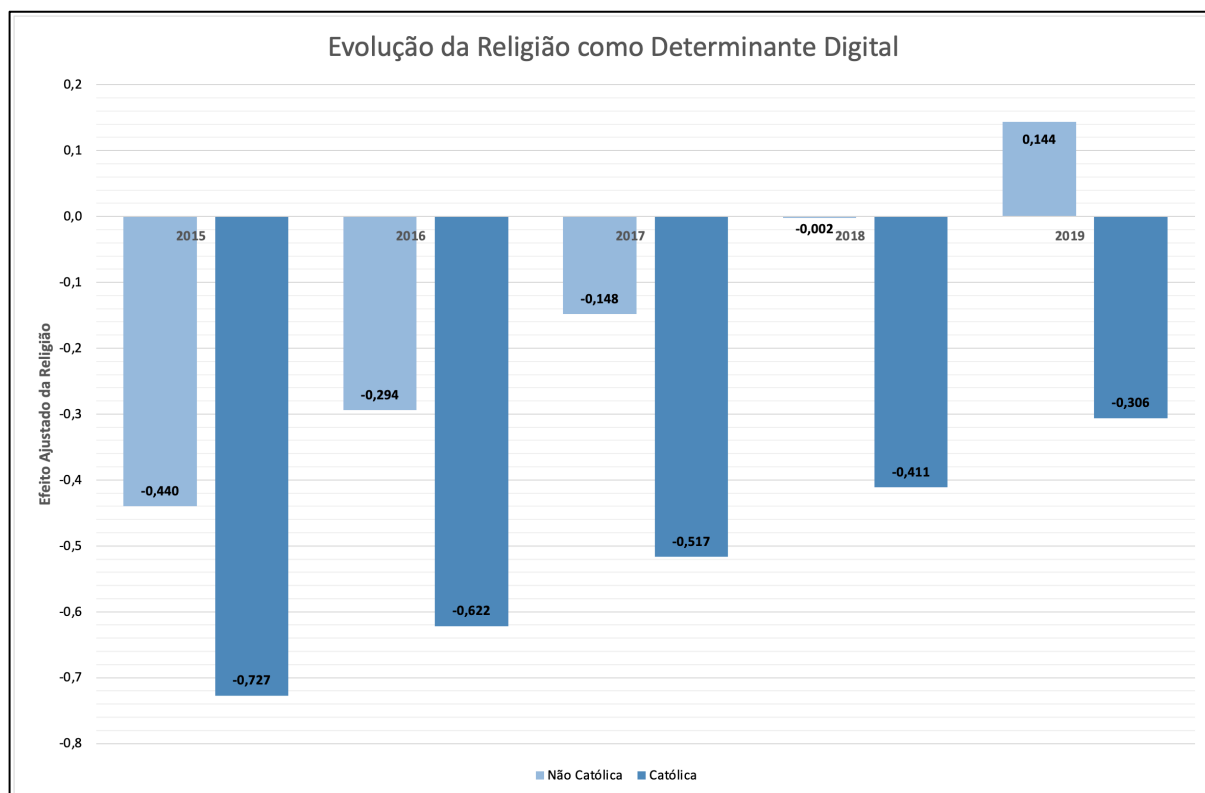
Ainda que tenha havido um aumento de usuários católicos em 14,34 pp (Tabela B-4), a adesão em taxas maiores por parte de adeptos de algumas religiões específicas (e.g. Umbanda (22,52 pp), Igreja Evangélica Luterana/Presbiteriana/Methodista/Batista (23,83 pp) e Outras Igrejas Evangélicas (16,89 pp)) contribui com a redução do percentual de usuários católicos de 56,74%, em 2015, para 52,57%, em 2019, na população de brasileiros conectados e, consequentemente, com a melhoria da igualdade digital sob a óptica da religião.

Do ponto de vista da literatura, a variável religião é classificada como um determinante cultural no artigo de revisão sistemática de literatura, segundo o qual, pesquisas a respeito de determinantes “sociais (por exemplo, suporte digital e voluntariado formal) e culturais (por exemplo, capital cultural e religião) precisam de mais atenção e podem fornecer melhores explicações de como os usuários da Internet obtêm (ou não) benefícios” (SCHEERDER; VAN DEURSEN; VAN DIJK, 2017, p. 1614, tradução nossa).

Em um dos artigos identificados na revisão de literatura, Dilmaghani (2018, p.1, tradução nossa) sugere, para a realidade canadense, que “a religiosidade se associa negativamente ao acesso e à atividade da Internet”. E Lissitsa e Chachashvili-Bolotin (2015, p. 47-48, tradução nossa) relembram que, a respeito do abismo digital em Israel, “estudos indicam uma taxa crescente de acesso à Internet e variedade de usos digitais à medida que o nível de religiosidade diminui”. Os achados desse estudo, no entanto, não corroboram essa perspectiva negativa da religião para a realidade brasileira.

Por fim, outras análises com relação à religião foram verificadas (e.g. cristão/não cristão), mas, somente, o binômio católico/não católico apresentou relevância estatística na adoção da Internet no Brasil durante os anos de 2015-2019.

Figura 9 – Evolução da religião como determinante digital



Fonte: Elaborado pelo autor.

4.1.6 Área residencial

Os resultados presentes na Tabela 5, confirmam, com significância estatística (e.g. $\beta = 0,5549$, EP = 0,05, valor $p < 0,001$), o impacto positivo de se viver em área urbana na adoção da Internet no Brasil.

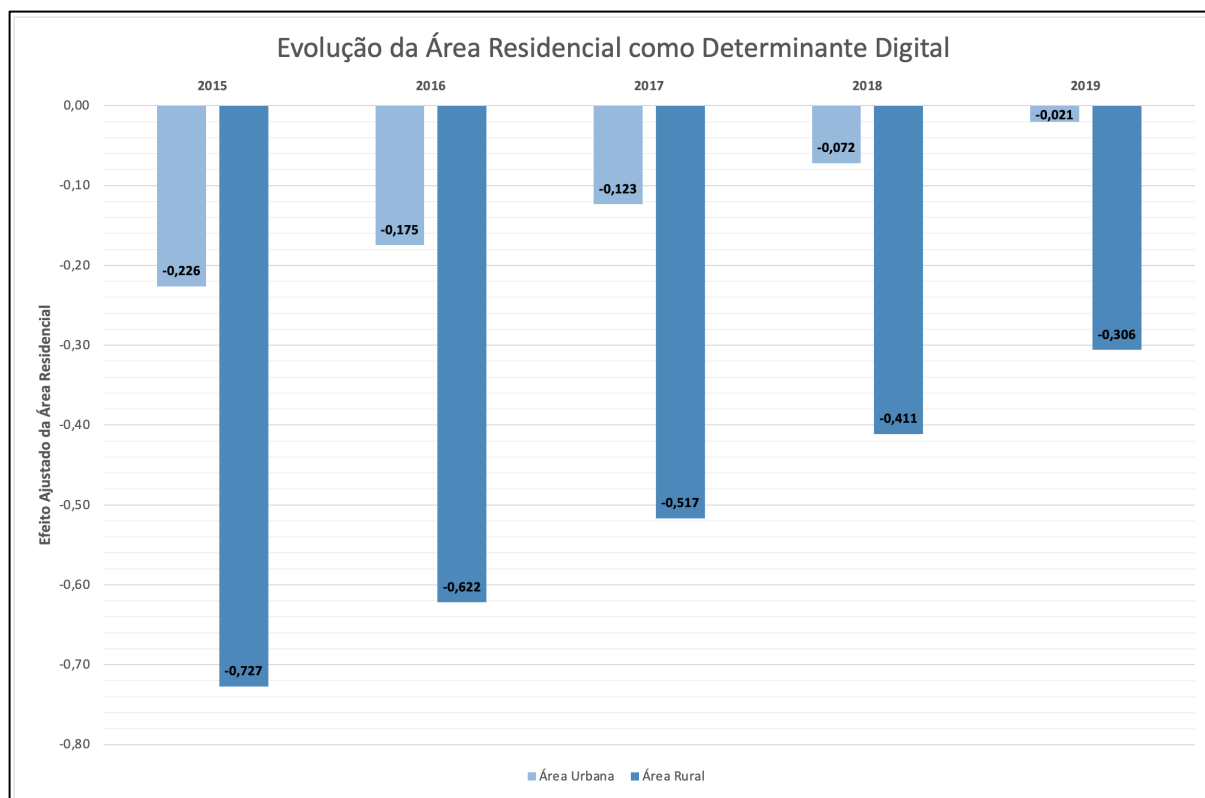
Nishijima, Ivanauskas e Sarti (2017, p. 21, tradução nossa) argumentam, com relação à realidade brasileira, que “indivíduos residentes em áreas rurais têm probabilidade reduzida de acesso à Internet, possivelmente devido à falta de infraestrutura” e projetam que “a localização geográfica (incluindo áreas estaduais e urbanas / rurais) mostra uma tendência crescente como determinante das desigualdades no acesso à Internet”.

A

Figura 10 apresenta a evolução do efeito ajustado deste determinante entre os anos de 2015-2019, cujos dados demonstram o aumento de relevância deste determinante. No entanto,

é possível constatar que as diferenças entre campo e cidade, para fins de adoção da Internet, diminuam, caracterizando uma melhoria, no período estudado, do abismo digital.

Figura 10 – Evolução da área residencial como determinante digital



Fonte: Elaborado pelo autor.

No primeiro ano estudado, as probabilidades de um morador de centros urbanos e um morador de áreas rurais fazerem uso da Internet eram de, respectivamente, 44,37% e 32,58%. Em 2019, esses valores foram de, respectivamente, 49,49% e 42,41%. Entre os anos de 2015 e 2019, a chance de um usuário rural fazer uso da Internet, em relação a um usuário urbano, saltou de 0,6059:1 (37,73%) para 0,7517:1 (42,91%). Ao final do período, 68,77% dos moradores de áreas urbanas, no Brasil, faziam uso da Internet e, pela primeira vez, as áreas rurais continham mais pessoas conectadas (50,75%) do que não conectadas, levando à redução de 6,17 pp da lacuna de acesso (Tabela B-4).

Os desafios de se levar a conectividade às comunidades no meio rural ou áreas remotas são complexos e, inevitavelmente, de alto custo. Segundo Haight, Quan-Haase e Corbett (2014, p. 505) sugerem, ao tratar do tema no contexto canadense, o número limitado de incentivos econômicos e a baixa densidade demográfica em face à enorme área geográfica de um país podem não ser atrativos para que operadoras de telecomunicações levem serviços de banda

larga para as localidades remotas - ainda que, nelas, possam existir pessoas com condições de arcar com esse serviço (ZHANG, 2013).

Zhang (2013, p. 526) argumenta que governos devem estabelecer políticas públicas para estimular ou subsidiar investimentos em infraestrutura onde os altos custos desestimulam as empresas de telecomunicações. Willies e Tranter (2006, p. 46) ratificam o entendimento de foco no acesso e qualidade de serviços de telecomunicações – inclusive da privatização de estatais de telecomunicações, como no caso da Austrália.

4.2 Análise de robustez e desempenho do modelo

O modelo teórico, utilizado neste estudo, tem como base o uso de regressão logística e busca identificar, dentre as variáveis de naturezas socioeconômicas, demográficas e culturais presentes na pesquisa TIC Domicílios 2015-2019, aquelas que tenham impacto no uso da Internet por qualquer brasileiro e que, por consequência, possam ser consideradas como determinantes do abismo digital no Brasil.

A expectativa com a utilização deste modelo não é de que ele seja o mais correto ou completo para tal fim, mas que possa ter um grau aceitável de desempenho de predição e que possa ser mais confiável do que simples adivinhações ou suposições.

Para identificar se o modelo adotado neste estudo tem tal característica, foram utilizadas recomendações de verificação de eficiência da regressão logística, que contemplam: “(a) avaliação geral do modelo, (b) teste de significância estatística de cada variável independente, (c) estatísticas de adequação e (d) validações de previsões de probabilidades (PENG; LEE; INGERSOLL, 2002, p. 5, tradução nossa).

Com relação à (a) avaliação geral do modelo e às (c) estatísticas de adequação, segundo (PENG; LEE; INGERSOLL, 2002, p. 5, tradução nossa), “um modelo logístico fornece um melhor ajuste aos dados se demonstrar uma melhoria em relação ao modelo nulo”, assim entendido, como um modelo sem variáveis independentes.

A Tabela 5 apresenta os resultados do modelo estatístico proposto. Esses sugerem, com base em evidências estatísticas significativas (valor p (LLR): 0,000), um melhor resultado do modelo proposto com relação ao modelo nulo. O valor do Pseudo R^2 de 0,4370 sugere um desempenho aceitável do modelo proposto neste estudo (TREHAN; JOSHI, 2018, p.38). O Pseudo R^2 de McFadden estabelece uma comparação do “*log likelihood*” (PENG; LEE; INGERSOLL, 2002, p. 6) do modelo estudado e do modelo nulo e é utilizado para averiguar o ajuste do modelo aos dados (LAITILA, 1993; MCFADDEN, 1973).

As análises dos (b) níveis de significância das variáveis independentes demonstram evidências estatísticas, resultantes do Teste Z, de que essas são relevantes para determinar o uso da Internet e, consequentemente, refutam a hipótese nula. (Valor $p < 0,001$ para constante e as variáveis Idade (e variável de interação), Grau de Instrução (e variável de interação Grau de Instrução²), Classe Econômicas A, B e C (e variáveis de interação), População Economicamente Ativa, Religião e Área Residencial. Valor $p < 0,05$ para as variáveis Ano, Idade² e variáveis de interação referentes à População Economicamente Ativa, Religião e Área Residencial.)

Com relação às (d) validações de previsões de probabilidades, ainda que o objetivo deste estudo não seja prever, com base nos determinantes digitais, se um indivíduo fará uso da Internet, foi realizado este procedimento, para a obtenção das estatísticas que completam a análise de desempenho do modelo.

A Tabela 6 apresenta as informações relativas a esta avaliação. As métricas de desempenho de predição dão uma noção do quão confiáveis são as previsões do modelo de que uma pessoa fará o uso da Internet, com base nos coeficientes associados aos impactos das variáveis independentes, e, consequentemente, da viabilidade do seu uso para o estudo dos impactos que os determinantes digitais têm no abismo digital (TREHAN; JOSHI, 2018, p.39).

O modelo adotado possui acurácia de 83,07% ao prever corretamente se indivíduos são (ou não) usuários da Internet.

O indicador de sensibilidade informa a acurácia das previsões corretas de usuários em relação ao universo de usuários observados – em outras palavras, levando-se em consideração os resultados categorizados como verdadeiros positivos e falsos negativos. O indicador de especificidade refere-se à acurácia das previsões de não usuários em relação ao universo de não usuários observados – levando-se em consideração os resultados categorizados como verdadeiros negativos e falsos positivos. Os desempenhos do modelo de regressão logística para os indicadores de sensibilidade e especificidade foram respectivamente 88,61% e 75,11%.

O indicador de confiabilidade positiva informa a acurácia das previsões corretas de usuários em relação a todas as previsões de usuários. E o indicador de confiabilidade negativa informa a acurácia das previsões corretas de não usuários em relação a todas as previsões de não usuários. Os desempenhos do modelo de regressão logística para os indicadores de confiabilidade positiva e confiabilidade negativa foram respectivamente 83,66% e 82,09%.

Os resultados de avaliação de desempenho do modelo apontam resultados satisfatórios para identificar se uma pessoa faz uso da Internet. Alguns indicadores sugerem, no entanto, que o modelo tenha um melhor comportamento para identificar usuários do que não usuários e que

essa característica vai sendo reforçada ao longo dos anos estudados. Possivelmente, a ampliação da diferença na distribuição de usuários e não usuários, que é caracterizada pelo aumento anual da adoção de Internet no país, responde esta questão.

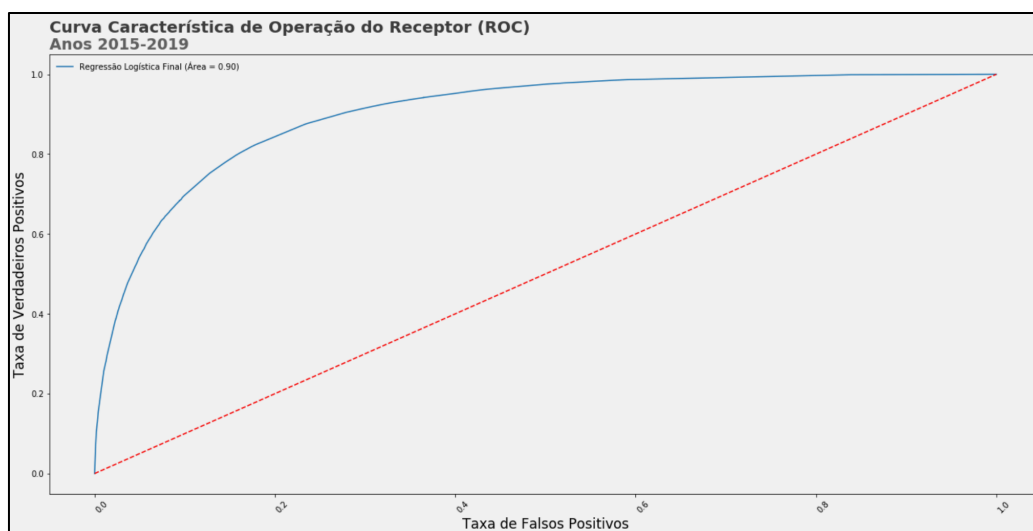
Tabela 6 – Métricas de desempenho de predição

Número de predições	89.399
Número de Verdadeiros Positivos	46.727 52,27%
Número de Verdadeiros Negativos	27.537 30,80%
Número de Falsos Positivos	9.126 10,21%
Número de Falsos Negativos	6.009 6,72%
Acurácia	83,07%
Taxa de Erro	16,93%
Sensibilidade	88,61%
Especificidade	75,11%
Confiabilidade Positiva	83,66%
Confiabilidade Negativa	82,09%

Fonte: Elaborado pelo autor.

A Curva ROC é um método gráfico utilizado para avaliação de desempenho de modelos de classificação binária. A área abaixo da Curva ROC, que varia, em valores absolutos, de 0,50 a 1,00, está relacionada à capacidade do modelo em prever as classes da variável dependente dicotômica, sugerindo que quanto maior a área, melhor o desempenho (TREHAN; JOSHI, 2018, p.40).

Figura 11 – Curva Característica de Operação de Receptor (ROC)



Fonte: Elaborado pelo autor.

A Figura 11 apresenta a Curva Característica de Operação do Receptor (ROC). O valor de 0,90 para a área abaixo da curva ratifica o bom desempenho do modelo deste estudo.

Levando-se em consideração os resultados das análises das Curvas ROC, das métricas de desempenho de predição, do Teste Z e do Pseudo R^2 , o modelo adotado neste estudo tem um bom ajuste aos dados e produz resultados, respaldados estatisticamente, no efeito das características socioeconômicas, demográficas e culturais dos indivíduos que participaram da pesquisa TIC Domicílios com a possibilidade de um indivíduo fazer uso da Internet e, conseqüentemente, utilizá-lo para identificar os determinantes digitais relacionados às características socioeconômicas, demográficas e culturais de brasileiros e analisar as suas influências no processo do abismo digital no país.

5 CONCLUSÃO

O tema do abismo digital é alvo de pesquisas há quase 30 anos. Ao longo deste período, muitas abordagens foram propostas para buscar a compreensão e o estabelecimento de um arcabouço teórico que pudesse explicar este fenômeno e estabelecer um padrão para a realização de pesquisas complementares e comparações entre realidades distintas (e.g. localidades, regiões, nações e, até mesmo, continentes).

Os distintos enfoques, que, ao longo desses anos, foram sendo aprimorados, buscaram analisar a desigualdade digital de diversos pontos de vista (e.g. do acesso, do uso, das capacidades necessárias, das motivações e dos benefícios do uso da Internet).

Nesse interim, oferta e consumo mudaram. Novas tecnologias de acesso surgiram, o perfil do usuário mudou, a mobilidade e uso de dispositivos pessoais, diferentes dos tradicionais computadores de mesa, correspondem ao padrão de uso atual e a importância da rede se tornou cada vez maior na vida das pessoas, das empresas e da administração pública.

Enquanto as sociedades e suas instituições vão, paulatinamente, substituindo os serviços analógicos por suas versões digitais, disponíveis na Internet, respaldadas pela expectativa de redução de custos, maior inclusão e maior efetividade, aumenta-se a preocupação com aqueles marginalizados pelo processo de digitalização e com a possibilidade de que eles, por não serem cidadãos conectados, não tenham as mesmas oportunidades e benefícios advindos da transformação digital e o acesso a serviços públicos que os usuários da Internet possuem.

A realidade do Brasil acompanha as mudanças globais e, ano após ano, a Internet vem ganhando mais usuários – o que ratifica a preocupação, no caso brasileiro, com aqueles que foram sendo esquecidos. Apesar de ter havido, nos últimos anos, iniciativas, públicas e privadas, que buscaram contribuir para a universalização do acesso e redução do abismo digital no país (e.g. crescimento do acesso em áreas rurais, classes econômicas inferiores, etc.), há ainda segmentos específicos da população brasileira que merecem atenção especial (e.g. 40 milhões de pessoas com ensino fundamental não fazem uso da Internet).

A pandemia da COVID-19 tornou explícitas não somente as possibilidades do uso de tecnologia para proporcionar, às pessoas, às organizações, à academia e aos governos, um meio de se adaptarem, derrubando barreiras e preconceitos, a um “novo normal”¹⁹, mas, demonstrou, também, indícios de que este pode não estar acessível a todos os cidadãos brasileiros (e.g. fim

¹⁹ Termo cunhado no período e amplamente utilizado pela imprensa e meios de comunicações para fazer referência às consequências, temporárias ou permanentes, que a pandemia trouxe às realidades de indivíduos, empresas e sociedade geral.

das atividades no ensino básico na rede público e continuação das mesmas na rede privada, por meio de aulas *online*, durante a pandemia). Atribui-se à crise sanitária um caráter de catalisador da adoção tecnológica, mas, sob a ótica de políticas públicas e governo, é necessário esclarecer se ela ocorreu de forma inclusiva e justa, ou seja, é vital compreender o impacto da doença com relação às desigualdades digitais, não, somente, enquanto acesso à Internet, mas, principalmente, aos benefícios que ele traz e aos conhecimentos e às habilidades exigidas para tal.

Os dados secundários utilizados nesta pesquisa não contribuem para responder esta questão – visto que sua coleta foi encerrada em março de 2020, mês em que foi declarada a pandemia global e estabelecidas as regras sanitárias de combate ao surto do coronavírus e as adaptações pertinentes ao dia a dia das pessoas e das instituições (e.g. isolamento social e teletrabalho). No entanto, eles contribuem com o objetivo desta dissertação de diagnosticar a situação anterior à pandemia, identificando os fatores que contribuem com o abismo digital no Brasil e analisando como suas influências neste processo evoluíram ao longo dos 5 anos que antecederam a crise sanitária.

Os achados deste estudo apontam que, dentre os indicadores socioeconômicos, demográficos e culturais presentes na pesquisa TIC Domicílios, idade, grau de instrução, classe econômica, população economicamente ativa, área residencial e religião impactaram, com relevância estatística, o uso da Internet no Brasil ao longo do período analisado e, por isso, apresentam características de determinantes digitais – ainda que, para o caso da religião, faz-se necessário compreender melhor as razões pelas quais o fato de não ser católico tenha apresentado impacto expressivo. Os indicadores de sexo, etnia e regionalidade não apresentaram relação estatisticamente relevante com o uso da rede no país nos anos analisados. E a renda familiar foi desconsiderada pela correlação com a classe econômica.

Dentre os determinantes do modelo proposto, o grau de instrução, classe econômica, ser economicamente ativo, não ser católico e morar em centros urbanos têm influência positiva na adoção da Internet, enquanto, a idade apresentou influência negativa.

Com relação à evolução dos determinantes nos anos de 2015-2019, a idade perdeu relevância no estabelecimento das desigualdades digitais. O grau de instrução, por sua vez, apresentou queda da influência dos ensinos médio e superior (12 anos de estudos ou mais) e aumento no caso de analfabetismo e início da vida escolar, minimizando, assim, as consequências indesejadas desse determinante para o estabelecimento do abismo digital e beneficiando o público com menor nível educacional.

Com relação aos indicadores econômicos, a importância de ser economicamente ativo perdeu relevância ao longo do período, chegando a ser equiparada, em 2019, pelos efeitos de não ser membro dessa população. Esse achado refuta a hipótese originalmente estabelecida para este estudo.

Do ponto de vista da classe econômica, o impacto negativo associado aos estratos sociais com menor poder de compra diminuiu, contribuindo para o aumento da probabilidade dos membros das classes C, D e E fazerem uso da Internet e reduzindo o abismo digital associado a este determinante.

Com relação aos aspectos culturais, no período analisado, não ser católico apresentou uma crescente influência para determinar se um indivíduo será usuário da rede. No entanto, pesquisas mais aprofundadas são sugeridas, para a compreensão deste tema, cujas razões podem não estar totalmente associadas ao campo religioso.

E, por fim, o aumento da quantidade de pessoas conectadas no meio rural, apresentado nos anos de 2015-2019, fez com que o efeito positivo de viver em centros urbanos perdesse força, contribuindo para a redução das desigualdades digitais entre campo e cidade.

A identificação, a análise e o monitoramento da evolução dos determinantes do abismo digital são aspectos importantes para o estabelecimento e a avaliação de resultados de políticas públicas que tratem a desigualdade social em uma nação que oferece serviços públicos e oportunidades econômicas, educacionais e culturais via Internet.

A história recente do Brasil possui exemplos promissores de políticas que visam à inclusão digital, com expectativas de resultados relevantes a médio e longo prazos, mas que precisariam ser intensificados e priorizados para atender a urgência demandada a esta questão, inclusive por parte de governos federais e estaduais que buscam, por meio da adoção de tecnologias digitais (e.g. Internet, dispositivos móveis, Inteligência Artificial, etc.) reduzir custos, desburocratizar o Estado e melhorar o atendimento, em termos de rapidez e precisão, à população.

Marcos regulatórios como a Lei nº 13.879, de 03 de outubro de 2019, que alterou a modalidade de outorga dos serviços de telecomunicações e a Lei nº 14.109, de 16 de dezembro de 2020, que trata da finalidade, destinação de recursos e gestão do Fundo de Universalização dos Serviços de Telecomunicações (FUST), buscam desonerar o setor de obrigações relativas ao Sistema Telefônico Fixo Comutado, destravar investimentos e promover o aumento de cobertura e velocidade do acesso Internet no território brasileiro, minimizando, assim, os impacto negativo do determinante relativo à área de residência.

A reativação da estatal Telecomunicações Brasileiras S.A. (Telebras), o lançamento do Satélite Geoestacionário de Defesa e Comunicações Estratégicas (SGDC), programas como Governo Eletrônico - Serviço de Atendimento ao Cidadão (Gesac) que levam “gratuitamente, conexão à internet em banda larga - por via terrestre e satélite - a telecentros, escolas, unidades de saúde, aldeias indígenas, postos de fronteira e quilombos” (BRASIL, 2020i) e o estabelecimento de áreas públicas (e.g. praças, estacionamentos, estações de metrô, etc.) com acesso gratuito à Internet (e.g. Programa Wi-Fi Brasil) foram ações propostas para ajudar no combate ao abismo digital de acesso.

E subsídios de planos de dados (“*zero rating*”) voltados para serviços públicos básicos e universais ajudam, não só com a questão do acesso, mas contribuem positivamente com aspectos dos níveis II e III do abismo digital (e.g. motivação e benefícios). Um exemplo recente de tal política é o caso da Secretaria de Educação do Estado de São Paulo, que iniciou, em 2021, a entrega de *chips* de Internet móvel (3G) para que alunos da rede pública de ensino possam acessar, gratuitamente, os conteúdos do Centro de Mídias de São Paulo (CMSP) e de outras plataformas por até uma hora e quarenta e cinco minutos por dia. O órgão espera entregar um total de 500 mil chips para alunos do “6º ao 9º ano do Ensino Fundamental e do Ensino Médio das escolas regulares (incluindo EEI [Educação Escolar Indígena] – Indígena, Quilombo, Área de Assentamento e alunos do noturno regular das PEIs [Programa de Ensino Integral])” (SÃO PAULO (ESTADO), 2021).

As iniciativas de políticas públicas precisam focar seus esforços e recursos nos determinantes digitais identificados – buscando conter os seus impactos negativos e minimizar as suas consequências. Ações fora deste escopo são louváveis e justificáveis, particularmente nos contextos específicos de comunidades locais. No entanto, elas não encontram amparo na racionalidade econômica, que considera a análise de custo de oportunidade e ganhos marginais, buscando maximizar os resultados durante as tomadas de decisão.

Com relação às limitações e às possibilidades de continuidade desse estudo, uma contribuição para o campo de pesquisa poderia ser feita por meio da utilização de outras metodologias e instrumentos estatísticos que possam validar os achados apresentados nesta dissertação.

Uma proposição para tal poderia ser a análise da difusão desigual da Internet, e de seus determinantes, com base em informações relativas ao país e não, somente, de usuários (e.g. Produto Interno Bruto per capita, Índice de Desigualdade Social (Índice Gini), Velocidade de Adoção de Países Similares, etc. (ZHANG, 2013).

Outra possibilidade seria a utilização de informações de bases distribuídas (e.g. operadoras de telecomunicações e provedores de acesso, Agência Nacional de Telecomunicações, Instituto Brasileiro de Geografia e Estatística, Ministérios, etc.), ao invés de microdados de uma pesquisa amostral, como é o caso da TIC Domicílios, e a utilização de tecnologias, como Big Data, que facilitam o processamento de um grande volume de dados (especialmente, no caso de dados não estruturados) e a aplicação de algoritmos distintos da regressão logística.

Finalmente, a contribuição que este estudo pretende deixar, além de dar continuidade a pesquisas anteriores e utilizar-se de variáveis específicas, como o Critério Brasil 2015 e a religião no contexto do abismo digital no Brasil, é de prover informações relevantes às futuras discussões e implementações de políticas públicas que possam ajudar a reduzir as lacunas de conhecimentos associadas à desigualdade digital no país e permitir a inclusão digital do maior número de brasileiros. Adicionalmente, dadas as circunstâncias nas quais este estudo foi realizado, os seus resultados estabelecem um ponto de referência em relação à situação prévia à pandemia da COVID-19 e pode ser utilizado como contraponto aos desdobramentos da crise sanitária em avaliações do seu impacto sobre o abismo digital no país.

REFERÊNCIA BIBLIOGRÁFICA

AGARWAL, R.; ANIMESH, A.; PRASAD, K. Social interactions and the “digital divide”: Explaining variations in internet use. **Information Systems Research**, v. 20, n. 2, p. 277–294, 2009.

ARIA, M.; CUCCURULLO, C. bibliometrix: An R-tool for comprehensive science mapping analysis. **Journal of Informetrics**, v. 11, n. 4, p. 959–975, 2017.

BARZILAI-NAHON, K. Gaps and bits: Conceptualizing measurements for digital divide/s. **Information Society**, v. 22, n. 5, p. 269–278, 2006.

BONFADELLI, H. The Internet and knowledge gaps - A theoretical and empirical investigation. **European Journal of Communication**, v. 17, n. 1, p. 65–84, mar. 2002.

BRANDTZAEG, P. B.; HEIM, J.; KARAHASANOVIC, A. Understanding the new digital divide-A typology of Internet users in Europe. **International Journal of Human-Computer Studies**, v. 69, n. 3, p. 123–138, mar. 2011.

BRASIL. Decreto nº 7.175, de 12 maio de 2010. Institui o Programa Nacional de Banda Larga - PNBL. **Diário Oficial da União**, Brasília, DF, 13 maio 2010. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/decreto/d7175.htm>. Acesso em: 2 jun. 2021. 2010.

BRASIL. Resolução Administrativa nº 1499, de 1º de fevereiro de 2012. Regulamenta o teletrabalho no âmbito do Tribunal Superior do Trabalho e dá outras providências. **Diário da Justiça eletrônico (DJe)**, Brasília, DF, 03 fev. 2012. Disponível em: <http://www.tst.jus.br/legislacao-pessoal/-/document_library/kT11/view_file/3717965>. Acesso em: 2 jun. 2021. 2012.

BRASIL. Lei nº 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. **Diário Oficial da União**, Brasília, DF, 24 abr. 2014. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato20112014/2014/lei/12_965.htm>. Acesso em: 2 jun. 2021. 2014.

BRASIL. Projeto de Lei (PL) 2723/15. Acrescenta o §3º ao art. 19 da Lei nº 8.112, de 11 de dezembro de 1990, que dispõe sobre o regime jurídico dos servidores públicos civis da União, das autarquias e das fundações públicas federais, para autorizar a implantação. 2015. Disponível em: <<https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=1672572>>. Acesso em: 2 jun. 2021. 2015.

BRASIL. Decreto nº 9.204 de 23 de novembro de 2017. Institui o Programa de Inovação Educação Conectada e dá outras providências. **Diário Oficial da União**, Brasília, DF, 24 Nov

2017, Disponível em: <<https://www2.camara.leg.br/legin/fed/decret/2017/decreto-9204-23-no-vembro-2017-785784-publicacaooriginal-154288-pe.html>>. Acesso em: 2 jun. 2021. 2017a.

BRASIL. MEC lança Programa Nordeste Conectado, que poderá beneficiar 16 milhões de pessoas. Ministério da Educação, Brasília, DF. Disponível em: <http://portal.mec.gov.br/ultimas-noticias/12-acoes-programas-e-projetos-637152388/5_0881-mec-lanca-programa-nordeste-conectado-que-podera-beneficiar-16-milhoes-de-pessoas>. Acesso em: 2 jun. 2021. 2017b.

BRASIL. Estratégia de Governança Digital - EGD. Transformação Digital: Cidadania e Governo. 2016 – 2019. Brasília, DF. Disponível em: <<https://www.gov.br/governodigital/pt-br/estrategia-de-governanca-digital/historico>>. Acesso em: 2 jun. 2021. 2018.

BRASIL. Lei nº 13.879 de 03 de outubro de 2019. Altera a Lei nº 9.472, de 16 de julho de 1997, para permitir a adaptação da modalidade de outorga de serviço de telecomunicações de concessão para autorização. **Diário Oficial da União**, Brasília, DF, 04 out. 2019. Disponível em: <<https://legis.senado.leg.br/norma/31534957>>. Acesso em: 2 jun. 2021. 2019a.

BRASIL. Decreto nº 9.854, de 25 de junho de 2019. Institui o Plano Nacional de Internet das Coisas e dispõe sobre a Câmara de Gestão e Acompanhamento do Desenvolvimento de Sistemas de Comunicação Máquina a Máquina e Internet das Coisas. **Diário Oficial da União**, Brasília, DF, 26 jun. 2019. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Decreto/D9854.htm>. Acesso em: 2 jun. 2021. 2019b.

BRASIL. Lei nº 13.989, de 15 de abril de 2020. Dispõe sobre o uso da telemedicina durante a crise causada pelo coronavírus (SARS-CoV-2). **Diário Oficial da União**, Brasília, DF, 04 de abr. 2020. Disponível em: <<https://legis.senado.leg.br/norma/31534957>>. Acesso em: 2 jun. 2021. 2020a.

BRASIL. Resolução nº 329 de 30 de julho de 2020. Regulamenta e estabelece critérios para a realização de audiências e outros atos processuais por videoconferência, em processos penais e de execução penal, durante o estado de calamidade pública, reconhecido pelo D. **Diário da Justiça eletrônico (DJe)**, Brasília, DF, 31 jul. 2020. Disponível em: <<https://atos.cnj.jus.br/atos/detalhar/3400>>. Acesso em: 2 jun. 2021. 2020b.

BRASIL. Decreto nº 10.332, de 28 de abril de 2020. Institui a Estratégia de Governo Digital para o período de 2020 a 2022, no âmbito dos órgãos e das entidades da administração pública federal direta, autárquica e fundacional e dá outras providências. **Diário Oficial da União**, Brasília, DF, 29 abr. 2020. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/decreto/D10332.htm>. Acesso em: 2 jun. 2021. 2020c.

BRASIL. Decreto nº 10.222, de 5 de fevereiro de 2020. Aprova a Estratégia Nacional de Segurança Cibernética. **Diário Oficial da União**, Brasília, DF, 06 fev. 2020. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/decreto/D10222.htm>. Acesso

em: 2 jun. 2021. 2020d.

BRASIL. Governo trabalha para digitalizar todos serviços públicos. Disponível em: <<https://www.gov.br/pt-br/noticias/financas-impostos-e-gestao-publica/2020/07/governo-trabalha-para-digitalizar-todos-servicos-publicos>>. Acesso em: 2 jun. 2021. 2020e.

BRASIL. Instrução Normativa nº 65, de 30 de julho de 2020. Estabelece orientações, critérios e procedimentos gerais a serem observados pelos órgãos e entidades integrantes do Sistema de Pessoal Civil da Administração Federal - SIPEC relativos à implementação de Programa de Gestão. **Diário Oficial da União**, Brasília, DF, 31 jun. 2020. Disponível em: <<https://www.in.gov.br/en/web/dou/-/instrucao-normativa-n-65-de-30-de-julho-de-2020-269669395>>. Acesso em: 2 jun. 2021. 2020f.

BRASIL. Portaria nº 343, de 17 de março de 2020. Dispõe sobre a substituição das aulas presenciais por aulas em meios digitais enquanto durar a situação de pandemia do Novo Coronavírus - COVID-19. **Diário Oficial da União**, Brasília, DF, 18 mar. 2020. Disponível em: <<https://www.in.gov.br/en/web/dou/-/portaria-n-343-de-17-de-marco-de-2020-248564376>>. Acesso em: 2 jun. 2021. 2020g.

BRASIL. Portaria nº 61. Institui a plataforma emergencial de videoconferência para realização de audiências e sessões de julgamento nos órgãos do Poder Judiciário, no período de isolamento social, decorrente da pandemia Covid-19. **Diário da Justiça eletrônico (DJe)**, Brasília, DF, 31 mar. 2020. Disponível em: <<https://atos.cnj.jus.br/atos/detalhar/3266>>. Acesso em: 2 jun. 2021. 2020h.

BRASIL. **Obter conexão de internet GESAC (GESAC)**. Sítio Governo do Brasil, Brasília, DF, 29 out. 2020. Disponível em: <<https://www.gov.br/pt-br/servicos/obter-conexao-de-internet-gesac>>. Acesso em: 2 jun. 2021. 2020i.

BRASIL. **Programa Wi-Fi Brasil chega à marca de 13,2 mil pontos de internet e mais de 8,5 milhões de beneficiados**. Sítio Governo do Brasil, Brasília, DF, 16 mar. 2021. Disponível em: <<https://www.gov.br/mcom/pt-br/noticias/2021/marco/programa-wi-fi-brasil-chega-a-marca-de-13-2-mil-pontos-de-internet-e-mais-de-8-5-milhoes-de-beneficiados>>. Acesso em: 2 jun. 2021. 2021a.

BRASIL. **Leilão 5G. Conselho da Anatel é unânime quanto à rede privativa para o Governo Federal**. Sítio Governo do Brasil, Brasília, DF, 02 fev. 2021, Disponível em: <<https://www.gov.br/pt-br/noticias/transito-e-transportes/2021/02/conselho-da-anatel-e-unanime-quanto-a-rede-privativa-para-o-governo-federal>>. Acesso em: 2 jun. 2021. 2021b.

BRASIL. Decreto nº 10.661, de 26 de março de 2021. Regulamenta a Medida Provisória nº 1.039, de 18 de março de 2021, que institui o Auxílio Emergencial 2021 para o enfrentamento da emergência de saúde pública de importância internacional decorrente do coronavírus. **Diário Oficial da União**, Brasília, DF, 26 mar. 2021. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/decreto/D10661.htm>. Acesso em: 2 jun. 2021. 2021c.

COMISSÃO ECONÔMICA PARA A AMÉRICA LATINA (CEPAL). Universalizar el acceso a las tecnologías digitales para enfrentar los efectos del COVID-19. Disponível em: < <https://www.cepal.org/es/publicaciones/45938-universalizar-acceso-tecnologias-digitales-enfrentar-efectos-covid-19>>. Acesso em: 2 jun. 2021. 2020.

CENTRO REGIONAL DE ESTUDOS PARA O DESENVOLVIMENTO DA SOCIEDADE DA INFORMAÇÃO (CETIC.BR). **TIC Domicílios 2015**. Pesquisa Sobre o Uso das Tecnologias de Informação e Comunicação nos Domicílios Brasileiros. Disponível em: <<https://cetic.br/pt/publicacao/pesquisa-sobre-o-uso-das-tecnologias-de-informacao-e-comunicacao-nos-domicilios-brasileiros-tic-domicilios-2015/>>. Acesso em: 2 jun. 2021. 2016.

CENTRO REGIONAL DE ESTUDOS PARA O DESENVOLVIMENTO DA SOCIEDADE DA INFORMAÇÃO (CETIC.BR). **TIC Domicílios 2016**. Pesquisa Sobre o Uso das Tecnologias de Informação e Comunicação nos Domicílios Brasileiros. Disponível em: <<https://cetic.br/pt/publicacao/pesquisa-sobre-o-uso-das-tecnologias-de-informacao-e-comunicacao-nos-domicilios-brasileiros-tic-domicilios-2016/>>. Acesso em: 2 jun. 2021. 2017.

CENTRO REGIONAL DE ESTUDOS PARA O DESENVOLVIMENTO DA SOCIEDADE DA INFORMAÇÃO (CETIC.BR). **TIC Domicílios 2017**. Pesquisa Sobre o Uso das Tecnologias de Informação e Comunicação nos Domicílios Brasileiros. **Grappa Marketing Editorial**. Disponível em: <<https://cetic.br/pt/publicacao/pesquisa-sobre-o-uso-das-tecnologias-de-informacao-e-comunicacao-nos-domicilios-brasileiros-tic-domicilios-2017/>>. Acesso em: 2 jun. 2021. 2018.

CENTRO REGIONAL DE ESTUDOS PARA O DESENVOLVIMENTO DA SOCIEDADE DA INFORMAÇÃO (CETIC.BR). **TIC Domicílios 2018**. Pesquisa Sobre o Uso das Tecnologias de Informação e Comunicação nos Domicílios Brasileiros. **Grappa Marketing Editorial**. Disponível em: <<https://cetic.br/pt/publicacao/pesquisa-sobre-o-uso-das-tecnologias-de-informacao-e-comunicacao-nos-domicilios-brasileiros-tic-domicilios-2018/>>. Acesso em: 2 jun. 2021. 2019.

CENTRO REGIONAL DE ESTUDOS PARA O DESENVOLVIMENTO DA SOCIEDADE DA INFORMAÇÃO (CETIC.BR). **TIC Domicílios 2019**. Pesquisa Sobre o Uso das Tecnologias de Informação e Comunicação nos Domicílios Brasileiros. **Grappa Marketing Editorial**. Disponível em: <<https://cetic.br/pt/publicacao/pesquisa-sobre-o-uso-das-tecnologias-de-informacao-e-comunicacao-nos-domicilios-brasileiros-tic-domicilios-2019/>>. Acesso em: 2 jun. 2021. 2020.

CHAUDHURI, A.; FLAMM, K. S.; HERRIGAN, J. An analysis of the determinants of internet access. **Telecommunications Policy**, v. 29, n. 9–10, p. 731–755, 2005.

CHOI, N. G.; DINITTO, D. M. The Digital Divide Among Low-Income Homebound Older Adults: Internet Use Patterns, eHealth Literacy, and Attitudes Toward Computer/Internet Use. **Journal of Medical Internet Research**, v. 15, n. 5, maio 2013.

CULLEN, R. Addressing the digital divide. **Online Information Review**, v. 25, n. 5, p. 311–320, 2001.

D'HAENENS, L.; KOEMAN, J.; SAEYS, F. Digital citizenship among ethnic minority youths in the Netherlands and Flanders. **New Media & Society**, v. 9, n. 2, p. 278–299, abr. 2007.

DAVIS, R. A. A cognitive-behavioral model of pathological Internet use. **Computers in Human Behavior**, v. 17, n. 2, p. 187–195, mar. 2001.

DILMAGHANI, M. Religiosity and the digital divide in Canada. **The Communication Review**, v. 21, n. 3, p. 181–211, 2018.

DIMAGGIO, P. et al. Social implications of the Internet. **Annual Review of Sociology**, v. 27, p. 307–336, 2001.

DISTRITO FEDERAL. Decreto no 40.546, de 20 de março de 2020. Dispõe sobre o teletrabalho, em caráter excepcional e provisório, para os órgãos da administração pública direta, indireta, autárquica e fundacional do Distrito Federal, a partir de 23 de março de 2020, como medida necessária à continuidade do funcionamento da administração pública distrital, em virtude da atual situação de emergência em saúde pública e pandemia declarada pela Organização Mundial de Saúde (OMS), em decorrência do coronavírus (COVID-19). Diário Oficial do Distrito Federal, Brasília, DF, 20 mar. 2020. Disponível em: <<http://www.pg.df.gov.br/wp-conteudo/uploads/2020/03/40.546.pdf>>. Acesso em: 2 jun. 2021. 2020.

FERRO, E.; HELBIG, N. C.; GIL-GARCIA, J. R. The role of IT literacy in defining digital divide policy needs. **Government Information Quarterly**, v. 28, n. 1, p. 3–10, 2011.

FILGUEIRAS, F.; FLÁVIO, C.; PALOTTI, P. Digital Transformation and Public Service Delivery in Brazil. **Latin American Policy**, v. 10, n. 2, p. 195–219, 2019.

GIL-GARCIA, J. R.; HELBIG., N. C.; FERRO, E. **Is it only about Internet access? An empirical test of a multi-dimensional digital divide**. (K. Wimmer, MA and Scholl, HJ and Gronlund, A e Andersen, Ed.). Berlim, Alemanha: Springer-Verlag Berlin, 2006.

GILLEARD, C.; HIGGS, P. Internet use and the digital divide in the English longitudinal study of ageing. **European Journal of Ageing**, v. 5, n. 3, p. 233–239, set. 2008.

GOLDFARB, A.; PRINCE, J. Internet adoption and usage patterns are different: Implications for the digital divide. **Information Economics & Policy**, v. 20, n. 1, p. 2–15, mar. 2008.

GÓMEZ, D. C. The third digital divide and Bourdieu: Bidirectional conversion of economic,

cultural, and social capital to (and from) digital capital among young people in Madrid. **New Media & Society**, p. 1–20, 2020.

HAIGHT, M.; QUAN-HAASE, A.; CORBETT, B. A. Revisiting the digital divide in Canada: the impact of demographic factors on access to the internet, level of online activity, and social networking site usage. **Information Communication & Society**, v. 17, n. 4, SI, p. 503–519, abr. 2014.

HAMARI, J.; SJOKLINT, M.; UKKONEN, A. The sharing economy: Why people participate in collaborative consumption. **Journal of The Association for Information Science and Technology**, v. 67, n. 9, p. 2047–2059, set. 2016.

HARGITTAL, E. Whose Space? Differences Among Users and Non-Users of Social Network Sites. **Journal of Computer-Mediated Communication**, v. 13, n. 1, p. 276–297, out. 2007.

HARGITTAL, E.; HINNANT, A. Digital inequality - Differences in young adults' use of the Internet. **Communication Research**, v. 35, n. 5, p. 602–621, out. 2008.

HOPF, H. et al. Fake science and the knowledge crisis: ignorance can be fatal. **Royal Society Open Science**, v. 6, n. 5, maio 2019.

INTERNATIONAL TELECOMMUNICATION UNION (ITU). Manual for Measuring ICT Access and Use by Households and Individuals. p. 207, 2014.

IYENGAR, S.; MASSEY, D. S. Scientific communication in a post-truth society. **Proceedings of The National Academy of Sciences of The United States of America**, v. 116, n. 16, p. 7656–7661, abr. 2019.

JOHNSON, J. **Worldwide digital population as of January 2021**. Disponível em: <<https://www.statista.com/statistics/617136/digital-population-worldwide/>>. Acesso em: 4 jan. 2021. 2021.

KATZ, J.; ASPDEN, P. Motivations for and barriers to Internet usage: Results of a national public opinion survey. **Internet Research-Electronic Networking Applications and Policy**, v. 7, n. 3, p. 170, 1997.

KHONDKER, H. H. Role of the New Media in the Arab Spring. **Globalizations**, v. 8, n. 5, p. 675–679, 2011.

KING, A. L. S. et al. Nomophobia: Dependency on virtual environments or social phobia? **Computers in Human Behavior**, v. 29, n. 1, p. 140–144, jan. 2013.

KUSS, D. J.; GRIFFITHS, M. D. Online Social Networking and Addiction-A Review of the Psychological Literature. **International Journal of Environmental Research and Public Health**, v. 8, n. 9, p. 3528–3552, set. 2011.

LAITILA, T. A pseudo-R² measure for limited and qualitative dependent variable models. **Journal of Econometrics**, v. 56, n. 3, p. 341–356, abr. 1993.

LIBERATI, A. et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. **BMJ-British Medical Journal**, v. 339, jul. 2009.

LISSITSA, S.; CHACHASHVILI-BOLOTIN, S. Does the wind of change blow in late adulthood? Adoption of ICT by senior citizens during the past decade. **Poetics**, v. 52, p. 44–63, out. 2015.

LOO, B. P. Y.; NGAN, Y. L. Developing mobile telecommunications to narrow digital divide in developing countries? Some lessons from China. **Telecommunications Policy**, v. 36, n. 10–11, SI, p. 888–900, 2012.

LYSENKO, V. V; DESOUSA, K. C. Charting the coevolution of cyberprotest and counteraction: The case of former Soviet Union states from 1997 to 2011. **Convergence-The International Journal of Research into New Media Technologies**, v. 20, n. 2, p. 176–200, maio 2014.

MARR, B. **How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read** *Forbes*, 2018. Disponível em: <<https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/?sh=4ea96f6660ba>>. Acesso em: 4 jan. 2021. 2021.

MATO GROSSO DO SUL. Decreto nº 15.395, de 19 de março de 2020. Institui o Regime Excepcional de Teletrabalho no âmbito da Administração Pública do Estado de Mato Grosso do Sul, para a prevenção da transmissão e da proliferação da COVID-19 no território sul-mato-grossense. **Diário Oficial Eletrônico de Mato Grosso do Sul**, Campo Grande, MS, 20 mar. 2020. Disponível em: <<https://www.tat.ms.gov.br/decreto-no-15-395-de-19-de-marco-de-2020/>>. Acesso em: 4 jan. 2021. 2020.

MCFADDEN, D. **Conditional Logit Analysis of Qualitative Choice Behaviour**. New York, NY, USA: Academic Press New York, 1973.

MOHER, D. et al. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. **Plos Medicine.**, v. 6, n. 7, jul. 2009.

NEGROPONTE, N. **Being Digital**. New York: Alfred A. Knopf, Inc., 1995.

NGUYEN, M. H.; HARGITTAI, E.; MARLER, W. Digital inequality in communication during a time of physical distancing: The case of COVID-19. **Computers in Human Behavior**, v.

120, 2021.

NISHIJIMA, M.; IVANAUSKAS, T. M.; SARTI, F. M. Evolution and determinants of digital divide in Brazil (2005-2013). **Telecommunications Policy**, v. 41, n. 1, p. 12–24, fev. 2017.

OECD. **Making the Most of Technology for Learning and Training in Latin America** Paris, França OECD Skills Studies, OECD Publishing, 2020.

ONO, H.; ZAVODNY, M. Digital inequality: A five country comparison using microdata. **Social Science Research**, v. 36, n. 3, p. 1135–1155, set. 2007.

PAMPEL, F. C. **Logistic regression**. Thousand Oaks, Calif: Sage Publications Inc, 2000.

PENG, C. Y. J.; LEE, K. L.; INGERSOLL, G. M. An introduction to logistic regression analysis and reporting. **Journal of Education Research**, v. 96, n. 1, p. 3–14, 2002.

RAMSETTY, A.; ADAMS, C. Impact of the digital divide in the age of COVID-19. **Journal of the American Medical Informatics Association**, v. 27, n. 7, p. 1147–1148, 2020.

RICE, R. E.; KATZ, J. E. Comparing internet and mobile phone usage: digital divides of usage, adoption, and dropouts. **Telecommunications Policy**, v. 27, n. 8–9, p. 597–623, 2003.

RODICIO-GARCÍA, M. L. et al. La Brecha Digital en Estudiantes Españoles ante la Crisis de la Covid-19. **Revista Internacional de Educación para la Justicia Social**, v. 9, n. 3, p. 103–125, 2020.

SÃO PAULO (ESTADO). **Primeiros chips de internet começam a ser entregues para alunos da rede estadual de SP**. Portal do Governo do Estado de São Paulo. São Paulo, 2021. Disponível em: <<https://www.saopaulo.sp.gov.br/ultimas-noticias/educacao/primeiros-chips-de-internet-comecam-a-ser-entregues-para-alunos-da-rede-estadual-de-sp/>>. Acesso em: 2 jun. 2021

SCHEERDER, A.; VAN DEURSEN, A.; VAN DIJK, J. Determinants of Internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. **Telematics and Informatics**, v. 34, n. 8, p. 1607–1624, dez. 2017.

SELWYN, N. **Defining the “Digital Divide”: Developing a Theoretical Understanding of Inequalities in the Information Age**. Cardiff.: Cardiff University School of Social Sciences, 2002.

SHAH, D. V.; KWAK, N.; HOLBERT, R. L. “Connecting” and “disconnecting” with civic life: Patterns of Internet use and the production of social capital. **Political Communication**, v. 18, n. 2, p. 141–162, 2001.

STOLTERMAN, E.; FORS, A. C. Information technology and the good life. **IFIP Advances in Information and Communication Technology**, v. 143, n. January 2014, p. 687–692, 2004.

TICHENOR, P. J.; DONOHUE, G. A.; OLIEN, C. N. Mass Media Flow and Differential Growth in Knowledge. **Public Opinion Quarterly**, v. 34, n. 2, p. 159–, 1970.

TOKUNAGA, R. S. Following you home from school: A critical review and synthesis of research on cyberbullying victimization. **Computers in Human Behavior**, v. 26, n. 3, p. 277–287, maio 2010.

TREHAN, S.; JOSHI, R. M. Building and Evaluating Logistic Regression Models for Explaining the Choice to Adopt MOOCs in India. **International Journal of Education and Development using Information and Communication Technology**, v. 14, n. 1, p. 33–51, 2018.

VAN DEURSEN, A.J.A.M. AND HELSPER, E. J. “The Third-Level Digital Divide: Who Benefits Most from Being Online?” **Communication and Information Technologies Annual (Studies in Media and Communications)**, v. 10, p. 29–52, 2015.

VAN DEURSEN, A. J. A. M. Digital inequality during a pandemic: Quantitative study of differences in COVID-19-related internet uses and outcomes among the general population. **Journal of Medical Internet Research**, v. 22, n. 8, p. 1–13, 2020.

VAN DEURSEN, A. J. A. M.; VAN DIJK, J. A. G. M. Internet skills and the digital divide. **New Media & Society**, v. 13, n. 6, p. 893–911, set. 2011.

VAN DEURSEN, A. J. A. M.; VAN DIJK, J. A. G. M. The digital divide shifts to differences in usage. **New Media & Society**, v. 16, n. 3, p. 507–526, maio 2014.

VAN DEURSEN, A. J. A. M.; VAN DIJK, J. A. G. M. The first-level digital divide shifts from inequalities in physical access to inequalities in material access. **New Media & Society**, v. 21, n. 2, p. 354–375, fev. 2019.

WARSCHAUER, M. Dissecting the “digital divide”: A case study in Egypt. **Information Society**, v. 19, n. 4, p. 297–304, 2003.

WATTS, G. COVID-19 and the digital divide in the UK. **The Lancet Digital Health**, v. 2, n. 8, p. e395–e396, 2020.

WEI, K.-K. et al. Conceptualizing and Testing a Social Cognitive Model of the Digital Divide. **Information Systems Research**, v. 22, n. 1, p. 170–187, mar. 2011.

WILLIS, S.; TRANTER, B. Beyond the “digital divide” - Internet diffusion and inequality in Australia. **Journal of Sociology**, v. 42, n. 1, p. 43–59, mar. 2006.

WILSON, K. R.; WALLIN, J. S.; REISER, C. Social stratification and the Digital Divide. **Social Science Computer Review**, v. 21, n. 2, p. 133–143, 2003.

YILDIRIM, C.; CORREIA, A.-P. Exploring the dimensions of nomophobia: Development and validation of a self-reported questionnaire. **Computers in Human Behavior**, v. 49, p. 130–137, ago. 2015.

ZHANG, X. Income disparity and digital divide: The Internet Consumption Model and cross-country empirical research. **Telecommunications Policy**, v. 37, n. 6–7, p. 515–529, 2013.

APÊNDICE A - INFORMAÇÕES BIBLIOMÉTRICAS

Seguem, abaixo, os achados, referentes à bibliometria para a produção científica a respeito do abismo digital, identificados durante a revisão de literatura desta dissertação de mestrado.

Artigos de Revisão Sistemática de Literatura ou Meta Análise

Dentre os resultados encontrados, foram identificados 15 artigos que tem natureza de revisão sistemática de literatura (“*systematic review*”) e/ou meta análise (“*meta analysis*”) e foi verificado se estavam de acordo com *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) (LIBERATI et al., 2009; MOHER et al., 2009).

Tabela A-1 – Artigos de revisão sistemática de literatura e/ou meta análise

Título do Artigo	Autores	Publicação	Ano	Influência (5 anos)	PRISMA
A Systematic Review of Technology-Based Prevention and Treatment Interventions for Perinatal Depression and Anxiety in Latina and African American Women	Lara-Cinisomo, S; Olarte, AR; Rosales, M; Barrera, AZ	MATERNAL AND CHILD HEALTH JOURNAL	2021	2.311	
Applying the Electronic Health Literacy Lens: Systematic Review of Electronic Health Interventions Targeted at Socially Disadvantaged Groups	Cheng, C; Beauchamp, A; Elsworth, GR; Osborne, RH	JOURNAL OF MEDICAL INTERNET RESEARCH	2020	5.996	
The educational value of internet use in the home for school children: A systematic review of literature	Daoud, R; Starkey, L; Eppel, E; Vo, TD; Sylvester, A	JOURNAL OF RESEARCH ON TECHNOLOGY IN EDUCATION	2020 (early access)	1.585 (2019)	
Factors that affect the use of electronic personal health records among patients: A systematic review	Abd-alrazaq, AA; Bewick, BM; Farragher, T; Gardner, P	INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS	2019	3.525	
The Use of Cancer-Specific Patient-Centered Technologies Among Underserved Populations in the United States: Systematic Review	Tarver, WL; Haggstrom, DA	JOURNAL OF MEDICAL INTERNET RESEARCH	2019	5.996	
A Meta-Analysis of the Effects of Sociodemographic Factors on Social Media Adoption	Feng, GC; Zhang, YT; Lin, ZL	INTERNATIONAL JOURNAL OF COMMUNICATION JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION	2019	1.64	
Assessment of eHealth behaviors in national surveys: a systematic review of instruments	Hong, YA; Cho, J	JOURNAL OF TELEMEDICINE AND TELECare	2018	4.327	
Enablers and barriers to using two-way information technology in the management of adults with diabetes: A descriptive systematic review	Macdonald, EM; Perrin, BM; Kingsley, MIC	TELEMATICS AND INFORMATICS	2017	4.454	
Determinants of Internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide	Scheerder, A; van Deursen, A; van Dijk, J				

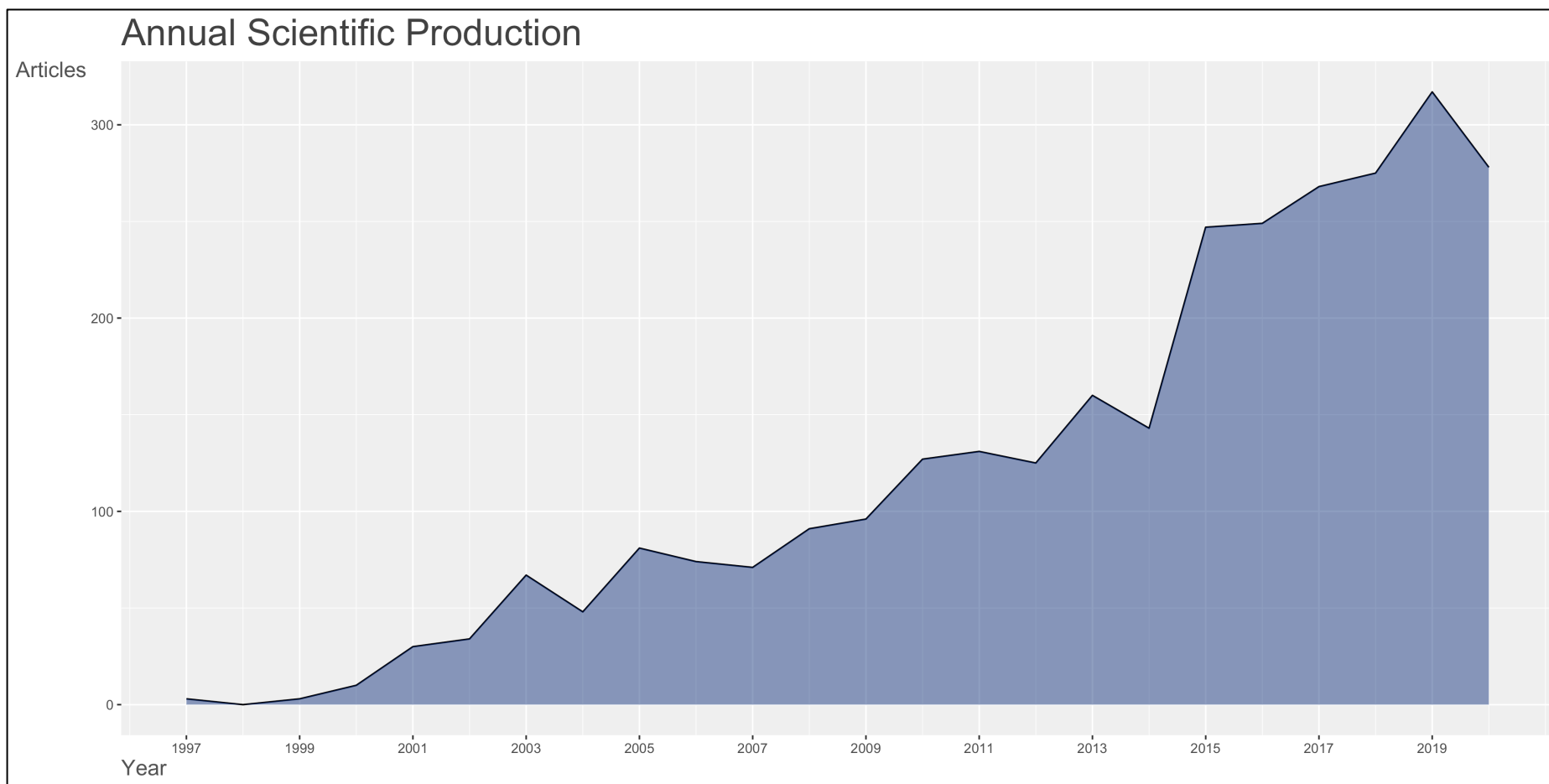
Health literacy in the eHealth era: A systematic review of the literature	Kim, H; Xie, B	PATIENT EDUCATION AND COUNSELING 23rd Americas Conference on Information Systems (AMCIS)	2017	3.408
Digital Divide in Information Systems Research: A Meta-Analysis and Framework	Palvia, Prashant; Oguz, Abdullah; Huang, Ziyue; et al.		2017	-
Work system barriers to patient, provider, and caregiver use of personal health records: A systematic review	Thompson, Morgan J.; Reilly, Jeremiah D.; Valdez, Rupa S.	APPLIED ERGONOMICS	2016	3.054
Navigating the digital divide: A systematic review of eHealth literacy in underserved populations in the United States	Chesser, Amy; Burke, Anne; Reyes, Jared; et al.	INFORMATICS FOR HEALTH & SOCIAL CARE	2016	1.73
Low Health Literacy and Evaluation of Online Health Information: A Systematic Review of the Literature	Diviani, Nicola; van den Putte, Bas; Giani, Stefano; et al.	JOURNAL OF MEDICAL INTERNET RESEARCH	2015	5.996
Towards a unified Media-User Typology (MUT): A meta-analysis and review of the research literature on media-user typologies	Brandtzaeg, Petter Bae	COMPUTERS IN HUMAN BEHAVIOR	2010	5.696

Fonte: Web of Science.

Produção Anual

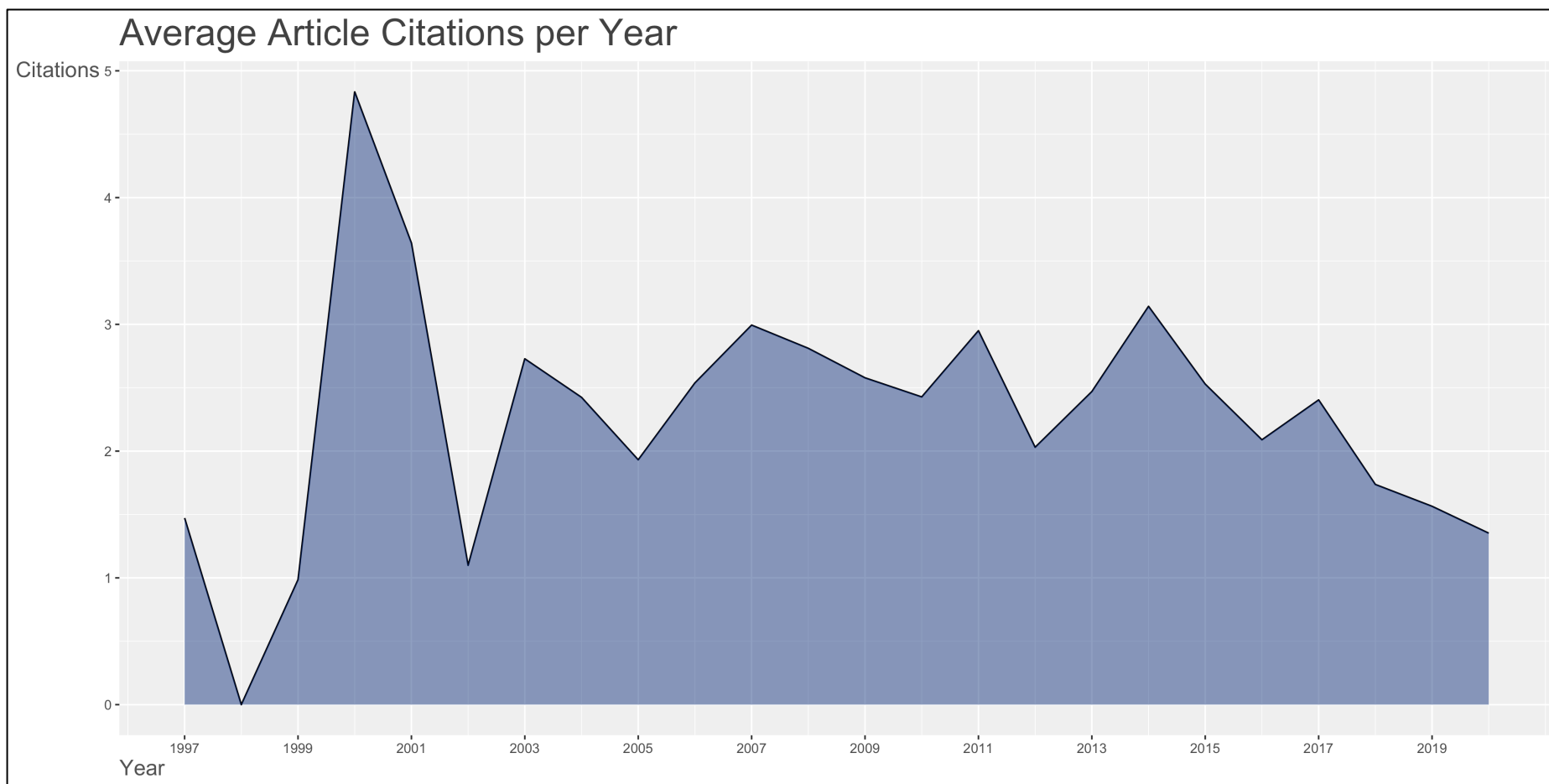
Com relação à atividade da produção científica mundial na área estudada, é observado, segundo Figura A1, que o interesse pelo tema vem aumentando a uma taxa média de 22,86% ao ano de 1997 a 2020. A Figura A2 apresenta a quantidade média de citações de artigos no mesmo período.

Figura A-1 – Histórico de produção científica a respeito do abismo digital



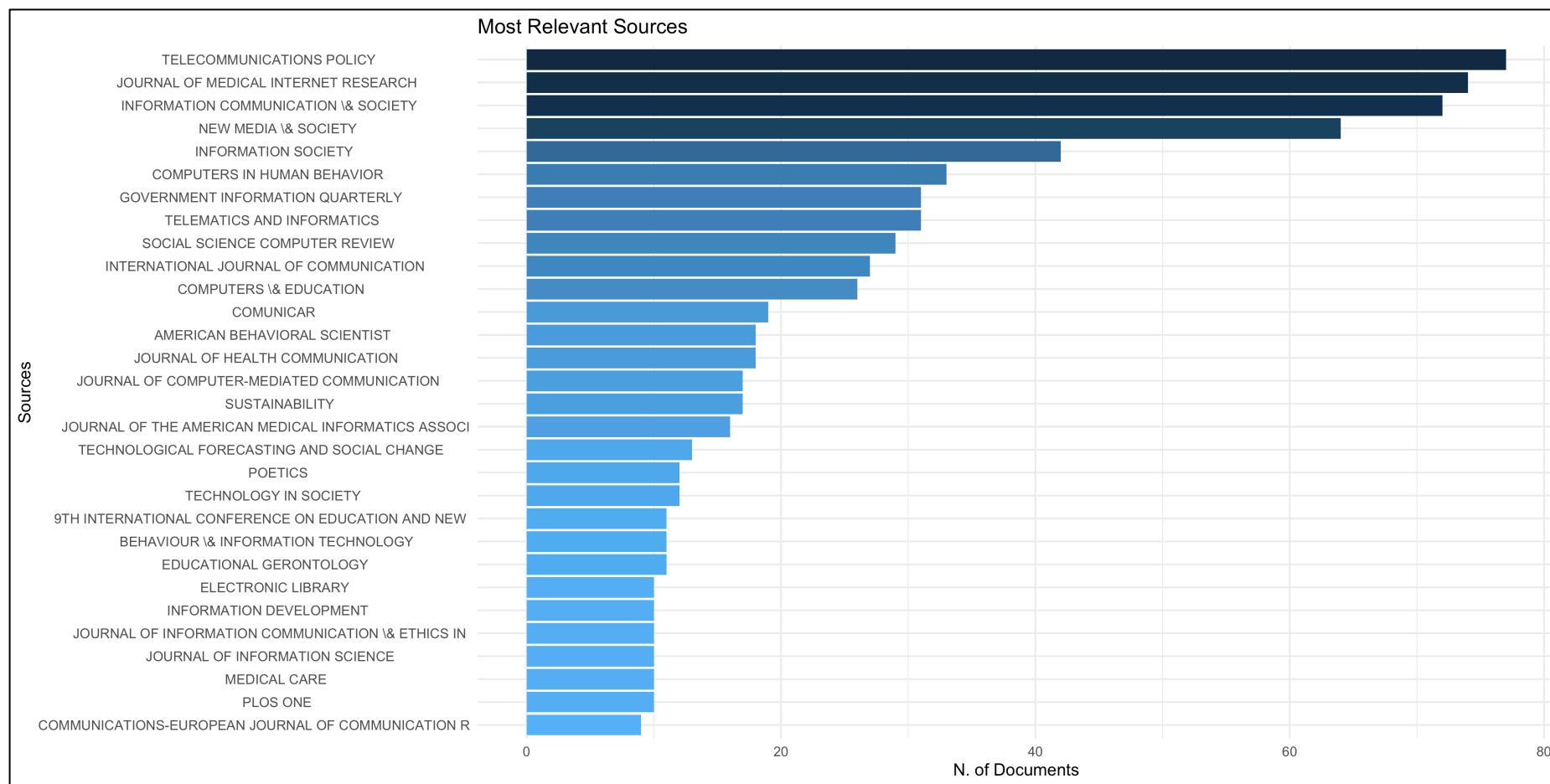
Fonte: Web of Science. Ferramenta: Bibliometrix

Figura A-2 – Histórico de citações de artigos que tratam do abismo digital



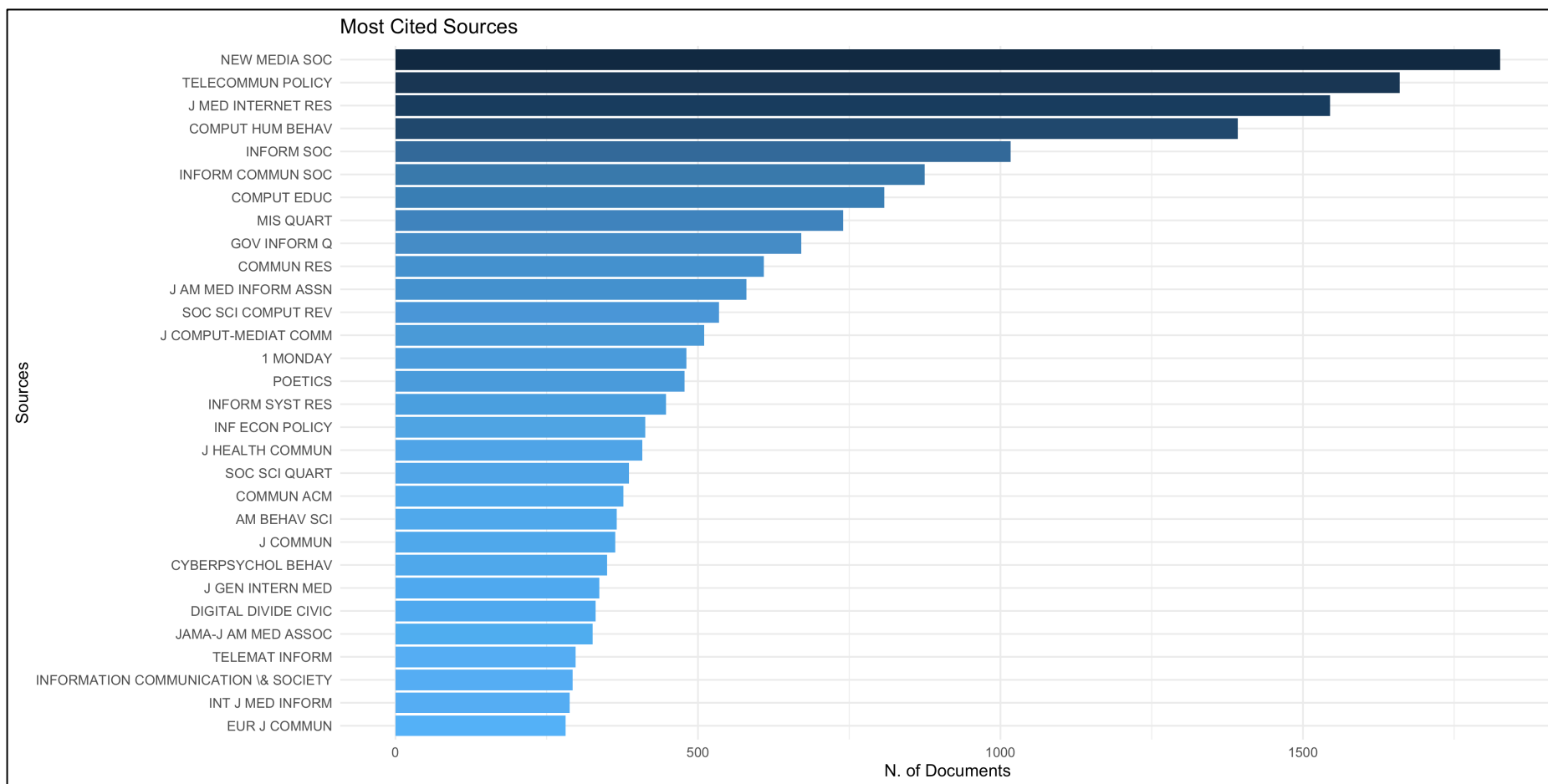
Fonte: Web of Science. Ferramenta: Bibliometrix

Figura A-3 – Gráfico com relação das 30 fontes mais relevantes a respeito do abismo digital



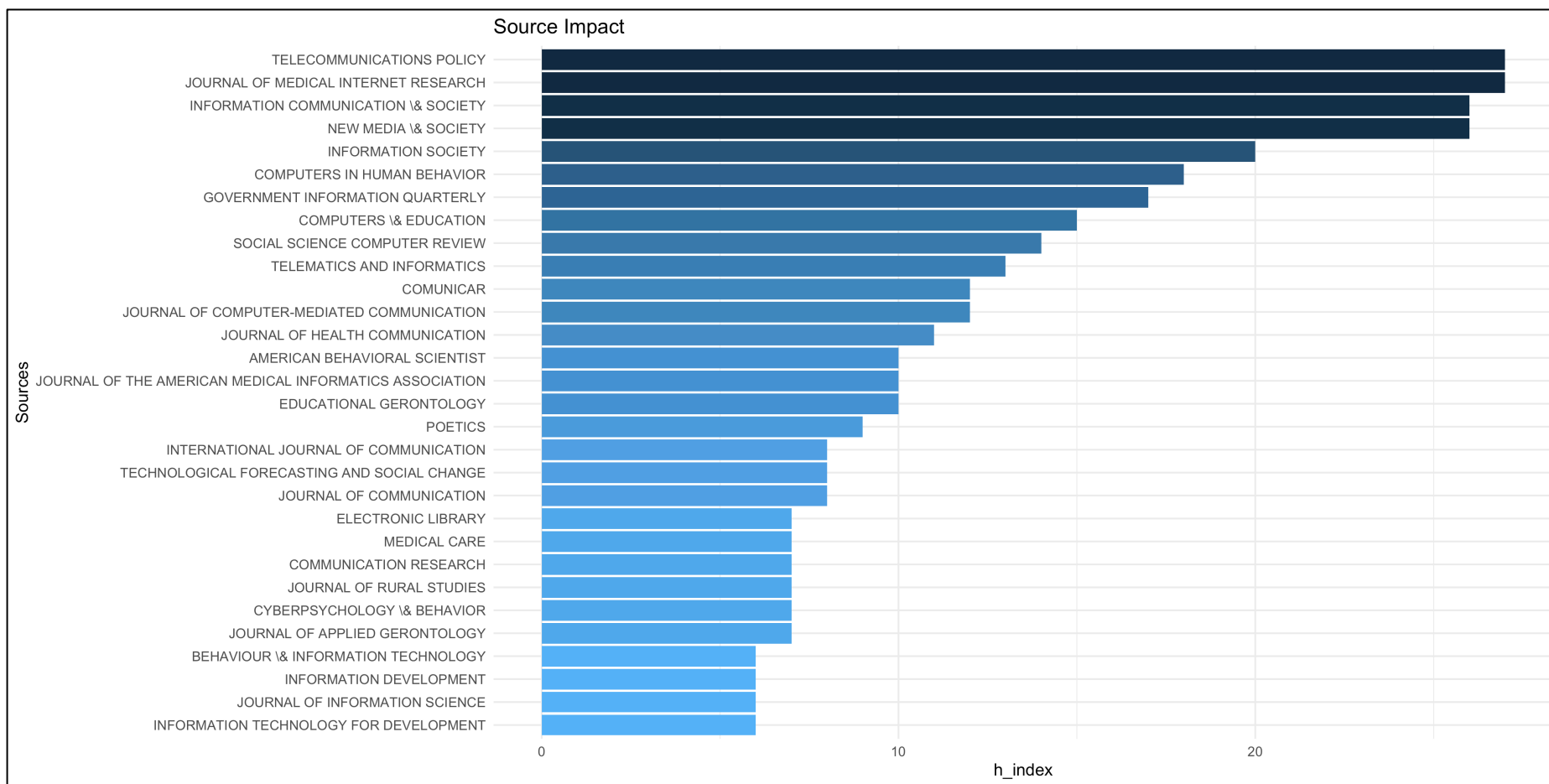
Fonte: Web of Science. Ferramenta: Bibliometrix

Figura A-4 – Gráfico com relação das 30 fontes mais citadas a respeito do abismo digital



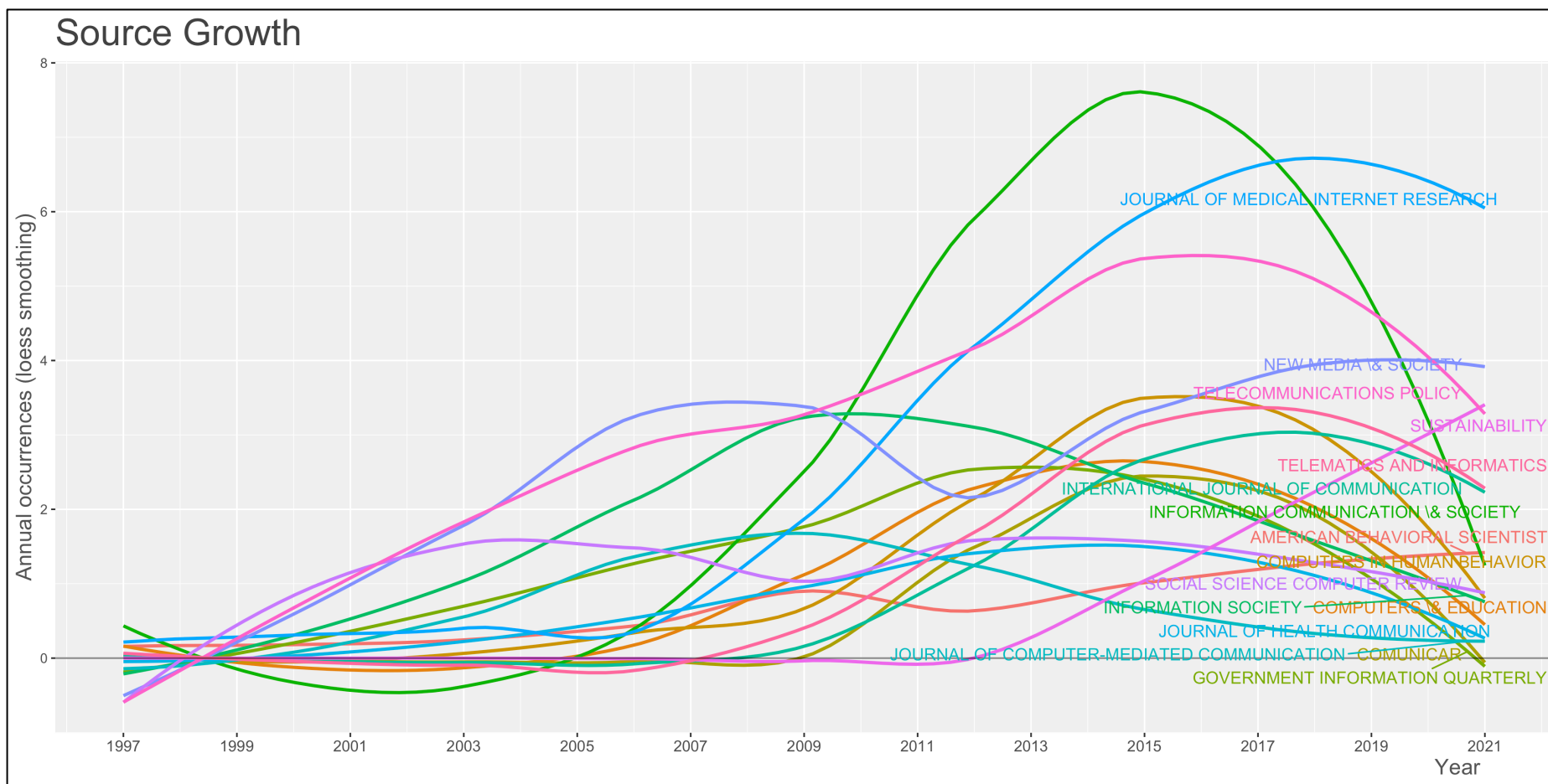
Fonte: Web of Science. Ferramenta: Bibliometrix

Figura A-5 – Gráfico com relação das 30 fontes de abismo digital com maior impacto



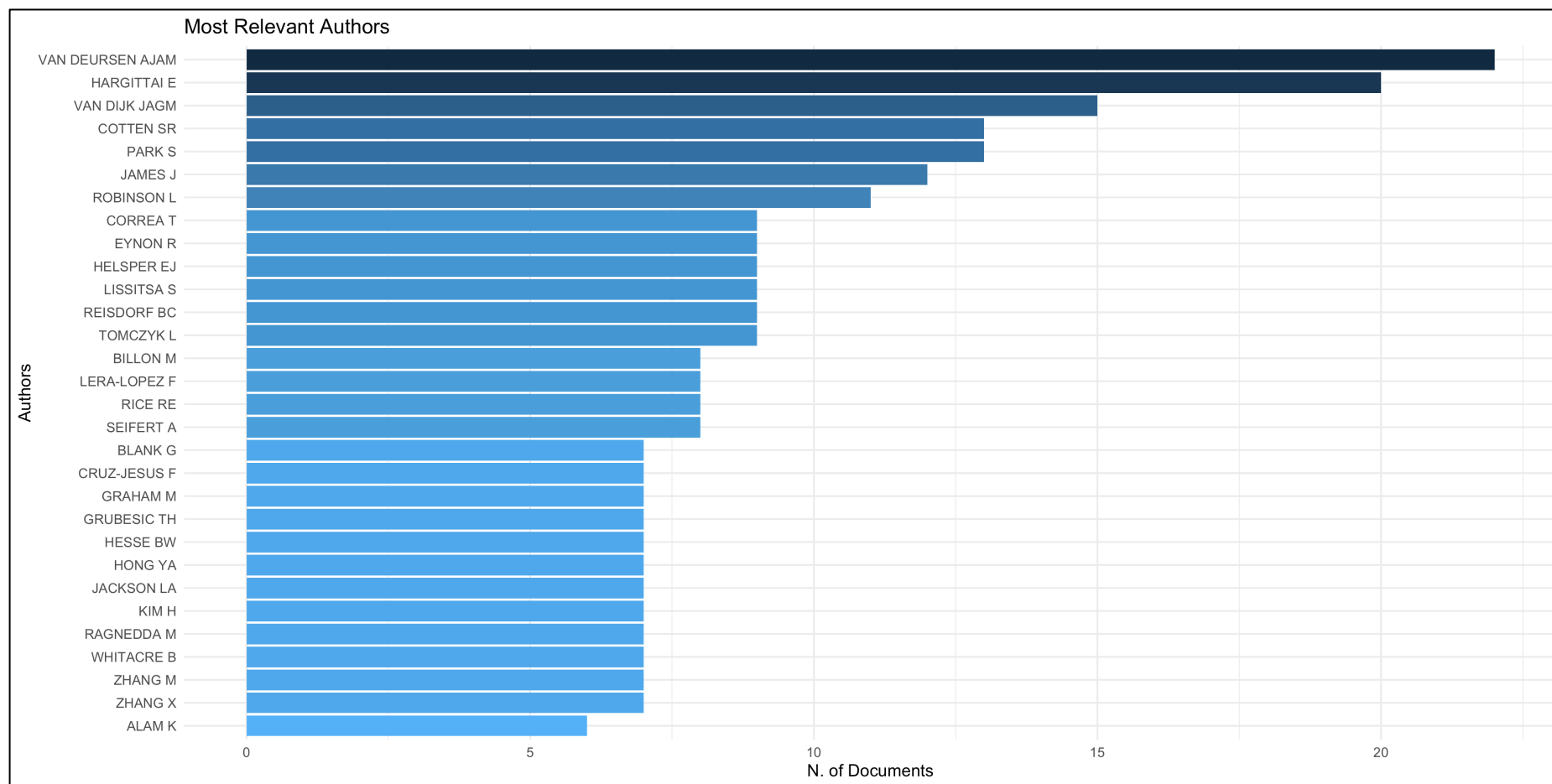
Fonte: Web of Science. Ferramenta: Bibliometrix

Figura A-6 – Histórico de produção anual das 15 fontes mais relevantes



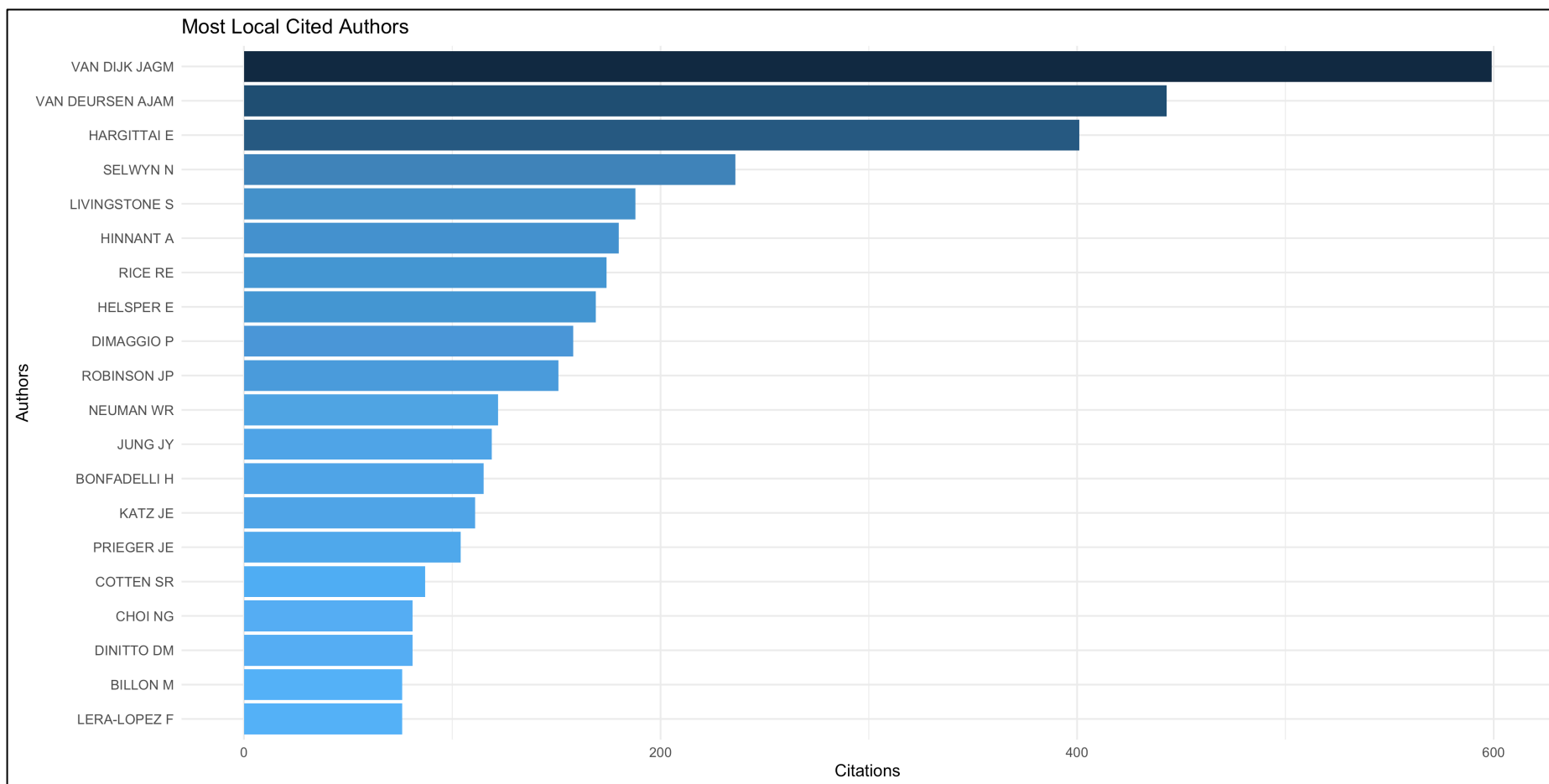
Fonte: Web of Science. Ferramenta: Bibliometrix

Figura A-7 – Gráfico com produção total dos 30 autores mais relevantes



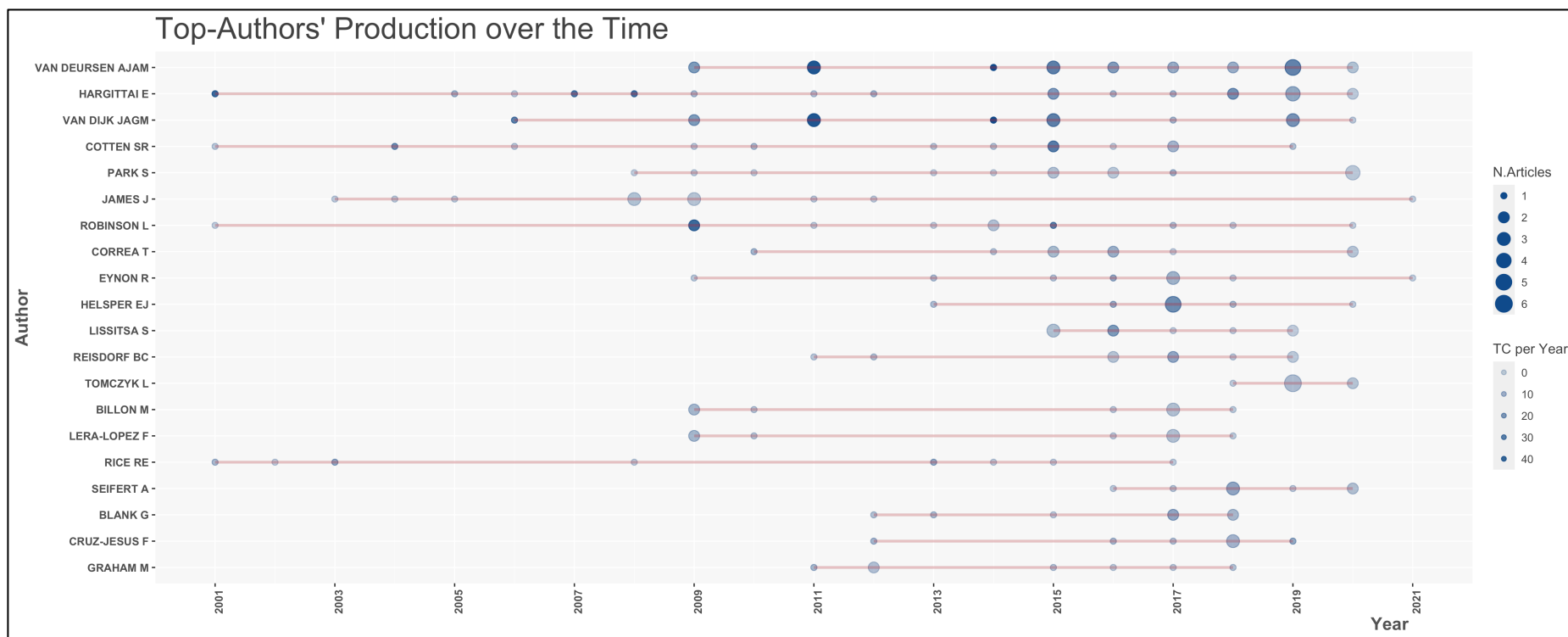
Fonte: Web of Science. Ferramenta: Bibliometrix

Figura A-8 – Gráfico com produção total dos 15 autores mais citados

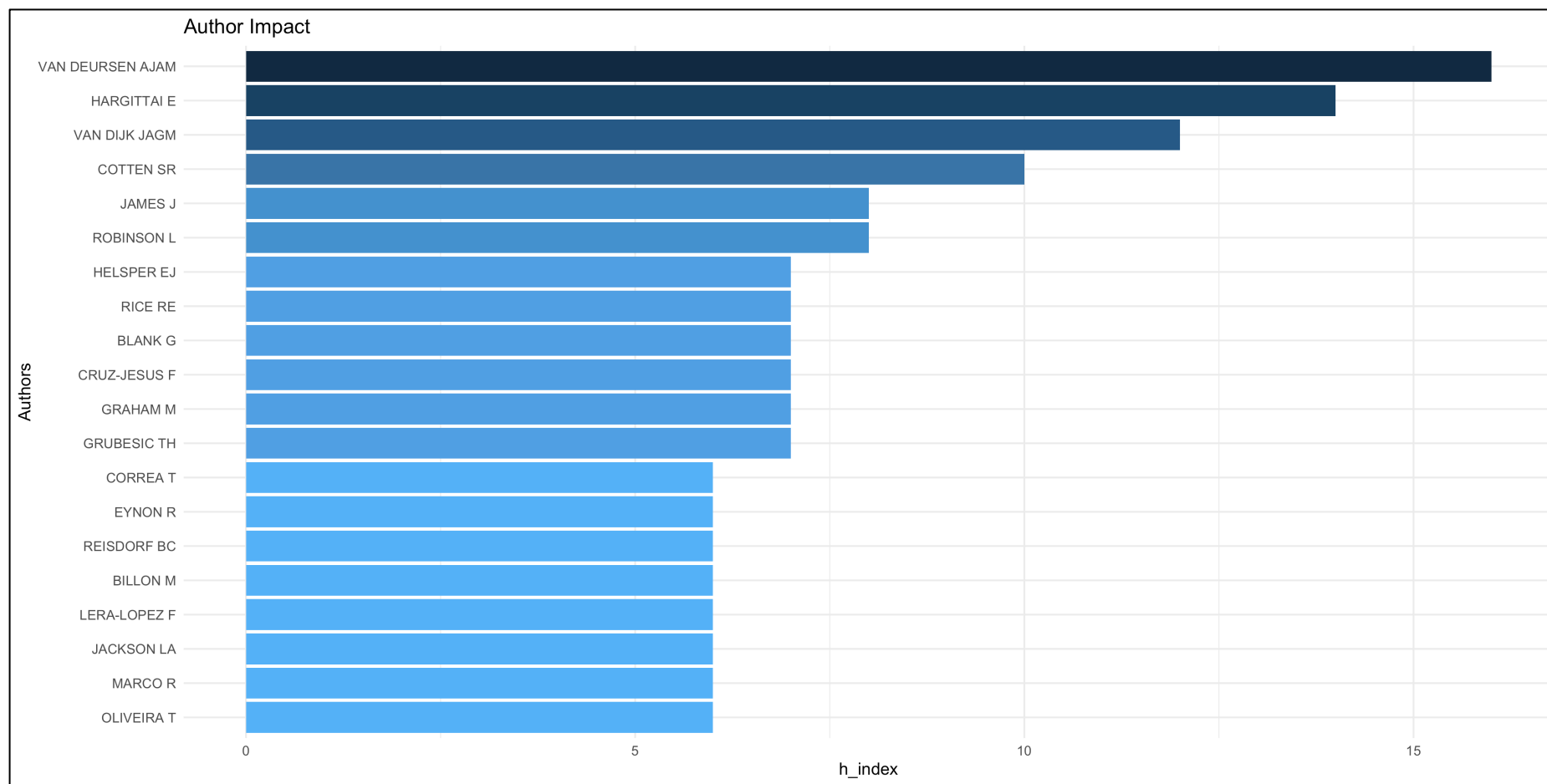


Fonte: Web of Science. Ferramenta: Bibliometrix

Figura A-9 – Histórico de produção anual dos 15 autores mais importantes

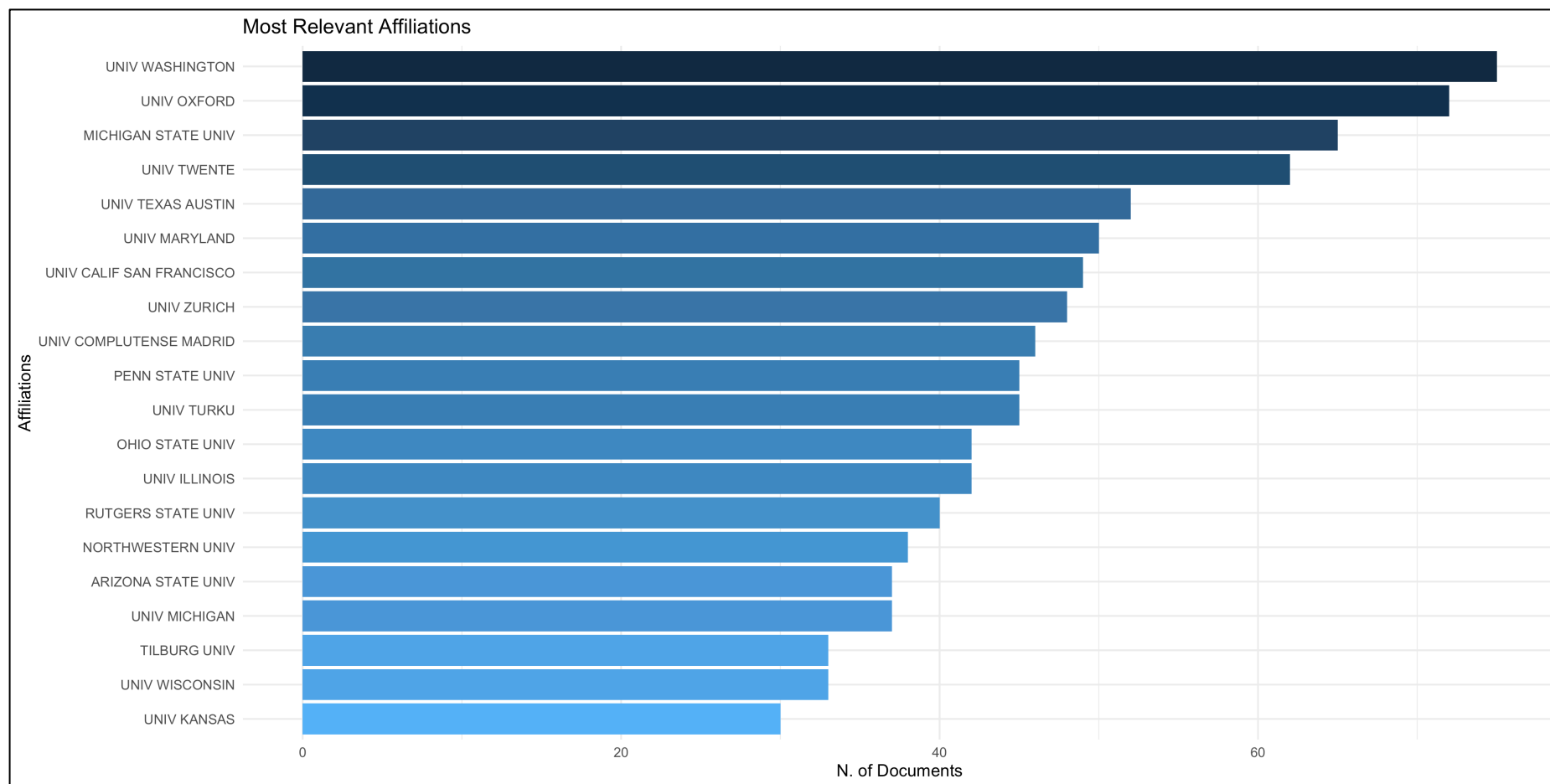


Fonte: Web of Science. Ferramenta: Bibliometrix

Figura A-10 – Relação dos 15 autores de maior impacto

Fonte: Web of Science. Ferramenta: Bibliometrix

Figura A-11– Relação das 20 instituições acadêmica mais relevantes com relação ao estudo do abismo digital



Fonte: Web of Science. Ferramenta: Bibliometria

APÊNDICE B - INFORMAÇÕES ESTATÍSTICAS

Tabela B-1– Informações originais de indicadores demográficos, socioeconômicos e culturais da Pesquisa TIC Domicílios (2015-2019)

Variáveis	2015	2016	2017	2018	2019	Total	Média	Desvio Padrão	Erro Padrão	Margem de Erro
Registros	20.397	20.722	20.490	20.544	20.536	102.689	Referentes aos Percentuais Anuais			
Sexo										
Masculino	9.399 46,08%	9.472 45,71%	9.324 45,51%	9.488 46,18%	9.211 44,85%	46.894 45,666%	0,457	0,005	0,002	0,005
Feminino	10.998 53,92%	11.250 54,29%	11.166 54,49%	11.056 53,82%	11.325 55,15%	55.795 54,334%	0,543	0,005	0,002	0,005
Idade (Anos)										
Média	42,694	42,692	43,029	43,410	44,177					
Desvio Padrão	19,646	19,454	19,628	19,767	20,021					
Mínima	10	10	10	10	10					
25%	26	26	26	26	26					
50%	41	41	42	42	43,5					
75%	59	58	59	59	60					
Máxima	99	99	100	100	100					
Faixa Etária										
De 10 a 15 anos	1.371 6,72%	1.334 6,44%	1.408 6,87%	1.360 6,62%	1.305 6,35%	6.778 6,601%	0,066	0,002	0,001	0,002
De 16 a 24 anos	3.040 14,90%	3.102 14,97%	3.013 14,70%	3.073 14,96%	3.051 14,86%	15.279 14,879%	0,149	0,001	0,000	0,001
De 25 a 34 anos	3.915 19,19%	3.880 18,72%	3.678 17,95%	3.533 17,20%	3.380 16,46%	18.386 17,905%	0,179	0,011	0,005	0,010
De 35 a 44 anos	2.752 13,49%	3.029 14,62%	2.896 14,13%	2.895 14,09%	2.746 13,37%	14.318 13,943%	0,139	0,005	0,002	0,004
De 45 a 59 anos	4.488 22,00%	4.566 22,03%	4.537 22,14%	4.550 22,15%	4.583 22,32%	22.724 22,129%	0,221	0,001	0,001	0,001
60 anos ou mais	4.831 23,68%	4.811 23,22%	4.958 24,20%	5.133 24,99%	5.471 26,64%	25.204 24,544%	0,245	0,013	0,006	0,012
Grau de instrução (completa)										
Analfabeto	2.137 10,48%	2.348 11,33%	2.358 11,51%	2.626 12,78%	2.563 12,57%	12.032 11,717%	0,117	0,009	0,004	0,008
Sabe ler/escrever, mas não cursou escola	264 1,29%	230 1,11%	186 0,91%	177 0,86%	143 0,70%	1.000 0,974%	0,010	0,002	0,001	0,002
Até pré-escola incompleto	61 0,30%	85 0,41%	42 0,20%	36 0,18%	39 0,19%	263 0,256%	0,003	0,001	0,000	0,001
Pré-escola completo	26 0,13%	28 0,14%	25 0,12%	19 0,09%	11 0,05%	109 0,106%	0,001	0,000	0,000	0,000
1a até 3a série/1a até 3a série	1.570 7,70%	1.424 6,87%	1.382 6,74%	1.340 6,52%	1.278 6,27%	6.994 6,811%	0,068	0,005	0,002	0,005
4a série/4a até 5a série	2.691 13,19%	2.659 12,83%	2.480 12,10%	2.502 12,18%	2.146 10,52%	12.478 12,151%	0,122	0,010	0,005	0,009
5a até 7a série/1a até 3a série	2.578 12,64%	2.545 12,28%	2.595 12,66%	2.441 11,88%	2.810 13,78%	12.969 12,629%	0,126	0,007	0,003	0,006
8a série/4a série	2.075 10,17%	2.011 9,70%	2.073 10,12%	2.063 10,04%	2.020 9,90%	10.242 9,974%	0,100	0,002	0,001	0,002
1a e 2a série/1a e 2a série	2.139 10,49%	2.361 11,39%	2.221 10,84%	2.005 9,76%	1.783 8,74%	10.509 10,234%	0,102	0,010	0,005	0,009
3a série/vestibular	4.438 21,76%	4.594 22,17%	4.637 22,63%	4.956 24,12%	5.205 25,52%	23.830 23,206%	0,232	0,016	0,007	0,014
Superior incompleto	941 4,61%	932 4,50%	861 4,20%	829 4,04%	845 4,14%	4.408 4,293%	0,043	0,002	0,001	0,002
Superior completo	1.477 7,24%	1.505 7,26%	1.630 7,96%	1.550 7,54%	1.693 8,30%	7.855 7,649%	0,077	0,005	0,002	0,004
Grau de instrução (reduzido)										
Analfabeto / Educação infantil	2.488 12,20%	2.691 12,99%	2.611 12,74%	2.858 13,91%	2.756 13,42%	13.404 13,053%	0,131	0,007	0,003	0,006
Fundamental	8.914 43,70%	8.639 41,69%	8.530 41,63%	8.346 40,63%	8.254 40,19%	42.683 41,565%	0,416	0,014	0,006	0,012
Médio	6.577 32,24%	6.955 33,56%	6.858 33,47%	6.961 33,88%	6.988 34,03%	34.339 33,440%	0,334	0,007	0,003	0,006
Superior	2.418 11,85%	2.437 11,76%	2.491 12,16%	2.379 11,58%	2.538 12,36%	12.263 11,942%	0,119	0,003	0,001	0,003

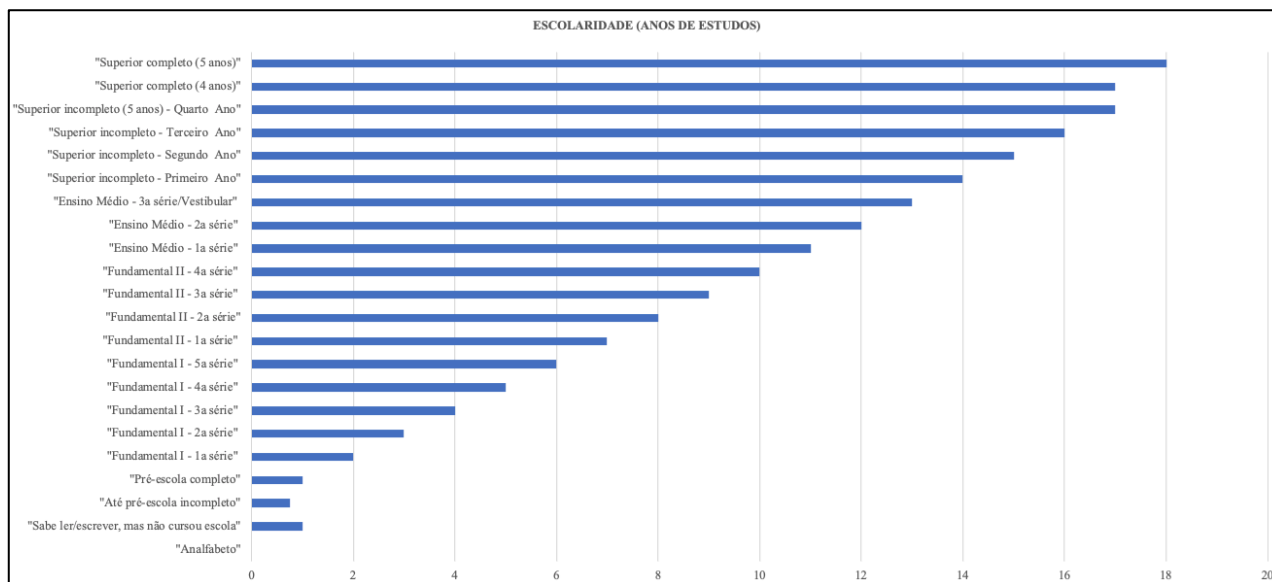
Classe Econômica										
Classe A	363 1,78%	326 1,57%	292 1,43%	259 1,26%	206 1,00%	1.446 1,408%	0,014	0,003	0,001	0,003
Classe B	3.469 17,01%	3.247 15,67%	3.322 16,21%	3.068 14,93%	2.447 11,92%	15.553 15,146%	0,151	0,020	0,009	0,017
Classe C	8.743 42,86%	8.880 42,85%	8.814 43,02%	8.865 43,15%	9.347 45,52%	44.649 43,480%	0,435	0,011	0,005	0,010
Classes DE	7.822 38,35%	8.269 39,90%	8.062 39,35%	8.352 40,65%	8.536 41,57%	41.041 39,966%	0,400	0,012	0,005	0,011
Renda Pessoal										
Até R\$ 788,00	6.875 33,71%	7.761 37,45%	8.762 42,76%	8.917 43,40%	8.974 43,70%	41.289 40,208%	0,402	0,044	0,020	0,039
De R\$ 788,01 até R\$ 1.576,00	5.661 27,75%	4.899 23,64%	4.250 20,74%	4.162 20,26%	4.194 20,42%	23.166 22,559%	0,226	0,032	0,014	0,028
De R\$ 1.576,01 até R\$ 2.364,00	1.748 8,57%	1.674 8,08%	1.441 7,03%	1.436 6,99%	1.410 6,87%	7.709 7,507%	0,075	0,008	0,003	0,007
De R\$ 2.364,01 até R\$ 3.940,00	875 4,29%	878 4,24%	822 4,01%	805 3,92%	830 4,04%	4.210 4,100%	0,041	0,002	0,001	0,001
De R\$ 3.940,01 até R\$ 7.880,00	444 2,18%	395 1,91%	381 1,86%	357 1,74%	368 1,79%	1.945 1,894%	0,019	0,002	0,001	0,001
De R\$ 7.880,01 até R\$ 15.760,00	129 0,63%	106 0,51%	91 0,44%	81 0,39%	89 0,43%	496 0,483%	0,005	0,001	0,000	0,001
De R\$ 15.760,01 até R\$ 23.640,00	24 0,12%	17 0,08%	18 0,09%	16 0,08%	17 0,08%	92 0,090%	0,001	0,000	0,000	0,000
Mais de R\$ 23.640,00	4 0,02%	8 0,04%	6 0,03%	15 0,07%	8 0,04%	41 0,040%	0,000	0,000	0,000	0,000
Não tem renda	3.744 18,36%	3.829 18,48%	3.743 18,27%	3.729 18,15%	3.638 17,72%	18.683 18,194%	0,182	0,003	0,001	0,003
Não sabe	235 1,15%	318 1,53%	321 1,57%	305 1,48%	275 1,34%	1.454 1,416%	0,014	0,002	0,001	0,001
Não respondeu	658 3,23%	837 4,04%	655 3,20%	721 3,51%	733 3,57%	3.604 3,510%	0,035	0,003	0,002	0,003
Renda Familiar										
Até R\$ 937,00	4.946 24,25%	5.782 27,90%	6.397 31,22%	6.853 33,36%	6.626 32,27%	30.604 29,803%	0,298	0,037	0,017	0,033
De R\$ 937,01 até R\$ 1.874,00	6.507 31,90%	6.061 29,25%	5.779 28,20%	5.545 26,99%	5.507 26,82%	29.399 28,629%	0,286	0,021	0,009	0,018
De R\$ 1.874,01 até R\$ 2.811,00	3.570 17,50%	3.131 15,11%	2.900 14,15%	2.811 13,68%	2.890 14,07%	15.302 14,901%	0,149	0,015	0,007	0,014
De R\$ 2.811,01 até R\$ 4.685,00	2.204 10,81%	2.096 10,11%	2.051 10,01%	1.948 9,48%	1.992 9,70%	10.291 10,022%	0,100	0,005	0,002	0,004
De R\$ 4.685,01 até R\$ 9.370,00	1.120 5,49%	1.004 4,85%	963 4,70%	951 4,63%	977 4,76%	5.015 4,884%	0,049	0,003	0,002	0,003
De R\$ 9.370,01 até R\$ 18.740,00	349 1,71%	285 1,38%	271 1,32%	246 1,20%	272 1,32%	1.423 1,386%	0,014	0,002	0,001	0,002
De R\$ 18.740,01 até R\$ 28.110,00	81 0,40%	61 0,29%	40 0,20%	42 0,20%	43 0,21%	267 0,260%	0,003	0,001	0,000	0,001
Mais de R\$ 28.110,00	17 0,08%	27 0,13%	19 0,09%	26 0,13%	28 0,14%	117 0,114%	0,001	0,000	0,000	0,000
Não tem renda	0 0,00%	444 2,14%	389 1,90%	374 1,82%	450 2,19%	1.657 1,614%	0,016	0,009	0,004	0,008
Não sabe	839 4,11%	925 4,46%	962 4,69%	909 4,42%	934 4,55%	4.569 4,449%	0,044	0,002	0,001	0,002
Não respondeu	764 3,75%	906 4,37%	719 3,51%	839 4,08%	817 3,98%	4.045 3,939%	0,039	0,003	0,001	0,003
Renda Familiar (SM)										
Até 1 Salário Mínimo (SM)	4.946 24,25%	5.782 27,90%	6.397 31,22%	6.853 33,36%	6.626 32,27%	30.604 29,803%	0,298	0,037	0,017	0,033
Mais de 1 SM até 2 SM	6.507 31,90%	6.061 29,25%	5.779 28,20%	5.545 26,99%	5.507 26,82%	29.399 28,629%	0,286	0,021	0,009	0,018
Mais de 2 SM até 3 SM	3.570 17,50%	3.131 15,11%	2.900 14,15%	2.811 13,68%	2.890 14,07%	15.302 14,901%	0,149	0,015	0,007	0,014
Mais de 3 SM até 5 SM	2.204 10,81%	2.096 10,11%	2.051 10,01%	1.948 9,48%	1.992 9,70%	10.291 10,022%	0,100	0,005	0,002	0,004
Mais de 5 SM até 10 SM	1.120 5,49%	1.004 4,85%	963 4,70%	951 4,63%	977 4,76%	5.015 4,884%	0,049	0,003	0,002	0,003
Mais de 10 SM	447 2,19%	373 1,80%	330 1,61%	314 1,53%	343 1,67%	1.807 1,760%	0,018	0,003	0,001	0,002
Não tem renda	0 0,00%	444 2,14%	389 1,90%	374 1,82%	450 2,19%	1.657 1,614%	0,016	0,009	0,004	0,008
Não sabe	839 4,11%	925 4,46%	962 4,69%	909 4,42%	934 4,55%	4.569 4,449%	0,044	0,002	0,001	0,002

Não respondeu	764 3,75%	906 4,37%	719 3,51%	839 4,08%	817 3,98%	4.045 3,939%	0,039	0,003	0,001	0,003
População Economicamente Ativa										
Sim	13.856 67,93%	13.945 67,30%	13.871 67,70%	13.853 67,43%	14.102 68,67%	69.627 67,804%	0,678	0,005	0,002	0,005
Não	6.541 32,07%	6.777 32,70%	6.619 32,30%	6.691 32,57%	6.434 31,33%	33.062 32,196%	0,322	0,005	0,002	0,005
PEA Detalhado										
Trabalha em atividade remunerada (PEA)	10.826 53,08%	10.762 51,94%	10.335 50,44%	10.115 49,24%	10.151 49,43%	52.189 50,822%	0,508	0,017	0,007	0,014
Trabalha em atividade não remunerada, como ajudante (PEA)	1.449 7,10%	1.584 7,64%	1.880 9,18%	1.921 9,35%	2.216 10,79%	9.050 8,813%	0,088	0,015	0,007	0,013
Trabalha mas está afastado (PEA)	479 2,35%	442 2,13%	463 2,26%	623 3,03%	588 2,86%	2.595 2,527%	0,025	0,004	0,002	0,003
Tomou providência para conseguir trabalho nos últimos 30 dias (PEA)	1.102 5,40%	1.157 5,58%	1.193 5,82%	1.194 5,81%	1.147 5,59%	5.793 5,641%	0,056	0,002	0,001	0,002
Não trabalha e não procurou trabalho nos últimos 30 dias (NÃO PEA)	6.541 32,07%	6.777 32,70%	6.619 32,30%	6.691 32,57%	6.434 31,33%	33.062 32,196%	0,322	0,005	0,002	0,005
Religião										
Católica Apostólica Romana/ Ortodoxa	12.533 61,45%	12.289 59,30%	11.931 58,23%	11.896 57,90%	11.722 57,08%	60.371 58,790%	0,588	0,017	0,008	0,015
Igreja Evangélica Luterana/ Presbiteriana/Metodista/Batista	1.767 8,66%	648 3,13%	567 2,77%	583 2,84%	536 2,61%	4.101 3,994%	0,040	0,026	0,012	0,023
Outras Igrejas Evangélicas	3.263 16,00%	4.526 21,84%	4.855 23,69%	4.794 23,34%	4.782 23,29%	22.220 21,638%	0,216	0,032	0,014	0,028
Outras religiosidades cristãs	310 1,52%	420 2,03%	486 2,37%	513 2,50%	511 2,49%	2.240 2,181%	0,022	0,004	0,002	0,004
Islamismo / Muçulmano	4 0,02%	6 0,03%	7 0,03%	8 0,04%	2 0,01%	27 0,026%	0,000	0,000	0,000	0,000
Espírita	475 2,33%	429 2,07%	404 1,97%	372 1,81%	406 1,98%	2.086 2,031%	0,020	0,002	0,001	0,002
Umbanda	121 0,59%	98 0,47%	120 0,59%	100 0,49%	131 0,64%	570 0,555%	0,006	0,001	0,000	0,001
Candomblé	46 0,23%	47 0,23%	51 0,25%	35 0,17%	59 0,29%	238 0,232%	0,002	0,000	0,000	0,000
Judaísmo	6 0,03%	10 0,05%	5 0,02%	3 0,01%	2 0,01%	26 0,025%	0,000	0,000	0,000	0,000
Hinduismo	0 0,00%	3 0,01%	0 0,00%	0 0,00%	2 0,01%	5 0,005%	0,000	0,000	0,000	0,000
Budismo	29 0,14%	18 0,09%	28 0,14%	28 0,14%	21 0,10%	124 0,121%	0,001	0,000	0,000	0,000
Sem religião	949 4,65%	1.103 5,32%	1.048 5,11%	1.046 5,09%	1.304 6,35%	5.450 5,307%	0,053	0,006	0,003	0,006
Agnóstico	210 1,03%	219 1,06%	182 0,89%	211 1,03%	163 0,79%	985 0,959%	0,010	0,001	0,001	0,001
Ateu	245 1,20%	357 1,72%	256 1,25%	365 1,78%	247 1,20%	1.470 1,432%	0,014	0,003	0,001	0,003
Outras religiões	200 0,98%	252 1,22%	236 1,15%	244 1,19%	286 1,39%	1.218 1,186%	0,012	0,001	0,001	0,001
Não respondeu	239 1,17%	297 1,43%	314 1,53%	346 1,68%	362 1,76%	1.558 1,517%	0,015	0,002	0,001	0,002
Etnia										
Branca	7.758 38,04%	7.597 36,66%	7.205 35,16%	7.173 34,92%	7.199 35,06%	36.932 35,965%	0,360	0,014	0,006	0,012
Preta	2.957 14,50%	3.217 15,52%	3.111 15,18%	3.023 14,71%	3.086 15,03%	15.394 14,991%	0,150	0,004	0,002	0,004
Amarela	8.326	8.312	8.517	8.724	8.656	42.535	0,414	0,010	0,004	0,008

	40,82%	40,11%	41,57%	42,46%	42,15%	41,421%				
Parda	541 2,65%	554 2,67%	560 2,73%	543 2,64%	529 2,58%	2.727 2,656%	0,027	0,001	0,000	0,000
Indígena	344 1,69%	442 2,13%	424 2,07%	392 1,91%	358 1,74%	1.960 1,909%	0,019	0,002	0,001	0,002
Não respondeu	471 2,31%	600 2,90%	673 3,28%	689 3,35%	708 3,45%	3.141 3,059%	0,031	0,005	0,002	0,004
Região										
Sudeste	7.509 36,81%	7.700 37,16%	7.436 36,29%	7.313 35,60%	7.601 37,01%	37.559 36,575%	0,366	0,006	0,003	0,006
Nordeste	5.789 28,38%	5.977 28,84%	5.892 28,76%	6.074 29,57%	5.977 29,10%	29.709 28,931%	0,289	0,004	0,002	0,004
Sul	3.253 15,95%	3.356 16,20%	3.342 16,31%	3.365 16,38%	3.222 15,69%	16.538 16,105%	0,161	0,003	0,001	0,002
Norte	2.036 9,98%	2.020 9,75%	2.122 10,36%	2.071 10,08%	1.962 9,55%	10.211 9,944%	0,099	0,003	0,001	0,003
Centro-Oeste	1.810 8,87%	1.669 8,05%	1.698 8,29%	1.721 8,38%	1.774 8,64%	8.672 8,445%	0,084	0,003	0,001	0,003
Área										
Urbana	18.608 91,23%	18.424 88,91%	18.470 90,14%	18.365 89,39%	18.369 89,45%	92.236 89,821%	0,898	0,009	0,004	0,008
Rural	1.789 8,77%	2.298 11,09%	2.020 9,86%	2.179 10,61%	2.167 10,55%	10.453 10,179%	0,102	0,009	0,004	0,008
Acesso Internet										
Há menos de 3 meses	10.320 50,60%	11.050 53,32%	12.117 59,14%	12.708 61,86%	13.332 64,92%	59.527 57,968%	0,580	0,059	0,027	0,052
Entre 3 meses e 12 meses	929 4,55%	919 4,43%	797 3,89%	684 3,33%	631 3,07%	3.960 3,856%	0,039	0,007	0,003	0,006
Mais de 12 meses atrás	789 3,87%	846 4,08%	719 3,51%	688 3,35%	738 3,59%	3.780 3,681%	0,037	0,003	0,001	0,003
Não se aplica	8.359 40,98%	7.907 38,16%	6.857 33,47%	6.464 31,46%	5.835 28,41%	35.422 34,494%	0,345	0,051	0,023	0,044

Fonte: Elaborado pelo autor com base em dados de pesquisa TIC Domicílios 2015-2019

Figura B-1 – Informações a respeito de grau de escolaridade e anos de estudos



Fonte: Elaborado pelo autor com base em Lei nº 11.274, de 6 de fevereiro de 2006

Tabela B-2 – Formatação final de variáveis independentes para regressão logística

Variável	Descrição	ID	Códigos Originais	Reestruturação	2015	2016	2017	2018	2019
SEXO (REFERÊNCIA: FEMININO)	Masculino	SEXO_1	1	Dummy					
	Feminino	SEXO_2	2	Referência					
IDADE	Valor em anos	-	Numérico	Média	42,694	42,692	43,029	43,410	44,177
				Desvio Padrão	19,646	19,454	19,628	19,767	20,021
				Mínima	10	10	10	10	10

			0,25		26	26	26	26	26
			0,5		41	41	42	42	43,5
			0,75		59	58	59	59	60
			Máxima		99	99	100	100	100
GRAU DE INSTRUÇÃO (DETALHADO)	"Analfabeto"	GRAU_INST RUÇÃO	1="Analfabeto"	0					
	"Sabe ler/escrever, mas não cursou escola"	GRAU_INST RUÇÃO	2="Sabe ler/escrever, mas não cursou escola"	1 ano					
	"Até pré-escola incompleto"	GRAU_INST RUÇÃO	3="Até pré-escola incompleto"	0,75 ano					
	"Pré-escola completo"	GRAU_INST RUÇÃO	4="Pré-escola completo"	1 ano					
	"1a até 3a série/1a até 3a série"	GRAU_INST RUÇÃO	5="1a até 3a série/1a até 3a série"	3 anos					
	"4a série/4a até 5a série"	GRAU_INST RUÇÃO	6="4a série/4a até 5a série"	5,5 anos					
	"5a até 7a série/1a até 3a série"	GRAU_INST RUÇÃO	7="5a até 7a série/1a até 3a série"	8 anos					
	"8a série/4a série"	GRAU_INST RUÇÃO	8="8a série/4a série"	10 anos					
	"1a e 2a série/1a e 2a série"	GRAU_INST RUÇÃO	9="1a e 2a série/1a e 2a série"	11,5 anos					
	"3a série/vestibular"	GRAU_INST RUÇÃO	10="3a série/ vestibular"	13 anos					
	"Superior incompleto"	GRAU_INST RUÇÃO	11="Superior incompleto"	15,5 anos					
	"Superior completo"	GRAU_INST RUÇÃO	12="Superior completo"	17,5 anos					
CLASSE ECONÔMICA (REFERÊNCIA: CLASSES DE)	Classe 'A'	CLASSE_E CONÔMIC A_1	1="A"	Dummy					
	Classe 'B'	CLASSE_E CONÔMIC A_2	2="B"	Dummy					
	Classe 'C'	CLASSE_E CONÔMIC A_3	3="C"	Dummy					
	Classes 'DE'	CLASSE_E CONÔMIC A_4	4="DE"	Referência					
RENDA FAMILIAR	SEM RENDA	-	9 = "Não tem renda"	0	0,000	0,000	0,000	0,000	0,000
	1 SALÁRIO MÍNIMO	-	1 = "primeira faixa"	média dos valores da faixa	457,524	487,730	506,573	498,585	499,000
	1-2 SALÁRIOS MÍNIMOS	-	2 = "segunda faixa"	média dos valores da faixa	1.372,577	1.463,196	1.519,723	1.495,759	1.497,005
	2-3 SALÁRIOS MÍNIMOS	-	3 = "terceira faixa"	média dos valores da faixa	2.287,624	2.438,656	2.532,868	2.492,928	2.495,005
	3-5 SALÁRIOS MÍNIMOS	-	4 = "quarta faixa"	média dos valores da faixa	3.660,195	3.901,846	4.052,586	3.988,682	3.992,005
	5-10 SALÁRIOS MÍNIMOS	-	5 = "quinta faixa"	média dos valores da faixa	6.862,861	7.360,296	7.598,594	7.478,775	7.485,005
	10-20 SALÁRIOS MÍNIMOS	-	6 = "sexta faixa"	média dos valores da faixa	13.725,716	14.676,248	15.197,182	14.957,544	14.970,005
	20-30 SALÁRIOS MÍNIMOS	-	7 = "sétima faixa"	média dos valores da faixa	22.876,190	24.386,511	25.328,633	24.929,237	24.950,005
	Média de renda > 30 SM	-	8 = "oitava faixa"	média dos valores da faixa	41.408,018	44.141,839	45.847,169	45.124,225	45.161,818
	Sim	PEA_2_1	1 = "PEA"	Dummy					

POPULAÇÃO ECONOMICAMENTE ATIVA (REFERÊNCIA: NÃO)	Não	PEA_2_2	2 = "Não PEA"	Referência
RELIGIÃO (REFERÊNCIA: CATÓLICA)	Católica Apostólica Romana/ Ortodoxa	CATÓLICA	1 = "Católica" 2 = "Luterana/Presb iteriana/Metodis ta/Batista" 3 = "Outras Igrejas Evangélicas" 4 = "Outras religiosidades cristãs" 5 = "Islamismo / Muçulmano" 6 = "Espírita" 7 = "Umbanda" 8 = "Candomblé" 9 = "Judaísmo" 10 = "Hinduísmo" 11 = "Budismo" 12 = "Sem religião" 13 = "Agnóstico" 14 = "Ateu " 15 = "Outras religiões"	Referência Dummy
ETNIA (REFERÊNCIA: BRANCA)	Branca Preta Amarela Parda Indígena	RACA_1 RACA_2 RACA_3 RACA_4 RACA_5	1="Branca" 2="Preta" 3="Amarela" 4="Parda" 5="Indígena"	Referência Dummy Dummy Dummy Dummy
REGIÃO (REFERÊNCIA: SUL)	Sudeste Nordeste Sul Norte Centro-Oeste	REGIAO_1 REGIAO_2 REGIAO_3 REGIAO_4 REGIAO_5	1 = "Sudeste" 2 = "Nordeste" 3 = "Sul" 4 = "Norte" 5 = "Centro- Oeste"	Dummy Dummy Referência Dummy Dummy
ÁREA (REFERÊNCIA: RURAL)	Urbana Rural	AREA_1 AREA_2	1="Urbana" 2="Rural"	Dummy Referência

Fonte: Elaborado pelo autor com base em dados de pesquisa TIC Domicílios 2015-2019

Tabela B-3 – Informações modificadas de indicadores demográficos, socioeconômicos e culturais da Pesquisa TIC Domicílios para este estudo (2015-2019)

Variáveis	2015	2016	2017	2018	2019	Total	Média	Desvio Padrão	Erro Padrão	Margem de Erro
Registros	18.083	17.936	17.803	17.766	17.811	89.399	Referentes aos Percentuais Anuais			
Sexo										
Masculino	8367 46,27%	8.250 46,00%	8.112 45,57%	8.223 46,29%	8.038 45,13%	40.990 45,851%	0,458	0,005	0,002	0,004
Feminino	9.716 53,73%	9.686 54,00%	9.691 54,43%	9.543 53,71%	9.773 54,87%	48.409 54,149%	0,542	0,005	0,002	0,004
Idade (Anos)										
Média	42,486	42,401	42,628	42,865	43,506					
Desvio Padrão	19,384	19,105	19,320	19,448	19,628					
Mínima	10	10	10	10	10					
25%	26	27	26	26	26					
50%	40	40	41	41	42					

75%	58	58	58	59	59					
Máxima	99	99	100	100	100					
Grau de instrução (reduzido)										
Analfabeto / Educação infantil	2.063 11,41%	2.161 12,05%	2.072 11,64%	2.292 12,90%	2.128 11,95%	10.716 11,987%	0,120	0,006	0,003	0,005
Fundamental	7.945 43,94%	7.523 41,94%	7.427 41,72%	7.208 40,57%	7.150 40,14%	37.253 41,670%	0,417	0,015	0,007	0,013
Médio	5.934 32,82%	6.106 34,04%	6.128 34,42%	6.162 34,68%	6.275 35,23%	30.605 34,234%	0,342	0,009	0,004	0,008
Superior	2.141 11,84%	2.146 11,96%	2.176 12,22	2.104 11,84	2.258 12,68%	10.825 12,109%	0,121	0,004	0,002	0,003
Classe Econômica										
Classe A	293 1,62%	269 1,50%	232 1,30%	210 1,18%	177 0,99%	1.181 1,321%	0,013	0,002	0,001	0,002
Classe B	2.996 16,57%	2.737 15,26%	2.827 15,88%	2.613 14,71%	2.113 11,86%	13.286 14,861%	0,149	0,018	0,008	0,016
Classe C	7.819 43,24%	7.811 43,55%	7.780 43,70%	7.698 43,33%	8.189 45,98%	39.297 43,957%	0,440	0,011	0,005	0,010
Classes DE	6.975 38,57%	7.119 39,69%	6.964 39,12%	7.245 40,78%	7.332 41,17%	35.635 39,861%	0,399	0,011	0,005	0,010
Renda Familiar										
Média	2,283	2,239	2,209	2,138	2,185					
Desvio Padrão	2,894	3,079	2,928	3,020	3,077					
Mínima	0,458	0,000	0,000	0,000	0,000					
25%	0,458	0,488	0,507	0,499	0,499					
50%	1,373	1,463	1,520	1,496	1,497					
75%	2,288	2,439	2,533	2,493	2,495					
Máxima	41,408	44,142	45,847	45,124	45,162					
População Economicamente Ativa										
Sim	12.519 69,23%	12.378 69,01%	12.289 69,03%	12.231 68,84%	12.476 70,05%	61.893 69,232%	0,692	0,005	0,002	0,004
Não	5.564 30,77%	5.558 30,99%	5.514 30,97%	5.535 31,16%	5.335 29,95%	27.506 30,768%	0,308	0,005	0,002	0,004
Religião										
Católica Apostólica Romana/ Ortodoxa	11.289 62,43%	10.816 60,30%	10.816 59,20%	10.816 58,93%	10.306 57,86%	53.419 59,753%	0,597	0,017	0,008	0,015
Não-Católica	6.794 37,57%	7.120 39,70%	7.120 40,80%	7.120 41,07%	7.505 42,14%	35.980 40,247%	0,403	0,017	0,008	0,015
Etnia										
Branca	6.936 38,36%	6.700 37,36%	6.310 35,44%	6.314 35,54%	6.367 35,75%	32.627 36,496%	0,365	0,013	0,006	0,011
Preta	2.707 14,97%	2.877 16,04%	2.832 15,91%	2.738 15,41%	2.756 15,47%	13.910 15,559%	0,156	0,004	0,002	0,004
Amarela	7.628 42,18%	7.482 41,71%	7.754 43,55%	7.879 44,35%	7.907 44,39%	38.650 43,233%	0,432	0,012	0,006	0,011
Parda	500 2,77%	475 2,65%	517 2,90%	475 2,67%	461 2,59%	2.428 2,716%	0,027	0,001	0,001	0,001
Indígena	312 1,73%	402 2,24%	390 2,19%	360 2,03%	320 1,80%	1.784 1,996%	0,020	0,002	0,001	0,002
Região										
Sudeste	6.665 36,86%	6.561 36,58%	6.348 35,66%	6.199 34,89%	6.486 36,42%	32.259 36,084%	0,361	0,008	0,004	0,007
Nordeste	5.149 28,47%	5.313 29,62%	5.205 29,24%	5.324 29,97%	5.162 28,98%	26.153 29,254%	0,293	0,006	0,003	0,005
Sul	2.852 15,77%	3.007 16,77%	2.907 16,33%	2.922 16,45%	2.783 15,63%	14.471 16,187%	0,162	0,005	0,002	0,004
Norte	1.816 10,04%	1.653 9,22%	1.897 10,66%	1.830 10,30%	1.782 10,01%	8.978 10,043%	0,100	0,005	0,002	0,005
Centro-Oeste	1.601 8,85%	1.402 7,82%	1.446 8,12%	1.491 8,39%	1.598 8,97%	7.538 8,432%	0,084	0,005	0,002	0,004
Área										
Urbana	16.496 91,22%	15.952 88,94%	16.034 90,06%	15.859 89,27%	15.933 89,46%	80.274 89,793%	0,898	0,009	0,004	0,008
Rural	1.587 8,78%	1.984 11,06%	1.769 9,94%	1.907 10,73%	1.878 10,54%	9.125 10,207%	0,102	0,009	0,004	0,008
Acesso Internet										
Há menos de 3 meses	9.233 51,06%	9.681 53,98%	10.705 60,13%	11.207 63,08%	11.910 66,87%	52.736 58,989%	0,590	0,065	0,029	0,057

Fonte: Elaborado pelo autor com base em dados de pesquisa TIC Domicílios 2015-2019

Tabela B-4 – Lacunas digitais identificadas nos microdados da TIC Domicílios (2015-2019)

Variáveis	2015	2016	2017	2018	2019	Variação
Sexo						
Masculino	52,04%	54,48%	60,37%	62,74%	65,82%	13,79 pp
Feminino	50,22%	53,54%	59,93%	63,38%	67,73%	17,51 pp
Lacuna (Gap)	1,82 pp	0,94 pp	0,44 pp	0,64 pp	1,90 pp	-0,08 pp
Faixa Etária						
De 10 a 15 anos	78,46%	79,96%	85,75%	84,96%	86,52%	8,07 pp
De 16 a 24 anos	83,93%	84,50%	87,44%	88,48%	91,81%	7,88 pp
De 25 a 34 anos	75,03%	77,53%	83,50%	85,43%	87,99%	12,95 pp
De 35 a 44 anos	60,07%	62,80%	71,63%	75,51%	80,55%	20,48 pp
De 45 a 59 anos	35,20%	40,70%	50,52%	55,77%	61,41%	26,21 pp
60 anos ou mais	11,37%	12,99%	18,41%	22,97%	29,60%	18,22 pp
Lacuna (Gap)	72,55 pp	71,51 pp	69,03 pp	65,51 pp	62,21 pp	-10,34 pp
Grau de instrução (completa)						
Analfabeto	3,13%	4,28%	8,86%	12,22%	14,74%	11,61 pp
Sabe ler/escrever, mas não cursou escola	3,04%	6,08%	8,33%	15,38%	23,15%	20,10 pp
Até pré-escola incompleto	9,09%	9,68%	12,50%	6,67%	34,29%	25,19 pp
Pré-escola completo	21,05%	16,00%	39,13%	12,50%	27,27%	6,22 pp
1a até 3a série/1a até 3a série	8,88%	12,32%	16,77%	21,72%	29,49%	20,62 pp
4a série/4a até 5a série	18,02%	23,90%	28,97%	38,64%	41,30%	23,28 pp
5a até 7a série/1a até 3a série	46,17%	50,20%	57,60%	62,73%	64,05%	17,88 pp
8a série/4a série	52,74%	57,93%	66,01%	70,30%	72,74%	20,01 pp
1a e 2a série/1a e 2a série	74,08%	75,44%	81,62%	82,16%	84,50%	10,42 pp
3a série/vestibular	78,64%	78,78%	84,47%	86,23%	88,13%	9,49 pp
Superior incompleto	94,63%	93,70%	96,64%	96,94%	97,21%	2,58 pp
Superior completo	91,43%	92,35%	93,37%	94,23%	95,22%	3,79 pp
Lacuna (Gap)	91,59 pp	89,43 pp	88,31 pp	90,27 pp	82,47 pp	-9,12 pp
Grau de instrução (reduzido)						
Analfabeto / Educação infantil	3,44%	4,72%	9,22%	12,35%	15,55%	12,11 pp
Fundamental	32,70%	37,80%	44,93%	51,04%	55,12%	22,42 pp
Médio	77,17%	77,66%	83,57%	85,07%	87,22%	10,05 pp
Superior	92,71%	92,87%	94,49%	95,20%	95,88%	3,17 pp
Lacuna (Gap)	89,27 pp	88,15 pp	85,27 pp	82,85 pp	80,33 pp	-8,95 pp
Classe Econômica						
Classe A	95,22%	91,45%	94,83%	96,67%	96,05%	0,82 pp
Classe B	84,28%	87,18%	90,13%	91,04%	91,81%	7,53 pp
Classe C	59,60%	64,09%	69,34%	73,79%	76,09%	16,49 pp
Classes DE	25,36%	28,70%	36,50%	40,65%	48,68%	23,32 pp
Lacuna (Gap)	69,86 pp	62,75 pp	58,33 pp	56,02 pp	47,37 pp	-22,49 pp
Renda Pessoal						
Sem renda	64,23%	66,48%	71,55%	72,30%	75,22%	10,99 pp
1 salário mínimo	34,26%	36,62%	45,06%	48,72%	54,50%	20,24 pp
1-2 salários mínimos	50,20%	57,77%	66,50%	71,69%	73,67%	23,47 pp
2-3 salários mínimos	67,48%	71,69%	77,98%	80,34%	82,82%	15,34 pp
3-5 salários mínimos	75,83%	77,03%	84,24%	87,01%	85,37%	9,54 pp
5-10 salários mínimos	86,28%	88,65%	92,29%	89,71%	92,37%	6,09 pp
10-20 salários mínimos	84,25%	82,35%	88,76%	92,00%	89,66%	5,40 pp
20-30 salários mínimos	78,26%	82,35%	87,50%	73,33%	82,35%	4,09 pp
> 30 salários mínimos	75,00%	71,43%	40,00%	75,00%	83,33%	8,33 pp
Lacuna (Gap)	52,02 pp	52,03 pp	52,29 pp	43,28 pp	37,87 pp	-14,15 pp
Renda Familiar						
Sem renda	-	49,88%	51,17%	53,13%	53,32%	3,44 pp
1 salário mínimo	32,92%	35,64%	43,98%	47,22%	52,71%	19,79 pp
1-2 salários mínimos	44,46%	50,39%	58,15%	62,23%	66,16%	21,70 pp
2-3 salários mínimos	60,44%	64,74%	70,92%	75,69%	78,25%	17,81 pp
3-5 salários mínimos	70,87%	76,58%	81,80%	84,81%	84,73%	13,86 pp
5-10 salários mínimos	82,06%	84,84%	88,95%	89,31%	90,14%	8,08 pp
10-20 salários mínimos	88,72%	87,55%	91,25%	93,31%	91,73%	3,01 pp
20-30 salários mínimos	84,81%	86,44%	86,84%	90,00%	92,86%	8,05 pp
> 30 salários mínimos	76,47%	84,62%	82,35%	84,62%	80,77%	4,30 pp
Lacuna (Gap)	55,80 pp	51,90 pp	47,28 pp	46,08 pp	40,14 pp	-15,66 pp
Renda Familiar (SM)						
Não tem renda	-	49,88%	51,17%	53,13%	53,32%	3,44 pp
Até 1 Salário Mínimo (SM)	32,92%	35,64%	43,98%	47,22%	52,71%	19,79 pp
Mais de 1 SM até 2 SM	44,46%	50,39%	58,15%	62,23%	66,16%	21,70 pp

Mais de 2 SM até 3 SM	60,44%	64,74%	70,92%	75,69%	78,25%	17,81 pp
Mais de 3 SM até 5 SM	70,87%	76,58%	81,80%	84,81%	84,73%	13,86 pp
Mais de 5 SM até 10 SM	82,06%	84,84%	88,95%	89,31%	90,14%	8,08 pp
Mais de 10 SM	87,53%	87,15%	90,25%	92,13%	91,02%	3,49 pp
Lacuna (Gap)	54,61 pp	51,51 pp	46,28 pp	44,91 pp	38,31 pp	-16,30 pp
População Economicamente Ativa						
Sim	57,61%	60,18%	66,21%	69,64%	72,50%	14,89 pp
Não	36,32%	40,16%	46,59%	48,58%	53,70%	17,38 pp
Lacuna (Gap)	21,29 pp	20,02 pp	19,62 pp	21,06 pp	18,80 pp	-2,49 pp
PEA Detalhado						
Trabalha em atividade remunerada (PEA)	57,53%	60,45%	66,72%	70,99%	73,83%	16,30 pp
Trabalha em atividade não remunerada, como ajudante (PEA)	50,19%	50,55%	56,21%	58,61%	62,24%	12,05 pp
Trabalha mas está afastado (PEA)	58,47%	56,88%	64,79%	64,72%	70,41%	11,94 pp
Tomou providência para conseguir trabalho nos últimos 30 dias (PEA)	68,02%	71,88%	77,62%	77,92%	81,05%	13,03 pp
Não trabalha e não procurou trabalho nos últimos 30 dias (NÃO PEA)	36,32%	40,16%	46,59%	48,58%	53,70%	17,38 pp
Lacuna (Gap)	31,69 pp	31,73 pp	31,03 pp	29,34 pp	27,35 pp	-4,35 pp
Religião						
Católica Apostólica Romana/ Ortodoxa	46,41%	48,19%	54,07%	57,17%	60,75%	14,34 pp
Igreja Evangélica Luterana/ Presbiteriana/ Metodista/ Batista	53,63%	58,52%	68,88%	68,96%	77,46%	23,83 pp
Outras Igrejas Evangélicas	55,55%	58,93%	64,73%	69,02%	72,44%	16,89 pp
Outras religiosidades cristãs	61,25%	64,77%	74,94%	71,30%	75,65%	14,40 pp
Islamismo / Muçulmano	50,00%	83,33%	66,67%	100,00%	50,00%	0,00 pp
Espírita	66,44%	69,84%	74,00%	79,56%	80,51%	14,07 pp
Umbanda	64,55%	66,29%	77,68%	86,02%	87,07%	22,52 pp
Candomblé	50,00%	56,82%	58,82%	72,73%	81,82%	31,82 pp
Judaísmo	66,67%	62,50%	100,00%	100,00%	100,00%	33,33 pp
Hinduismo	-	66,67%	-	-	50,00%	-16,67 pp
Budismo	66,67%	83,33%	61,54%	81,48%	61,11%	-5,56 pp
Sem religião	67,56%	70,38%	76,03%	76,00%	78,68%	11,12 pp
Agnóstico	75,14%	76,17%	83,54%	85,79%	85,03%	9,90 pp
Ateu	67,40%	73,87%	77,33%	74,10%	80,45%	13,05 pp
Outras religiões	63,64%	64,32%	82,41%	72,25%	79,46%	15,82 pp
Lacuna (Gap)	28,73 pp	35,15 pp	45,93 pp	42,83 pp	50,00 pp	21,27 pp
Etnia						
Branca	51,54%	55,57%	60,98%	63,21%	65,45%	13,90 pp
Preta	46,77%	47,55%	54,13%	58,98%	62,92%	16,15 pp
Amarela	51,99%	55,67%	62,32%	65,11%	70,10%	18,11 pp
Parda	55,20%	57,05%	58,22%	61,89%	64,86%	9,66 pp
Indígena	48,08%	38,31%	48,97%	49,17%	52,19%	4,11 pp
Lacuna (Gap)	8,43 pp	18,74 pp	13,34 pp	15,94 pp	17,91 pp	9,48 pp
Região						
Sudeste	53,05%	58,74%	64,16%	66,32%	69,63%	16,57 pp
Nordeste	47,95%	45,91%	54,37%	57,55%	63,72%	15,76 pp
Sul	51,51%	53,87%	60,23%	64,96%	66,58%	15,08 pp
Norte	51,21%	57,35%	59,83%	65,03%	70,76%	19,55 pp
Centro-Oeste	51,78%	58,49%	63,35%	63,31%	62,02%	10,23 pp
Lacuna (Gap)	5,10 pp	12,83 pp	9,79 pp	8,77 pp	8,75 pp	3,65 pp
Área						
Urbana	53,18%	56,44%	62,25%	65,22%	68,77%	15,59 pp
Rural	28,99%	34,17%	40,93%	45,31%	50,75%	21,76 pp
Lacuna (Gap)	24,20 pp	22,26 pp	21,32 pp	19,91 pp	18,02 pp	-6,17 pp

Fonte: Elaborado pelo autor com base em dados de pesquisa TIC Domicílios 2015-2019

Tabela B-5 – Matriz de correlação de variáveis independentes e dependentes

	INTERNET_1	IDADE	GRAU_INSTRUÇÃO	CLASSE_ECONÔMICA_1	CLASSE_ECONÔMICA_2	CLASSE_ECONÔMICA_3	PEA_2_1	NÃO_CATÓLICA	ÁREA_1
INTERNET_1	1.000000	-0.521656	0.584388	0.083924	0.251946	0.173700	0.189693	0.144533	0.127990
IDADE	-0.521656	1.000000	-0.423567	-0.001593	-0.068495	-0.075190	-0.171910	-0.147651	-0.000998
GRAU_INSTRUÇÃO	0.584388	-0.423567	1.000000	0.137444	0.339935	0.180819	0.269201	0.107819	0.147831
CLASSE_ECONÔMICA_1	0.083924	-0.001593	0.137444	1.000000	-0.048341	-0.102470	0.021093	-0.005057	0.035127
CLASSE_ECONÔMICA_2	0.251946	-0.068495	0.339935	-0.048341	1.000000	-0.370015	0.081885	0.015572	0.112194
CLASSE_ECONÔMICA_3	0.173700	-0.075190	0.180819	-0.102470	-0.370015	1.000000	0.048820	0.047028	0.148220
PEA_2_1	0.189693	-0.171910	0.269201	0.021093	0.081885	0.048820	1.000000	0.043749	0.027095
NÃO_CATÓLICA	0.144533	-0.147651	0.107819	-0.005057	0.015572	0.047028	0.043749	1.000000	0.072443
ÁREA_1	0.127990	-0.000998	0.147831	0.035127	0.112194	0.148220	0.027095	0.072443	1.000000

Fonte: Elaborado pelo autor

Figura B-2 – Resultado de regressão logística (2015-2019)

Logit Regression Results						
Dep. Variable:	INTERNET_1	No. Observations:	89399			
Model:	Logit	Df Residuals:	89379			
Method:	MLE	Df Model:	19			
Date:	Mon, 12 Jul 2021	Pseudo R-squ.:	0.4370			
Time:	15:44:20	Log-Likelihood:	-34068.			
converged:	True	LL-Null:	-60514.			
Covariance Type:	nonrobust	LLR p-value:	0.000			
	coef	std err	z	P> z	[0.025	0.975]
Intercept	-0.8324	0.107	-7.760	0.000	-1.043	-0.622
ANO	0.1053	0.042	2.506	0.012	0.023	0.188
I(IDADE ** 2)	-0.0001	6.08e-05	-2.358	0.018	-0.000	-2.42e-05
IDADE	-0.0587	0.005	-11.488	0.000	-0.069	-0.049
I(GRAU_INSTRUÇÃO ** 2)	0.0039	0.000	8.025	0.000	0.003	0.005
GRAU_INSTRUÇÃO	0.1539	0.008	19.372	0.000	0.138	0.170
CLASSE_ECONÔMICA_1	2.7706	0.150	18.445	0.000	2.476	3.065
CLASSE_ECONÔMICA_2	2.2719	0.064	35.494	0.000	2.146	2.397
CLASSE_ECONÔMICA_3	1.1662	0.038	30.567	0.000	1.091	1.241
PEA_2_1	0.2183	0.041	5.291	0.000	0.137	0.299
NÃO_CATÓLICA	0.2469	0.035	7.112	0.000	0.179	0.315
ÁREA_1	0.5549	0.058	9.596	0.000	0.442	0.668
ANO:I(IDADE ** 2)	-0.0001	2.33e-05	-6.141	0.000	-0.000	-9.73e-05
ANO:IDADE	0.0172	0.002	8.549	0.000	0.013	0.021
ANO:I(GRAU_INSTRUÇÃO ** 2)	-0.0008	0.000	-7.760	0.000	-0.001	-0.001
ANO:CLASSE_ECONÔMICA_2	-0.1037	0.029	-3.620	0.000	-0.160	-0.048
ANO:CLASSE_ECONÔMICA_3	-0.0536	0.015	-3.495	0.000	-0.084	-0.024
ANO:PEA_2_1	-0.0423	0.016	-2.591	0.010	-0.074	-0.010
ANO:NÃO_CATÓLICA	0.0405	0.014	2.852	0.004	0.013	0.068
ANO:ÁREA_1	-0.0539	0.023	-2.384	0.017	-0.098	-0.010

Fonte: Elaborado pelo autor

Tabela B-6– Odds Ratio de Regressão Logística (2015-2019)

	0.025	Odds_Ratio	0.975
Intercept	0.3525	0.4350	0.5368
ANO	1.0232	1.1110	1.2064
I(IDADE ** 2)	0.9997	0.9999	1.0000
IDADE	0.9336	0.9430	0.9525
I(GRAU_INSTRUÇÃO ** 2)	1.0029	1.0039	1.0048
GRAU_INSTRUÇÃO	1.1484	1.1664	1.1847
CLASSE_ECONÔMICA_1	11.8959	15.9681	21.4344
CLASSE_ECONÔMICA_2	8.5541	9.6975	10.9936
CLASSE_ECONÔMICA_3	2.9784	3.2097	3.4589
PEA_2_1	1.1473	1.2440	1.3487
NÃO_CATÓLICA	1.1959	1.2801	1.3702
ÁREA_1	1.5551	1.7417	1.9507
ANO:I(IDADE ** 2)	0.9998	0.9999	0.9999
ANO:IDADE	1.0133	1.0173	1.0213
ANO:I(GRAU_INSTRUÇÃO ** 2)	0.9990	0.9992	0.9994
ANO:CLASSE_ECONÔMICA_2	0.8523	0.9015	0.9536
ANO:CLASSE_ECONÔMICA_3	0.9197	0.9478	0.9767
ANO:PEA_2_1	0.9283	0.9586	0.9897
ANO:NÃO_CATÓLICA	1.0127	1.0413	1.0707
ANO:ÁREA_1	0.9065	0.9475	0.9905

Fonte: Elaborado pelo autor

APÊNDICE C - TRATAMENTO DE TABELAS EXPERIMENTAIS

Tabela C-1 – Achados de processo de verificação e normalização das tabelas experimentais

Ano(s)	Achado	Normalização
2015-2019	Discrepância dos nomes das colunas (variáveis) das tabelas experimentais	Padronização dos nomes das colunas das tabelas experimentais
2015	O código para “CLASSE_ECONÔMICA” é: 1="A", 3="B", 5="C", 6="DE". No entanto, no dicionário de variáveis de 2015 e nos demais anos, o código é: 1="A", 2="B", 3="C", 4="DE"	Padronização de codificação como: 1="A" 2="B" 3="C" 4="DE"
2015	A variável “RELIGIÃO” contém as opções "Católica Apostólica Romana" e "Católica Ortodoxa". A partir de 2016, elas foram consolidadas.	Consolidação das opções "Católica Apostólica Romana" e "Católica Ortodoxa" para o ano de 2015, conforme os anos de 2016-2019.
2019	O código para a variável “FAIXA_ETÁRIA” é 0="De 10 a 15 anos", 1="De 16 a 24 anos", 2="De 25 a 34 anos", 3="De 35 a 44 anos", 4="De 45 a 59 anos", 5="60 anos ou mais". Para os demais anos, ela é: 1="De 10 a 15 anos", 2="De 16 a 24 anos", 3="De 25 a 34 anos", 4="De 35 a 44 anos", 5="De 45 a 59 anos", 6="60 anos ou mais"	Padronização de codificação como: 1="De 10 a 15 anos" 2="De 16 a 24 anos" 3="De 25 a 34 anos" 4="De 35 a 44 anos" 5="De 45 a 59 anos" 6="60 anos ou mais"
2015-2019	A variável “RELIGIÃO” possui código "98" para "Não respondeu".	Linhas com código "98" foram descartadas das tabelas experimentais.
2015-2019	A variável “REND_FAMILIAR” possui código "97" para "Não sabe" e código "98" para "Não respondeu".	Linhas com códigos "97" e "98" foram descartadas das tabelas experimentais.
2016-2019	A variável “REND_FAMILIAR” possui código "9" para "Não tem renda". O ano de 2015 não possui linha com este código.	Linhas com código "97" tiveram seus códigos alterados para "0".
2015-2019	A variável “REND_FAMILIAR” possui categorias relativas a faixas de renda.	As faixas de renda foram transformadas em variável contínua com o valor médio de cada categoria.

2015-2019	A variável “RENDA_FAMILIAR_2” possui código "97" para "Não sabe" e código "98" para "Não respondeu".	Linhas com códigos "97" e "98" foram descartadas das tabelas experimentais.
2016-2019	A variável “RENDA_FAMILIAR_2” possui código "9" para "Não tem renda". O ano de 2015 não possui linha com este código.	Linhas com código "97" tiveram seus códigos alterados para "0".
2015-2019	A variável “RENDA_FAMILIAR_2” possui categorias relativas a faixas de renda.	As faixas de renda foram transformadas em variável contínua com o valor médio de cada categoria.
2015-2019	A variável “RENDA_PESSOAL” possui código "97" para "Não sabe" e código "98" para "Não respondeu".	Linhas com códigos "97" e "98" foram descartadas das tabelas experimentais.
2016-2019	A variável “RENDA_PESSOAL” possui código "9" para "Não tem renda". O ano de 2015 não possui linha com este código.	Linhas com códigos "97" tiveram seus códigos alterados para "0".
2015-2019	A variável “RENDA_PESSOAL” possui categorias relativas a faixas de renda.	As faixas de renda foram transformadas em variável contínua com o valor médio de cada categoria.
2015-2019	A variável “ETNIA” possui código "98" para "Não respondeu".	Linhas com código "98" foram descartadas das tabelas experimentais.
2015	A variável “ETNIA” possui o código: 1="Branca", 2="Preta", 3="Amarela", 4="Parda", 5="Indígena". Para os demais anos, o código é: 1="Branca", 2="Preta", 3="Parda", 4="Amarela", 5="Indígena"	Padronização de codificação como: 1="Branca" 2="Preta" 3="Parda" 4="Amarela" 5="Indígena"
2015-2019	Os códigos da variável “RENDA_FAMILIAR” para todos os anos de pesquisa fazem referência ao valor do salário mínimo do ano de 2017.	Os códigos da variável “RENDA_FAMILIAR” foram atualizados para os valores do salário mínimo respectivos aos anos de 2015-2019 conforme orientação do Cetic.br via contato.

Fonte: Elaborado pelo autor com base na pesquisa TIC Domicílios 2015-2019.

APÊNDICE D - PROGRAMAÇÃO EM PYTHON

Programação em Python para estudo de determinantes do abismo digital no Brasil:

Ativação e configuração de ambiente para o estudo

```
##### Importação das bibliotecas de Python com as extensões de funções necessárias para estudo
import pandas as pd
import numpy as np
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, accuracy_score
import statsmodels.api as sm
import statsmodels.formula.api as smf
from sklearn.metrics import roc_curve, auc
import matplotlib.patches as mpatches
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.stats.outliers_influence import variance_inflation_factor
pd.options.mode.chained_assignment = None # default='warn'
import matplotlib.style as style
```

```
##### Carregamento de resultados da pesquisa TIC Domicílios (Indivíduos) para aos anos 2015-2019
df2015_ind_org = pd.read_excel('TIC_DOMICILIOS_2015_INDIVIDUOS.xls')
df2016_ind_org = pd.read_excel('TIC_DOMICILIOS_2016_INDIVIDUOS.xls')
df2017_ind_org = pd.read_excel('TIC_DOMICILIOS_2017_INDIVIDUOS.xls')
df2018_ind_org = pd.read_excel('TIC_DOMICILIOS_2018_INDIVIDUOS.xls')
df2019_ind_org = pd.read_excel('TIC_DOMICILIOS_2019_INDIVIDUOS.xls')
```

```
##### Confirmação de carregamento dos resultados da pesquisa TIC Domicílios para aos anos 2015-2019
df2015_ind_org
df2016_ind_org
df2017_ind_org
df2018_ind_org
df2019_ind_org
```

- Identificação de variáveis (determinantes) e seleção de colunas das tabelas originais

```
### Comando para agrupamento de campo específico. Utilizado para consolidar dados de tabelas específicas.
#
# Referência de Variáveis Independentes e Dependentes
#
# - Tabelas:
#   df2015: TIC Domicílios 2015
#   df2016: TIC Domicílios 2016
#   df2017: TIC Domicílios 2017
#   df2018: TIC Domicílios 2018
#   df2019: TIC Domicílios 2019
#
#- Variáveis:
# - Exploratórias:
#   SEXO: 'sexo' (df2015/df2016), 'SEXO' (df2017/df2018/df2019)
#   IDADE: 'idade' (df2015/df2016), 'IDADE' (df2017/df2018/df2019)
#   FAIXA ETÁRIA: 'faixa_etaria' (df2015/df2016), 'FAIXA_ETARIA' (df2017/df2018/df2019)
#   GRAU DE INSTRUÇÃO: 'grau_instrucao' (df2015/df2016), 'GRAU_INSTRUCAO' (df2017/df2018/df2019)
#   'GRAU_INSTRUCAO_2' (df2015/df2016/df2017/df2018/df2019)
#   RENDA PESSOAL: 'RENDA_PESSOAL' (df2015/df2016/df2017/df2018/df2019)
#   RENDA FAMILIAR: 'RENDA_FAMILIAR' (df2015/df2016/df2017/df2018/df2019)
#   'RENDA_FAMILIAR_2' (df2015/df2016/df2017/df2018/df2019)
#   RELIGIÃO: 'RELIGIAO' (df2015/df2016/df2017/df2018/df2019)
#   ETNIA: 'ETNIA' (df2015/df2016/df2017/df2018/df2019)
#   REGIÃO: 'COD_REGIAO_2' (df2015/df2016/df2017/df2018/df2019)
#   ÁREA RURAL/URBANA: 'area' (df2015), 'AREA' (df2016/df2017/df2018/df2019)
#   CLASSE ECONÔMICA: 'CLASSE_2015' (df2015/df2016/df2017/df2018/), 'CLASSE_CB2015' (df2019)
#   PAE: 'pea' (df2015/df2016), 'PEA' (df2017/df2018/df2019)
```



```
# 'PEA_2' (df2015/df2016/df2017/df2018/df2019)
####
```

Correções, normalizações e estruturação de variáveis

```
# Lista de colunas das tabelas originais que serão utilizadas no experimento
TODAS_VARIAVEIS_2015_ORG = ['sexo', 'idade', 'faixa_etaria', 'grau_instrucao', 'GRAU_INSTRUCAO_2', \
    'CLASSE_2015', 'REND_A_PESSOAL', 'REND_A_FAMILIAR', 'REND_A_FAMILIAR_2', 'PEA_2', 'pea', \
    'RELIGIAO', 'RACA', 'COD_REGIAO_2', 'area']
TODAS_VARIAVEIS_2016_ORG = ['sexo', 'idade', 'faixa_etaria', 'grau_instrucao', 'GRAU_INSTRUCAO_2', \
    'CLASSE_2015', 'REND_A_PESSOAL', 'REND_A_FAMILIAR', 'REND_A_FAMILIAR_2', 'PEA_2', 'pea', \
    'RELIGIAO', 'RACA', 'COD_REGIAO_2', 'AREA']
TODAS_VARIAVEIS_2017_ORG = ['SEXO', 'IDADE', 'FAIXA_ETARIA', 'GRAU_INSTRUCAO', \
    'GRAU_INSTRUCAO_2', 'CLASSE_2015', 'REND_A_PESSOAL', 'REND_A_FAMILIAR', 'REND_A_FAMILIAR_2', \
    'PEA_2', 'PEA', 'RELIGIAO', 'RACA', 'COD_REGIAO_2', 'AREA']
TODAS_VARIAVEIS_2018_ORG = ['SEXO', 'IDADE', 'FAIXA_ETARIA', 'GRAU_INSTRUCAO', \
    'GRAU_INSTRUCAO_2', 'CLASSE_2015', 'REND_A_PESSOAL', 'REND_A_FAMILIAR', 'REND_A_FAMILIAR_2', \
    'PEA_2', 'PEA', 'RELIGIAO', 'RACA', 'COD_REGIAO_2', 'AREA']
TODAS_VARIAVEIS_2019_ORG = ['SEXO', 'IDADE', 'FAIXA_ETARIA', 'GRAU_INSTRUCAO', \
    'GRAU_INSTRUCAO_2', 'CLASSE_CB2015', 'REND_A_PESSOAL', 'REND_A_FAMILIAR', \
    'REND_A_FAMILIAR_2', 'PEA_2', 'PEA', 'RELIGIAO', 'RACA', 'COD_REGIAO_2', 'AREA']

df2015_ind_exp = df2015_ind_org[TODAS_VARIAVEIS_2015_ORG]
df2016_ind_exp = df2016_ind_org[TODAS_VARIAVEIS_2016_ORG]
df2017_ind_exp = df2017_ind_org[TODAS_VARIAVEIS_2017_ORG]
df2018_ind_exp = df2018_ind_org[TODAS_VARIAVEIS_2018_ORG]
df2019_ind_exp = df2019_ind_org[TODAS_VARIAVEIS_2019_ORG]

#### Padronização de nomes de colunas
VARS_IND = ['SEXO', 'IDADE', 'FAIXA_ETARIA', 'GRAU_INSTRUÇÃO', 'GRAU_INSTRUÇÃO_2', \
    'CLASSE_ECONÔMICA', 'REND_A_PESSOAL', 'REND_A_FAMILIAR', 'REND_A_FAMILIAR_2', \
    'PEA_2', 'PEA', 'RELIGIÃO', 'ETNIA', 'REGIÃO', 'ÁREA']
df2015_ind_exp.columns = VARS_IND
df2016_ind_exp.columns = VARS_IND
df2017_ind_exp.columns = VARS_IND
df2018_ind_exp.columns = VARS_IND
df2019_ind_exp.columns = VARS_IND

#### Normalização de respostas
##### Normalização (Em 2015, os valores para a classe econômica (CLASSE_2015) não são 1,2,3,4; mas 1,3,5,6)
df2015_ind_exp.loc[(df2015_ind_exp['CLASSE_ECONÔMICA'] == 3), 'CLASSE_ECONÔMICA'] = 2
df2015_ind_exp.loc[(df2015_ind_exp['CLASSE_ECONÔMICA'] == 5), 'CLASSE_ECONÔMICA'] = 3
df2015_ind_exp.loc[(df2015_ind_exp['CLASSE_ECONÔMICA'] == 6), 'CLASSE_ECONÔMICA'] = 4

##### Normalização (Em 2015, Religião tinha as opções 'Católica Apostólica Romana' e 'Católica Ortodoxa'.
##### Em 2016, elas foram consolidadas.)
df2015_ind_exp.loc[(df2015_ind_exp['RELIGIÃO'] == 1), 'RELIGIÃO'] = 2
df2015_ind_exp['RELIGIÃO'] = df2015_ind_exp['RELIGIÃO'] - 1

##### Normalização (Retirada de opção 98 (Não respondeu))
df2015_ind_exp.drop(df2015_ind_exp[df2015_ind_exp['RELIGIÃO'] > 15].index, inplace = True)
df2016_ind_exp.drop(df2016_ind_exp[df2016_ind_exp['RELIGIÃO'] > 15].index, inplace = True)
df2017_ind_exp.drop(df2017_ind_exp[df2017_ind_exp['RELIGIÃO'] > 15].index, inplace = True)
df2018_ind_exp.drop(df2018_ind_exp[df2018_ind_exp['RELIGIÃO'] > 15].index, inplace = True)
df2019_ind_exp.drop(df2019_ind_exp[df2019_ind_exp['RELIGIÃO'] > 15].index, inplace = True)

##### Normalização (Em 2019, os valores para faixa etária eram 0,1,2,3,4,5 e não 1,2,3,4,5,6)
df2019_ind_exp['FAIXA_ETARIA'] = df2019_ind_exp['FAIXA_ETARIA'] + 1

##### Alteração de ETNIA (Retirada de Opção 98 (Não respondeu))
df2015_ind_exp.drop(df2015_ind_exp[df2015_ind_exp['ETNIA'] > 5].index, inplace = True)
df2016_ind_exp.drop(df2016_ind_exp[df2016_ind_exp['ETNIA'] > 5].index, inplace = True)
df2017_ind_exp.drop(df2017_ind_exp[df2017_ind_exp['ETNIA'] > 5].index, inplace = True)
df2018_ind_exp.drop(df2018_ind_exp[df2018_ind_exp['ETNIA'] > 5].index, inplace = True)
df2019_ind_exp.drop(df2019_ind_exp[df2019_ind_exp['ETNIA'] > 5].index, inplace = True)

##### Alteração de ETNIA (Em 2015, valores para pardo e amarelo estão invertidos)
```

```
df2015_ind_exp.loc[(df2015_ind_exp['ETNIA'] == 3), 'ETNIA'] = -4
df2015_ind_exp.loc[(df2015_ind_exp['ETNIA'] == 4), 'ETNIA'] = -3
df2015_ind_exp.loc[(df2015_ind_exp['ETNIA'] == -3), 'ETNIA'] = 3
df2015_ind_exp.loc[(df2015_ind_exp['ETNIA'] == -4), 'ETNIA'] = 4
```

Alteração de RENDA FAMILIAR. Opção 9 (Não tem renda) vai para o início.

Conversão para média de valores e não faixas. E descarte de opções 97 (Não sabe) e 98 (Não respondeu)

```
df2015_ind_exp.drop(df2015_ind_exp[df2015_ind_exp['RENDA_FAMILIAR'] > 9].index, inplace = True)
df2016_ind_exp.drop(df2016_ind_exp[df2016_ind_exp['RENDA_FAMILIAR'] > 9].index, inplace = True)
df2017_ind_exp.drop(df2017_ind_exp[df2017_ind_exp['RENDA_FAMILIAR'] > 9].index, inplace = True)
df2018_ind_exp.drop(df2018_ind_exp[df2018_ind_exp['RENDA_FAMILIAR'] > 9].index, inplace = True)
df2019_ind_exp.drop(df2019_ind_exp[df2019_ind_exp['RENDA_FAMILIAR'] > 9].index, inplace = True)
```

Não tem df2015

```
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR'] == 9), 'RENDA_FAMILIAR'] = 0
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR'] == 9), 'RENDA_FAMILIAR'] = 0
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR'] == 9), 'RENDA_FAMILIAR'] = 0
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR'] == 9), 'RENDA_FAMILIAR'] = 0
```

```
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_FAMILIAR'] == 9), 'RENDA_FAMILIAR'] = 0
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_FAMILIAR'] == 1), 'RENDA_FAMILIAR'] = 394.000
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_FAMILIAR'] == 2), 'RENDA_FAMILIAR'] = 1182.005
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_FAMILIAR'] == 3), 'RENDA_FAMILIAR'] = 1970.005
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_FAMILIAR'] == 4), 'RENDA_FAMILIAR'] = 3152.005
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_FAMILIAR'] == 5), 'RENDA_FAMILIAR'] = 5910.005
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_FAMILIAR'] == 6), 'RENDA_FAMILIAR'] = 11820.005
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_FAMILIAR'] == 7), 'RENDA_FAMILIAR'] = 19700.005
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_FAMILIAR'] == 8), 'RENDA_FAMILIAR'] = 23640.00
```

```
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR'] == 9), 'RENDA_FAMILIAR'] = 0
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR'] == 1), 'RENDA_FAMILIAR'] = 440.000
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR'] == 2), 'RENDA_FAMILIAR'] = 1320.005
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR'] == 3), 'RENDA_FAMILIAR'] = 2200.005
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR'] == 4), 'RENDA_FAMILIAR'] = 3520.005
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR'] == 5), 'RENDA_FAMILIAR'] = 6640.005
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR'] == 6), 'RENDA_FAMILIAR'] = 13240.005
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR'] == 7), 'RENDA_FAMILIAR'] = 22000.005
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR'] == 8), 'RENDA_FAMILIAR'] = 26400.00
```

```
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR'] == 9), 'RENDA_FAMILIAR'] = 0
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR'] == 1), 'RENDA_FAMILIAR'] = 468.500
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR'] == 2), 'RENDA_FAMILIAR'] = 1405.505
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR'] == 3), 'RENDA_FAMILIAR'] = 2342.505
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR'] == 4), 'RENDA_FAMILIAR'] = 3748.005
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR'] == 5), 'RENDA_FAMILIAR'] = 7027.505
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR'] == 6), 'RENDA_FAMILIAR'] = 14055.005
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR'] == 7), 'RENDA_FAMILIAR'] = 23425.005
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR'] == 8), 'RENDA_FAMILIAR'] = 28110.00
```

```
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR'] == 9), 'RENDA_FAMILIAR'] = 0
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR'] == 1), 'RENDA_FAMILIAR'] = 477.000
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR'] == 2), 'RENDA_FAMILIAR'] = 1431.005
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR'] == 3), 'RENDA_FAMILIAR'] = 2385.005
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR'] == 4), 'RENDA_FAMILIAR'] = 3816.005
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR'] == 5), 'RENDA_FAMILIAR'] = 7155.005
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR'] == 6), 'RENDA_FAMILIAR'] = 14310.005
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR'] == 7), 'RENDA_FAMILIAR'] = 23850.005
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR'] == 8), 'RENDA_FAMILIAR'] = 28620.00
```

```
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR'] == 9), 'RENDA_FAMILIAR'] = 0
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR'] == 1), 'RENDA_FAMILIAR'] = 499.000
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR'] == 2), 'RENDA_FAMILIAR'] = 1497.005
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR'] == 3), 'RENDA_FAMILIAR'] = 2495.005
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR'] == 4), 'RENDA_FAMILIAR'] = 3992.005
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR'] == 5), 'RENDA_FAMILIAR'] = 7485.005
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR'] == 6), 'RENDA_FAMILIAR'] = 14970.005
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR'] == 7), 'RENDA_FAMILIAR'] = 24950.005
```

```
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR'] == 8), 'RENDA_FAMILIAR'] = 29940.00
```

```
##### Alteração de RENDA FAMILIAR 2. Opção 7 (Não tem renda) vai para o início.
```

```
##### Conversão para média de valores e não faixas. E descarte de opções 97 (Não sabe) e 98 (Não respondeu)
```

```
df2015_ind_exp.drop(df2015_ind_exp[df2015_ind_exp['RENDA_FAMILIAR_2'] > 7].index, inplace = True)
```

```
df2016_ind_exp.drop(df2016_ind_exp[df2016_ind_exp['RENDA_FAMILIAR_2'] > 7].index, inplace = True)
```

```
df2017_ind_exp.drop(df2017_ind_exp[df2017_ind_exp['RENDA_FAMILIAR_2'] > 7].index, inplace = True)
```

```
df2018_ind_exp.drop(df2018_ind_exp[df2018_ind_exp['RENDA_FAMILIAR_2'] > 7].index, inplace = True)
```

```
df2019_ind_exp.drop(df2019_ind_exp[df2019_ind_exp['RENDA_FAMILIAR_2'] > 7].index, inplace = True)
```

```
##### Não tem df2015
```

```
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR_2'] == 7), 'RENDA_FAMILIAR_2'] = 0
```

```
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR_2'] == 7), 'RENDA_FAMILIAR_2'] = 0
```

```
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR_2'] == 7), 'RENDA_FAMILIAR_2'] = 0
```

```
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR_2'] == 7), 'RENDA_FAMILIAR_2'] = 0
```

```
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_FAMILIAR_2'] == 1), 'RENDA_FAMILIAR_2'] = (1)/2
```

```
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_FAMILIAR_2'] == 2), 'RENDA_FAMILIAR_2'] = (1 + 2)/2
```

```
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_FAMILIAR_2'] == 3), 'RENDA_FAMILIAR_2'] = (2 + 3)/2
```

```
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_FAMILIAR_2'] == 4), 'RENDA_FAMILIAR_2'] = (3 + 5)/2
```

```
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_FAMILIAR_2'] == 5), 'RENDA_FAMILIAR_2'] = (5 + 10)/2
```

```
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_FAMILIAR_2'] == 6), 'RENDA_FAMILIAR_2'] = (10)
```

```
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR_2'] == 1), 'RENDA_FAMILIAR_2'] = (1)/2
```

```
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR_2'] == 2), 'RENDA_FAMILIAR_2'] = (1 + 2)/2
```

```
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR_2'] == 3), 'RENDA_FAMILIAR_2'] = (2 + 3)/2
```

```
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR_2'] == 4), 'RENDA_FAMILIAR_2'] = (3 + 5)/2
```

```
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR_2'] == 5), 'RENDA_FAMILIAR_2'] = (5 + 10)/2
```

```
df2016_ind_exp.loc[(df2016_ind_exp['RENDA_FAMILIAR_2'] == 6), 'RENDA_FAMILIAR_2'] = (10)
```

```
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR_2'] == 1), 'RENDA_FAMILIAR_2'] = (1)/2
```

```
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR_2'] == 2), 'RENDA_FAMILIAR_2'] = (1 + 2)/2
```

```
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR_2'] == 3), 'RENDA_FAMILIAR_2'] = (2 + 3)/2
```

```
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR_2'] == 4), 'RENDA_FAMILIAR_2'] = (3 + 5)/2
```

```
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR_2'] == 5), 'RENDA_FAMILIAR_2'] = (5 + 10)/2
```

```
df2017_ind_exp.loc[(df2017_ind_exp['RENDA_FAMILIAR_2'] == 6), 'RENDA_FAMILIAR_2'] = (10)
```

```
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR_2'] == 1), 'RENDA_FAMILIAR_2'] = (1)/2
```

```
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR_2'] == 2), 'RENDA_FAMILIAR_2'] = (1 + 2)/2
```

```
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR_2'] == 3), 'RENDA_FAMILIAR_2'] = (2 + 3)/2
```

```
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR_2'] == 4), 'RENDA_FAMILIAR_2'] = (3 + 5)/2
```

```
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR_2'] == 5), 'RENDA_FAMILIAR_2'] = (5 + 10)/2
```

```
df2018_ind_exp.loc[(df2018_ind_exp['RENDA_FAMILIAR_2'] == 6), 'RENDA_FAMILIAR_2'] = (10)
```

```
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR_2'] == 1), 'RENDA_FAMILIAR_2'] = (1)/2
```

```
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR_2'] == 2), 'RENDA_FAMILIAR_2'] = (1 + 2)/2
```

```
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR_2'] == 3), 'RENDA_FAMILIAR_2'] = (2 + 3)/2
```

```
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR_2'] == 4), 'RENDA_FAMILIAR_2'] = (3 + 5)/2
```

```
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR_2'] == 5), 'RENDA_FAMILIAR_2'] = (5 + 10)/2
```

```
df2019_ind_exp.loc[(df2019_ind_exp['RENDA_FAMILIAR_2'] == 6), 'RENDA_FAMILIAR_2'] = (10)
```

```
##### Alteração de RENDA PESSOAL. Opção 9 (Não tem renda) vai para o início.
```

```
##### Conversão para média de valores e não faixas. E descarte de opções 97 (Não sabe) e 98 (Não respondeu)
```

```
df2015_ind_exp.drop(df2015_ind_exp[df2015_ind_exp['RENDA_PESSOAL'] > 9].index, inplace = True)
```

```
df2016_ind_exp.drop(df2016_ind_exp[df2016_ind_exp['RENDA_PESSOAL'] > 9].index, inplace = True)
```

```
df2017_ind_exp.drop(df2017_ind_exp[df2017_ind_exp['RENDA_PESSOAL'] > 9].index, inplace = True)
```

```
df2018_ind_exp.drop(df2018_ind_exp[df2018_ind_exp['RENDA_PESSOAL'] > 9].index, inplace = True)
```

```
df2019_ind_exp.drop(df2019_ind_exp[df2019_ind_exp['RENDA_PESSOAL'] > 9].index, inplace = True)
```

```
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_PESSOAL'] == 9), 'RENDA_PESSOAL'] = 0
```

```
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_PESSOAL'] == 1), 'RENDA_PESSOAL'] = (788.00)/2
```

```
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_PESSOAL'] == 2), 'RENDA_PESSOAL'] = (788.01 + 1576.00)/2
```

```
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_PESSOAL'] == 3), 'RENDA_PESSOAL'] = (1576.01 + 2364.00)/2
```

```
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_PESSOAL'] == 4), 'RENDA_PESSOAL'] = (2364.01 + 3940.00)/2
```

```
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_PESSOAL'] == 5), 'RENDA_PESSOAL'] = (3940.01 + 7880.00)/2
```

```
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_PESSOAL'] == 6), 'RENDA_PESSOAL'] = 11820.005
```

```
df2015_ind_exp.loc[(df2015_ind_exp['RENDA_PESSOAL'] == 7), 'RENDA_PESSOAL'] = (15760.01 + 23640.00)/2
```

df2015_ind_exp.loc[(df2015_ind_exp['RENDA_PESSOAL'] == 8), 'RENDA_PESSOAL'] = 23640.00

df2016_ind_exp.loc[(df2016_ind_exp['RENDA_PESSOAL'] == 9), 'RENDA_PESSOAL'] = 0
 df2016_ind_exp.loc[(df2016_ind_exp['RENDA_PESSOAL'] == 1), 'RENDA_PESSOAL'] = (880.00)/2
 df2016_ind_exp.loc[(df2016_ind_exp['RENDA_PESSOAL'] == 2), 'RENDA_PESSOAL'] = (880.01 + 1760.00)/2
 df2016_ind_exp.loc[(df2016_ind_exp['RENDA_PESSOAL'] == 3), 'RENDA_PESSOAL'] = (1760.01 + 2640.00)/2
 df2016_ind_exp.loc[(df2016_ind_exp['RENDA_PESSOAL'] == 4), 'RENDA_PESSOAL'] = (2640.01 + 4400.00)/2
 df2016_ind_exp.loc[(df2016_ind_exp['RENDA_PESSOAL'] == 5), 'RENDA_PESSOAL'] = (4400.01 + 8880.00)/2
 df2016_ind_exp.loc[(df2016_ind_exp['RENDA_PESSOAL'] == 6), 'RENDA_PESSOAL'] = (8880.01 + 17600.00)/2
 df2016_ind_exp.loc[(df2016_ind_exp['RENDA_PESSOAL'] == 7), 'RENDA_PESSOAL'] = (17600.01 + 26400.00)/2
 df2016_ind_exp.loc[(df2016_ind_exp['RENDA_PESSOAL'] == 8), 'RENDA_PESSOAL'] = 26400.00

df2017_ind_exp.loc[(df2017_ind_exp['RENDA_PESSOAL'] == 9), 'RENDA_PESSOAL'] = 0
 df2017_ind_exp.loc[(df2017_ind_exp['RENDA_PESSOAL'] == 1), 'RENDA_PESSOAL'] = (937.00)/2
 df2017_ind_exp.loc[(df2017_ind_exp['RENDA_PESSOAL'] == 2), 'RENDA_PESSOAL'] = (937.01 + 1874.00)/2
 df2017_ind_exp.loc[(df2017_ind_exp['RENDA_PESSOAL'] == 3), 'RENDA_PESSOAL'] = (1874.01 + 2811.00)/2
 df2017_ind_exp.loc[(df2017_ind_exp['RENDA_PESSOAL'] == 4), 'RENDA_PESSOAL'] = (2811.01 + 4685.00)/2
 df2017_ind_exp.loc[(df2017_ind_exp['RENDA_PESSOAL'] == 5), 'RENDA_PESSOAL'] = (4685.01 + 9370.00)/2
 df2017_ind_exp.loc[(df2017_ind_exp['RENDA_PESSOAL'] == 6), 'RENDA_PESSOAL'] = (9370.01 + 18740.00)/2
 df2017_ind_exp.loc[(df2017_ind_exp['RENDA_PESSOAL'] == 7), 'RENDA_PESSOAL'] = (18740.01 + 28110.00)/2
 df2017_ind_exp.loc[(df2017_ind_exp['RENDA_PESSOAL'] == 8), 'RENDA_PESSOAL'] = 28110.00

df2018_ind_exp.loc[(df2018_ind_exp['RENDA_PESSOAL'] == 9), 'RENDA_PESSOAL'] = 0
 df2018_ind_exp.loc[(df2018_ind_exp['RENDA_PESSOAL'] == 1), 'RENDA_PESSOAL'] = (954.00)/2
 df2018_ind_exp.loc[(df2018_ind_exp['RENDA_PESSOAL'] == 2), 'RENDA_PESSOAL'] = (954.01 + 1908.00)/2
 df2018_ind_exp.loc[(df2018_ind_exp['RENDA_PESSOAL'] == 3), 'RENDA_PESSOAL'] = (1908.01 + 2862.00)/2
 df2018_ind_exp.loc[(df2018_ind_exp['RENDA_PESSOAL'] == 4), 'RENDA_PESSOAL'] = (2862.01 + 4770.00)/2
 df2018_ind_exp.loc[(df2018_ind_exp['RENDA_PESSOAL'] == 5), 'RENDA_PESSOAL'] = (4770.01 + 9540.00)/2
 df2018_ind_exp.loc[(df2018_ind_exp['RENDA_PESSOAL'] == 6), 'RENDA_PESSOAL'] = (9540.01 + 19080.00)/2
 df2018_ind_exp.loc[(df2018_ind_exp['RENDA_PESSOAL'] == 7), 'RENDA_PESSOAL'] = (19080.01 + 28620.00)/2
 df2018_ind_exp.loc[(df2018_ind_exp['RENDA_PESSOAL'] == 8), 'RENDA_PESSOAL'] = 28620.00

df2019_ind_exp.loc[(df2019_ind_exp['RENDA_PESSOAL'] == 9), 'RENDA_PESSOAL'] = 0
 df2019_ind_exp.loc[(df2019_ind_exp['RENDA_PESSOAL'] == 1), 'RENDA_PESSOAL'] = (998.00)/2
 df2019_ind_exp.loc[(df2019_ind_exp['RENDA_PESSOAL'] == 2), 'RENDA_PESSOAL'] = (998.01 + 1996.00)/2
 df2019_ind_exp.loc[(df2019_ind_exp['RENDA_PESSOAL'] == 3), 'RENDA_PESSOAL'] = (1996.01 + 2994.00)/2
 df2019_ind_exp.loc[(df2019_ind_exp['RENDA_PESSOAL'] == 4), 'RENDA_PESSOAL'] = (2994.01 + 4990.00)/2
 df2019_ind_exp.loc[(df2019_ind_exp['RENDA_PESSOAL'] == 5), 'RENDA_PESSOAL'] = (4990.01 + 9980.00)/2
 df2019_ind_exp.loc[(df2019_ind_exp['RENDA_PESSOAL'] == 6), 'RENDA_PESSOAL'] = (9980.01 + 19960.00)/2
 df2019_ind_exp.loc[(df2019_ind_exp['RENDA_PESSOAL'] == 7), 'RENDA_PESSOAL'] = (19960.01 + 29940.00)/2
 df2019_ind_exp.loc[(df2019_ind_exp['RENDA_PESSOAL'] == 8), 'RENDA_PESSOAL'] = 29940.00

Alteração de GRAU_INSTRUÇÃO para anos de estudos

df2015_ind_exp.loc[(df2015_ind_exp['GRAU_INSTRUÇÃO'] == 1), 'GRAU_INSTRUÇÃO'] = 0
 df2015_ind_exp.loc[(df2015_ind_exp['GRAU_INSTRUÇÃO'] == 2), 'GRAU_INSTRUÇÃO'] = 1
 df2015_ind_exp.loc[(df2015_ind_exp['GRAU_INSTRUÇÃO'] == 3), 'GRAU_INSTRUÇÃO'] = 0.75
 df2015_ind_exp.loc[(df2015_ind_exp['GRAU_INSTRUÇÃO'] == 4), 'GRAU_INSTRUÇÃO'] = 1
 df2015_ind_exp.loc[(df2015_ind_exp['GRAU_INSTRUÇÃO'] == 5), 'GRAU_INSTRUÇÃO'] = 3
 df2015_ind_exp.loc[(df2015_ind_exp['GRAU_INSTRUÇÃO'] == 6), 'GRAU_INSTRUÇÃO'] = 5.5

Ordem invertida para não dar problema

df2015_ind_exp.loc[(df2015_ind_exp['GRAU_INSTRUÇÃO'] == 12), 'GRAU_INSTRUÇÃO'] = 17.5
 df2015_ind_exp.loc[(df2015_ind_exp['GRAU_INSTRUÇÃO'] == 11), 'GRAU_INSTRUÇÃO'] = 15.5
 df2015_ind_exp.loc[(df2015_ind_exp['GRAU_INSTRUÇÃO'] == 10), 'GRAU_INSTRUÇÃO'] = 13
 df2015_ind_exp.loc[(df2015_ind_exp['GRAU_INSTRUÇÃO'] == 9), 'GRAU_INSTRUÇÃO'] = 11.5
 df2015_ind_exp.loc[(df2015_ind_exp['GRAU_INSTRUÇÃO'] == 8), 'GRAU_INSTRUÇÃO'] = 10
 df2015_ind_exp.loc[(df2015_ind_exp['GRAU_INSTRUÇÃO'] == 7), 'GRAU_INSTRUÇÃO'] = 8

df2016_ind_exp.loc[(df2016_ind_exp['GRAU_INSTRUÇÃO'] == 1), 'GRAU_INSTRUÇÃO'] = 0
 df2016_ind_exp.loc[(df2016_ind_exp['GRAU_INSTRUÇÃO'] == 2), 'GRAU_INSTRUÇÃO'] = 1
 df2016_ind_exp.loc[(df2016_ind_exp['GRAU_INSTRUÇÃO'] == 3), 'GRAU_INSTRUÇÃO'] = 0.75
 df2016_ind_exp.loc[(df2016_ind_exp['GRAU_INSTRUÇÃO'] == 4), 'GRAU_INSTRUÇÃO'] = 1
 df2016_ind_exp.loc[(df2016_ind_exp['GRAU_INSTRUÇÃO'] == 5), 'GRAU_INSTRUÇÃO'] = 3
 df2016_ind_exp.loc[(df2016_ind_exp['GRAU_INSTRUÇÃO'] == 6), 'GRAU_INSTRUÇÃO'] = 5.5

Ordem invertida para não dar problema

df2016_ind_exp.loc[(df2016_ind_exp['GRAU_INSTRUÇÃO'] == 12), 'GRAU_INSTRUÇÃO'] = 17.5
 df2016_ind_exp.loc[(df2016_ind_exp['GRAU_INSTRUÇÃO'] == 11), 'GRAU_INSTRUÇÃO'] = 15.5
 df2016_ind_exp.loc[(df2016_ind_exp['GRAU_INSTRUÇÃO'] == 10), 'GRAU_INSTRUÇÃO'] = 13

```
df2016_ind_exp.loc[(df2016_ind_exp['GRAU_INSTRUÇÃO'] == 9), 'GRAU_INSTRUÇÃO'] = 11.5
df2016_ind_exp.loc[(df2016_ind_exp['GRAU_INSTRUÇÃO'] == 8), 'GRAU_INSTRUÇÃO'] = 10
df2016_ind_exp.loc[(df2016_ind_exp['GRAU_INSTRUÇÃO'] == 7), 'GRAU_INSTRUÇÃO'] = 8
```

```
df2017_ind_exp.loc[(df2017_ind_exp['GRAU_INSTRUÇÃO'] == 1), 'GRAU_INSTRUÇÃO'] = 0
df2017_ind_exp.loc[(df2017_ind_exp['GRAU_INSTRUÇÃO'] == 2), 'GRAU_INSTRUÇÃO'] = 1
df2017_ind_exp.loc[(df2017_ind_exp['GRAU_INSTRUÇÃO'] == 3), 'GRAU_INSTRUÇÃO'] = 0.75
df2017_ind_exp.loc[(df2017_ind_exp['GRAU_INSTRUÇÃO'] == 4), 'GRAU_INSTRUÇÃO'] = 1
df2017_ind_exp.loc[(df2017_ind_exp['GRAU_INSTRUÇÃO'] == 5), 'GRAU_INSTRUÇÃO'] = 3
df2017_ind_exp.loc[(df2017_ind_exp['GRAU_INSTRUÇÃO'] == 6), 'GRAU_INSTRUÇÃO'] = 5.5
```

Ordem invertida para não dar problema

```
df2017_ind_exp.loc[(df2017_ind_exp['GRAU_INSTRUÇÃO'] == 12), 'GRAU_INSTRUÇÃO'] = 17.5
df2017_ind_exp.loc[(df2017_ind_exp['GRAU_INSTRUÇÃO'] == 11), 'GRAU_INSTRUÇÃO'] = 15.5
df2017_ind_exp.loc[(df2017_ind_exp['GRAU_INSTRUÇÃO'] == 10), 'GRAU_INSTRUÇÃO'] = 13
df2017_ind_exp.loc[(df2017_ind_exp['GRAU_INSTRUÇÃO'] == 9), 'GRAU_INSTRUÇÃO'] = 11.5
df2017_ind_exp.loc[(df2017_ind_exp['GRAU_INSTRUÇÃO'] == 8), 'GRAU_INSTRUÇÃO'] = 10
df2017_ind_exp.loc[(df2017_ind_exp['GRAU_INSTRUÇÃO'] == 7), 'GRAU_INSTRUÇÃO'] = 8
```

```
df2018_ind_exp.loc[(df2018_ind_exp['GRAU_INSTRUÇÃO'] == 1), 'GRAU_INSTRUÇÃO'] = 0
df2018_ind_exp.loc[(df2018_ind_exp['GRAU_INSTRUÇÃO'] == 2), 'GRAU_INSTRUÇÃO'] = 1
df2018_ind_exp.loc[(df2018_ind_exp['GRAU_INSTRUÇÃO'] == 3), 'GRAU_INSTRUÇÃO'] = 0.75
df2018_ind_exp.loc[(df2018_ind_exp['GRAU_INSTRUÇÃO'] == 4), 'GRAU_INSTRUÇÃO'] = 1
df2018_ind_exp.loc[(df2018_ind_exp['GRAU_INSTRUÇÃO'] == 5), 'GRAU_INSTRUÇÃO'] = 3
df2018_ind_exp.loc[(df2018_ind_exp['GRAU_INSTRUÇÃO'] == 6), 'GRAU_INSTRUÇÃO'] = 5.5
```

Ordem invertida para não dar problema

```
df2018_ind_exp.loc[(df2018_ind_exp['GRAU_INSTRUÇÃO'] == 12), 'GRAU_INSTRUÇÃO'] = 17.5
df2018_ind_exp.loc[(df2018_ind_exp['GRAU_INSTRUÇÃO'] == 11), 'GRAU_INSTRUÇÃO'] = 15.5
df2018_ind_exp.loc[(df2018_ind_exp['GRAU_INSTRUÇÃO'] == 10), 'GRAU_INSTRUÇÃO'] = 13
df2018_ind_exp.loc[(df2018_ind_exp['GRAU_INSTRUÇÃO'] == 9), 'GRAU_INSTRUÇÃO'] = 11.5
df2018_ind_exp.loc[(df2018_ind_exp['GRAU_INSTRUÇÃO'] == 8), 'GRAU_INSTRUÇÃO'] = 10
df2018_ind_exp.loc[(df2018_ind_exp['GRAU_INSTRUÇÃO'] == 7), 'GRAU_INSTRUÇÃO'] = 8
```

```
df2019_ind_exp.loc[(df2019_ind_exp['GRAU_INSTRUÇÃO'] == 1), 'GRAU_INSTRUÇÃO'] = 0
df2019_ind_exp.loc[(df2019_ind_exp['GRAU_INSTRUÇÃO'] == 2), 'GRAU_INSTRUÇÃO'] = 1
df2019_ind_exp.loc[(df2019_ind_exp['GRAU_INSTRUÇÃO'] == 3), 'GRAU_INSTRUÇÃO'] = 0.75
df2019_ind_exp.loc[(df2019_ind_exp['GRAU_INSTRUÇÃO'] == 4), 'GRAU_INSTRUÇÃO'] = 1
df2019_ind_exp.loc[(df2019_ind_exp['GRAU_INSTRUÇÃO'] == 5), 'GRAU_INSTRUÇÃO'] = 3
df2019_ind_exp.loc[(df2019_ind_exp['GRAU_INSTRUÇÃO'] == 6), 'GRAU_INSTRUÇÃO'] = 5.5
```

Ordem invertida para não dar problema

```
df2019_ind_exp.loc[(df2019_ind_exp['GRAU_INSTRUÇÃO'] == 12), 'GRAU_INSTRUÇÃO'] = 17.5
df2019_ind_exp.loc[(df2019_ind_exp['GRAU_INSTRUÇÃO'] == 11), 'GRAU_INSTRUÇÃO'] = 15.5
df2019_ind_exp.loc[(df2019_ind_exp['GRAU_INSTRUÇÃO'] == 10), 'GRAU_INSTRUÇÃO'] = 13
df2019_ind_exp.loc[(df2019_ind_exp['GRAU_INSTRUÇÃO'] == 9), 'GRAU_INSTRUÇÃO'] = 11.5
df2019_ind_exp.loc[(df2019_ind_exp['GRAU_INSTRUÇÃO'] == 8), 'GRAU_INSTRUÇÃO'] = 10
df2019_ind_exp.loc[(df2019_ind_exp['GRAU_INSTRUÇÃO'] == 7), 'GRAU_INSTRUÇÃO'] = 8
```

Criação de tabelas dummies

Índice de variáveis dummies

```
VARS_IND_DUM = ['SEXO', 'GRAU_INSTRUÇÃO_2', 'CLASSE_ECONÔMICA', 'PEA_2', 'RELIGIÃO', 'RAÇA',\
'REGIÃO', 'ÁREA', 'PEA', 'FAIXA_ETARIA']
```

```
VARS_IND_NODUM = ['IDADE', 'GRAU_INSTRUÇÃO', 'RENDA_PESSOAL', 'RENDA_FAMILIAR',\
'RENDA_FAMILIAR_2']
```

Criação de tabelas dummies para variáveis dependentes

```
INTERNET_2015 = pd.get_dummies(df2015_ind_org['C3'], prefix='INTERNET')
```

```
INTERNET_2016 = pd.get_dummies(df2016_ind_org['C3'], prefix='INTERNET')
```

```
INTERNET_2017 = pd.get_dummies(df2017_ind_org['C3'], prefix='INTERNET')
```

```
INTERNET_2018 = pd.get_dummies(df2018_ind_org['C3'], prefix='INTERNET')
```

```
INTERNET_2019 = pd.get_dummies(df2019_ind_org['C3'], prefix='INTERNET')
```

Inclusão de coluna INTERNET_1 nas tabelas Experimentais(sem dummies)

```
df2015_ind_exp.insert(0, 'INTERNET_1', INTERNET_2015['INTERNET_1'])
```

```
df2016_ind_exp.insert(0, 'INTERNET_1', INTERNET_2016['INTERNET_1'])
```

```
df2017_ind_exp.insert(0, 'INTERNET_1', INTERNET_2017['INTERNET_1'])
```

```
df2018_ind_exp.insert(0, 'INTERNET_1', INTERNET_2018['INTERNET_1'])
```

```
df2019_ind_exp.insert(0, 'INTERNET_1', INTERNET_2019['INTERNET_1'])
```

```

### Variáveis dummies para variáveis independentes
df2015_ind_dum = pd.concat([pd.get_dummies(df2015_ind_exp[col], prefix=col) for col in VARS_IND_DUM], axis=1,
sort=False)
df2016_ind_dum = pd.concat([pd.get_dummies(df2016_ind_exp[col], prefix=col) for col in VARS_IND_DUM], axis=1,
sort=False)
df2017_ind_dum = pd.concat([pd.get_dummies(df2017_ind_exp[col], prefix=col) for col in VARS_IND_DUM], axis=1,
sort=False)
df2018_ind_dum = pd.concat([pd.get_dummies(df2018_ind_exp[col], prefix=col) for col in VARS_IND_DUM], axis=1,
sort=False)
df2019_ind_dum = pd.concat([pd.get_dummies(df2019_ind_exp[col], prefix=col) for col in VARS_IND_DUM], axis=1,
sort=False)

### Inclusão de coluna INTERNET_1 nas tabelas dummies (variáveis independentes)
df2015_ind_dum.insert(0, 'INTERNET_1', INTERNET_2015['INTERNET_1'])
df2016_ind_dum.insert(0, 'INTERNET_1', INTERNET_2016['INTERNET_1'])
df2017_ind_dum.insert(0, 'INTERNET_1', INTERNET_2017['INTERNET_1'])
df2018_ind_dum.insert(0, 'INTERNET_1', INTERNET_2018['INTERNET_1'])
df2019_ind_dum.insert(0, 'INTERNET_1', INTERNET_2019['INTERNET_1'])

###
df2015_ind_dum[VARS_IND_NODUM] = df2015_ind_exp[VARS_IND_NODUM]
df2016_ind_dum[VARS_IND_NODUM] = df2016_ind_exp[VARS_IND_NODUM]
df2017_ind_dum[VARS_IND_NODUM] = df2017_ind_exp[VARS_IND_NODUM]
df2018_ind_dum[VARS_IND_NODUM] = df2018_ind_exp[VARS_IND_NODUM]
df2019_ind_dum[VARS_IND_NODUM] = df2019_ind_exp[VARS_IND_NODUM]

### União de dummies de Religião para criar variável Católico e Não Católico
df2015_ind_dum['NÃO_CATÓLICA'] = df2015_ind_dum['RELIGIÃO_2'] + df2015_ind_dum['RELIGIÃO_3'] + \
df2015_ind_dum['RELIGIÃO_4'] + df2015_ind_dum['RELIGIÃO_5'] + df2015_ind_dum['RELIGIÃO_6'] + \
df2015_ind_dum['RELIGIÃO_7'] + df2015_ind_dum['RELIGIÃO_8'] + df2015_ind_dum['RELIGIÃO_9'] + \
df2015_ind_dum['RELIGIÃO_11'] + df2015_ind_dum['RELIGIÃO_12'] + df2015_ind_dum['RELIGIÃO_13'] + \
df2015_ind_dum['RELIGIÃO_14'] + df2015_ind_dum['RELIGIÃO_15']

df2016_ind_dum['NÃO_CATÓLICA'] = df2016_ind_dum['RELIGIÃO_2'] + df2016_ind_dum['RELIGIÃO_3'] + \
df2016_ind_dum['RELIGIÃO_4'] + df2016_ind_dum['RELIGIÃO_5'] + df2016_ind_dum['RELIGIÃO_6'] + \
df2016_ind_dum['RELIGIÃO_7'] + df2016_ind_dum['RELIGIÃO_8'] + df2016_ind_dum['RELIGIÃO_9'] + \
df2016_ind_dum['RELIGIÃO_10'] + df2016_ind_dum['RELIGIÃO_11'] + df2016_ind_dum['RELIGIÃO_12'] + \
df2016_ind_dum['RELIGIÃO_13'] + df2016_ind_dum['RELIGIÃO_14'] + df2016_ind_dum['RELIGIÃO_15']

df2017_ind_dum['NÃO_CATÓLICA'] = df2017_ind_dum['RELIGIÃO_2'] + df2017_ind_dum['RELIGIÃO_3'] + \
df2017_ind_dum['RELIGIÃO_4'] + df2017_ind_dum['RELIGIÃO_5'] + df2017_ind_dum['RELIGIÃO_6'] + \
df2017_ind_dum['RELIGIÃO_7'] + df2017_ind_dum['RELIGIÃO_8'] + df2017_ind_dum['RELIGIÃO_9'] + \
df2017_ind_dum['RELIGIÃO_11'] + df2017_ind_dum['RELIGIÃO_12'] + df2017_ind_dum['RELIGIÃO_13'] + \
df2017_ind_dum['RELIGIÃO_14'] + df2017_ind_dum['RELIGIÃO_15']

df2018_ind_dum['NÃO_CATÓLICA'] = df2018_ind_dum['RELIGIÃO_2'] + df2018_ind_dum['RELIGIÃO_3'] + \
df2018_ind_dum['RELIGIÃO_4'] + df2018_ind_dum['RELIGIÃO_5'] + df2018_ind_dum['RELIGIÃO_6'] + \
df2018_ind_dum['RELIGIÃO_7'] + df2018_ind_dum['RELIGIÃO_8'] + df2018_ind_dum['RELIGIÃO_9'] + \
df2018_ind_dum['RELIGIÃO_11'] + df2018_ind_dum['RELIGIÃO_12'] + df2018_ind_dum['RELIGIÃO_13'] + \
df2018_ind_dum['RELIGIÃO_14'] + df2018_ind_dum['RELIGIÃO_15']

df2019_ind_dum['NÃO_CATÓLICA'] = df2019_ind_dum['RELIGIÃO_2'] + df2019_ind_dum['RELIGIÃO_3'] + \
df2019_ind_dum['RELIGIÃO_4'] + df2019_ind_dum['RELIGIÃO_5'] + df2019_ind_dum['RELIGIÃO_6'] + \
df2019_ind_dum['RELIGIÃO_7'] + df2019_ind_dum['RELIGIÃO_8'] + df2019_ind_dum['RELIGIÃO_9'] + \
df2019_ind_dum['RELIGIÃO_10'] + df2019_ind_dum['RELIGIÃO_11'] + df2019_ind_dum['RELIGIÃO_12'] + \
df2019_ind_dum['RELIGIÃO_13'] + df2019_ind_dum['RELIGIÃO_14'] + df2019_ind_dum['RELIGIÃO_15']

df2015_ind_dum['CATÓLICA'] = df2015_ind_dum['RELIGIÃO_1']
df2016_ind_dum['CATÓLICA'] = df2016_ind_dum['RELIGIÃO_1']
df2017_ind_dum['CATÓLICA'] = df2017_ind_dum['RELIGIÃO_1']
df2018_ind_dum['CATÓLICA'] = df2018_ind_dum['RELIGIÃO_1']
df2019_ind_dum['CATÓLICA'] = df2019_ind_dum['RELIGIÃO_1']

### Remoção de colunas referentes a variáveis dummies não utilizadas
df2015_ind_dum.drop(['SEXO_2', 'PEA_2_2', 'RELIGIÃO_1', 'RELIGIÃO_2', 'RELIGIÃO_3', 'RELIGIÃO_4', \
'RELIGIÃO_5', 'RELIGIÃO_6', 'RELIGIÃO_7', 'RELIGIÃO_8', 'RELIGIÃO_9', 'RELIGIÃO_11', 'RELIGIÃO_12', \

```

```

'RELIGIÃO_13', 'RELIGIÃO_14', 'RELIGIÃO_15', 'RAÇA_1', 'REGIÃO_3', 'ÁREA_2', 'PEA_5'], axis=1, inplace=True)
df2016_ind_dum.drop(['SEXO_2', 'PEA_2_2', 'RELIGIÃO_1', 'RELIGIÃO_2', 'RELIGIÃO_3', 'RELIGIÃO_4',
'RELIGIÃO_5', 'RELIGIÃO_6', 'RELIGIÃO_7', 'RELIGIÃO_8', 'RELIGIÃO_9', 'RELIGIÃO_10', 'RELIGIÃO_11',
'RELIGIÃO_12', 'RELIGIÃO_13', 'RELIGIÃO_14', 'RELIGIÃO_15', 'RAÇA_1', 'REGIÃO_3', 'ÁREA_2', 'PEA_5'], axis=1,
inplace=True)
df2017_ind_dum.drop(['SEXO_2', 'PEA_2_2', 'RELIGIÃO_1', 'RELIGIÃO_2', 'RELIGIÃO_3', 'RELIGIÃO_4',
'RELIGIÃO_5', 'RELIGIÃO_6', 'RELIGIÃO_7', 'RELIGIÃO_8', 'RELIGIÃO_9', 'RELIGIÃO_11', 'RELIGIÃO_12',
'RELIGIÃO_13', 'RELIGIÃO_14', 'RELIGIÃO_15', 'RAÇA_1', 'REGIÃO_3', 'ÁREA_2', 'PEA_5'], axis=1, inplace=True)
df2018_ind_dum.drop(['SEXO_2', 'PEA_2_2', 'RELIGIÃO_1', 'RELIGIÃO_2', 'RELIGIÃO_3', 'RELIGIÃO_4',
'RELIGIÃO_5', 'RELIGIÃO_6', 'RELIGIÃO_7', 'RELIGIÃO_8', 'RELIGIÃO_9', 'RELIGIÃO_11', 'RELIGIÃO_12',
'RELIGIÃO_13', 'RELIGIÃO_14', 'RELIGIÃO_15', 'RAÇA_1', 'REGIÃO_3', 'ÁREA_2', 'PEA_5'], axis=1, inplace=True)
df2019_ind_dum.drop(['SEXO_2', 'PEA_2_2', 'RELIGIÃO_1', 'RELIGIÃO_2', 'RELIGIÃO_3', 'RELIGIÃO_4',
'RELIGIÃO_5', 'RELIGIÃO_6', 'RELIGIÃO_7', 'RELIGIÃO_8', 'RELIGIÃO_9', 'RELIGIÃO_10', 'RELIGIÃO_11',
'RELIGIÃO_12', 'RELIGIÃO_13', 'RELIGIÃO_14', 'RELIGIÃO_15', 'RAÇA_1', 'REGIÃO_3', 'ÁREA_2', 'PEA_5'],
axis=1, inplace=True)

```

Matrix de correlação e Fator de Inflação de Variância (VIF)

```

### Matrix de Correlação
corMatrix2015_ind_dum = pd.DataFrame.corr(df2015_ind_dum[['INTERNET_1', 'CLASSE_ECONÔMICA_1',
'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_3', 'PEA_2_1', 'ÁREA_1', 'IDADE', 'GRAU_INSTRUÇÃO',
'NÃO_CATÓLICA']])
corMatrix2016_ind_dum = pd.DataFrame.corr(df2016_ind_dum[['INTERNET_1', 'CLASSE_ECONÔMICA_1',
'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_3', 'PEA_2_1', 'ÁREA_1', 'IDADE', 'GRAU_INSTRUÇÃO',
'NÃO_CATÓLICA']])
corMatrix2017_ind_dum = pd.DataFrame.corr(df2017_ind_dum[['INTERNET_1', 'CLASSE_ECONÔMICA_1',
'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_3', 'PEA_2_1', 'ÁREA_1', 'IDADE', 'GRAU_INSTRUÇÃO',
'NÃO_CATÓLICA']])
corMatrix2018_ind_dum = pd.DataFrame.corr(df2018_ind_dum[['INTERNET_1', 'CLASSE_ECONÔMICA_1',
'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_3', 'PEA_2_1', 'ÁREA_1', 'IDADE', 'GRAU_INSTRUÇÃO',
'NÃO_CATÓLICA']])
corMatrix2019_ind_dum = pd.DataFrame.corr(df2019_ind_dum[['INTERNET_1', 'CLASSE_ECONÔMICA_1',
'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_3', 'PEA_2_1', 'ÁREA_1', 'IDADE', 'GRAU_INSTRUÇÃO',
'NÃO_CATÓLICA']])

```

```

### Matrix de correlação
fig, ax = plt.subplots(figsize=(25, 15))
colormap = sns.diverging_palette(230, 15, as_cmap = True)
tabel = corMatrix2015_ind_dum
dropvals = np.zeros_like(tabel)
dropvals[np.triu_indices_from(dropvals)] = True
sns.heatmap(tabel, cmap=colormap, linewidths = .5, annot = True, fmt = ".2f", mask = dropvals, annot_kws={'size':12})

```

```

### Cálculo de Fator de Inflação de Variância (Variance Inflation Factor)
# VIF_DUM = ['SEXO_1', 'GRAU_INSTRUÇÃO_2_1', 'GRAU_INSTRUÇÃO_2_2',
# 'GRAU_INSTRUÇÃO_2_3', 'CLASSE_ECONÔMICA_2',
# 'CLASSE_ECONÔMICA_3', 'CLASSE_ECONÔMICA_4', 'PEA_2_1', 'ETNIA_2',
# 'ETNIA_3', 'ETNIA_4', 'ETNIA_5', 'REGIÃO_1', 'REGIÃO_2', 'REGIÃO_4',
# 'REGIÃO_5', 'ÁREA_1', 'PEA_1', 'PEA_2', 'PEA_3', 'PEA_4',
# 'IDADE', 'GRAU_INSTRUÇÃO',
# 'REDA_PESSOAL', 'REDA_FAMILIAR', 'REDA_FAMILIAR_2', 'NÃO_CATÓLICA',
# ]

```

```

VIF_DUM = ['IDADE',
'GRAU_INSTRUÇÃO', 'CLASSE_ECONÔMICA_1', 'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_3',
'PEA_2_1', 'NÃO_CATÓLICA', 'ÁREA_1', 'REDA_FAMILIAR']

```

```

### Validar para cada ano
vif2015_ind_dum = df2015_ind_dum[VIF_DUM]
vif = pd.DataFrame()
vif["VARIÁVEIS"] = vif2015_ind_dum.columns
vif["VIF"] = [variance_inflation_factor(vif2015_ind_dum.values, i) for i in range(0, vif2015_ind_dum.shape[1])]
vif

```

Regressão Logística (Anos individuais)

```

# TABELA EXPERIMENTO
X_2015 = df2015_ind_dum[VIF_DUM].to_numpy()
X_2016 = df2016_ind_dum[VIF_DUM].to_numpy()
X_2017 = df2017_ind_dum[VIF_DUM].to_numpy()

```

```

X_2018 = df2018_ind_dum[VIF_DUM].to_numpy()
X_2019 = df2019_ind_dum[VIF_DUM].to_numpy()

Y_2015 = df2015_ind_dum['INTERNET_1'].to_numpy()
Y_2016 = df2016_ind_dum['INTERNET_1'].to_numpy()
Y_2017 = df2017_ind_dum['INTERNET_1'].to_numpy()
Y_2018 = df2018_ind_dum['INTERNET_1'].to_numpy()
Y_2019 = df2019_ind_dum['INTERNET_1'].to_numpy()

X_2015 = X_2015.astype(np.int64)
X_2016 = X_2016.astype(np.int64)
X_2017 = X_2017.astype(np.int64)
X_2018 = X_2018.astype(np.int64)
X_2019 = X_2019.astype(np.int64)

X_2015 = sm.add_constant(X_2015)
X_2016 = sm.add_constant(X_2016)
X_2017 = sm.add_constant(X_2017)
X_2018 = sm.add_constant(X_2018)
X_2019 = sm.add_constant(X_2019)

Y_2015 = Y_2015.astype(np.int64)
Y_2016 = Y_2016.astype(np.int64)
Y_2017 = Y_2017.astype(np.int64)
Y_2018 = Y_2018.astype(np.int64)
Y_2019 = Y_2019.astype(np.int64)

X_train_2015, X_test_2015, y_train_2015, y_test_2015 = train_test_split(X_2015, Y_2015, train_size=0.80, \
    random_state=0)
X_train_2016, X_test_2016, y_train_2016, y_test_2016 = train_test_split(X_2016, Y_2016, train_size=0.80, \
    random_state=0)
X_train_2017, X_test_2017, y_train_2017, y_test_2017 = train_test_split(X_2017, Y_2017, train_size=0.80, \
    random_state=0)
X_train_2018, X_test_2018, y_train_2018, y_test_2018 = train_test_split(X_2018, Y_2018, train_size=0.80, \
    random_state=0)
X_train_2019, X_test_2019, y_train_2019, y_test_2019 = train_test_split(X_2019, Y_2019, train_size=0.80, \
    random_state=0)

### Cálculo dos modelos
model_2015 = sm.Logit(Y_2015, X_2015).fit()
model_2016 = sm.Logit(Y_2016, X_2016).fit()
model_2017 = sm.Logit(Y_2017, X_2017).fit()
model_2018 = sm.Logit(Y_2018, X_2018).fit()
model_2019 = sm.Logit(Y_2019, X_2019).fit()

model_2015_probit = sm.Probit(Y_2015, X_2015).fit()
model_2016_probit = sm.Probit(Y_2016, X_2016).fit()
model_2017_probit = sm.Probit(Y_2017, X_2017).fit()
model_2018_probit = sm.Probit(Y_2018, X_2018).fit()
model_2019_probit = sm.Probit(Y_2019, X_2019).fit()

link_logit = sm.families.links.logit
family_logit = sm.families.Binomial(link_logit)

link_probit = sm.families.links.probit
family_probit = sm.families.Binomial(link_probit)

link_cloglog = sm.families.links.cloglog
family_cloglog = sm.families.Binomial(link_cloglog)

model_2015_GLM_logit = sm.GLM(Y_2015, X_2015, family_logit).fit()
model_2016_GLM_logit = sm.GLM(Y_2016, X_2016, family_logit).fit()
model_2017_GLM_logit = sm.GLM(Y_2017, X_2017, family_logit).fit()
model_2018_GLM_logit = sm.GLM(Y_2018, X_2018, family_logit).fit()
model_2019_GLM_logit = sm.GLM(Y_2019, X_2019, family_logit).fit()

model_2015_GLM_probit = sm.GLM(Y_2015, X_2015, family_probit).fit()

```



```

model_2016_GLM_probit = sm.GLM(Y_2016, X_2016,family_probit).fit()
model_2017_GLM_probit = sm.GLM(Y_2017, X_2017,family_probit).fit()
model_2018_GLM_probit = sm.GLM(Y_2018, X_2018,family_probit).fit()
model_2019_GLM_probit = sm.GLM(Y_2019, X_2019,family_probit).fit()

model_2015_GLM_cloglog = sm.GLM(Y_2015, X_2015,family_cloglog).fit()
model_2016_GLM_cloglog = sm.GLM(Y_2016, X_2016,family_cloglog).fit()
model_2017_GLM_cloglog = sm.GLM(Y_2017, X_2017,family_cloglog).fit()
model_2018_GLM_cloglog = sm.GLM(Y_2018, X_2018,family_cloglog).fit()
model_2019_GLM_cloglog = sm.GLM(Y_2019, X_2019,family_cloglog).fit()

#### Impressão de tabela de regressão
VIF_NOME_2015 = [
    'Intercept', 'IDADE',\
    'GRAU_INSTRUÇÃO', 'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_4',\
    'RENDA_FAMILIAR', 'PEA_2_1', 'NÃO_CATÓLICA',\
    'ÁREA_1', 'RENDA_FAMILIAR'
]

VIF_NOME_2016 = [
    'Intercept', 'IDADE',\
    'GRAU_INSTRUÇÃO', 'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_4',\
    'RENDA_FAMILIAR', 'PEA_2_1', 'NÃO_CATÓLICA',\
    'ÁREA_1', 'RENDA_FAMILIAR'
]

VIF_NOME_2017 = [
    'Intercept', 'IDADE',\
    'GRAU_INSTRUÇÃO', 'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_4',\
    'RENDA_FAMILIAR', 'PEA_2_1', 'NÃO_CATÓLICA',\
    'ÁREA_1', 'RENDA_FAMILIAR'
]

VIF_NOME_2018 = [
    'Intercept', 'IDADE',\
    'GRAU_INSTRUÇÃO', 'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_4',\
    'RENDA_FAMILIAR', 'PEA_2_1', 'NÃO_CATÓLICA',\
    'ÁREA_1', 'RENDA_FAMILIAR'
]

VIF_NOME_2019 = [
    'Intercept', 'IDADE',\
    'GRAU_INSTRUÇÃO', 'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_4',\
    'RENDA_FAMILIAR', 'PEA_2_1', 'NÃO_CATÓLICA',\
    'ÁREA_1', 'RENDA_FAMILIAR'
]

print(model_2015.summary(xname=VIF_NOME_2015))

# Curva Sigmoidal
fig, ax = plt.subplots(figsize=(22, 10), facecolor='#F0F0F0')

plt.scatter(np.arange(1, len(df2015_ind_dum['INTERNET_1']) + 1, 1).reshape(-1,1),
df2015_ind_dum['INTERNET_1'].sort_values(ascending=True), color='red')
#plt.scatter(np.arange(1, len(pd.Series(model_2015.predict(X_2015)))) + 1, 1).reshape(-1,1),
pd.Series(model_2015.predict(X_2015)).sort_values(ascending=True))

sns.regplot(np.arange(1, len(pd.Series(model_2015.predict(X_2015)))) + 1, 1).reshape(-1,1),
pd.Series(model_2015.predict(X_2015)).sort_values(ascending=True), logistic=True, ci=95)

#### Calculando Odds Ratio com Intervalo de Confiança
CI_2015 = model_2015.conf_int()
CI_2016 = model_2016.conf_int()
CI_2017 = model_2017.conf_int()
CI_2018 = model_2018.conf_int()
CI_2019 = model_2019.conf_int()

```

```
Odds_Ratio_2015 = [(np.exp(CI_2015[i][0]).round(4), np.exp(model_2015.params[i]).round(4),\
np.exp(CI_2015[i][1]).round(4)) for i in range(0, len(CI_2015))]
Odds_Ratio_2016 = [(np.exp(CI_2016[i][0]).round(4), np.exp(model_2016.params[i]).round(4),\
np.exp(CI_2016[i][1]).round(4)) for i in range(0, len(CI_2016))]
Odds_Ratio_2017 = [(np.exp(CI_2017[i][0]).round(4), np.exp(model_2017.params[i]).round(4),\
np.exp(CI_2017[i][1]).round(4)) for i in range(0, len(CI_2017))]
Odds_Ratio_2018 = [(np.exp(CI_2018[i][0]).round(4), np.exp(model_2018.params[i]).round(4),\
np.exp(CI_2018[i][1]).round(4)) for i in range(0, len(CI_2018))]
Odds_Ratio_2019 = [(np.exp(CI_2019[i][0]).round(4), np.exp(model_2019.params[i]).round(4),\
np.exp(CI_2019[i][1]).round(4)) for i in range(0, len(CI_2019))]
```

```
Odds_Ratio_2015 = pd.DataFrame(Odds_Ratio_2015, index=VIF_NOME_2015, columns=['0.025','Odds_Ratio','0.975'])
Odds_Ratio_2016 = pd.DataFrame(Odds_Ratio_2016, index=VIF_NOME_2016, columns=['0.025','Odds_Ratio','0.975'])
Odds_Ratio_2017 = pd.DataFrame(Odds_Ratio_2017, index=VIF_NOME_2017, columns=['0.025','Odds_Ratio','0.975'])
Odds_Ratio_2018 = pd.DataFrame(Odds_Ratio_2018, index=VIF_NOME_2018, columns=['0.025','Odds_Ratio','0.975'])
Odds_Ratio_2019 = pd.DataFrame(Odds_Ratio_2019, index=VIF_NOME_2019, columns=['0.025','Odds_Ratio','0.975'])
```

Regressão Logística para os 05 anos

CRIAÇÃO DE NOVAS TABELAS

```
df2015_ind_reg = df2015_ind_dum[VIF_DUM]
df2016_ind_reg = df2016_ind_dum[VIF_DUM]
df2017_ind_reg = df2017_ind_dum[VIF_DUM]
df2018_ind_reg = df2018_ind_dum[VIF_DUM]
df2019_ind_reg = df2019_ind_dum[VIF_DUM]
```

```
df2015_ind_reg.insert(loc=0, column='IDADE2', value=df2015_ind_reg['IDADE'] ** 2)
df2016_ind_reg.insert(loc=0, column='IDADE2', value=df2016_ind_reg['IDADE'] ** 2)
df2017_ind_reg.insert(loc=0, column='IDADE2', value=df2017_ind_reg['IDADE'] ** 2)
df2018_ind_reg.insert(loc=0, column='IDADE2', value=df2018_ind_reg['IDADE'] ** 2)
df2019_ind_reg.insert(loc=0, column='IDADE2', value=df2019_ind_reg['IDADE'] ** 2)
```

```
df2015_ind_reg.insert(loc=0, column='GRAU_INSTRUÇÃO2', value=df2015_ind_reg['GRAU_INSTRUÇÃO'] ** 2)
df2016_ind_reg.insert(loc=0, column='GRAU_INSTRUÇÃO2', value=df2016_ind_reg['GRAU_INSTRUÇÃO'] ** 2)
df2017_ind_reg.insert(loc=0, column='GRAU_INSTRUÇÃO2', value=df2017_ind_reg['GRAU_INSTRUÇÃO'] ** 2)
df2018_ind_reg.insert(loc=0, column='GRAU_INSTRUÇÃO2', value=df2018_ind_reg['GRAU_INSTRUÇÃO'] ** 2)
df2019_ind_reg.insert(loc=0, column='GRAU_INSTRUÇÃO2', value=df2019_ind_reg['GRAU_INSTRUÇÃO'] ** 2)
```

CORREÇÃO DE VALORES DE ACORDO COM IPCA

```
df2015_ind_reg.loc[(df2015_ind_reg['RENDIA_FAMILIAR'] == 394.000), 'RENDIA_FAMILIAR'] = 0.457524
df2015_ind_reg.loc[(df2015_ind_reg['RENDIA_FAMILIAR'] == 1182.005), 'RENDIA_FAMILIAR'] = 1.372577
df2015_ind_reg.loc[(df2015_ind_reg['RENDIA_FAMILIAR'] == 1970.005), 'RENDIA_FAMILIAR'] = 2.287624
df2015_ind_reg.loc[(df2015_ind_reg['RENDIA_FAMILIAR'] == 3152.005), 'RENDIA_FAMILIAR'] = 3.660195
df2015_ind_reg.loc[(df2015_ind_reg['RENDIA_FAMILIAR'] == 5910.005), 'RENDIA_FAMILIAR'] = 6.862861
df2015_ind_reg.loc[(df2015_ind_reg['RENDIA_FAMILIAR'] == 11820.005), 'RENDIA_FAMILIAR'] = 13.725716
df2015_ind_reg.loc[(df2015_ind_reg['RENDIA_FAMILIAR'] == 19700.005), 'RENDIA_FAMILIAR'] = 22.876190
df2015_ind_reg.loc[(df2015_ind_reg['RENDIA_FAMILIAR'] == 23640.000), 'RENDIA_FAMILIAR'] = 41.408018
```

```
df2016_ind_reg.loc[(df2016_ind_reg['RENDIA_FAMILIAR'] == 440.000), 'RENDIA_FAMILIAR'] = 0.487730
df2016_ind_reg.loc[(df2016_ind_reg['RENDIA_FAMILIAR'] == 1320.005), 'RENDIA_FAMILIAR'] = 1.463196
df2016_ind_reg.loc[(df2016_ind_reg['RENDIA_FAMILIAR'] == 2200.005), 'RENDIA_FAMILIAR'] = 2.438656
df2016_ind_reg.loc[(df2016_ind_reg['RENDIA_FAMILIAR'] == 3520.005), 'RENDIA_FAMILIAR'] = 3.901846
df2016_ind_reg.loc[(df2016_ind_reg['RENDIA_FAMILIAR'] == 6640.005), 'RENDIA_FAMILIAR'] = 7.360296
df2016_ind_reg.loc[(df2016_ind_reg['RENDIA_FAMILIAR'] == 13240.005), 'RENDIA_FAMILIAR'] = 14.676248
df2016_ind_reg.loc[(df2016_ind_reg['RENDIA_FAMILIAR'] == 22000.005), 'RENDIA_FAMILIAR'] = 24.386511
df2016_ind_reg.loc[(df2016_ind_reg['RENDIA_FAMILIAR'] == 26400.000), 'RENDIA_FAMILIAR'] = 44.141839
```

```
df2017_ind_reg.loc[(df2017_ind_reg['RENDIA_FAMILIAR'] == 468.500), 'RENDIA_FAMILIAR'] = 0.506573
df2017_ind_reg.loc[(df2017_ind_reg['RENDIA_FAMILIAR'] == 1405.505), 'RENDIA_FAMILIAR'] = 1.519723
df2017_ind_reg.loc[(df2017_ind_reg['RENDIA_FAMILIAR'] == 2342.505), 'RENDIA_FAMILIAR'] = 2.532868
df2017_ind_reg.loc[(df2017_ind_reg['RENDIA_FAMILIAR'] == 3748.005), 'RENDIA_FAMILIAR'] = 4.052586
df2017_ind_reg.loc[(df2017_ind_reg['RENDIA_FAMILIAR'] == 7027.505), 'RENDIA_FAMILIAR'] = 7.598594
df2017_ind_reg.loc[(df2017_ind_reg['RENDIA_FAMILIAR'] == 14055.005), 'RENDIA_FAMILIAR'] = 15.197182
df2017_ind_reg.loc[(df2017_ind_reg['RENDIA_FAMILIAR'] == 23425.005), 'RENDIA_FAMILIAR'] = 25.328633
df2017_ind_reg.loc[(df2017_ind_reg['RENDIA_FAMILIAR'] == 28110.000), 'RENDIA_FAMILIAR'] = 45.847169
```

```
df2018_ind_reg.loc[(df2018_ind_reg['RENDIA_FAMILIAR'] == 477.000), 'RENDIA_FAMILIAR'] = 0.498585
```

```

df2018_ind_reg.loc[(df2018_ind_reg['REND_A_FAMILIAR'] == 1431.005), 'REND_A_FAMILIAR'] = 1.495759
df2018_ind_reg.loc[(df2018_ind_reg['REND_A_FAMILIAR'] == 2385.005), 'REND_A_FAMILIAR'] = 2.492928
df2018_ind_reg.loc[(df2018_ind_reg['REND_A_FAMILIAR'] == 3816.005), 'REND_A_FAMILIAR'] = 3.988682
df2018_ind_reg.loc[(df2018_ind_reg['REND_A_FAMILIAR'] == 7155.005), 'REND_A_FAMILIAR'] = 7.478775
df2018_ind_reg.loc[(df2018_ind_reg['REND_A_FAMILIAR'] == 14310.005), 'REND_A_FAMILIAR'] = 14.957544
df2018_ind_reg.loc[(df2018_ind_reg['REND_A_FAMILIAR'] == 23850.005), 'REND_A_FAMILIAR'] = 24.929237
df2018_ind_reg.loc[(df2018_ind_reg['REND_A_FAMILIAR'] == 28620.000), 'REND_A_FAMILIAR'] = 45.124225

df2019_ind_reg.loc[(df2019_ind_reg['REND_A_FAMILIAR'] == 499.000), 'REND_A_FAMILIAR'] = 0.499000
df2019_ind_reg.loc[(df2019_ind_reg['REND_A_FAMILIAR'] == 1497.005), 'REND_A_FAMILIAR'] = 1.497005
df2019_ind_reg.loc[(df2019_ind_reg['REND_A_FAMILIAR'] == 2495.005), 'REND_A_FAMILIAR'] = 2.495005
df2019_ind_reg.loc[(df2019_ind_reg['REND_A_FAMILIAR'] == 3992.005), 'REND_A_FAMILIAR'] = 3.992005
df2019_ind_reg.loc[(df2019_ind_reg['REND_A_FAMILIAR'] == 7485.005), 'REND_A_FAMILIAR'] = 7.485005
df2019_ind_reg.loc[(df2019_ind_reg['REND_A_FAMILIAR'] == 14970.005), 'REND_A_FAMILIAR'] = 14.970005
df2019_ind_reg.loc[(df2019_ind_reg['REND_A_FAMILIAR'] == 24950.005), 'REND_A_FAMILIAR'] = 24.950005
df2019_ind_reg.loc[(df2019_ind_reg['REND_A_FAMILIAR'] == 29940.000), 'REND_A_FAMILIAR'] = 45.161818

df2015_ind_reg.insert(loc=0, column='REND_A_FAMILIAR2', value=df2015_ind_reg['REND_A_FAMILIAR'] ** 2)
df2016_ind_reg.insert(loc=0, column='REND_A_FAMILIAR2', value=df2016_ind_reg['REND_A_FAMILIAR'] ** 2)
df2017_ind_reg.insert(loc=0, column='REND_A_FAMILIAR2', value=df2017_ind_reg['REND_A_FAMILIAR'] ** 2)
df2018_ind_reg.insert(loc=0, column='REND_A_FAMILIAR2', value=df2018_ind_reg['REND_A_FAMILIAR'] ** 2)
df2019_ind_reg.insert(loc=0, column='REND_A_FAMILIAR2', value=df2019_ind_reg['REND_A_FAMILIAR'] ** 2)

##### CRIAÇÃO DE COLUNA ANO
df2015_ind_reg['ANO'] = 0
df2016_ind_reg['ANO'] = 1
df2017_ind_reg['ANO'] = 2
df2018_ind_reg['ANO'] = 3
df2019_ind_reg['ANO'] = 4

df2015_ind_reg['ANO_IDADE2'] = df2015_ind_reg['IDADE2'] * df2015_ind_reg['ANO']
df2016_ind_reg['ANO_IDADE2'] = df2016_ind_reg['IDADE2'] * df2016_ind_reg['ANO']
df2017_ind_reg['ANO_IDADE2'] = df2017_ind_reg['IDADE2'] * df2017_ind_reg['ANO']
df2018_ind_reg['ANO_IDADE2'] = df2018_ind_reg['IDADE2'] * df2018_ind_reg['ANO']
df2019_ind_reg['ANO_IDADE2'] = df2019_ind_reg['IDADE2'] * df2019_ind_reg['ANO']

df2015_ind_reg['ANO_IDADE'] = df2015_ind_reg['IDADE'] * df2015_ind_reg['ANO']
df2016_ind_reg['ANO_IDADE'] = df2016_ind_reg['IDADE'] * df2016_ind_reg['ANO']
df2017_ind_reg['ANO_IDADE'] = df2017_ind_reg['IDADE'] * df2017_ind_reg['ANO']
df2018_ind_reg['ANO_IDADE'] = df2018_ind_reg['IDADE'] * df2018_ind_reg['ANO']
df2019_ind_reg['ANO_IDADE'] = df2019_ind_reg['IDADE'] * df2019_ind_reg['ANO']

df2015_ind_reg['ANO_GRAU_INSTRUÇÃO'] = df2015_ind_reg['GRAU_INSTRUÇÃO'] * df2015_ind_reg['ANO']
df2016_ind_reg['ANO_GRAU_INSTRUÇÃO'] = df2016_ind_reg['GRAU_INSTRUÇÃO'] * df2016_ind_reg['ANO']
df2017_ind_reg['ANO_GRAU_INSTRUÇÃO'] = df2017_ind_reg['GRAU_INSTRUÇÃO'] * df2017_ind_reg['ANO']
df2018_ind_reg['ANO_GRAU_INSTRUÇÃO'] = df2018_ind_reg['GRAU_INSTRUÇÃO'] * df2018_ind_reg['ANO']
df2019_ind_reg['ANO_GRAU_INSTRUÇÃO'] = df2019_ind_reg['GRAU_INSTRUÇÃO'] * df2019_ind_reg['ANO']

df2015_ind_reg['ANO_GRAU_INSTRUÇÃO2'] = df2015_ind_reg['GRAU_INSTRUÇÃO2'] * df2015_ind_reg['ANO']
df2016_ind_reg['ANO_GRAU_INSTRUÇÃO2'] = df2016_ind_reg['GRAU_INSTRUÇÃO2'] * df2016_ind_reg['ANO']
df2017_ind_reg['ANO_GRAU_INSTRUÇÃO2'] = df2017_ind_reg['GRAU_INSTRUÇÃO2'] * df2017_ind_reg['ANO']
df2018_ind_reg['ANO_GRAU_INSTRUÇÃO2'] = df2018_ind_reg['GRAU_INSTRUÇÃO2'] * df2018_ind_reg['ANO']
df2019_ind_reg['ANO_GRAU_INSTRUÇÃO2'] = df2019_ind_reg['GRAU_INSTRUÇÃO2'] * df2019_ind_reg['ANO']

df2015_ind_reg['ANO_CLASSE_ECONÔMICA_1'] = df2015_ind_reg['CLASSE_ECONÔMICA_1'] * \
df2015_ind_reg['ANO']
df2016_ind_reg['ANO_CLASSE_ECONÔMICA_1'] = df2016_ind_reg['CLASSE_ECONÔMICA_1'] * \
df2016_ind_reg['ANO']
df2017_ind_reg['ANO_CLASSE_ECONÔMICA_1'] = df2017_ind_reg['CLASSE_ECONÔMICA_1'] * \
df2017_ind_reg['ANO']
df2018_ind_reg['ANO_CLASSE_ECONÔMICA_1'] = df2018_ind_reg['CLASSE_ECONÔMICA_1'] * \
df2018_ind_reg['ANO']
df2019_ind_reg['ANO_CLASSE_ECONÔMICA_1'] = df2019_ind_reg['CLASSE_ECONÔMICA_1'] * \
df2019_ind_reg['ANO']

df2015_ind_reg['ANO_CLASSE_ECONÔMICA_2'] = df2015_ind_reg['CLASSE_ECONÔMICA_2'] * \
df2015_ind_reg['ANO']

```

```

df2016_ind_reg['ANO_CLASSE_ECONÔMICA_2'] = df2016_ind_reg['CLASSE_ECONÔMICA_2'] * \
    df2016_ind_reg['ANO']
df2017_ind_reg['ANO_CLASSE_ECONÔMICA_2'] = df2017_ind_reg['CLASSE_ECONÔMICA_2'] * \
    df2017_ind_reg['ANO']
df2018_ind_reg['ANO_CLASSE_ECONÔMICA_2'] = df2018_ind_reg['CLASSE_ECONÔMICA_2'] * \
    df2018_ind_reg['ANO']
df2019_ind_reg['ANO_CLASSE_ECONÔMICA_2'] = df2019_ind_reg['CLASSE_ECONÔMICA_2'] * \
    df2019_ind_reg['ANO']

df2015_ind_reg['ANO_CLASSE_ECONÔMICA_3'] = df2015_ind_reg['CLASSE_ECONÔMICA_3'] * \
    df2015_ind_reg['ANO']
df2016_ind_reg['ANO_CLASSE_ECONÔMICA_3'] = df2016_ind_reg['CLASSE_ECONÔMICA_3'] * \
    df2016_ind_reg['ANO']
df2017_ind_reg['ANO_CLASSE_ECONÔMICA_3'] = df2017_ind_reg['CLASSE_ECONÔMICA_3'] * \
    df2017_ind_reg['ANO']
df2018_ind_reg['ANO_CLASSE_ECONÔMICA_3'] = df2018_ind_reg['CLASSE_ECONÔMICA_3'] * \
    df2018_ind_reg['ANO']
df2019_ind_reg['ANO_CLASSE_ECONÔMICA_3'] = df2019_ind_reg['CLASSE_ECONÔMICA_3'] * \
    df2019_ind_reg['ANO']

df2015_ind_reg['ANO_RENDA_FAMILIAR'] = df2015_ind_reg['RENDA_FAMILIAR'] * df2015_ind_reg['ANO']
df2016_ind_reg['ANO_RENDA_FAMILIAR'] = df2016_ind_reg['RENDA_FAMILIAR'] * df2016_ind_reg['ANO']
df2017_ind_reg['ANO_RENDA_FAMILIAR'] = df2017_ind_reg['RENDA_FAMILIAR'] * df2017_ind_reg['ANO']
df2018_ind_reg['ANO_RENDA_FAMILIAR'] = df2018_ind_reg['RENDA_FAMILIAR'] * df2018_ind_reg['ANO']
df2019_ind_reg['ANO_RENDA_FAMILIAR'] = df2019_ind_reg['RENDA_FAMILIAR'] * df2019_ind_reg['ANO']

df2015_ind_reg['ANO_RENDA_FAMILIAR2'] = df2015_ind_reg['RENDA_FAMILIAR2'] * df2015_ind_reg['ANO']
df2016_ind_reg['ANO_RENDA_FAMILIAR2'] = df2016_ind_reg['RENDA_FAMILIAR2'] * df2016_ind_reg['ANO']
df2017_ind_reg['ANO_RENDA_FAMILIAR2'] = df2017_ind_reg['RENDA_FAMILIAR2'] * df2017_ind_reg['ANO']
df2018_ind_reg['ANO_RENDA_FAMILIAR2'] = df2018_ind_reg['RENDA_FAMILIAR2'] * df2018_ind_reg['ANO']
df2019_ind_reg['ANO_RENDA_FAMILIAR2'] = df2019_ind_reg['RENDA_FAMILIAR2'] * df2019_ind_reg['ANO']

df2015_ind_reg['ANO_PEA_2_1'] = df2015_ind_reg['PEA_2_1'] * df2015_ind_reg['ANO']
df2016_ind_reg['ANO_PEA_2_1'] = df2016_ind_reg['PEA_2_1'] * df2016_ind_reg['ANO']
df2017_ind_reg['ANO_PEA_2_1'] = df2017_ind_reg['PEA_2_1'] * df2017_ind_reg['ANO']
df2018_ind_reg['ANO_PEA_2_1'] = df2018_ind_reg['PEA_2_1'] * df2018_ind_reg['ANO']
df2019_ind_reg['ANO_PEA_2_1'] = df2019_ind_reg['PEA_2_1'] * df2019_ind_reg['ANO']

df2015_ind_reg['ANO_NÃO_CATÓLICA'] = df2015_ind_reg['NÃO_CATÓLICA'] * df2015_ind_reg['ANO']
df2016_ind_reg['ANO_NÃO_CATÓLICA'] = df2016_ind_reg['NÃO_CATÓLICA'] * df2016_ind_reg['ANO']
df2017_ind_reg['ANO_NÃO_CATÓLICA'] = df2017_ind_reg['NÃO_CATÓLICA'] * df2017_ind_reg['ANO']
df2018_ind_reg['ANO_NÃO_CATÓLICA'] = df2018_ind_reg['NÃO_CATÓLICA'] * df2018_ind_reg['ANO']
df2019_ind_reg['ANO_NÃO_CATÓLICA'] = df2019_ind_reg['NÃO_CATÓLICA'] * df2019_ind_reg['ANO']

df2015_ind_reg['ANO_ÁREA_1'] = df2015_ind_reg['ÁREA_1'] * df2015_ind_reg['ANO']
df2016_ind_reg['ANO_ÁREA_1'] = df2016_ind_reg['ÁREA_1'] * df2016_ind_reg['ANO']
df2017_ind_reg['ANO_ÁREA_1'] = df2017_ind_reg['ÁREA_1'] * df2017_ind_reg['ANO']
df2018_ind_reg['ANO_ÁREA_1'] = df2018_ind_reg['ÁREA_1'] * df2018_ind_reg['ANO']
df2019_ind_reg['ANO_ÁREA_1'] = df2019_ind_reg['ÁREA_1'] * df2019_ind_reg['ANO']

df2015_ind_reg['INTERNET_1'] = df2015_ind_dum['INTERNET_1']
df2016_ind_reg['INTERNET_1'] = df2016_ind_dum['INTERNET_1']
df2017_ind_reg['INTERNET_1'] = df2017_ind_dum['INTERNET_1']
df2018_ind_reg['INTERNET_1'] = df2018_ind_dum['INTERNET_1']
df2019_ind_reg['INTERNET_1'] = df2019_ind_dum['INTERNET_1']

### Só para ter as regressões por ano com as alterações.... Não usar ANO porque a regressão é para cada ano.
REG = ['IDADE', 'GRAU_INSTRUÇÃO', 'CLASSE_ECONÔMICA_1', 'CLASSE_ECONÔMICA_2', \
    'CLASSE_ECONÔMICA_3', 'RENDA_FAMILIAR', 'PEA_2_1', 'NÃO_CATÓLICA', 'ÁREA_1']

X_2015_reg = df2015_ind_reg[REG].to_numpy()
X_2016_reg = df2016_ind_reg[REG].to_numpy()
X_2017_reg = df2017_ind_reg[REG].to_numpy()
X_2018_reg = df2018_ind_reg[REG].to_numpy()
X_2019_reg = df2019_ind_reg[REG].to_numpy()

Y_2015_reg = df2015_ind_reg['INTERNET_1'].to_numpy()

```

```

Y_2016_reg = df2016_ind_reg['INTERNET_1'].to_numpy()
Y_2017_reg = df2017_ind_reg['INTERNET_1'].to_numpy()
Y_2018_reg = df2018_ind_reg['INTERNET_1'].to_numpy()
Y_2019_reg = df2019_ind_reg['INTERNET_1'].to_numpy()

#### Se tirar roda a regressão e melhora o Pseudo R^2
X_2015_reg = X_2015_reg.astype(np.int64)
X_2016_reg = X_2016_reg.astype(np.int64)
X_2017_reg = X_2017_reg.astype(np.int64)
X_2018_reg = X_2018_reg.astype(np.int64)
X_2019_reg = X_2019_reg.astype(np.int64)

X_2015_reg = sm.add_constant(X_2015_reg)
X_2016_reg = sm.add_constant(X_2016_reg)
X_2017_reg = sm.add_constant(X_2017_reg)
X_2018_reg = sm.add_constant(X_2018_reg)
X_2019_reg = sm.add_constant(X_2019_reg)

Y_2015_reg = Y_2015_reg.astype(np.int64)
Y_2016_reg = Y_2016_reg.astype(np.int64)
Y_2017_reg = Y_2017_reg.astype(np.int64)
Y_2018_reg = Y_2018_reg.astype(np.int64)
Y_2019_reg = Y_2019_reg.astype(np.int64)

model_2015_reg = sm.Logit(Y_2015_reg, X_2015_reg).fit()
model_2016_reg = sm.Logit(Y_2016_reg, X_2016_reg).fit()
model_2017_reg = sm.Logit(Y_2017_reg, X_2017_reg).fit()
model_2018_reg = sm.Logit(Y_2018_reg, X_2018_reg).fit()
model_2019_reg = sm.Logit(Y_2019_reg, X_2019_reg).fit()

#### REGRESSÃO COM VARIÁVEIS QUE APRESENTARAM RELEVÂNCIA ESTATÍSTICA.
REG = ['ANO', 'IDADE2', 'IDADE', 'GRAU_INSTRUÇÃO2', 'GRAU_INSTRUÇÃO', 'CLASSE_ECONÔMICA_1',
       'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_3', 'PEA_2_1', 'NÃO_CATÓLICA', 'ÁREA_1',
       'ANO_IDADE2', 'ANO_IDADE', 'ANO_GRAU_INSTRUÇÃO2', 'ANO_CLASSE_ECONÔMICA_2',
       'ANO_CLASSE_ECONÔMICA_3', 'ANO_PEA_2_1', 'ANO_NÃO_CATÓLICA', 'ANO_ÁREA_1']
df_ind_reg = df2015_ind_reg
df_ind_reg = df_ind_reg.append([df2016_ind_reg, df2017_ind_reg, df2018_ind_reg, df2019_ind_reg], ignore_index = True)

X_reg = df_ind_reg[REG].to_numpy()
Y_reg = df_ind_reg['INTERNET_1'].to_numpy()

#### Se tirar roda a regressão e melhora o Pseudo R^2
#X_reg = X_reg.astype(np.int64)
X_reg = sm.add_constant(X_reg)
Y_reg = Y_reg.astype(np.int64)

model_reg = sm.Logit(Y_reg, X_reg).fit()
print(model_reg.summary(xname=['Intercept', 'ANO', 'IDADE2', 'IDADE', 'GRAU_INSTRUÇÃO2', 'GRAU_INSTRUÇÃO',
                               'CLASSE_ECONÔMICA_1', 'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_3', 'PEA_2_1', 'NÃO_CATÓLICA',
                               'ÁREA_1', 'ANO_IDADE2', 'ANO_IDADE', 'ANO_GRAU_INSTRUÇÃO2', 'ANO_CLASSE_ECONÔMICA_2',
                               'ANO_CLASSE_ECONÔMICA_3', 'ANO_PEA_2_1', 'ANO_NÃO_CATÓLICA', 'ANO_ÁREA_1']))

CI = model_reg.conf_int()
Odds_Ratio = [(np.exp(CI[i][0]).round(4), np.exp(model_reg.params[i]).round(4), np.exp(CI[i][1]).round(4)) for i in
               range(0, len(CI))]
Odds_Ratio = pd.DataFrame(Odds_Ratio, index=REG, columns=['0.025', 'Odds_Ratio', '0.975'])

#### REGRESSÃO COM VARIÁVEIS QUE APRESENTARAM RELEVÂNCIA ESTATÍSTICA (OUTRA
BIBLIOTECA).
model_reg_2 = smf.logit(formula='INTERNET_1 ~ ANO + I(IDADE**2) + IDADE + I(GRAU_INSTRUÇÃO**2) +\
    GRAU_INSTRUÇÃO + CLASSE_ECONÔMICA_1 + CLASSE_ECONÔMICA_2 + CLASSE_ECONÔMICA_3 +\
    PEA_2_1 + NÃO_CATÓLICA + ÁREA_1 + ANO:I(IDADE**2) + ANO:IDADE + ANO:I(GRAU_INSTRUÇÃO**2) +\
    ANO:CLASSE_ECONÔMICA_2 + ANO:CLASSE_ECONÔMICA_3 + ANO:CLASSE_ECONÔMICA_3 +\
    ANO:PEA_2_1 + ANO:NÃO_CATÓLICA + ANO:ÁREA_1 + 1', data=df_ind_reg).fit()
print(model_reg_2.summary())

CI = model_reg_2.conf_int()

```

```

Odds_Ratio = [(np.exp(CI[0][i]).round(4), np.exp(model_reg_2.params[i]).round(4), np.exp(CI[1][i]).round(4)) for i in\
range(0, len(CI))]
Odds_Ratio = pd.DataFrame(Odds_Ratio, index=model_reg_2.params.index, columns=['0.025','Odds_Ratio','0.975'])

### MATRIX DE CORRELAÇÃO
corMatrix_ind_dum = pd.DataFrame.corr(df_ind_reg[['INTERNET_1', 'IDADE', 'GRAU_INSTRUÇÃO',\
'CLASSE_ECONÔMICA_1', 'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_3', 'PEA_2_1', 'NÃO_CATÓLICA',\
'ÁREA_1']])
corMatrix_ind_dum

### Cálculo de Fator de Inflação de Variância (Variance Inflation Factor)

VIF_DUM = ['IDADE', 'GRAU_INSTRUÇÃO', 'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_3',\
'CLASSE_ECONÔMICA_4', 'PEA_2_1', 'NÃO_CATÓLICA', 'ÁREA_1']

vif_ind_dum = df_ind_reg[VIF_DUM]
vif = pd.DataFrame()
vif["VARIÁVEIS"] = vif_ind_dum.columns
vif["VIF"] = [variance_inflation_factor(vif_ind_dum.values, i) for i in range(0, vif_ind_dum.shape[1])]
vif

### INFORMAÇÕES DO MODELO DE REGRESSÃO ÚNICA PARA 05 ANOS
Confusion_Matrix2 = model_reg.pred_table()
df_Confusion_Matrix2 = pd.DataFrame(Confusion_Matrix2, columns=['Positive', 'Negative'], index=['Positive', 'Negative'])
Confusion_Matrix2

#### Campos da Matrix são:
### [Verdadeiro Negativo, Falso Positivo]
### [Falso Negativo, Verdadeiro Positivo]

Verdadeiro_Positivo = Confusion_Matrix2[1,1]
Verdadeiro_Negativo = Confusion_Matrix2[0,0]
Falso_Positivo = Confusion_Matrix2[0,1]
Falso_Negativo = Confusion_Matrix2[1,0]
Número_Predições = Verdadeiro_Positivo + Verdadeiro_Negativo + Falso_Positivo + Falso_Negativo

Acuracidade = (Verdadeiro_Positivo + Verdadeiro_Negativo)/Número_Predições
Erro = (Falso_Positivo + Falso_Negativo)/Número_Predições
Sensibilidade = Verdadeiro_Positivo / (Verdadeiro_Positivo + Falso_Negativo)
Especificidade = Verdadeiro_Negativo / (Verdadeiro_Negativo + Falso_Positivo)
Confiabilidade_Positiva = Verdadeiro_Positivo / (Verdadeiro_Positivo + Falso_Positivo)
Confiabilidade_Negativa = Verdadeiro_Negativo / (Verdadeiro_Negativo + Falso_Negativo)

print('Número de Predições: %i' %(Número_Predições))
print('Número de Verdadeiros Positivos: %i' %(Verdadeiro_Positivo))
print('Número de Verdadeiros Negativos: %i' %(Verdadeiro_Negativo))
print('Número de Falsos Positivos: %i' %(Falso_Positivo))
print('Número de Falsos Negativos: %i' %(Falso_Negativo))
print("")
print('Acuracidade: %.4f' %(Acuracidade))
print('Erro: %.4f' %(Erro))
print('Sensibilidade: %.4f' %(Sensibilidade))
print('Especificidade: %.4f' %(Especificidade))
print('Confiabilidade Positiva: %.4f' %(Confiabilidade_Positiva))
print('Confiabilidade Negativa: %.4f' %(Confiabilidade_Negativa))

### CURVA CARACTERÍSTICA DE OPERAÇÃO DO RECEPTOR - REGRESSÃO ÚNICA PARA 05 ANOS
REG = ['Intercept', 'ANO', 'IDADE2', 'IDADE', 'GRAU_INSTRUÇÃO2', 'GRAU_INSTRUÇÃO',\
'CLASSE_ECONÔMICA_1', 'CLASSE_ECONÔMICA_2', 'CLASSE_ECONÔMICA_3', 'PEA_2_1', 'NÃO_CATÓLICA',\
'ÁREA_1', 'ANO_IDADE2', 'ANO_IDADE', 'ANO_GRAU_INSTRUÇÃO2', 'ANO_CLASSE_ECONÔMICA_2',\
'ANO_CLASSE_ECONÔMICA_3', 'ANO_PEA_2_1', 'ANO_NÃO_CATÓLICA', 'ANO_ÁREA_1']

df_ind_reg['Intercept'] = -0.8324
X_reg_final = df_ind_reg[REG].to_numpy().astype(np.int64)
Y_reg_final = df_ind_reg['INTERNET_1'].to_numpy().astype(np.int64)

y_pred_final = model_reg.predict(X_reg_final)

```

```

fpr_final, tpr_final, thresholds_final = roc_curve(Y_reg_final, y_pred_final)
roc_auc_final = auc(fpr_final, tpr_final)

fig, ax = plt.subplots(figsize=(22, 10), facecolor='#F0F0F0')

plt.suptitle('Curva Característica de Operação do Receptor (ROC)', x=0.313, y=0.94, fontsize = 20, weight = 'bold', \
            alpha = .75, va='top')
plt.title('Anos 2015-2019', fontsize = 19, weight = 'bold', alpha = .60, loc='left')
plt.plot(fpr_final, tpr_final, label='Regressão Logística Final (Área = %0.2f)' % roc_auc_final)

plt.plot([0, 1], [0, 1], 'r--')

ax.set_ylabel(u'Taxa de Verdadeiros Positivos', fontsize = 17)
ax.set_xlabel('Taxa de Falsos Positivos', fontsize = 17)
ax.set_facecolor('#F0F0F0')
ax.legend(frameon=False)
plt.setp(ax.xaxis.get_majorticklabels(), rotation=45, ha="left")
plt.autoscale(tight=False)
style.use('fast')
plt.show()

```