

Carolina Lourenço de Sousa

**Análise dos fatores de risco associados à
depressão no Brasil, no ano de 2019.**

Niterói - RJ, Brasil

01 de agosto de 2022

Carolina Lourenço de Sousa

**Análise dos fatores de risco
associados à depressão no Brasil, no
ano de 2019.**

Trabalho de Conclusão de Curso

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

Orientador(a): Prof. Dr. Ana Beatriz Monteiro Fonseca

Niterói - RJ, Brasil

01 de agosto de 2022

Carolina Lourenço de Sousa

**Análise dos fatores de risco associados à
depressão no Brasil, no ano de 2019.**

Monografia de Projeto Final de Graduação sob o título “*Análise dos fatores de risco associados à depressão no Brasil, no ano de 2019.*”, defendida por Carolina Lourenço de Sousa e aprovada em 01 de agosto de 2022, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

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

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

Profa. Dra. Jessica Quintanilha Kubrusly
Departamento de Estatística - UFF

Niterói, 01 de agosto de 2022

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

S725a Sousa, Carolina Lourenço de
Análise dos fatores de risco associados à depressão no Brasil, no ano de 2019. / Carolina Lourenço de Sousa ; Ana Beatriz Monteiro Fonseca, orientadora. Niterói, 2022.
60 f.

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

1. Modelos de Regressão Logística. 2. Amostra Complexa. 3. Pesquisa Nacional de Saúde. 4. Depressão. 5. Produção intelectual. I. Fonseca, Ana Beatriz Monteiro, orientadora. II. Universidade Federal Fluminense. Instituto de Matemática e Estatística. III. Título.

CDD -

Resumo

A depressão é um transtorno mental que coloca em risco a qualidade de vida do indivíduo, podendo prejudicar a vida pessoal, profissional e acadêmica de qualquer pessoa. É considerada, pela Organização Mundial de Saúde, a principal causa de incapacitação e está entre as dez principais causas de patologia, a nível mundial. O presente trabalho tem, então, o objetivo de conhecer o perfil da população brasileira maior de 14 anos que foi diagnosticada com depressão no ano de 2019, e identificar fatores possam estar a ela associados. A partir dos dados da Pesquisa Nacional de Saúde, do ano de 2019, foi levantado que perfil dos brasileiros acima de 14 anos é composto por de mulheres (52,95%), pessoas brancas e pardas (totalizando, aproximadamente, 88%), solteiros (47%), com média da idade de $43,24 \pm 18,12$ anos, com o ensino médio completo e faixa de rendimento domiciliar *per capita* entre $\frac{1}{2}$ até 2 salários mínimos (aproximadamente 57,2%). Na análise de fatores associados à depressão, pelo teste qui-quadrado, foi concluído que, ao nível de significância de 5%, não há associação da variável relacionada à exercícios físicos ou esporte com o diagnóstico da depressão, enquanto que para todas as demais a associação foi considerada presente e altamente significativa ($p - \text{valor} < 0,01$). A partir desses resultados foi construído um modelo de regressão logística para explicar a chance de ocorrência de depressão em função das variáveis: sexo, idade, cor ou raça, estado civil, curso mais elevado que frequentou, frequência de problemas no sono, consumo de bebida alcoólica, prática de exercício físico, consumo de tabaco, renda domiciliar, renda domiciliar *per capita*, faixa de rendimento *per capita* e o nível de instrução mais elevado alcançado, que obteve ótimas qualidades de ajuste (Pseudo- R^2 próximo de 0,89) e de previsão (taxa de acerto global próxima de 90%). A partir do modelo estimado, observou-se que mulher, pessoa de cor ou raça branca, ser divorciado(a), ter problemas frequentes para dormir e hábito de fumar são fatores de risco para o diagnóstico da depressão. Além disso, o modelo estimou que, à medida que o indivíduo envelhece um ano, a chance de ser diagnosticado com depressão aumenta em 0,4%. Por outro lado, identificou-se, através do modelo estimado, que à medida que o grau de instrução aumenta, a chance de ser diagnosticado também aumenta. Por fim, ao contrário do esperado, consumir bebida alcoólica foi considerado fator de proteção para a doença.

Palavras-chave: Modelos de Regressão Logística. Amostra Complexa. Pesquisa Nacional de Saúde. Depressão.

Dedicatória

Dedico este trabalho às minhas duas mães, com todo amor e gratidão. Minha mãe, Vera Lúcia (*in memoriam*) e minha avó, Amélia (*in memoriam*).

Agradecimentos

Primeiramente, quero agradecer a minha mãe Vera Lúcia (*in memoriam*), que sempre lutou para me dar uma educação de qualidade e, por fim, chegar ao fim da minha graduação.

À professora Dra. Ana Beatriz Monteiro Fonseca por aceitar me orientar, pela paciência e dedicação neste trabalho.

Aos professores, Dr. Rafael Santos Erbisti e Dra. Jéssica Quintanilha Kubrusly, por aceitarem participar da banca deste trabalho.

Aos amigos que conheci na faculdade, em especial: Beatriz Loureiro, Francisco Sebastian, Gabriela de Barros, Marlon Alves, Simone Galdino e Pedro Mauricio.

À minha melhor amiga, Alessandra Asevedo, por estar sempre comigo em todos momentos.

À minha família, em especial, minhas primas que além de família são minhas amigas e estão sempre comigo.

Obrigada à todos que sempre estiveram comigo.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 12
1.1	Depressão	p. 12
1.2	Depressão no mundo	p. 13
1.3	Depressão no Brasil	p. 15
1.4	Objetivos	p. 16
1.4.1	Objetivo Geral	p. 16
1.4.2	Objetivos Específicos	p. 16
1.5	Organização	p. 17
2	Materiais e Métodos	p. 18
2.1	Pesquisa Nacional de Saúde (PNS)	p. 18
2.2	Fonte de dados	p. 19
2.3	Análises Estatísticas Preliminares	p. 24
2.3.1	Estimação pontual	p. 25
2.3.2	Teste Qui-Quadrado de Independência	p. 26
2.3.3	Coefficiente de Assimetria	p. 28
2.4	Modelos Lineares Generalizados (MLG)	p. 28
2.4.1	Família Exponencial	p. 29
2.4.2	Modelo de Regressão Logística	p. 30

2.4.2.1	Função de Ligação - Logit	p. 31
2.4.2.2	Estimação dos parâmetros do modelo	p. 31
2.4.3	Modelo Logístico para dados amostrais complexos	p. 32
2.4.3.1	Estimação dos parâmetros por Máxima Pseudo-verossimilhança	p. 32
2.4.3.2	Razão de Chances	p. 33
2.4.3.3	Avaliação da capacidade preditiva do modelo	p. 35
3	Resultados	p. 37
3.1	Análises preliminares	p. 37
3.1.1	Perfil Demográfico Populacional	p. 37
3.1.2	Identificação dos fatores associados à depressão	p. 43
3.2	Modelos de Regressão Logística	p. 48
3.2.1	Modelos Propostos	p. 49
3.2.2	Comparação de modelos	p. 53
3.2.3	Modelo selecionado	p. 54
3.2.4	Discussão dos resultados	p. 55
4	Conclusões	p. 57
	Referências	p. 59

Lista de Figuras

1	Distribuição por sexo e por Cor ou raça	p. 38
2	Distribuição por Estado Civil	p. 38
3	Histograma da Idade	p. 39
4	Distribuição percentual da população por Escolaridade através da variável D00901	p. 39
5	Distribuição percentual da população por Instrução através da variável VDD004A	p. 40
6	Histogramas da renda domiciliar e renda domiciliar <i>per capita</i>	p. 41
7	Distribuição da faixa de rendimento <i>per capita</i>	p. 41
8	Distribuição de cor ou raça por grau de instrução	p. 42
9	Distribuição de cor ou raça por faixa de renda <i>per capita</i>	p. 42
10	Distribuição de diagnóstico de depressão por sexo e e Cor ou raça	p. 45
11	Distribuição de diagnóstico de depressão por estado civil	p. 45
12	Distribuição de diagnóstico de depressão por grupos de escolaridade	p. 46
13	Distribuição de diagnóstico de depressão por grupos de grau de instrução	p. 46
14	Distribuição de diagnóstico de depressão em relação a frequência de problemas com sono e ao consumo de bebida alcoólica	p. 47
15	Distribuição de diagnóstico de depressão em relação ao consumo de tabaco e a faixa de rendimento <i>per capita</i>	p. 47

Lista de Tabelas

1	Descrição das variáveis encontradas dicionário da PNS	p. 20
2	Nomes resumidos de cada variável	p. 23
3	Agregação das categorias da variável C009 - Cor ou raça	p. 23
4	Agregação das categorias da variável D00901 - Escolaridade	p. 24
5	Agregação das categorias da variável P027 - Álcool	p. 24
6	Agregação das categorias da variável P050 - Tabaco	p. 24
7	Exemplo de tabela de contingência para as variáveis X e Y	p. 26
8	Distribuição das unidades segundo valor previsto e observado de Y . . .	p. 35
9	Estatísticas descritivas da Idade (em anos)	p. 38
10	Estatísticas descritivas do rendimento domiciliar total e <i>per capita</i> (em reais)	p. 41
11	Distribuição de frequências dos fatores sociodemográficos por existência de diagnóstico de depressão e teste qui-quadrado para independência .	p. 43
12	Intervalo de média \pm desvio padrão para as variáveis idade, renda domiciliar e renda domiciliar <i>per capita</i> por ocorrência de diagnóstico de depressão	p. 48
13	Combinação das variáveis de escolaridade e renda em cada modelo proposto.	p. 48
14	Categorias de referência para as variáveis qualitativas	p. 49
15	Estimativas do Modelo inicial 1 completo	p. 50
16	Estimativas do Modelo inicial 2 completo	p. 50
17	Estimativas do Modelo inicial 3 completo	p. 51
18	Estimativas do Modelo inicial 4 completo	p. 51

19	Estimativas do Modelo inicial 5 completo	p. 52
20	Estimativas do Modelo inicial 6 completo	p. 52
21	Variáveis excluídas de cada modelo	p. 53
22	Métricas dos modelos reduzidos	p. 53
23	Estimativas de parâmetros para o modelo selecionado, OR e intervalos de confiança	p. 54

1 Introdução

1.1 Depressão

A depressão é um transtorno mental que coloca em risco a qualidade de vida do indivíduo, podendo prejudicar a vida pessoal, profissional e acadêmica de qualquer pessoa (COUTINHO, 2013). Classificada como uma Doença Crônica Não Transmissível (DCNT), a depressão é considerada, pela Organização Mundial de Saúde (OMS), a principal causa de incapacitação e está entre as dez principais causas de patologia, a nível mundial. (WHO, 2001).

Para seu diagnóstico são considerados os aspectos psíquicos, fisiológicos e evidências comportamentais (PORTO, 1999). Através da interação de fatores biológicos, psicológicos e sociais ocorre o desenvolvimento de Transtornos Mentais e do Comportamento (TMC) (GUIMARÃES; GRUBITS, 2004). Dentre os fatores biológicos que podem desencadear a depressão, a genética é um fator significativo em um quadro depressivo (SANTOS, 2021).

De acordo com a Classificação Internacional de Doenças (CID-10), a quantidade de sintomas e a sua gravidade determinam o grau da depressão, podendo ser: leve, moderado e grave (SAÚDE, 1994). Os sintomas utilizados para o diagnóstico, de acordo com o Manual Diagnóstico e Estatístico de Transtornos Mentais, o DSM-V, são: humor deprimido na maior parte do dia, diminuição no interesse ou no prazer de realizar atividades, perda ou ganho excessivo de peso, insônia ou hipersonia, fadiga, perda de concentração, sensação de inutilidade e culpa excessiva, e pensamentos suicidas (ASSOCIATION et al., 2014). No pior dos casos, a doença pode levar ao suicídio (PAHO, 2021).

Segundo Paho (2021), a depressão é diferente das flutuações usuais de humor e respostas emocionais de curta duração referentes aos desafios da vida cotidiana. Pessoas que passaram por situações de luto, desemprego ou algum trauma psicológico estão mais propensas a adquirirem a doença. O tratamento deste transtorno de humor é feito através de medicamentos antidepressivos ou terapias psicológicas, como terapia cognitivo-

comportamental e psicoterapia interpessoal, dependendo do grau da doença em que o paciente se encontra (PAHO, 2021). Segundo Quevedo, Nardi e Silva (2018), cerca de um terço dos pacientes conseguem atingir a remissão da doença.

1.2 Depressão no mundo

Conhecida como o “mal do século” pela OMS, a depressão afeta aproximadamente 280 milhões de pessoas ao redor do mundo. Ela é o segundo maior problema de saúde mental, com cerca de 20% de pessoas de países desenvolvidos apresentam a doença em algum momento da vida (CRUWYS et al., 2014). Segundo OMS (2021), a depressão é considerada uma doença comum, que afeta cerca de 3,8% da população mundial geral, 5,0% dos adultos e 5,7% dos adultos com mais de 60 anos.

Apesar de existirem inúmeros tratamentos para a doença, 75% da população de países subdesenvolvidos e em desenvolvimento não recebem tratamento para depressão. Os principais motivos são: a falta de acesso a profissionais de saúde devidamente treinados, a falta de recursos e o estigma associado a transtornos mentais. Em diferentes países, foram registradas diversas ocasiões em que ocorre o erro de diagnóstico, com avaliações imprecisas. Pessoas são diagnosticadas erroneamente com depressão sem apresentar, realmente, a doença e pessoas que têm depressão não são identificadas como tal (OMS, 2021). Estima-se que, em 2030, a depressão será o fator mais associado a comorbidades - quando duas ou mais doenças estão relacionadas entre si - em todo o mundo. (WHO, 2012).

A OMS da Região Européia revelou que cerca de 4,3% daquela população foi afetada pela depressão, ou seja, cerca de 40 milhões de pessoas. A OMS trabalha para conscientizar a população sobre as consequências da depressão e automutilação, para diminuir o estigma e a discriminação, e melhorar o acesso aos cuidados de saúde (WHO, 2018). A denominada “depressão maior” é caracterizada por pelo menos duas semanas com humor deprimido ou perda de interesse ou prazer em realizar quase todas as atividades, associado a pelo menos quatro sintomas de depressão (PARKER; BROTCHE, 2009). Menos de 50% das pessoas diagnosticadas com depressão maior recebem ajuda profissional (COPPENS et al., 2013). O tratamento é particularmente pouco aplicado a homens, pessoas de faixas etárias mais avançadas ou mais jovens, naqueles que têm renda mais baixa, com níveis educacionais mais baixos ou, ainda, para pessoas casadas (COPPENS et al., 2013).

O resultado da pesquisa de Coppens et al. (2013) revela que, em geral, há um grau moderado de estigmatização pessoal em relação à depressão, com cerca de um terço dos

entrevistados concordando com expressões do tipo: “A depressão é sinal de fraqueza pessoal”, “As pessoas podem sair dessa (depressão) se quiserem”, “Pessoas com depressão são imprevisíveis” e “Eu não votaria em um político se eu soubesse que ele (ou ela) está deprimido”. Apesar desse estigma alto em relação à doença, cerca de 70% dos respondentes concordam que é necessário buscar ajuda profissional. Todavia, mais de 70% dos entrevistados também acham admiráveis pessoas que lidam com seus problemas emocionais sozinhas ao invés de recorrer a ajuda profissional.

Estudos foram realizados por um dos maiores multicentros, o SABE - Saúde, Bem Estar e Envelhecimento, em seis capitais da América Latina, usando uma escala de depressão geriátrica e observou-se uma prevalência de depressão entre 16,5% e 30,1% em mulheres, e de 11,8% a 19,6% em homens. Estes resultados foram consistentes com duas estimativas de depressão tardia na Europa: a SHARE e a EURODEP (GUERRA et al., 2016).

Segundo Ren et al. (2020) estima-se que cerca de 173 milhões de adultos na China são afetados por doenças mentais. Em 2017, esse país tinha cerca de 56,36 milhões de casos de depressão, ou cerca de 21,3% dos casos do mundo. Ali, a prevalência da depressão é superior à global, porém inferior a países como Estados Unidos e Inglaterra. Esta condição pode estar relacionado ao crescimento da economia, aumento na industrialização do país, rápida modernização e aumento drástico na taxa de urbanização, levando ao crescimento rápido da população. O número de divórcios na China aumentou drasticamente e a disfunção familiar tornou-se um problema que agrava o quadro depressivo. Por último, o estudo afirma que a China é o maior produtor e consumidor de tabaco do mundo. Lopez et al. (2011) afirma que o consumo de álcool e de derivados tabaco também está associado à depressão.

Em Ettman et al. (2020), foi estudada a prevalência de sintomas de depressão em adultos nos Estados Unidos, antes e durante, a pandemia de COVID-19, uma doença respiratória causada pelo vírus SARS-Cov-2, cuja propagação se iniciou em 2020 e dura até o presente momento. Seus resultados mostram que a prevalência de sintomas depressivos aumentou durante a pandemia, passando de 8,5% para 27,8%. Também segundo o mesmo estudo, pessoas com baixa renda estão mais propensas a desenvolver sintomas depressivos do que os de renda mais alta (49,9% contra 16,9%) e mulheres continuaram sendo mais propensas a ter sintomas depressivos: antes da pandemia a prevalência era de 10,1% e passa para 33,3%. Durante a pandemia, indivíduos casados obtiveram uma taxa de depressão menor (18,3%) do que os viúvos, divorciados ou separados (31,5%), porém pessoas que nunca foram casadas ou em união estável alcançaram os maiores percentuais

(39,8% e 37,7% respectivamente).

Em Fukita, Kawasaki e Yamasaki (2021) foi realizado um estudo para levantar fatores relacionados à depressão em pessoas de idades entre 40 e 59 anos, em um município rural no oeste do Japão. Foi aplicado um questionário levantando fatores psicológicos, comportamentais (relacionados à saúde), características demográficas e socioeconômicas. Neste trabalho foi feita um análise de regressão logística para identificar fatores associados à depressão. A pesquisa mostra que existia uma chance maior de depressão para indivíduos de sexo masculino, de alto nível de estresse cognitivo, baixa busca por ajuda, com qualidade de sono ruim, sem nenhum tipo de passatempo e com pequeno senso de coerência.

1.3 Depressão no Brasil

Diversos estudos foram realizados no Brasil, relacionando a depressão a alguns fatores. Gazalle et al. (2004) conduziram um estudo com idosos na cidade de Pelotas, Rio Grande de Sul, onde verificaram que a prevalência de sintomas depressivos neste grupo era muito alta. O estudo efetuou uma análise multivariável, e apontou uma maior chance de sintomas depressivos em mulheres, pessoas com idade mais avançada, baixa escolaridade, sem trabalho remunerado, fumantes e perderam algum ente querido.

Em Costa, Soares e Teixeira (2007) foi estudado o benefício da realização de atividades físicas na prevenção da depressão. Foi apontado que o condicionamento físico está ligado de forma positiva à saúde mental, pois acredita-se na hipótese de que as atividades físicas ajudam na produção de endorfinas. Segundo os autores, além do tratamento padrão da depressão, com remédios e psicoterapias, os exercícios físicos podem auxiliar no tratamento da doença e praticar exercícios regularmente acarreta benefícios psicológicos como melhora no bem-estar, na autoestima e no humor. Ribeiro (1998) também afirma que a atividade física pode atuar como fator de prevenção dos sintomas depressivos.

O estudo realizado por Cunha, Bastos e Duca (2012) teve como objetivo estimar a prevalência de depressão e os fatores demográficos e socioeconômicos associados em uma amostra de adultos e idosos de uma comunidade de baixa renda em Porto Alegre. As covariáveis investigadas foram: sexo, idade, estado civil, escolaridade e nível econômico, e as análises realizadas indicaram que a depressão está mais associada ao sexo feminino e que há tendência de aumento da chance de ter depressão com o avanço da idade.

De acordo com Stopa et al. (2015), o diagnóstico de depressão, em 2013, era predomi-

nantemente dado a mulheres (10,9%), a pessoas na faixa etária entre 60 e 64 anos (11,1%), a faixa etária com menor número de casos é a de 18 a 29 anos (3,9%) e prevalência de casos em brasileiros com ensino médio completo a nível superior incompleto foi de 6,4%, e com ensino superior completo foi de 8,7%. Avaliando a influência da raça ou cor, houve a predominância de casos de depressão em indivíduos brancos, com 9,0% e quanto ao local de residência, pessoas que moram em áreas urbanas têm maior número de diagnósticos de depressão com 8,0%. A região Sul do Brasil obteve o maior percentual de casos de depressão, com 12,6%, e a região Norte, obteve o menor, com 3,1%.

Em Gonçalves et al. (2018), avaliou-se a prevalência de depressão e fatores associados em mulheres de 20 a 59 anos de áreas cobertas pela Estratégia de Saúde da Família de municípios da Zona da Mata Mineira. O estudo incluiu variáveis como idade, escolaridade, cor/raça, religião, estado civil, renda, número de pessoas na residência, apoio por arte de parentes/ amigos, autoavaliação de saúde, prática de atividade física, consumo de álcool, fumo, frequência a cultos religiosos, presença de doença mental, uso de medicamentos para dormir e, por fim, busca por atendimento na UAPS (Unidades de Atenção Primária em Saúde). Dentre as mulheres avaliadas, 19,7% foram apontadas como depressivas, e neste grupo houve predominância de mulheres de 50 a 59 anos (22,42%). O uso de medicamentos para dormir, ter baixa renda, estar trabalhando, ter menor escolaridade, ser tabagista e ter procurado UAPS no último ano foram apontados como fatores de propensão à ocorrência de depressão. Em contrapartida, viver com um companheiro, ter apoio de parentes, praticar atividade física e ter autoavaliação de saúde “boa” ou “muito boa” foram fatores de proteção para a depressão.

1.4 Objetivos

1.4.1 Objetivo Geral

O objetivo do presente trabalho é estimar a prevalência de depressão autodeclarada no Brasil, no ano de 2019, e identificar fatores que possam estar a ela associados.

1.4.2 Objetivos Específicos

Os objetivos específicos do trabalho são:

- Efetuar análise exploratória sobre as variáveis da base de dados da Pesquisa Nacional de Saúde (PNS) e caracterizar o perfil dos respondentes;

- Obter estimativa da prevalência de depressão no Brasil, para o ano de 2019;
- Identificar fatores que estão associados à depressão a partir de testes estatísticos;
- Construir um modelo de regressão logística que relacione a chance de ocorrência de depressão (autodeclarada) a fatores que estão a ela associados.

1.5 Organização

No Capítulo 2 será apresentada a fonte de dados, as variáveis utilizadas no estudo, as ferramentas associadas à análise exploratória dos dados e os modelos lineares generalizados adotados na pesquisa, cuja estimação foi adaptada para amostras complexas. No Capítulo 3, serão apresentados os principais resultados do estudo, com a apresentação do perfil demográfico da população considerada e dos fatores associados à ocorrência de depressão previamente diagnosticada e autodeclarada. No Capítulo 4, serão apresentadas as conclusões do estudo e sugestões para estudos futuros.

2 Materiais e Métodos

Nesta seção será apresentada a Pesquisa Nacional de Saúde (PNS), de onde foi feita a extração da base de dados deste trabalho e feita a seleção das variáveis que serão investigadas na busca de fatores relacionados à depressão. Também serão exibidas ferramentas preliminares usadas na análise exploratória dos dados e os principais conceitos envolvidos na utilização de Modelos de Regressão Logística. Para a estimação e para a modelagem também, serão apresentadas as mudanças na metodologia considerando amostras complexas.

2.1 Pesquisa Nacional de Saúde (PNS)

A realização de inquéritos de saúde é importante para se obter a avaliação de sistemas de saúde sob a perspectiva do usuário. Desde a década 1960, eles são realizados em países de alta renda, mas em países de baixa e média renda, a prática é mais recente. No Brasil, o primeiro inquérito no âmbito da saúde foi realizado em 1974, na cidade de Ribeirão Preto, SP, sobre conjunto de relatos referente a problemas de saúde de um indivíduo (BARRETO; GRISI, 2010), a chamada morbidade referida, e utilização de serviços de saúde (STOPA et al., 2020).

Foram produzidas nos anos de 1998, 2003 e 2008, com apoio do Ministério Público, pesquisas relacionadas com a Saúde a partir da Pesquisa Nacional por Amostra de Domicílios (PNAD), do Instituto Brasileiro de Geografia e Estatística (IBGE). Por meio desta, foi possível comparar resultados e avaliar as condições de saúde dos brasileiros. Em 2013 foi criada a Pesquisa Nacional de Saúde (PNS), uma pesquisa de âmbito nacional, realizada em intervalos de 5 anos. Entretanto, no ano de 2018 não foi possível realizá-la, pois ocorreram problemas relacionados ao seu delineamento e execução, passando assim sua execução para o ano subsequente (IBGE, 2020).

A amostragem da PNS foi desenvolvida a partir da Amostra Mestra do Sistema Integrado de Pesquisa Domiciliares - SIPD, do IBGE, sendo a sua amostra da PNS, uma

subamostra da Amostra Mestra do SIPD/IBGE. O plano amostral utilizado foi o de Amostragem por Conglomerados em três estágios, sendo o primeiro estágio a estratificação das Unidades Primárias de Amostragem (UPAs), as quais são formadas por Setores Censitários ou por um conjunto de setores (SZWARCWALD et al., 2019), através de amostra aleatória simples (AAS).

O segundo estágio é feito através de AAS dentro de cada UPA selecionada no primeiro estágio, onde são selecionados de 12 a 18 domicílios construídos exclusivamente para moradia. São excluídos domicílios que são localizados em setores censitários especiais ou de população escassa, como bases militares, alojamentos, acampamentos, agrupamentos indígenas, penitenciárias, conventos, etc. A listagem de domicílios é retirados do Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE) (SZWARCWALD et al., 2019).

Por fim, no terceiro estágio, é feita uma entrevista em que se é obtida uma lista de moradores do domicílio. Após esse primeiro momento, seleciona-se um morador de 15 anos ou mais por AAS para responder o questionário por todos os demais habitantes do domicílio (IBGE, 2020).

O tamanho da amostra da PNS de 2019 contou com 8036 unidades primárias de amostragem, que representam cerca de 53% do total de UPAs do Brasil. Para as Unidades da Federação (UF) com maior quantitativo de UPAs, foi definido que o número de domicílios seria menor (12 domicílios). E, para estados com menor quantitativo de UPAs, definiu-se que tal quantidade seria maior (18 domicílios). As unidades que não se encaixassem em nenhum desses padrões teriam 15 domicílios por UPA (IBGE, 2019).

2.2 Fonte de dados

Nesse trabalho será utilizada a base de dados da Pesquisa Nacional de Saúde realizada no ano 2019. Foram selecionadas variáveis que foram identificadas como possíveis fatores associados à depressão.

A variável Q092 identificada como “*Algum médico ou profissional de saúde mental (como psiquiatra ou psicólogo) já lhe deu o diagnóstico de depressão?*” está ligada diretamente à identificação de pessoas que receberam o diagnóstico de depressão em algum momento. Na amostra de tamanho $n = 279.382$, cerca de 32,52% desta variável foi utilizada retirando os dados faltantes.

Os fatores testados na análise foram selecionados a partir da leitura dos artigos pre-

sententes no Capítulo 1. O banco de dados apresenta algumas variáveis que, aparentemente, têm significados semelhantes, e todas serão testadas para que o melhor modelo possa ser encontrado.

A Tabela 1 expõe as variáveis utilizadas no presente estudo, incluindo seu código, descrição e categorias definidos na base dentro da base de dados da PNS 2019.

Tabela 1: Descrição das variáveis encontradas dicionário da PNS

Código da variável	Descrição	Categoria
C006	Sexo	1 - Homem 2 - Mulher
C008	Idade do morador na data de referência	Numérica (em anos)
C009	Cor ou raça	1 - Branca 2 - Preta 3 - Amarela 4 - Parda 5 - Indígena 9 - Ignorado
C011	Qual é o estado civil?	1 - Casado(a) 2 - Divorciado(a) ou desquitado(a) ou separado(a) judicialmente 3 - Viúvo(a) 4 - Solteiro(a) 9 - Ignorado

Código da variável	Descrição	Categoria
D00901	Qual foi o curso mais elevado que frequentou?	01 - Creche 02 - Pré-escola 03 - Classe de alfabetização - CA 04 - Alfabetização de jovens e adultos 05 - Antigo primário (elementar) 06 - Antigo ginásial (médio 1º ciclo) 07 - Regular do ensino fundamental ou do 1º grau 08 - Educação de jovens e adultos (EJA) ou supletivo do ensino fundamental 09 - Antigo científico, clássico etc. (médio 2º ciclo) 10 - Regular do ensino médio ou do 2º grau 11 - Educação de jovens e adultos (EJA) ou supletivo do ensino médio 12 - Superior - graduação 13 - Especialização de nível superior (duração mínima de 360 horas) 14 - Mestrado 15 - Doutorado 99 - Ignorado
N010	Nas duas últimas semanas, com que frequência o(a) Sr(a) teve problemas no sono, como dificuldade para adormecer, acordar frequentemente à noite ou dormir mais do que de costume?	1 - Nenhum dia 2 - Menos da metade dos dias 3 - Mais da metade dos dias 4 - Quase todos dias 9 - Ignorado
P027	Com que frequência o(a) Sr(a) costuma consumir alguma bebida alcoólica?	1 - Não bebo nunca 2 - Menos de uma vez por mês 3 - Uma vez ou mais por mês 9 - Ignorado
P034	Nos últimos três meses, o(a) Sr(a) praticou algum tipo de exercício físico ou esporte?	1 - Sim 2 - Não 9 - Ignorado

Código da variável	Descrição	Categoria
P050	Atualmente, o(a) Sr(a) fuma algum produto do tabaco?	1 - Sim, diariamente 2 - Sim, menos que diariamente 3 - Não fumo atualmente 9 - Ignorado
VDF002	Rendimento domiciliar	Numérica (Em reais)
VDF003	Rendimento domiciliar <i>per capita</i>	Numérica (Em reais)
VDF004	Faixa de rendimento domiciliar <i>per capita</i>	1 - Até $\frac{1}{4}$ salário mínimo 2 - Mais de $\frac{1}{4}$ até $\frac{1}{2}$ salário mínimo 3 - Mais de $\frac{1}{2}$ até 1 salário mínimo 4 - Mais de 1 até 2 salários mínimos 5 - Mais de 2 até 3 salários mínimos 6 - Mais de 3 até 5 salários mínimos 7 - Mais de 5 salários mínimos
VDD004A	Nível de instrução mais elevado alcançado (pessoas de 5 anos ou mais de idade) padronizado para o Ensino Fundamental - SISTEMA DE 9 ANOS	1 - Sem instrução 2 - Fundamental incompleto ou equivalente 3 - Fundamental completo ou equivalente 4 - Médio incompleto ou equivalente 5 - Médio completo ou equivalente 6 - Superior incompleto ou equivalente 7 - Superior completo

Fonte: IBGE

Como as descrições das variáveis são muito extensas, para este trabalho foi necessário escolher uma palavra que resumiria a descrição de cada variável, essas palavras estão descritas na Tabela 2.

Tabela 2: Nomes resumidos de cada variável

Variável	Nome resumido
C006	Sexo
C008	Idade
C009	Etnia
C011	Estado civil
D00901	Escolaridade
N010	Sono
P027	Álcool
P034	Exercício
P050	Tabaco
VDF002	Renda
VDF003	Renda <i>per capita</i>
VDF004	Faixa de renda
VDD004A	Instrução

Foi necessária a agregação de algumas variáveis presentes na base de dados. São elas: C009 - “Cor ou raça”, D00901 - “Qual é o curso mais elevado que frequentou”, P027 - “Com que frequência o(a) Sr(a) costuma consumir alguma bebida alcoólica?” e P050 - “Atualmente, o(a) Sr(a) fuma algum produto do tabaco?”.

As agregações são demonstradas a partir das Tabelas 3, 4, 5 e 6.

Tabela 3: Agregação das categorias da variável C009 - Cor ou raça

Categoria inicial	Categoria após agregação
Branca	Branca
Preta	Preta
Parda	Parda
Amarela Indígena Ignorado	Outra

Tabela 4: Agregação das categorias da variável D00901 - Escolaridade

Categoria inicial	Categoria após agregação
Creche Pré-escola Classe de alfabetização – CA Antigo primário (elementar)	Até alfabetização ou primário
Antigo ginásial (médio 1º ciclo) Regular do ensino fundamental ou do 1º grau	Ensino fundamental regular ou ginásial
Antigo científico, clássico etc. (médio 2º ciclo) Regular do ensino médio ou do 2º grau	Ensino médio regular ou científico, clássico e etc
Alfabetização de jovens e adultos Educação de jovens e adultos (EJA) ou supletivo do ensino fundamental Educação de jovens e adultos (EJA) ou supletivo do ensino médio	Educação de jovens e Adultos (EJA)
Superior – graduação	Superior - graduação
Especialização de nível superior (duração mínima de 360 horas) Mestrado Doutorado	Pós-graduação

Tabela 5: Agregação das categorias da variável P027 - Álcool

Categoria inicial	Categoria após agregação
Não bebo nunca	Não bebo
Menos de uma vez por mês Uma vez ou mais por mês	Bebo

Tabela 6: Agregação das categorias da variável P050 - Tabaco

Categoria inicial	Categoria após agregação
Sim, diariamente Sim, menos que diariamente	Fumo
Não fumo atualmente	Não fumo

Algumas variáveis foram excluídas após a análise da quantidade de dados faltantes, são elas: D00301 - “Qual é o curso que frequenta”, P02801 - “Quantos dias por semana o(a) Sr(a) costuma consumir alguma bebida alcoólica?” e P035 - “Quantos dias por semana o(a) Sr(a) costuma (costumava) praticar exercício físico ou esporte?”.

2.3 Análises Estatísticas Preliminares

Nesta seção, serão apresentadas as ferramentas associadas à análise exploratória de dados, e que a complementam, como o teste qui-quadrado e o coeficiente de assime-

tria. Para isso, na análise univariada, as variáveis qualitativas serão descritas através das frequências relativas e representadas através de gráficos de barras e setores, e as variáveis C008, VDF002 e VDF003, que informa a idade dos respondentes, renda domiciliar e renda domiciliar per capita respectivamente, serão resumidas através de média, mediana, desvio padrão, mínimo, máximo e coeficiente de assimetria. Para a comparação entre os grupos de indivíduos com e sem depressão autodeclarada quanto a variáveis qualitativas, será utilizado um teste qui-quadrado.

2.3.1 Estimação pontual

O planejamento da PNS é feito a partir de uma amostragem complexa e, a partir da amostra coletada, pode-se inferir as características para a população de interesse.

Para uma população U com característica de interesse X , com (X_1, X_2, \dots, X_N) o vetor populacional e $i = 1, 2, \dots, N$, determinam-se o total populacional (τ) e média populacional (μ) como:

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

$$\mu = \frac{\tau}{N} = \sum_{i=1}^N \frac{X_i}{N} \quad (2.2)$$

Seja W_i o peso associado ao i –ésimo elemento amostral, n o tamanho da amostra s , com $s \in U$. Os estimadores não-viesados para τ e μ são:

$$\hat{\tau} = \sum_{i=1}^n W_i X_i \quad (2.3)$$

$$\hat{\mu} = \frac{\hat{\tau}}{n} = \frac{1}{n} \sum_{i=1}^n W_i X_i \quad (2.4)$$

Para este trabalho, a estimação da média é o valor esperado da distribuição Bernoulli, onde a média é igual a proporção de sucesso (p).

Diferentemente dos estimadores não-viesados para o total populacional $\hat{\tau}$ e média populacional $\hat{\mu}$, que consiste em incluir os pesos amostrais, estimar a variância não é um processo simples, pois em amostras complexas existem etapas de estratificação e de

trabalho com dados agregados em conglomerados. De modo geral, para estimar variâncias para um plano amostral complexo não existe expressão analítica melhor, mas existem maneiras de determinar o estimador adequado da variância dado um bom estimador.

Entre os mais comumente utilizados tem-se: Linearização de Taylor, Método Conglomerado Primário e Métodos de Replicação (Jackknife). Com a combinação desses métodos são desenvolvidos os algoritmos empregados pelos principais pacotes estatísticos especializados em estimação de variâncias de aleatorização (PESSOA; SILVA, 1998) e (MUIANGA, 2016).

Mais informações podem ser encontradas em Pessoa e Silva (1998) e Muianga (2016).

A partir de médias e desvios padrões estimados (sendo este definido pela raiz quadrada da variância), $\hat{\mu}$ e $\hat{\sigma}$, respectivamente, foi adotada, no presente trabalho, a notação de intervalo de referência para as variáveis quantitativas como $\hat{\mu} \pm \hat{\sigma}$.

2.3.2 Teste Qui-Quadrado de Independência

Este teste é utilizado em análises bivariadas em que se avalia se duas variáveis qualitativas estão associadas. De acordo com BUSSAB (2010), as hipóteses do teste qui-quadrado de independência são dadas por:

$$\begin{cases} H_0 : p_{ij} = p_{i \cdot} p_{\cdot j}, \text{ para todo par } (i, j), \\ H_1 : p_{ij} \neq p_{i \cdot} p_{\cdot j}, \text{ para algum par } (i, j) \end{cases}$$

onde p_{ij} é a probabilidade do indivíduo ser classificado na categoria i e j ao mesmo tempo, com $i = 1, \dots, r$ e $j = 1, \dots, s$. E $p_{i \cdot}$ e $p_{\cdot j}$ são probabilidades marginais, com $p_{i \cdot} = \sum_{j=1}^s p_{ij}$ e $p_{\cdot j} = \sum_{i=1}^r p_{ij}$.

Para realizar o teste é necessário construir uma tabela de contingência com os valores observados para variáveis X e Y , em que r e s são as quantidades categorias de resposta dentro dessas variáveis, respectivamente, conforme exemplificado na Tabela 7.

Tabela 7: Exemplo de tabela de contingência para as variáveis X e Y

$Y \setminus X$	X_1	...	X_s	Total
Y_1	o_{11}	...	o_{1s}	$n_{1 \cdot}$
...
Y_r	o_{r1}	...	o_{rs}	$n_{r \cdot}$
Total	$n_{\cdot 1}$...	$n_{\cdot s}$	n

em que:

- o_{ij} é a frequência observada de elementos pertencentes à i – ésima categoria de Y e j – ésima categoria de X ;
- $n_{i\cdot}$ o número de elementos na i – ésima categoria de Y ;
- $n_{\cdot j}$ o número de elementos da j – ésima categoria de X ;
- n é o número total de elementos da amostra

Sob a hipótese de independência, o cálculo de valores esperados das frequências é dado por:

$$e_{ij} = \frac{n_{i\cdot} \times n_{\cdot j}}{n} \quad (2.5)$$

com $i = 1, \dots, r$ e $j = 1, \dots, s$.

A estatística de teste está baseada na diferença entre os valores esperados sob H_0 e os observados na amostra, e é dada por:

$$\chi_{obs}^2 = \sum_{i=1}^r \sum_{j=1}^s \left[\frac{(o_{ij} - e_{ij})^2}{e_{ij}} \right] \quad (2.6)$$

A estatística de teste possui $(r - 1) \times (s - 1)$ graus de liberdade e o p – valor é encontrado através pela seguinte expressão:

$$p - valor = P[\chi_{(r-1) \times (s-1)}^2 > \chi_{obs}^2 | H_0] \quad (2.7)$$

A hipótese nula, H_0 , será rejeitada se o p – valor for menor ou igual ao nível de significância α . Caso contrário, H_0 não é rejeitada.

Para este trabalho será adotado o nível de significância $\alpha = 0,05$.

Para amostras complexas, precisa-se incorporar o peso amostral definida na Seção 2.3.1 para que o teste considere o plano amostral complexo. Neste caso, os valores da estatística de teste podem ser muito grandes e levar a conclusões equivocadas a respeito da rejeição (ou não) da hipótese nula. Maiores detalhes podem ser vistos em Pessoa e Silva (1998) e Pereira (2016).

2.3.3 Coeficiente de Assimetria

O Coeficiente de Assimetria é utilizado para avaliar o comportamento da distribuição dos dados da amostra x_1, x_2, \dots, x_n em relação a característica. O coeficiente de Assimetria de Bowley (B) verifica se o segundo quartil está igualmente distante do primeiro e do terceiro quartil, ou se está mais próximo de 1 deles e é calculado da seguinte maneira:

$$B = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1} \quad (2.8)$$

onde Q_1, Q_2 e Q_3 são os primeiro, segundo e terceiros quartis, calculados considerando os pesos amostrais para a população U .

A interpretação do coeficiente pode ser feita da seguinte maneira:

- $0 < B < 1$: Assimetria à direita
- $-1 < B < 0$: Assimetria à esquerda
- $B = 0$: Simetria
- $0 < |B| \leq 0,1$: Assimetria fraca
- $0,1 < |B| \leq 0,3$: Assimetria moderada
- $0,3 < |B| \leq 1$: Assimetria forte

2.4 Modelos Lineares Generalizados (MLG)

Criado por Nelder e Wedderburn (1972), o Modelo Linear Generalizado (MLG) é utilizado em estudos para investigar a relação entre p variáveis explicativas e a variável resposta Y , em casos onde essa variável não segue uma distribuição normal.

Cordeiro e Demétrio (1986) consideram o MLG uma extensão dos modelos clássicos de regressão e o definem como sendo composto por uma variável resposta univariada, variáveis explicativas independentes e uma amostra aleatória com n observações independentes, sendo:

1. o **componente aleatório** do modelo, também nomeado de variável resposta, que tem distribuição pertencente à família exponencial 2.4.1;

2. o **componente sistemático** do modelo, construído a partir de um conjunto de variáveis explicativas, de estrutura linear;
3. a **função de ligação**, que tem como objetivo relacionar os componente aleatório e sistemático.

A variável resposta (ou variável dependente) no MLG pode ser de natureza contínua, discreta ou dicotômica, e as variáveis independentes ou explicativas, podem ser tanto quantitativas (contínuas ou discretas) quanto qualitativas (ordinais ou dicotômicas).

2.4.1 Família Exponencial

Dobson e Barnett (2018) definem a família exponencial considerando uma variável aleatória Y com função de probabilidade dependendo apenas de um parâmetro Θ . A distribuição de Y pertencerá à família exponencial se ela puder ser escrita da seguinte forma:

$$f(y; \Theta) = [s(y)t(\Theta) \exp[a(y)b(\Theta)] \quad (2.9)$$

onde a , b , s e t são funções conhecidas.

Essa expressão pode ser reescrita da seguinte maneira:

$$f(y; \Theta) = \exp\{a(y)b(\Theta) + c(\Theta) + d(y)\} \quad (2.10)$$

em que $s(y)$ e $t(\Theta)$ são definidas como $\exp[d(y)]$ e $\exp[b(\Theta)]$, respectivamente.

Desta forma:

- Se $a(y) = y$, a distribuição está na forma canônica e $b(\theta)$ é o parâmetro natural da função.
- Se existirem outros parâmetros além de Θ , eles são considerados parâmetros de ruído e farão parte das funções a , b , c e d , sendo tratados como conhecidos.

Alguns exemplos de distribuições que pertencem à família exponencial são: Poisson, binomial, normal e gama.

2.4.2 Modelo de Regressão Logística

O MLG em que a variável resposta Y é binária, é chamado de Modelo de Regressão Logística. Essa variável aleatória binária é definida por:

$$Y = \begin{cases} 1, & \text{se o resultado é sucesso} \\ 0, & \text{caso contrário} \end{cases}$$

onde $P(Y = 1) = \pi$ (probabilidade de sucesso) e $P(Y = 0) = 1 - \pi$ (probabilidade de fracasso). Desta forma, $Y \sim \text{Bernoulli}(\pi)$ (DOBSON; BARNETT, 2018).

A função de probabilidade de Bernoulli, com parâmetro π , é dada por:

$$P(Y = y) = \pi^y(1 - \pi)^{1-y} \text{ para } y=1, 0. \quad (2.11)$$

Pode-se mostrar que $E(Y) = \pi$ e $Var(Y) = \pi(1 - \pi)$.

Se (Y_1, Y_2, \dots, Y_n) é um vetor aleatório independente de dimensão n onde cada Y_i segue uma distribuição de Bernoulli, com $P(Y_j = 1) = \pi_j$, sua função de probabilidade conjunta será dada por:

$$\prod_{j=1}^n P(Y_j = y_j) = \exp \left[\sum_{j=1}^n y_j \log \left(\frac{\pi_j}{1 - \pi_j} \right) + \sum_{j=1}^n \log(1 - \pi_j) \right] \quad (2.12)$$

dado que $\prod_{j=1}^n P(Y_j = y_j)$ pertence à família exponencial, pois pode ser escrita conforme a equação 2.10

Então, pode-se construir um modelo que relaciona a probabilidade π_j as covariáveis, definido para cada $i = 1, 2, \dots, n$, como:

$$g(\pi_j) = \mathbf{X}_j^T \boldsymbol{\beta} \quad (2.13)$$

em que \mathbf{X}_j é um vetor p de variáveis explicativas, $\boldsymbol{\beta}$ é o vetor de parâmetros e g é denominada de função de ligação. Vale lembrar que o vetor \mathbf{X}_j pode incluir um termo igual à unidade de forma a acomodar no modelo a presença de um intercepto, que seria um dos componentes do vetor $\boldsymbol{\beta}$.

Para garantir que os valores de π estão restritos no intervalo $[0,1]$, π pode ser modelado através de uma função de distribuição acumulada, fazendo:

$$\pi = g^{-1}(\mathbf{X}^T \boldsymbol{\beta}) = \int_{-\infty}^t f(s) ds \quad (2.14)$$

onde $f(s) \geq 0$, $\int_{-\infty}^{\infty} f(s) ds = 1$ e $f(s)$ é chamada de distribuição de tolerância.

2.4.2.1 Função de Ligação - Logit

Cada MLG necessita estabelecer uma função de ligação. No caso do o modelo de regressão logística, a função de ligação canônica é denominada *logit* e é representada pela seguinte equação:

$$g(\pi_i) = \text{logit}(\pi_i) = \log\left(\frac{\pi_j}{1 - \pi_i}\right) = \eta_i = \mathbf{X}_i^T \boldsymbol{\beta} \quad (2.15)$$

Sua função de distribuição de tolerância é dada por:

$$\pi_i = \int_{-\infty}^{\eta_i} \frac{\exp(s)}{[1 + \exp(s)]^2} ds \quad (2.16)$$

Então,

$$\pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \quad (2.17)$$

2.4.2.2 Estimação dos parâmetros do modelo

A estimação do vetor de parâmetros $\boldsymbol{\beta}$ e das probabilidades $\pi_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$ do modelo de regressão logística é, usualmente, feita através da maximização da função log-verossimilhança, dada por:

$$\ln(L(\boldsymbol{\beta})) = \sum_{i=1}^n \pi_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)] \quad (2.18)$$

Para maximizar os valores de $\boldsymbol{\beta}$, é necessário igualar a zero a derivada a equação 2.18 com respeito a cada um dos elementos do vetor $\boldsymbol{\beta}$. Porém não existe forma fechada de obter a estimação do vetor de parâmetros $\boldsymbol{\beta}$, sendo necessário outros métodos iterativos, como o de Newton-Raphson.

2.4.3 Modelo Logístico para dados amostrais complexos

A Pesquisa Nacional de Saúde tem plano amostral complexo. Portanto, é necessário considerar o peso das unidades amostrais na estimação dos parâmetros populacionais e, assim, obter medidas não viesadas (SILVA; PESSOA; LILA, 2002).

2.4.3.1 Estimação dos parâmetros por Máxima Pseudo-verossimilhança

De modo geral, a estimação dos parâmetros do modelo de regressão generalizado é realizada através do método da máxima verossimilhança. Entretanto, a estimação dos parâmetros do modelo proposto para amostras complexas é realizada por meio do método de máxima pseudo-verossimilhança (MPV), que consiste em considerar no processo de estimação o desenho e os pesos amostrais.

Para inferir os parâmetros do modelo é utilizado o princípio de modelagem de superpopulação, onde Y_i é variável de interesse de uma população finita e y_1, y_2, \dots, y_N seus valores. O valores de Y_i são considerados observações do vetor aleatório (Y_1, \dots, Y_N) , e supõe-se que são independentes e identicamente distribuídos com distribuição $f(y; \boldsymbol{\beta})$.

Para todos os elementos da população finita U , conhecidos, a equação de verossimilhança populacional é dada por:

$$l_U(\boldsymbol{\beta}) = \sum_{i \in U} \mathbf{u}_i(\boldsymbol{\beta}) = \mathbf{0} \quad (2.19)$$

onde

$$\mathbf{u}_i(\boldsymbol{\beta}) = \partial \log[f(\mathbf{y}_i; \boldsymbol{\beta})] / \partial \boldsymbol{\beta} \quad (2.20)$$

é o vetor $K \times 1$ escore do elemento i , com $i \in U$.

Seja $T = \sum_{i \in U} \mathbf{u}_i(\boldsymbol{\beta})$ a soma de vetores escores da população, em que T é um vetor de totais populacionais. Para a estimar o vetor de totais, pode-se utilizar um estimador ponderado da forma $\hat{T} = \sum_{i \in s} w_i \mathbf{u}_i(\boldsymbol{\beta})$ em que w_i são os pesos amostrais. Com isso, é possível obter um estimador $\boldsymbol{\beta}_U$ igualando o estimador \hat{T} do total T a zero (SILVA; PESSOA; LILA, 2002).

O estimador $\hat{\boldsymbol{\beta}}_{MPV}$ de $\boldsymbol{\beta}_U$ (e de $\boldsymbol{\beta}$) é

$$\hat{T}(\boldsymbol{\beta}_{MPV}) = \sum_{i \in s} w_i \mathbf{u}_i(\boldsymbol{\beta}) = \mathbf{0} \quad (2.21)$$

E, através da linearização de Taylor, obtêm-se a variância de aleatorização assintótica de $\hat{\boldsymbol{\beta}}_{MPV}$, a $V_p(\hat{\boldsymbol{\beta}}_{MPV})$ e o estimador correspondente, $\hat{V}_p(\hat{\boldsymbol{\beta}}_{MPV})$, dados por:

$$V_p(\hat{\boldsymbol{\beta}}_{MPV}) \simeq [J(\boldsymbol{\beta}_U)]^{-1} V_p \left[\sum_{i \in s} w_i \mathbf{u}_i(\boldsymbol{\beta}_U) \right] [J(\boldsymbol{\beta}_U)]^{-1} \quad (2.22)$$

e

$$\hat{V}_p(\hat{\boldsymbol{\beta}}_{MPV}) = [\hat{J}(\hat{\boldsymbol{\beta}}_{MPV})]^{-1} \hat{V}_p \left[\sum_{i \in s} w_i \mathbf{u}_i(\hat{\boldsymbol{\beta}}_{MPV}) \right] [\hat{J}(\hat{\boldsymbol{\beta}}_{MPV})]^{-1} \quad (2.23)$$

onde

$$J(\boldsymbol{\beta}_U) = \left. \frac{\partial T(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_U} = \sum_{i \in U} \left. \frac{\partial \mathbf{u}_i(\boldsymbol{\beta})}{\partial (\boldsymbol{\beta})} \right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_U} \quad (2.24)$$

e

$$\hat{J}(\hat{\boldsymbol{\beta}}_{MPV}) = \left. \frac{\partial \hat{T}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{MPV}} = \sum_{i \in s} w_i \left. \frac{\partial \mathbf{u}_i(\boldsymbol{\beta})}{\partial (\boldsymbol{\beta})} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{MPV}} \quad (2.25)$$

onde $V_p[\sum_{i \in s} w_i \mathbf{u}_i(\boldsymbol{\beta}_U)]$ é a matriz de variância de dimensão $P \times P$ do estimador do total populacional dos escores \hat{T} e $\hat{V}_p[\sum_{i \in s} w_i \mathbf{u}_i(\hat{\boldsymbol{\beta}}_{MPV})]$ é um estimador da matriz de variância-covariância dos totais amostrais dos escores ponderados.

O Erro Padrão (EP) de $\hat{\boldsymbol{\beta}}_{MPV}$ corresponde ao desvio padrão estimado para ele.

Para grandes amostras, Binder (1983) mostrou que a distribuição assintótica de $\hat{\boldsymbol{\beta}}_{MPV}$ é:

$$[\hat{V}_p(\hat{\boldsymbol{\beta}}_{MPV})]^{-1/2} (\hat{\boldsymbol{\beta}}_{MPV} - \boldsymbol{\beta}_U) \sim \text{NM}(\mathbf{0}, \mathbf{I}) \quad (2.26)$$

2.4.3.2 Razão de Chances

Para interpretar a associação entre as variáveis explicativas e a variável resposta de um modelo de regressão logística é, geralmente, utilizada a medida de razão de chances (*Odds Ratio*, ou simplesmente OR) que é a razão entre as chances de ocorrência de um

evento de interesse em 2 grupos que estão sendo comparados.

Segundo Montgomery, Peck e Vining (2012), para entender como obter a razão de chances no modelo de regressão logística em que se tem p variáveis explicativas x_i , sendo $i = 1, \dots, p$, e $\hat{\eta}(x_i)$ seu preditor linear ajustado, é preciso considerar o caso particular em que há apenas uma variável explicativa, proposto anteriormente. Seja assim:

$$\hat{\eta}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (2.27)$$

Então, o valor ajustado para $x_i + 1$ é:

$$\hat{\eta}(x_i + 1) = \hat{\beta}_0 + \hat{\beta}_1(x_i + 1) \quad (2.28)$$

e a diferença entre os dois valores previstos é

$$\hat{\eta}(x_i + 1) - \hat{\eta}(x_i) = \hat{\beta}_1 \quad (2.29)$$

Percebe-se que $\hat{\eta}(x_i)$ e $\hat{\eta}(x_i + 1)$, pela equação 2.15 são o log da razão de chances calculadas nos valores x_i e $x_i + 1$. Então, com base nas equações 2.28 e 2.29, e na relação que as chances π guardam com η (equação 2.15), a diferença entre os dois valores ajustados é:

$$\hat{\eta}(x_i + 1) - \hat{\eta}(x_i) = \ln \left(\frac{\hat{\pi}(x_i + 1)}{1 - \hat{\pi}(x_i + 1)} \right) - \ln \left(\frac{\hat{\pi}(x_i)}{1 - \hat{\pi}(x_i)} \right) = \hat{\beta}_1 \quad (2.30)$$

Aplicando-se a função exponencial aos 2 lados da igualdade tem-se, então, que:

$$\hat{OR}_i = \exp(\hat{\beta}_1) \quad (2.31)$$

Para interpretar a razão de chances:

- Se $OR = 1$, chance da ocorrência do evento na presença do fator ($X_i = 1$) é igual a chance de ocorrência do evento na ausência do fator ($X_i = 0$);
- Se $OR > 1$, a chance da ocorrência do evento na presença do fator ($X_i = 1$) é maior do que na ausência do fator ($X_i = 0$);
- Se $OR < 1$, a chance da ocorrência do evento na presença do fator ($X_i = 1$) é menor do que na ausência do fator ($X_i = 0$);

Para encontrar o intervalo de confiança (IC) para OR, é necessário primeiro encontrar o IC com nível de confiança $100(1 - \alpha)\%$ para a estatística $\hat{\beta}_{MPV}$ e a partir do estimador da variância $\hat{V}(\hat{\beta}_{MPV})$ dada pela equação 2.23. Sendo assim:

$$\left[\hat{\beta}_{MPV} - z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\beta}_{MPV})}; \hat{\beta}_{MPV} + z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\beta}_{MPV})} \right] \quad (2.32)$$

onde $z_{1-\alpha/2}$ é percentil da distribuição normal padrão e $P[-z_{1-\alpha/2} < Z < z_{1-\alpha/2}] = 1 - \alpha$.

Sabe-se que $\hat{OR}_i = \exp(\hat{\beta}_i)$, então o $IC_{1-\alpha/2}(OR)$ é:

$$\left[\exp \left(\hat{\beta}_{MPV} - z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\beta}_{MPV})} \right); \exp \left(\hat{\beta}_{MPV} + z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\beta}_{MPV})} \right) \right] \quad (2.33)$$

Se o IC da razão de chances incluir o valor 1, os dois grupos possuem a mesma chance de ocorrência do evento de interesse e, portanto, não existe efeito significativo da variável definidora dos grupos sobre a chance daquele evento ocorrer.

2.4.3.3 Avaliação da capacidade preditiva do modelo

Para medir o desempenho do modelo é necessário avaliar algumas medidas de qualidade do ajuste do modelo. Dentro do modelo de regressão logística, estima-se a probabilidade π_i de ocorrência de um evento de interesse para cada indivíduo da população ou amostra.

Para determinar se houve ocorrência do evento para cada indivíduo, segundo o modelo estimado, é necessário estipular um ponto de corte α nessas probabilidades, em que se $\pi_i \geq \alpha$ (teria ocorrido o evento de interesse) e, caso contrário, $Y = 0$ (ou seja, o evento de interesse não teria ocorrido para aquele indivíduo i). Na Tabela 8, pode-se observar como seriam consideradas os indivíduos em razão da comparação entre seu status original e o seu desfecho previsto pelo modelo.

Tabela 8: Distribuição das unidades segundo valor previsto e observado de Y

		Valor observado	
		Sucesso	Fracasso
Valor previsto	Sucesso	Verdadeiro positivo (VP)	Falso positivo (FP)
	Fracasso	Falso negativo (FN)	Verdadeiro negativo (VN)

A Taxa Global de acertos (TG) é a razão entre o número total acertos e o total populacional considerado, em que se é utilizada a amostra completa (sem repartição de

amostra treino e teste), é dada por:

$$TG = \left(\frac{VP + VN}{VP + VN + FP + FN} \right) \times 100 \quad (2.34)$$

TG , então, corresponde ao percentual classificações corretas no modelo. Quanto maior seu valor, melhor é a capacidade preditiva do modelo.

O coeficiente de determinação R^2 , em modelos clássicos de regressão linear, pode ser interpretado como o percentual de variação total dos dados que é explicado pelo modelo, estando sempre entre 0 e 1. Quanto mais próximo de 1, melhor o ajuste do modelo. Se $R^2 = 1$ o modelo tem ajuste perfeito.

No modelo de regressão logística, não existe estatística equivalente ao R^2 do modelo de regressão linear, uma vez que a variável resposta do modelo de regressão linear é de natureza contínua, e o modelo de regressão logística é categórica, ou seja, tem desfecho binário. Então, foi adotado como medida de qualidade de ajuste o pseudo- R^2 , que é parecido com o R^2 da regressão linear, mesmo não sendo interpretado da mesma forma (GONZALEZ, 2018).

O Pseudo- R^2 de Cox-Snell indica o ganho de informação do modelo em relação ao modelo nulo, aquele que possui apenas o intercepto, e considera o tamanho da amostra, porém nunca atinge o valor 1.

Lumley e Scott (2017) propuseram uma forma de encontrar o valor do R^2 de Cox-Snell em modelos lineares generalizados oriundos de planos amostrais complexos por:

$$R_{CS}^2 = 1 - e \left\{ 2 \left[\frac{\ln L_S(\hat{\beta}^{(0)}) - \ln L_S(\hat{\beta})}{\sum_{i=1}^n w_i} \right] \right\} \quad (2.35)$$

onde $\ln L_S(\hat{\beta}^{(0)})$ é o logaritmo da função de pseudo-verossimilhança do modelo nulo (modelo que considera apenas o intercepto); $\ln L_S(\hat{\beta})$ é o logaritmo da função de pseudo-verossimilhança do modelo proposto e w_i é o peso amostral do i -ésimo indivíduo.

Para este trabalho, será utilizado o Pseudo- R^2 de Nagelkerke. Segundo Marôco (2018), ele é definido por:

$$R_N^2 = \frac{R_{CS}^2}{1 - e \left\{ \frac{2 \ln L_S(\hat{\beta}^{(0)})}{\sum_{i=1}^n w_i} \right\}} \quad (2.36)$$

3 Resultados

Neste capítulo serão apresentados os resultados do estudo, dividindo-os em duas seções. A primeira conterá as análises preliminares envolvendo o perfil demográfico populacional e a identificação dos fatores associados à depressão. A segunda seção é formada pelos resultados da estimação dos modelos de regressão logística, e está subdividida em três partes: modelos propostos, comparação de modelos e modelo selecionado.

3.1 Análises preliminares

Nessa seção será realizada uma análise exploratória dos dados, considerando a expansão dos resultados para a população através da inclusão dos pesos amostrais nas análises. Isto consiste em investigar e resumir características da população através de análises gráficas e de tabelas-resumo.

Primeiramente, será tratado o perfil demográfico populacional. Em um segundo momento, será verificada a existência de associação entre os fatores sociodemográficos e a ocorrência de depressão (autodeclarada), através de testes de hipóteses.

3.1.1 Perfil Demográfico Populacional

Nesta seção será realizada uma análise exploratória em relação à população maior de 14 anos. Serão estudadas variáveis relacionadas ao perfil demográfico da população do Brasil retratada a partir dos dados da PNS, como: sexo, cor ou raça, estado civil, idade, escolaridade e renda.

Conforme a Figura 1, verifica-se que os percentuais de pessoas do sexo masculino e do sexo feminino são próximos: 47,05% e 52,95%, respectivamente. Quanto à variável relacionada à etnia, nomeada de cor ou raça pela pesquisa, destacam-se as pessoas que se autodeclararam brancas (aproximadamente 43,96% da população) e pardas (aproximadamente 43,62% da população).

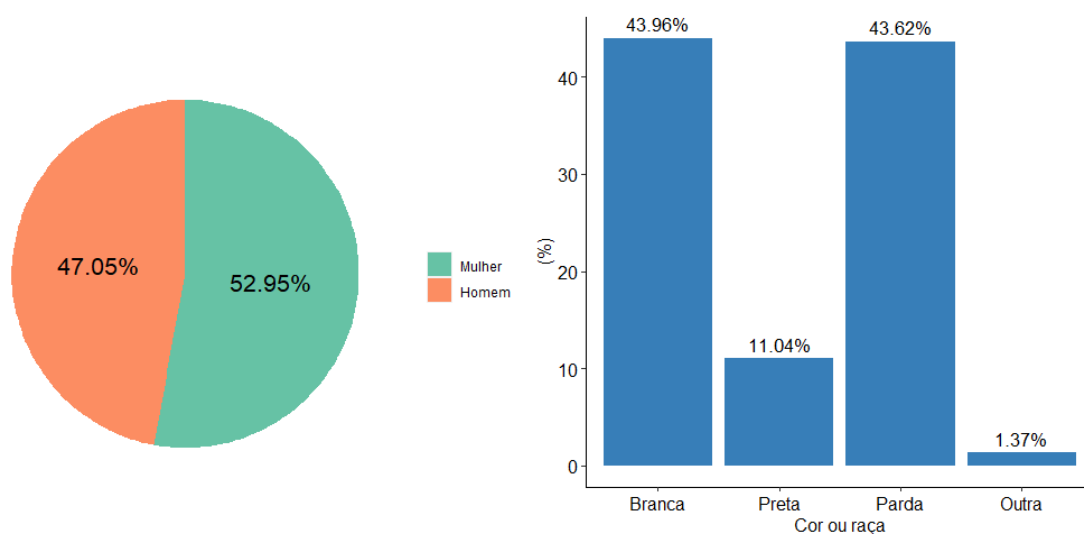


Figura 1: Distribuição por sexo e por Cor ou raça

A partir da Figura 2, é possível observar que a maioria da população é composta por pessoas solteiras, com aproximadamente 47,08%, seguida pelas casadas, com aproximadamente 40,21% da população.

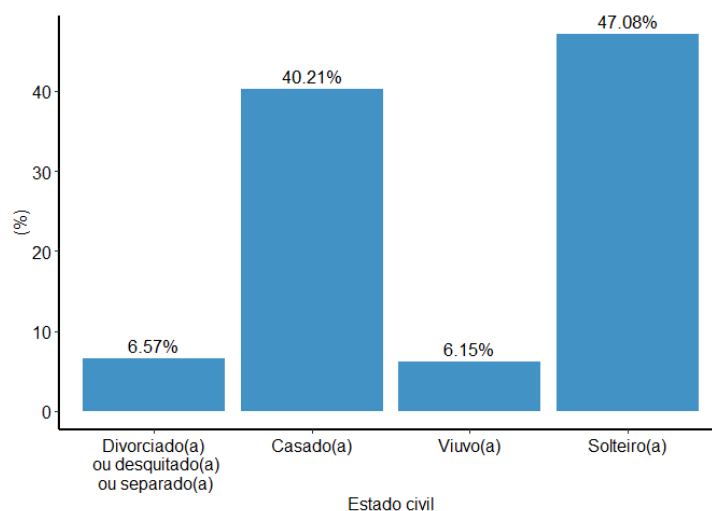


Figura 2: Distribuição por Estado Civil

Através da Figura 3 pode-se observar o histograma da variável Idade, por meio da Tabela 9, é possível verificar que a média de idade da população acima de 14 anos é de $43,24 \pm 18,12$ anos.

Tabela 9: Estatísticas descritivas da Idade (em anos)

Variável	Mínimo	Mediana	Média	Máximo	DP	Assimetria
Idade	15	42	43,24	112	18,12	0

DP: Desvio Padrão

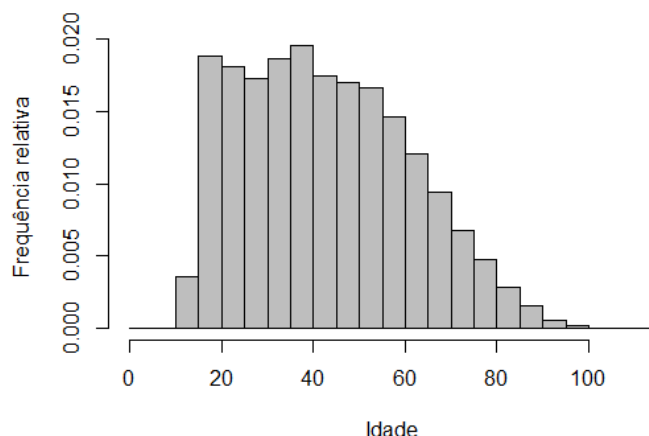


Figura 3: Histograma da Idade

Em relação às variáveis relacionadas à escolaridade, procurou-se na base de dados da PNS as variáveis que estivessem associadas a esta questão. Foram encontradas duas variáveis que poderiam indicar o grau de escolaridade do indivíduo. As variáveis D00901 e VDD004A, que são referentes às perguntas “Qual o curso mais elevado que frequentou?” e “Nível de instrução mais elevado alcançado (pessoas de 5 anos ou mais de idade) padronizado para o Ensino Fundamental - SISTEMA DE 9 ANOS”, respectivamente.

A partir da Figura 4, observa-se que a categoria predominante entre os brasileiros maiores de 14 anos, na variável D00901, é o ensino médio regular (37,24%), seguido de ensino fundamental ou ginásial (28,56%). A categoria de Educação de Jovens Adultos (EJA) obteve menor percentual, com 3,65%.

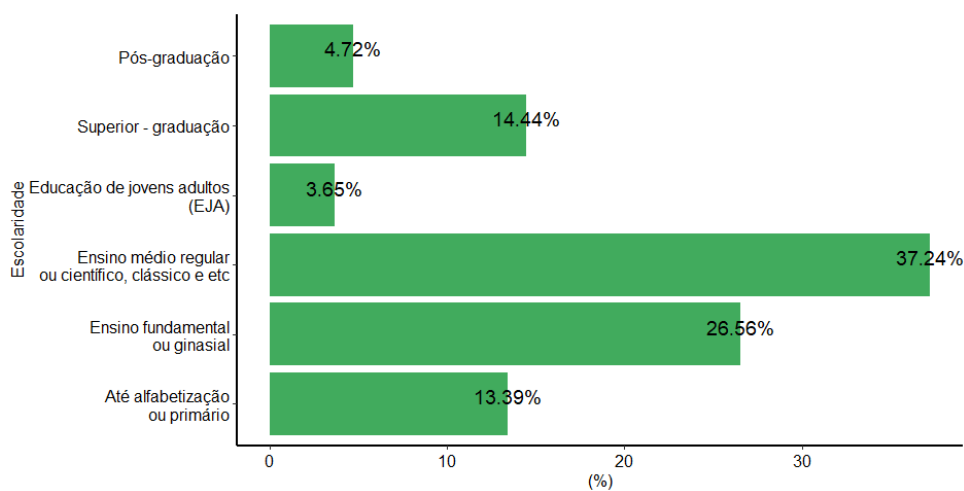


Figura 4: Distribuição percentual da população por Escolaridade através da variável D00901

Na variável VDD004A, na Figura 5, verifica-se a predominância em fundamental incompleto ou equivalente, e ensino médio completo ou equivalente, com percentuais de 28,28% e 28,22% respectivamente.

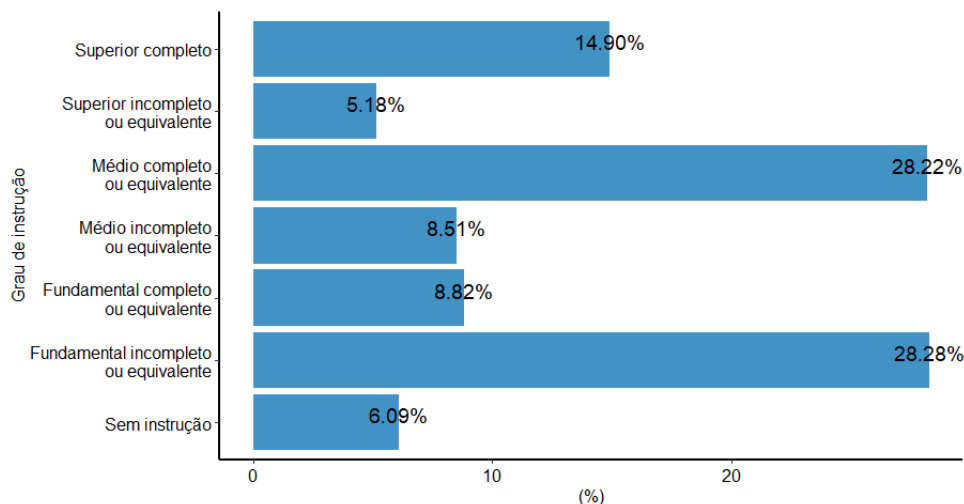


Figura 5: Distribuição percentual da população por Instrução através da variável VDD004A

Para variáveis relacionadas à renda foram encontradas três possíveis candidatas, são elas: VDF002 - “Rendimento domiciliar”, VDF003 - “Rendimento domiciliar *per capita*” e VDF004 - “Faixa de rendimento domiciliar *per capita*”. As duas primeiras são variáveis quantitativas e a última é variável qualitativa.

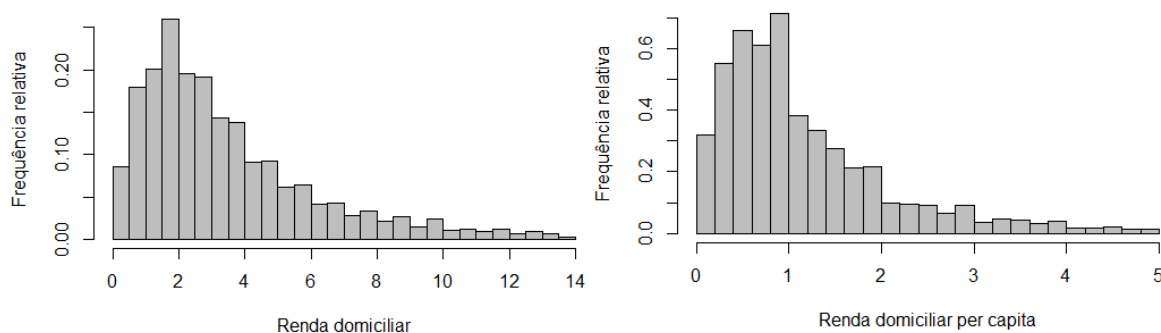
A Tabela 10 apresenta estatísticas descritivas referentes ao rendimento, tanto domiciliar quanto domiciliar *per capita*. Para o rendimento domiciliar, observa-se a renda mediana de R\$ 2.994,00; um rendimento médio de R\$ 4.621,50 e o valor máximo de R\$ 644.998,00. Para o rendimento domiciliar *per capita*, a mediana foi de R\$ 998,00 (que é igual ao salário mínimo de 2019) e a média foi igual a R\$ 1.661,50 e o valor máximo observado foi igual a R\$ 203.196,00. O desvio padrão para ambos os rendimentos foi que a média, indicando alta variabilidade e coeficientes de variação maiores que 100%. Através da avaliação da distribuição em termos de simetria, pode-se observar que as distribuições de renda domiciliar e renda *per capita* têm assimetria moderada à direita, indicando desigualdade da renda na população.

Na Figura 6, é possível observar os histogramas das variáveis quantitativas de renda domiciliar e renda domiciliar *per capita*. Para isso, foram retiradas 5% das rendas muito acima da população geral e as duas variáveis foram divididas por R\$ 1.000,00. Para a renda domiciliar, o limite foi em 13,8, ou seja, 95% da população tem a renda domiciliar até R\$ 13.800,00. E, para renda domiciliar *per capita*, o limite foi 5, ou seja, R\$ 5.000,00.

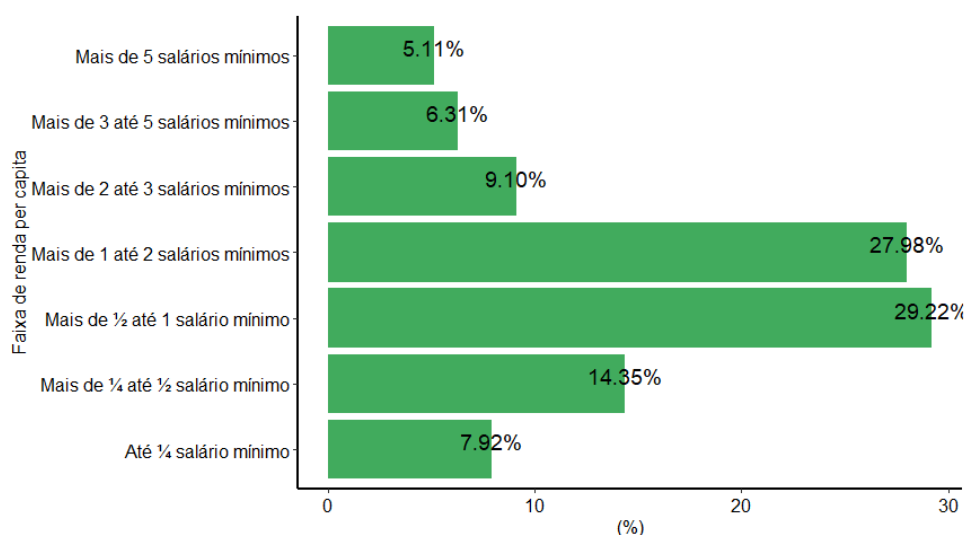
Tabela 10: Estatísticas descritivas do rendimento domiciliar total e *per capita* (em reais)

Variável	Mínimo	Mediana	Média	Máximo	DP	Assimetria
Rendimento domiciliar	0	2.994,00	4.621,50	644.998,00	6.996,00	0,22
Rendimento domiciliar <i>per capita</i>	0	998,00	1.611,50	203.196,00	2739,80	0,21

DP: Desvio Padrão

Figura 6: Histogramas da renda domiciliar e renda domiciliar *per capita*

Na Figura 7, percebe-se que a maioria da população está na faixa de rendimento domiciliar *per capita* de $\frac{1}{2}$ até 2 salários mínimos. São, também, destaques as categorias de $\frac{1}{2}$ até 1 salário mínimo e de 1 até 2 salários mínimos, com 29,22% e 27,98% da população, respectivamente.

Figura 7: Distribuição da faixa de rendimento *per capita*

Na Figura 8, é possível observar que a raça ou cor branca estão predominantemente em escolaridades mais elevadas e pessoas pardas prevalecem no grau de instrução “sem escolaridade”.

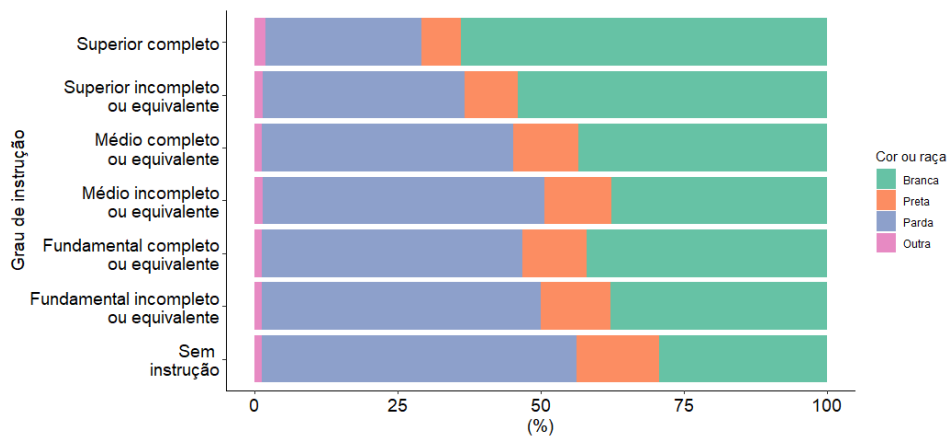


Figura 8: Distribuição de cor ou raça por grau de instrução

Na Figura 9, é interessante observar aumento de pessoas brancas a medida que a faixa de rendimento *per capita* aumenta. Entretanto, a quantidade de pessoas pardas diminuem a medida que a renda aumenta.

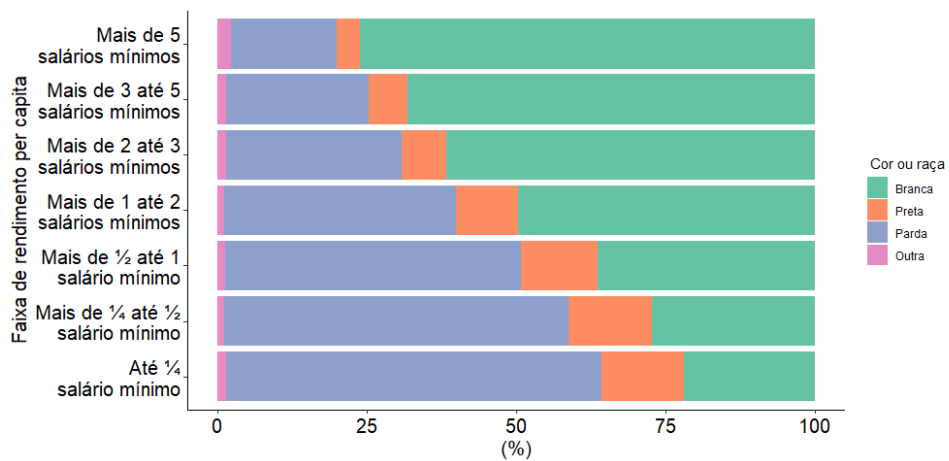


Figura 9: Distribuição de cor ou raça por faixa de renda *per capita*

3.1.2 Identificação dos fatores associados à depressão

Para a identificação dos fatores associados à depressão, foi realizada uma análise bivariada, através pelo teste qui-quadrado de independência, entre as variáveis de caracterização sociodemográfica e a variável de interesse, que nos informa se o indivíduo já teve (ou não) o diagnóstico de depressão feito por um profissional de saúde mental (psicólogo ou psiquiatra).

A partir da observação da Tabela 11, é possível notar que apenas a variável P034 - “Nos últimos três meses, o(a) Sr(a) praticou algum tipo de exercício físico ou esporte?” não está associada ao fato de ter sido diagnosticado (ou não) com depressão ($p - valor = 0,1748$). Para as demais variáveis, foi encontrada associação significativa com a ocorrência de depressão ($p - valor < 0,0001$ em todos os casos).

Tabela 11: Distribuição de frequências dos fatores sociodemográficos por existência de diagnóstico de depressão e teste qui-quadrado para independência

Variáveis	Diagnóstico de depressão?(%)		p-valor
	Não	Sim	
Sexo			
Homem	94,51	5,49	<0,0001
Mulher	84,66	15,34	
Cor ou raça			
Branca	87,07	12,93	<0,0001
Preta	91,42	8,58	
Parda	90,75	9,25	
Outra	92,47	7,53	
Estado Civil			
Divorciado(a)	81,83	18,17	<0,0001
Casado(a)	89,89	10,11	
Viúvo(a)	85,62	14,38	
Solteiro(a)	91,07	8,93	
Escolaridade			
Até alfabetização ou primário	87,69	12,31	<0,0001
Ensino fundamental regular ou ginásial	89,27	10,73	
Ensino médio regular ou científico, clássico e etc	90,61	9,39	
Educação de jovens e Adultos (EJA)	88,26	11,74	
Superior - graduação	87,98	12,02	
Pós-graduação	84,66	15,34	

Variáveis	Diagnóstico de depressão?(%)		p-valor
	Não	Sim	
Instrução			
Sem instrução	91,76	8,24	
Fundamental incompleto ou equivalente	88,19	11,81	
Fundamental completo ou equivalente	90,41	9,59	<0,0001
Médio incompleto ou equivalente	91,05	8,95	
Médio completo ou equivalente	90,59	9,41	
Superior incompleto ou equivalente	87,64	12,36	
Superior completo	86,98	13,02	
Sono			
Quase todos dias	71,09	28,91	
Mais da metade dos dias	75,80	24,20	<0,0001
Menos da metade dos dias	86,49	13,51	
Nenhum dia	95,01	4,99	
Álcool			
Bebo	91,19	8,81	<0,0001
Não bebo	87,83	12,17	
Exercício			
Não	88,96	11,04	0,1748
Sim	89,61	10,39	
Tabaco			
Fumo	87,59	12,41	<0,0001
Não fumo	89,47	10,53	
Faixa de renda			
Até $\frac{1}{4}$ salário mínimo	90,63	9,37	
Mais de $\frac{1}{4}$ até $\frac{1}{2}$ salário mínimo	90,79	9,21	
Mais de $\frac{1}{2}$ até 1 salário mínimo	89,82	10,18	
Mais de 1 até 2 salários mínimos	89,34	10,66	<0,0001
Mais de 2 até 3 salários mínimos	87,56	12,44	
Mais de 3 até 5 salários mínimos	87,50	12,50	
Mais de 5 salários mínimos	85,59	14,41	

A seguir, serão apresentados gráficos que ilustram as diferenças entre as categorias das variáveis que tiveram diferença significativa em relação à proporção de pessoas com depressão (autodeclarada).

Na Figura 10 é possível observar que mulheres foram proporcionalmente mais diagnosticadas com depressão. E, quanto à cor ou raça, pessoas brancas foram mais diagnosticadas com depressão do que os demais grupos da mesma variável.

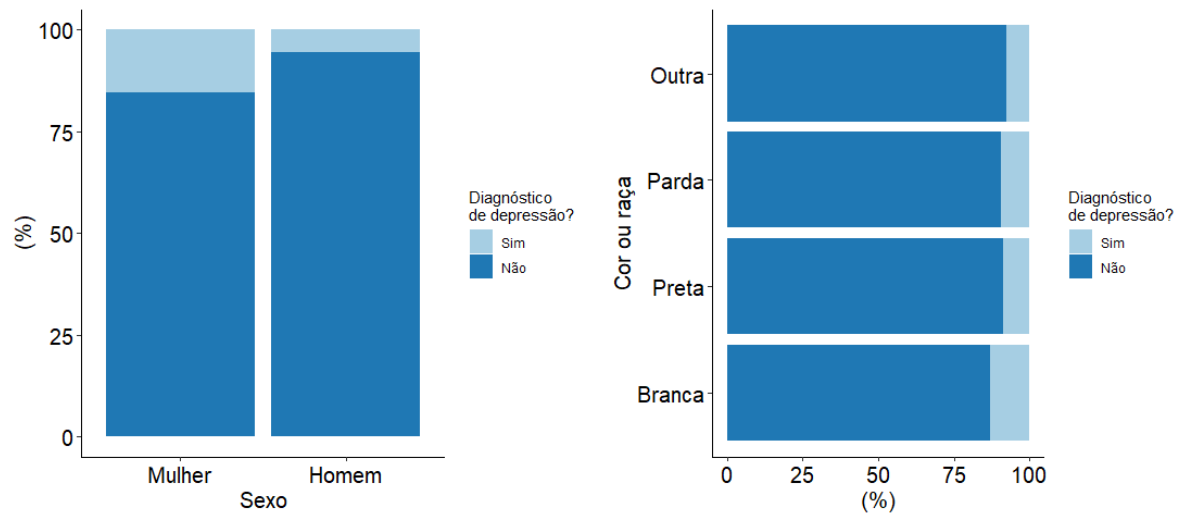


Figura 10: Distribuição de diagnóstico de depressão por sexo e e Cor ou raça

A Figura 11 mostra que pessoas casadas ou solteiras foram menos diagnosticadas com depressão.

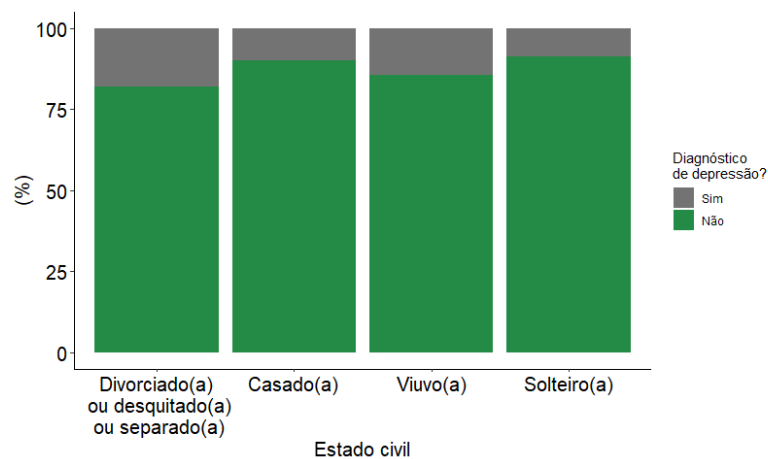


Figura 11: Distribuição de diagnóstico de depressão por estado civil

Conforme a Figura 12, na variável “Curso mais elevado que frequentou”, pode-se observar a predominância do diagnóstico de depressão em pessoas com escolaridade mais baixa, indo até a alfabetização, mas, também, em pessoas com pós-graduação.

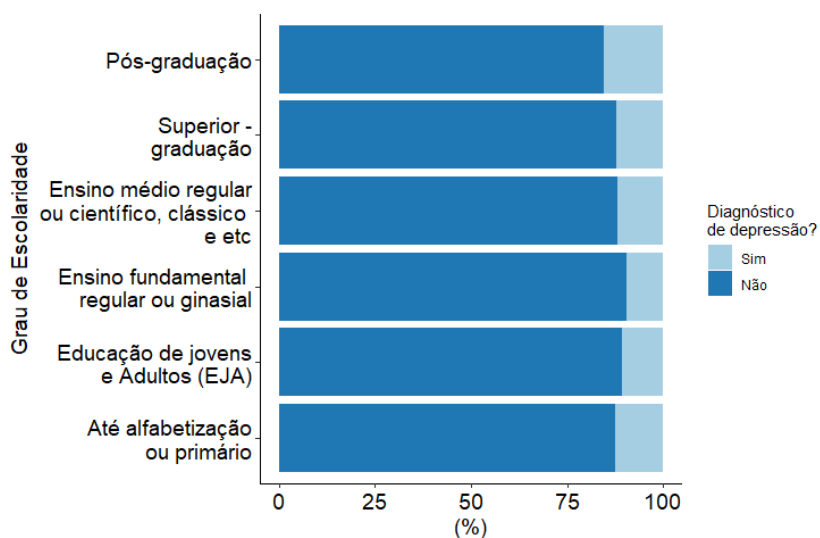


Figura 12: Distribuição de diagnóstico de depressão por grupos de escolaridade

Em relação “Nível de instrução mais alto alcançado”, pessoas com nível superior incompleto ou completo tiveram mais diagnósticos de depressão, conforme Figura 13.

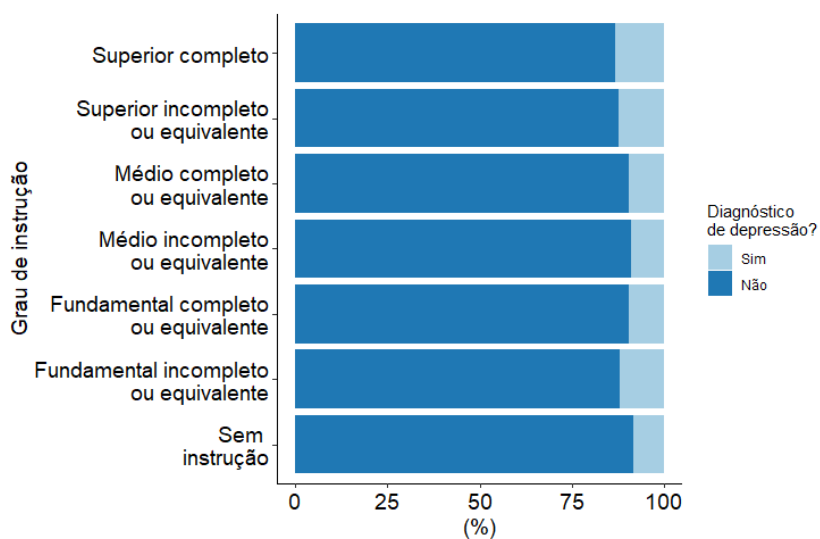


Figura 13: Distribuição de diagnóstico de depressão por grupos de grau de instrução

A partir da Figura 14 observa-se a tendência do diagnóstico de depressão em pessoas com maior frequência de problemas com sono e que não consomem bebida alcoólica.

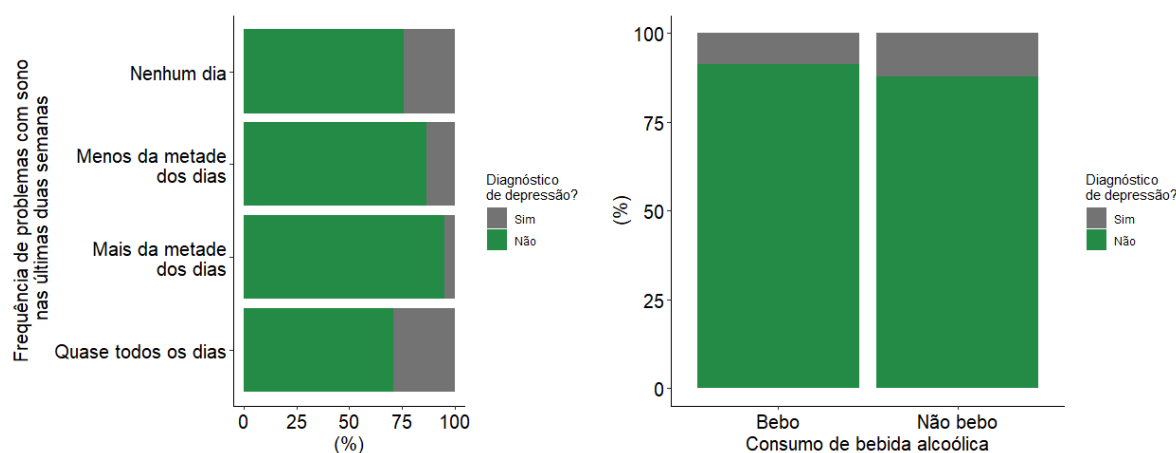


Figura 14: Distribuição de diagnóstico de depressão em relação a frequência de problemas com sono e ao consumo de bebida alcoólica

Por último, na Figura 15, observa-se que fumantes foram, proporcionalmente, mais frequentemente diagnosticados com depressão. É possível constatar que, quanto maior a faixa de rendimento domiciliar *per capita*, há maior presença de diagnósticos de depressão.

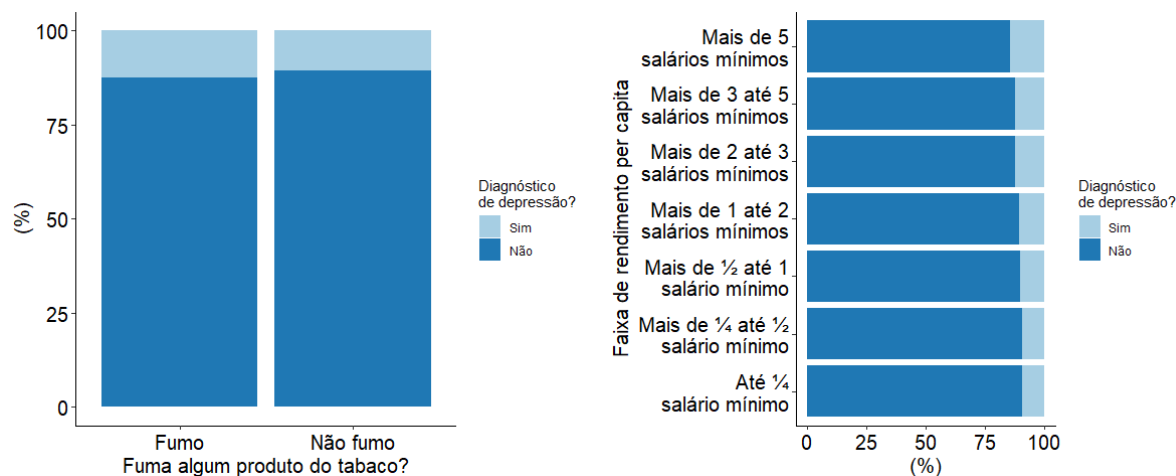


Figura 15: Distribuição de diagnóstico de depressão em relação ao consumo de tabaco e a faixa de rendimento *per capita*

Na Tabela 12, observa-se as médias das idades, rendas domiciliares e rendas domiciliares *per capita* entre os grupos de pessoas que já obtiveram o diagnóstico de depressão (ou não). Percebe-se que as médias foram maiores para todas as variáveis citadas para as pessoas que se autodeclararam como tendo sido diagnosticadas com depressão.

Tabela 12: Intervalo de média \pm desvio padrão para as variáveis idade, renda domiciliar e renda domiciliar *per capita* por ocorrência de diagnóstico de depressão

Variáveis	Diagnóstico de depressão?	
	Não	Sim
Idade (Em anos)	46,7 \pm 0, 128	50,2 \pm 0, 284
Rendimento domiciliar (Em reais)	4195 \pm 64, 1	4406 \pm 120
Rendimento domiciliar <i>per capita</i> (Em reais)	1753 \pm 31, 6	2007,00 \pm 69, 9

Greenland (1989) propõem que se utilize um nível de significância maior do que 5% para evitar que associações inicialmente que não sejam fortemente significativas possam ser testadas como candidatas à inclusão na modelagem conjunta realizada através de modelos de regressão linear ou logística. Com isso, todas as variáveis com $p - valor \leq 0,25$ no teste qui-quadrado foram inseridas no modelo de regressão logística, levando, então, à inserção de todas as variáveis levantadas no modelo completo.

3.2 Modelos de Regressão Logística

Nesta seção, serão apresentados os modelos com todas as variáveis selecionadas na Secção 2.2. Em cada modelo inicial, foi realizada uma combinação de variáveis que tinham significados semelhantes referentes à escolaridade e à renda, para evitar redundância e problemas de multicolinearidade no modelo. Na Tabela 13 é possível observar como foi feita esta combinação em função da escolha entre as variáveis D00901 e VDD004A, além da escolha entre as variáveis VDF002, VDF003 e VDF004.

Tabela 13: Combinação das variáveis de escolaridade e renda em cada modelo proposto.

Modelo	Variável	
	Escolaridade	Renda
1	D00901	VDF002
2	D00901	VDF003
3	D00901	VDF004
4	VDD004A	VDF002
5	VDD004A	VDF003
6	VDD004A	VDF004

A Tabela 14 apresenta a categoria de referência de cada variável qualitativa, selecionada a partir da revisão de literatura (apresentada na Introdução), e utilizada na etapa de modelagem.

Tabela 14: Categorias de referência para as variáveis qualitativas

Variável	Descrição (resumida)	Categoria de referência
C006	Sexo	Mulher
C009	Etnia	Branca
C011	Estado civil	Divorciado
D00901	Escolaridade	Até alfabetização ou primário
N010	Sono	Quase todos os dias
P027	Álcool	Bebo
P034	Exercício	Não
P050	Tabaco	Fumo
VDF004	Faixa de renda	Até $\frac{1}{4}$ salário mínimo
VDD004A	Instrução	Sem instrução

3.2.1 Modelos Propostos

A seguir serão apresentados todos os modelos propostos utilizando as variáveis selecionadas e, nos testes de hipóteses sobre os coeficientes, será utilizado o nível de significância de 5%. Após a estimação de cada um dos modelos, será feita uma avaliação da qualidade do ajuste para a escolha do modelo a ser adotado.

Os resultados da estimação do Modelo 1 foram apresentados na Tabela 15, utilizando as variáveis D00901 e VDF002 e é possível observar que os coeficientes das variáveis Exercício e Renda domiciliar (quantitativa) não foram significativos.

No Modelo 2 foram utilizadas as variáveis D00901 e VDF003, e seus resultados foram dispostos na Tabela 16, onde pode-se observar que, de forma semelhante ao ocorrido com o modelo 1, as variáveis Exercício e Renda *per capita* (quantitativa) também não foram consideradas significativas na predição da ocorrência de depressão.

Para o Modelo 3, utilizando as variáveis D00901 e VDF004, apenas o coeficiente da variável Exercício foi considerado não significativo, conforme Tabela 17.

No Modelo 4, conforme Tabela 18, mostrando a utilização das variáveis VDD004A e VDF002, verifica-se que os coeficientes das variáveis Exercício e Renda domiciliar não foram significativos no modelo.

Na Tabela 19, referente ao Modelo 5, com as variáveis VDD004A e VDF003, nota-se a não significância dos coeficientes das variáveis Exercício e Renda *per capita*.

Por último, o Modelo 6 que adota a combinação das variáveis VDD004A e VDF004, apresentado na Tabela 20, também apresentou coeficiente da variável Exercício como não significativo. E, diferentemente dos modelos 4 e 5, o coeficiente da variável Idade foi não

significativo neste último modelo.

A seguir são apresentadas Tabelas 15, 16, 17, 18, 19 e 20, representando os resultados dos modelos de 1 a 6.

Tabela 15: Estimativas do Modelo inicial 1 completo

Variáveis	β	EP (β)	p - valor
Intercepto	-0,39	0,14	0,01
Sexo (Homem)	-0,94	0,05	< 0,01
Idade	0,01	0,001	< 0,01
Etnia (Preta)	-0,38	0,08	< 0,01
Etnia (Parda)	-0,32	0,05	< 0,01
Etnia (Outra)	-0,56	0,17	< 0,01
Estado civil (Casado)	-0,40	0,07	< 0,01
Estado civil (Viúvo)	-0,51	0,08	< 0,01
Estado civil (Solteiro)	-0,48	0,07	< 0,01
Escolaridade (Ensino fund.regular ou ginásial)	0,14	0,07	0,06
Escolaridade (Ensino médio regular ou etc)	0,08	0,07	0,28
Escolaridade (Educação de jovens e adultos EJA)	0,14	0,11	0,17
Escolaridade (Superior - graduação)	0,24	0,08	< 0,01
Escolaridade (Pós-graduação)	0,49	0,11	< 0,01
Sono (Nenhum dia)	-1,87	0,06	< 0,01
Sono (Menos da metade dos dias)	-0,90	0,07	< 0,01
Sono (Mais da metade dos dias)	-0,17	0,08	0,03
Álcool I(Não bebo)	0,24	0,05	< 0,01
Exercício (Sim)	0,07	0,06	0,27
Tabaco (Não fumo)	-0,32	0,06	< 0,01
Renda	-3,181e-07	3,009e-06	0,92

Tabela 16: Estimativas do Modelo inicial 2 completo

Variável	β	EP (β)	p - valor
Intercepto	-0,38	0,14	0,01
Sexo (Homem)	-0,94	0,05	< 0,01
Idade	0,01	0,001	< 0,01
Etnia (Preta)	-0,38	0,07	< 0,01
Etnia (Parda)	-0,32	0,05	< 0,01
Etnia (Outra)	-0,56	0,17	< 0,01
Estado civil (Casado)	-0,40	0,07	< 0,01
Estado civil (Viúvo)	-0,51	0,08	< 0,01
Estado civil (Solteiro)	-0,48	0,07	< 0,01
Escolaridade (Ensino fund.regular ou ginásial)	0,14	0,07	0,07
Escolaridade (Ensino médio regular ou etc)	0,07	0,07	0,31
Escolaridade (Educação de jovens e adultos EJA)	0,14	0,11	0,18
Escolaridade (Superior - graduação)	0,23	0,08	< 0,01
Escolaridade (Pós-graduação)	0,46	0,11	< 0,01
Sono (Nenhum dia)	-1,87	0,06	< 0,01
Sono (Menos da metade dos dias)	-0,90	0,07	< 0,01
Sono (Mais da metade dos dias)	-0,17	0,08	0,03
Álcool (Não bebo)	0,24	0,05	< 0,01
Exercício (Sim)	0,07	0,06	0,28
Tabaco (Não fumo)	-0,32	0,06	< 0,01
Renda <i>per capita</i>	4,036e-06	6,226e-06	0,52

Tabela 17: Estimativas do Modelo inicial 3 completo

Variável	β	EP (β)	p - valor
Intercepto	-0,41	0,16	0,01
Sexo (Homem)	-0,95	0,05	< 0,01
Idade	0,004	0,002	0,02
Etnia (Preta)	-0,35	0,08	< 0,01
Etnia (Parda)	-0,29	0,05	< 0,01
Etnia (Outra)	-0,54	0,17	< 0,01
Estado civil (Casado)	-0,40	0,07	< 0,01
Estado civil (Viúvo)	-0,53	0,08	< 0,01
Estado civil (Solteiro)	-0,47	0,07	< 0,01
Escolaridade (Ensino fund.regular ou ginásial)	0,13	0,08	0,09
Escolaridade (Ensino médio regular ou etc)	0,03	0,08	0,66
Escolaridade (Educação de jovens e Adultos EJA)	0,13	0,11	0,23
Escolaridade (Superior - graduação)	0,13	0,09	0,12
Escolaridade (Pós-graduação)	0,34	0,12	< 0,01
Sono (Nenhum dia)	-1,87	0,06	< 0,01
Sono (Menos da metade dos dias)	-0,90	0,07	< 0,01
Sono (Mais da metade dos dias)	-0,17	0,08	0,03
Álcool (Não bebo)	0,26	0,05	< 0,01
Exercício (Sim)	0,05	0,06	0,43
Tabaco (Não fumo)	-0,33	0,06	< 0,01
Faixa de renda (Mais de $\frac{1}{4}$ até $\frac{1}{2}$ s.m)	0,02	0,09	0,81
Faixa de renda (Mais de $\frac{1}{2}$ até 1 s.m)	0,09	0,09	0,29
Faixa de renda (Mais de 1 até 2 s.m)	0,14	0,09	0,11
Faixa de renda (Mais de 2 até 3 s.m)	0,31	0,12	0,01
Faixa de renda (Mais de 3 até 5 s.m)	0,26	0,12	0,03
Faixa de renda (Mais de 5 s.m)	0,30	0,13	0,02

Tabela 18: Estimativas do Modelo inicial 4 completo

Variável	β	EP (β)	p - valor
Intercepto	-0,66	0,15	< 0,01
Sexo (Homem)	-0,92	0,05	< 0,01
Idade	0,005	0,001	< 0,01
Etnia (Preta)	-0,42	0,07	< 0,01
Etnia (Parda)	-0,30	0,04	< 0,01
Etnia (Outros)	-0,57	0,15	< 0,01
Estado civil (Casado)	-0,43	0,07	< 0,01
Estado civil (Viúvo)	-0,53	0,08	< 0,01
Estado civil (Solteiro)	-0,51	0,07	< 0,01
Instrução (Fundamental incomp. ou equiv.)	0,48	0,08	< 0,01
Instrução (Fundamental comp. ou equiv.)	0,39	0,11	< 0,01
Instrução (Médio incomp. ou equiv.)	0,49	0,12	< 0,01
Instrução (Médio comp. ou equiv.)	0,41	0,09	< 0,01
Instrução (Superior incomp. ou equiv.)	0,73	0,13	< 0,01
Instrução (Superior completo)	0,66	0,10	< 0,01
Sono (Nenhum dia)	-1,88	0,05	< 0,01
Sono (Menos da metade dos dias)	-0,89	0,06	< 0,01
Sono (Mais da metade dos dias)	-0,18	0,07	0,01
Álcool (Não bebo)	0,21	0,05	< 0,01
Exercício (Sim)	0,06	0,06	0,28
Tabaco (Não fumo)	-0,32	0,06	< 0,01
Renda	2,034e-06	2,687e-06	0,45

Tabela 19: Estimativas do Modelo inicial 5 completo

Variável	β	EP (β)	$p - valor$
Intercepto	-0,65	0,15	< 0,01
Sexo (Homem)	-0,92	0,05	< 0,01
Idade	0,004	0,001	< 0,01
Etnia (Preta)	-0,42	0,07	< 0,01
Etnia (Parda)	-0,30	0,04	< 0,01
Etnia (Outros)	-0,57	0,15	< 0,01
Estado civil (Casado)	-0,42	0,06	< 0,01
Estado civil (Viúvo)	-0,53	0,08	< 0,01
Estado civil (Solteiro)	-0,51	0,07	< 0,01
Instrução (Fund. Incomp. ou equiv.)	0,48	0,08	< 0,01
Instrução (Fund. comp. ou equiv.)	0,38	0,11	< 0,01
Instrução (Médio incomp. ou equiv.)	0,49	0,12	< 0,01
Instrução (Médio comp. ou equiv.)	0,40	0,09	< 0,01
Instrução (Superior incomp. ou equiv.)	0,72	0,13	< 0,01
Instrução (Superior completo)	0,64	0,10	< 0,01
Sono (Nenhum dia)	-1,88	0,05	< 0,01
Sono (Menos da metade dos dias)	-0,89	0,06	< 0,01
Sono (Mais da metade dos dias)	-0,18	0,07	0,01
Álcool(Não bebo nunca)	0,22	0,05	< 0,01
Exercício (Sim)	0,06	0,06	0,30
Tabaco (Não fumo atualmente)	-0,32	0,06	< 0,01
Renda <i>per capita</i>	1.031e-05	5.934e-06	0,08

Tabela 20: Estimativas do Modelo inicial 6 completo

Variável	β	EP (β)	$p - valor$
Intercepto	-0,63	0,16	< 0,01
Sexo (Homem)	-0,94	0,05	< 0,01
Idade	0,003	0,001	0,07
Etnia (Preta)	-0,39	0,07	< 0,01
Etnia (Parda)	-0,27	0,04	< 0,01
Etnia (Outra)	-0,55	0,15	< 0,01
Estado civil (Casado)	-0,42	0,06	< 0,01
Estado civil (Viúvo)	-0,55	0,08	< 0,01
Estado civil (Solteiro)	-0,51	0,07	< 0,01
Instrução (Fundamental incomp. ou equiv.)	0,45	0,08	< 0,01
Instrução (Fundamental comp. ou equiv.)	0,33	0,11	< 0,01
Instrução (Médio incomp. ou equiv.)	0,43	0,13	< 0,01
Instrução (Médio comp. ou equiv.)	0,32	0,09	< 0,01
Instrução (Superior incomp. ou equiv.)	0,60	0,13	< 0,01
Instrução (Superior completo)	0,48	0,10	< 0,01
Sono (Nenhum dia)	-1,89	0,05	< 0,01
Sono (Menos da metade dos dias)	-0,89	0,06	< 0,01
Sono (Mais da metade dos dias)	-0,18	0,07	0,01
Álcool (Não bebo)	0,24	0,05	< 0,01
Exercício (Sim)	0,04	0,06	0,51
Tabaco (Não fumo)	-0,33	0,06	< 0,01
Faixa de renda (Mais de $\frac{1}{4}$ até $\frac{1}{2}$ s.m)	0,01	0,08	0,89
Faixa de renda (Mais de $\frac{1}{2}$ até 1 s.m)	0,06	0,08	0,42
Faixa de renda (Mais de 1 até 2 s.m)	0,14	0,08	0,07
Faixa de renda (Mais de 2 até 3 s.m)	0,31	0,11	< 0,01
Faixa de renda (Mais de 3 até 5 s.m)	0,30	0,11	0,01
Faixa de renda (Mais de 5 s.m)	0,41	0,11	< 0,01

Resumidamente, a Tabela 21 apresenta as variáveis que foram excluídas de cada modelo. Observa-se que os modelos 1 e 2 são iguais e os modelos 4 e 5, também.

Tabela 21: Variáveis excluídas de cada modelo

Modelo	Variáveis excluídas
1	Exercício e Renda
2	Exercício e Renda <i>per capita</i>
3	Exercício
4	Exercício e Renda
5	Exercício e Renda <i>per capita</i>
6	Exercício e Idade

3.2.2 Comparação de modelos

Para avaliar os modelos, quanto à sua qualidade ajuste, foram utilizados os Pseudo- R^2 de Nagelkerke, e para comparar os modelos quanto à capacidade de predição, foi adotada a taxa global de acertos.

Na Tabela 22, é possível observar as métricas de comparação de todos os modelos reduzidos (após a exclusão das variáveis que não tiveram coeficientes significativamente diferentes de zero), apresentados anteriormente. Em negrito, pode-se observar a as melhores medidas dos modelos.

Tabela 22: Métricas dos modelos reduzidos

Modelos reduzidos	Pseudo- R^2 Nagelkerke	Taxa global de acertos
1 e 2	0,8944	89,3382
3	0,8946	89,3463
4 e 5	0,8409	89,4522
6	0,8413	89,4181

Todos os modelos têm valores próximos para as medidas de qualidade de ajuste. O critério adotado, então, para escolha do melhor modelo foi a taxa global de acertos. O modelo incluiu incluiu as variáveis: sexo, idade, cor ou raça, estado civil, instrução, sono, álcool e tabaco.

3.2.3 Modelo selecionado

Após a seleção do modelo de melhor ajuste segundo os critérios utilizados, é possível interpretar os coeficientes apresentados no modelo selecionado, representado pela Tabela 23.

Tabela 23: Estimativas de parâmetros para o modelo selecionado, OR e intervalos de confiança

Variável	β	EP(β)	OR	IC(95%)	$p - valor$
Intercepto	-0,64	0,15	0,53	[0,39 ; 0,71]	< 0,01
Sexo (Homem)	-0,92	0,05	0,40	[0,36 ; 0,44]	< 0,01
Idade	0,004	0,001	1,004	[1,002 ; 1,007]	< 0,01
Etnia (Preta)	-0,43	0,07	0,65	[0,57 ; 0,75]	< 0,01
Etnia (Parda)	-0,30	0,04	0,74	[0,68 ; 0,80]	< 0,01
Etnia (Outra)	-0,57	0,15	0,56	[0,42 ; 0,76]	< 0,01
Estado civil (Casado)	-0,43	0,06	0,65	[0,58 ; 0,74]	< 0,01
Estado civil (Viúvo)	-0,53	0,08	0,59	[0,51 ; 0,69]	< 0,01
Estado civil (Solteiro)	-0,51	0,07	0,60	[0,52 ; 0,68]	< 0,01
Instrução (Fund. incomp.o ou equiv.)	0,49	0,08	1,63	[1,39 ; 1,90]	< 0,01
Instrução (Fundamental comp. ou equiv.)	0,40	0,11	1,49	[1,21 ; 1,83]	< 0,01
Instrução (Médio incomp. ou equiv.)	0,50	0,12	1,65	[1,30 ; 2,11]	< 0,01
Instrução (Médio comp. ou equiv.)	0,42	0,09	1,52	[1,28 ; 1,81]	< 0,01
Instrução (Superior incomp. ou equiv.)	0,75	0,13	2,13	[1,65 ; 2,73]	< 0,01
Instrução (Superior completo)	0,70	0,09	2,01	[1,68 ; 2,40]	< 0,01
Sono (Nenhum dia)	-1,88	0,05	0,15	[0,14 ; 0,17]	< 0,01
Sono (Menos da metade dos dias)	-0,89	0,06	0,41	[0,36 ; 0,46]	< 0,01
Sono (Mais da metade dos dias)	-0,18	0,07	0,83	[0,72 ; 0,96]	0,01
Álcool (Não bebo)	0,20	0,05	1,23	[1,12 ; 1,34]	< 0,01
Tabaco (Não fumo)	-0,31	0,06	0,73	[0,66 ; 0,82]	< 0,01

Através das Razões de Chance (OR) encontradas é possível concluir que:

- Ser homem é fator de proteção para o diagnóstico de depressão. Homens têm chance de ser diagnosticados com depressão 60% menor do que as mulheres. O que é equivalente a dizer que a chance de uma mulher ser diagnosticada com depressão é $\frac{1}{0,4} = 2,5$ vezes maior que a chance de um homem ser diagnosticado com depressão. Ou seja, a mulher tem mais que o dobro da chance do homem de ter depressão.
- À medida que a pessoa envelhece um ano, a chance da pessoa ser diagnosticada com depressão aumenta em 0,4%. Com isso, pode-se afirmar também, que a cada 3 anos a chance da pessoa ser diagnosticada com depressão aumenta em 1%.
- Ser de etnia diferente da categoria “branca” se coloca como fator de proteção para a ocorrência de depressão. A chance de ter depressão é 35% menor em indivíduos de

cor ou raça preta, 26% menor nos indivíduos declaradamente pardos e 44% menor em outras etnias.

- Ser divorciado foi percebido como fator de risco para o diagnóstico de depressão. Pessoas casadas, viúvas e solteiras têm menor chance de serem diagnosticadas com depressão em relação as pessoas divorciadas, com percentuais de 35% , 41% e 40% respectivamente.
- Não ter instrução é fator de proteção para o diagnóstico de depressão. Em comparação ao grupo formado por pessoas sem instrução, os grupos populacionais com maior escolaridade igual a ensino fundamental incompleto, fundamental completo, médio incompleto, médio completo, superior incompleto e superior completo tiveram chances maiores de terem sido diagnosticadas com depressão (63%, 49%, 65%, 52%, 113% e 101%, respectivamente).
- Ter problema com sono com maior frequência, na categoria “quase todos dias” é fator de risco para a depressão. A chance de ser diagnosticado com depressão é menor nas categorias de “nenhum dia”, “menos da metade dos dias” e “mais da metade dos dias”, com percentuais de 85%, 59% e 17%, respectivamente.
- Não consumir bebida alcoólica é fator de risco para depressão. A pessoa que não bebia tem chance 23% maior de ser diagnosticado com depressão do que a pessoa que bebe.
- Não fumar é fator de proteção para o diagnóstico de depressão. Pessoas que não fumavam tinham 27% menos chance de serem diagnosticadas com esta doença quando comparado a pessoas que fumam.

3.2.4 Discussão dos resultados

O presente estudo avaliou fatores de risco e de proteção relacionados a pessoas que já haviam recebido o diagnóstico de depressão efetivamente dado por algum profissional da área de saúde mental (psicólogo ou psiquiatra). Além disso, o modelo final, proposto por esta pesquisa, foi estimado tendo por base a informação prestada por cada respondente a respeito do diagnóstico clínico da depressão, bem como sobre os fatores que foram testados. Neste sentido, algumas considerações precisam ser feitas para contextualizar os comentários dentro da realidade da população estudada.

Pode-se intuir que existam, na população, pessoas com depressão que não foram efetivamente diagnosticadas e que, portanto, não estariam entrando na amostra considerada no processo de estimação. A falta de acesso ao diagnóstico da doença é um importante fator que pode ter afetado o padrão de risco observado entre os grupos e pode ter favorecido a obtenção de um perfil demográfico prioritariamente formado por pessoas brancas, de escolaridades mais altas e com maiores rendas.

Por exemplo, o grupo "sem instrução" teve menor risco de ter tido diagnóstico de depressão do que os grupos com maior nível de instrução. Isso pode estar relacionado a um problema de acesso ou, mesmo, de desconhecimento individual a respeito da doença, sendo esperado, então, para este grupo uma menor proporção de indivíduos diagnosticados com a doença.

Por outro lado, pelo perfil demográfico apresentado, percebe-se que existe uma associação entre as variáveis cor ou raça e grau de instrução, com maior participação da população branca especialmente naqueles grupos com maior escolaridade. Da mesma forma, pelo mesmo perfil, percebeu-se que existe uma maior participação de pessoas de cor ou raça branca nas faixas de maior rendimento *per capita*, sugerindo que possa existir uma influência desse efeito sobre o acesso ao diagnóstico de depressão e, assim, à própria identificação da doença. Como visto na Tabela 11, o percentual de diagnóstico de depressão aumenta conforme a faixa de renda aumenta, onde há associação entre cor ou raça e a renda.

4 Conclusões

Segundo a OMS, a depressão é uma doença considerada frequente e que afeta cerca de 3,8% da população mundial. No Brasil, existem diversos estudos com objetivo de avaliar os fatores relacionados à depressão. O presente estudo teve, então, por objetivo principal, identificar, através da PNS de 2019, características populacionais que possam ser fatores de risco para esta doença no Brasil.

Com base em um levantamento das características da população que se autodeclarava como possuindo diagnóstico de depressão, percebeu-se que o perfil demográfico populacional dos brasileiros acima de 14 anos, era constituído, principalmente, por mulheres, de cor ou raça branca ou parda, com predomínio de pessoas solteiras ou casadas, com escolaridade inferior ao ensino médio completo e com rendimento domiciliar *per capita* entre $\frac{1}{2}$ até 2 salários mínimos. A prática de esportes ou exercício físico não foi considerado fator associado à ocorrência de depressão.

Após as análises preliminares através dos modelos de regressão logística, observou-se que as variáveis significativas foram: sexo, cor ou raça, estado civil, frequência de problemas no sono, consumo de álcool e consumo de tabaco. No modelo selecionado, as variáveis de exercício físico e renda não foram significativas. A partir dos coeficientes e Razões de Chance estimados para o modelo, é possível identificar que ser mulher, ser de raça ou cor branco(a) e divorciado(a) são fatores de risco para o diagnóstico para a depressão. Além disso, pessoas de mais idade também têm mais chance de serem diagnosticadas com depressão. A dificuldade para dormir e o hábito de fumar foram considerados relevantes para a ocorrência de depressão, mas, apesar de ser esperado que pessoas que consomem mais bebida alcoólica pudessem ter maior chance de depressão, o modelo indicou o contrário: a ausência do hábito de beber alguma bebida alcoólica como fator de risco.

O modelo proposto neste trabalho foi satisfatório, tanto em termos da medida de qualidade de ajuste e quanto para a capacidade de preditiva do modelo. Para pesquisas

futuras, poderia-se pensar em testar a inclusão, no modelo, de outras variáveis disponíveis na base de dados da PNS de 2019, especialmente outras relacionadas à qualidade de vida dos respondentes. Outra possibilidade seria refazer o estudo com grupos etários específicos ou, ainda, reavaliar futuramente a evolução da ocorrência de depressão a partir de novas edições da PNS.

Referências

- ASSOCIATION, A. P. et al. *DSM-5: Manual diagnóstico e estatístico de transtornos mentais*. [S.l.]: Artmed Editora, 2014.
- BARRETO, I. C. d. H. C.; GRISI, S. J. Morbidade referida e seus condicionantes em crianças de 5 a 9 anos em sobral, ce, brasil. *Revista Brasileira de Epidemiologia*, SciELO Public Health, v. 13, p. 35–48, 2010.
- BINDER, D. A. On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review/Revue Internationale de Statistique*, JSTOR, p. 279–292, 1983.
- BUSSAB, P. A. M. W. D. O. *Estatística Básica*. 6. ed. [S.l.]: Editora Saraiva, 2010.
- COPPENS, E. et al. Public attitudes toward depression and help-seeking in four european countries baseline survey prior to the ospi-europe intervention. *Journal of affective disorders*, Elsevier, v. 150, n. 2, p. 320–329, 2013.
- CORDEIRO, G. M.; DEMÉTRIO, C. *Modelos lineares generalizados*. [S.l.]: IMPA, 1986.
- COSTA, R. A.; SOARES, H. L. R.; TEIXEIRA, J. A. C. Benefícios da atividade física e do exercício físico na depressão. *Revista do Departamento de Psicologia. UFF*, SciELO Brasil, v. 19, n. 1, p. 273–274, 2007.
- COUTINHO, F. L. Evidências de validade de critério e convergente relacionada a outras variáveis da escala baptista de depressão para idosos-(ebadep-id). 2013.
- CRUWYS, T. et al. Depression and social identity: An integrative review. *Personality and Social Psychology Review*, v. 18, p. 215–238, 02 2014.
- CUNHA, R. V. d.; BASTOS, G. A. N.; DUCA, G. F. D. Prevalência de depressão e fatores associados em comunidade de baixa renda de porto alegre, rio grande do sul. *Revista Brasileira de Epidemiologia*, SciELO Brasil, v. 15, p. 346–354, 2012.
- DOBSON, A. J.; BARNETT, A. G. *An introduction to generalized linear models*. [S.l.]: CRC press, 2018.
- ETTMAN, C. K. et al. Prevalence of depression symptoms in us adults before and during the covid-19 pandemic. *JAMA network open*, American Medical Association, v. 3, n. 9, p. e2019686–e2019686, 2020.
- FUKITA, S.; KAWASAKI, H.; YAMASAKI, S. Comprehensive analysis of depression-related factors among middle-aged residents in japan, an eastern culture: A cross-sectional study. *Medicine*, Wolters Kluwer Health, v. 100, n. 19, 2021.

- GAZALLE, F. K. et al. Sintomas depressivos e fatores associados em população idosa no sul do Brasil. *Revista de Saúde Pública*, SciELO Public Health, v. 38, p. 365–371, 2004.
- GONÇALVES, A. M. C. et al. Prevalência de depressão e fatores associados em mulheres atendidas pela estratégia de saúde da família. *Jornal Brasileiro de Psiquiatria*, SciELO Brasil, v. 67, p. 101–109, 2018.
- GONZALEZ, L. d. A. Regressão logística e suas aplicações. Universidade Federal do Maranhão, 2018.
- GREENLAND, S. Modeling and variable selection in epidemiologic analysis. *American journal of public health*, American Public Health Association, v. 79, n. 3, p. 340–349, 1989.
- GUERRA, M. et al. A comparative cross-cultural study of the prevalence of late life depression in low and middle income countries. *Journal of affective disorders*, Elsevier, v. 190, p. 362–368, 2016.
- GUIMARÃES, L. A. M.; GRUBITS, S. *Série saúde mental e trabalho*. [S.l.]: Casa do Psicólogo, 2004. v. 3.
- IBGE. *IBGE - PNS - Conceitos e métodos*. 2019. <<https://www.ibge.gov.br/estatisticas/sociais/saude/9160-pesquisa-nacional-de-saude.html?=&t=conceitos-e-metodos>>. Acessado em 24/11/2021.
- IBGE, P. N. de S. *Percepção do estado de saúde, estilos de vida, doenças crônicas e saúde bucal*. [S.l.: s.n.], 2020.
- LOPEZ, M. R. A. et al. Depressão e qualidade de vida em jovens de 18 a 24 anos no sul do Brasil. *Revista de psiquiatria do Rio Grande do Sul*, SciELO Brasil, v. 33, p. 103–108, 2011.
- LUMLEY, T.; SCOTT, A. Fitting regression models to survey data. *Statistical Science*, JSTOR, p. 265–278, 2017.
- MARÔCO, J. *Análise Estatística com o SPSS Statistics.: 7ª edição*. [S.l.]: ReportNumber, Lda, 2018.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to linear regression analysis*. [S.l.]: John Wiley & Sons, 2012.
- MUIANGA, E. A. *Avaliação de métodos de estimação da variância em amostras complexas*. Dissertação (Mestrado) — Universidade de Évora, 2016.
- OMS, W. H. O. *Depression*. 2021. <<https://www.who.int/news-room/fact-sheets/detail/depression>>. Acessado em 16/11/2021.
- PAHO. *Depressão*. 2021. <<https://www.paho.org/pt/topicos/depressao>>. Acessado em 16/11/2021.
- PARKER, G.; BROTCHE, H. Depressão maior suscita questionamento maior. *Brazilian Journal of Psychiatry*, SciELO Brasil, v. 31, p. S3–S6, 2009.

- PEREIRA, G. R. *Uma revisão sobre o uso analítico de dados provenientes de amostras com estruturas complexas*. Tese (Doutorado) — Universidade de São Paulo, 2016.
- PESSOA, D. G. C.; SILVA, P. L. N. Análise de dados amostrais complexos. *São Paulo: Associação Brasileira de Estatística*, v. 112, 1998.
- PORTO, J. A. D. Conceito e diagnóstico. *Brazilian Journal of Psychiatry*, SciELO Brasil, v. 21, p. 06–11, 1999.
- QUEVEDO, J.; NARDI, A. E.; SILVA, A. G. da. *Depressão: Teoria e Clínica*. [S.l.]: Artmed Editora, 2018.
- REN, X. et al. Burden of depression in china, 1990–2017: findings from the global burden of disease study 2017. *Journal of affective disorders*, Elsevier, v. 268, p. 95–101, 2020.
- RIBEIRO, S. N. P. Atividade física e sua intervenção junto a depressão. *Revista Brasileira de Atividade Física & Saúde*, v. 3, n. 2, p. 73–79, 1998.
- SANTOS, M. C. Q. dos. Fatores de risco associados à depressão em idosos: uma revisão integrativa. 2021.
- SAÚDE, O. M. da. *CID-10: Classificação Estatística Internacional de Doenças com disquete Vol. 1*. [S.l.]: Edusp, 1994.
- SILVA, P. L. d. N.; PESSOA, D. G. C.; LILA, M. F. Análise estatística de dados da pnad: incorporando a estrutura do plano amostral. *Ciência & Saúde Coletiva*, SciELO Public Health, v. 7, p. 659–670, 2002.
- STOPA, S. R. et al. Prevalência do autorrelato de depressão no brasil: resultados da pesquisa nacional de saúde, 2013. *Revista Brasileira de Epidemiologia*, SciELO Public Health, v. 18, p. 170–180, 2015.
- STOPA, S. R. et al. Pesquisa nacional de saúde 2019: histórico, métodos e perspectivas. *Epidemiologia e Serviços de Saúde*, SciELO Brasil, v. 29, 2020.
- SZWARCWALD, C. L. et al. Exames laboratoriais da pesquisa nacional de saúde: metodologia de amostragem, coleta e análise dos dados. *Revista Brasileira de Epidemiologia*, SciELO Public Health, v. 22, p. E190004–SUPL, 2019.
- WHO. *The World Health Report 2001. Mental Health: New Understanding, New Hope*. [S.l.], 2001.
- WHO. *Depressão: Uma Crise Global - Jornada Mundial de Saúde Mental*. 2012. <<https://www.afro.who.int/pt/news/depressao-uma-crise-global-jornada-mundial-de-saude-mental>>. Acessado em 23/11/2021.
- WHO, W. H. O. R. O. for E. *Depression*. 2018. <<https://www.euro.who.int/en/health-topics/noncommunicable-diseases/mental-health/areas-of-work/depression>>. Acessado em 01/12/2021.