

ESTIMAÇÃO DA PREVALÊNCIA DA DEPRESSÃO VIA O  
MODELO DE CLASSES LATENTES UTILIZANDO A AMOSTRA DA  
PESQUISA NACIONAL DE SAÚDE, 2013

Rita de Cássia de Lima Idalino

Tese apresentada à Universidade Estadual  
Paulista “Júlio de Mesquita Filho” para ob-  
tenção do título de Doutor em Biometria.

BOTUCATU  
São Paulo - Brasil  
Fevereiro – 2019

**ESTIMAÇÃO DA PREVALÊNCIA DA DEPRESSÃO VIA O  
MODELO DE CLASSES LATENTES UTILIZANDO A AMOSTRA DA  
PESQUISA NACIONAL DE SAÚDE, 2013**

**Rita de Cássia de Lima Idalino**

Orientadora: Profa. Dra. **Luzia Aparecida Trinca**

Tese apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” para obtenção do título de Doutor em Biometria.

BOTUCATU  
São Paulo - Brasil  
Fevereiro – 2019



## Ficha Catalográfica

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.  
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP  
BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Idalino, Rita de Cássia de Lima.

Estimação da prevalência da depressão via o modelo de classes latentes utilizando a amostra da Pesquisa Nacional de Saúde, 2013 / Rita de Cássia de Lima Idalino. - Botucatu, 2019

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Luzia Aparecida Trinca

Capes: 90194000

1. Depressão. 2. Saúde mental. 3. Amostragem (Estatística). 4. Análise multivariada. 5. Variáveis latentes.

Palavras-chave: amostragem complexa; classes latentes; depressão; saúde mental.

## Dedicatória

A todos que lidam com a depressão, mas buscam constantemente por meios que podem trazer benefícios para a saúde mental.

## Agradecimentos

À Universidade Federal do Piauí por me conceder afastamento integral para que eu pudesse realizar o curso de doutorado.

À Professora Luzia Trinca por ter me recebido sob sua orientação e ter me apoiado muito além das questões previstas durante um curso de pós-graduação. Gratidão pela paciência e por compartilhar seu conhecimento de uma forma tão generosa.

A Raphael, por todas as horas dedicadas a mim, seja discutindo, ajudando, discordando (na maioria das vezes). Gratidão por você permanecer ao meu lado, independente das adversidades.

À Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP e ao Programa de Pós-Graduação em Biometria pela oportunidade de obter conhecimento, contribuindo para o meu crescimento pessoal e profissional.

A Djalma Pessoa, pela atenção e paciência para me ensinar a manusear os dados da PNS e por sempre me atender em meio as dúvidas que surgiam ao longo da pesquisa.

Aos membros da banca examinadora pelo tempo dedicado à leitura e as contribuições para melhoria do texto.

Aos funcionários do departamento de Bioestatística, Junior, Arthur, Luiz e Dona Selda, pela disposição em sempre ajudar e pelos momentos de descontração.

Agradeço aos amigos da sala 8 pela convivência, pelos momentos compartilhados de estudos, alegrias e angústias. Livia, Edijane e Farid, obrigada “colegas de classe”!

Aos que conviveram comigo nesse período trazendo experiências positivas a minha vida, em especial a Dennison, Lucas, Agda, Cássio, Clovis, Gisele, Aiara e Thiago.

A Maria Luisa e Fábio Abagabir pelo suporte clínico-psicológico.

Aos amigos que conheci em Botucatu, Eduardo, Mariana, Fabiane, Luan, Jefferson e Vitor por me proporcionarem momentos ímpares de alegria. Nos veremos em breve!

A minha mãe (Roneide), irmã (Maraíza) e sobrinho-afilhado (Heitor) que são fonte de amor na minha vida e aos demais familiares que estão sempre torcendo por mim.

A Josimar, amigo, padrinho e grande exemplo de luta e dedicação. Obrigada por me encorajar muito antes de entrarmos juntos na UFRN em 2004 para fazer o curso de Estatística.

Aos amigos que permanecem constantemente próximos independente da distância geográfica, agradeço a Danielle (Paixão), Marcelo, Julianne, Daniele, Juscilésio, Erica e Rafael. De modo especial, agradeço aos amigos que estão em Teresina-PI e foram sempre disponíveis para ajudar a resolver os problemas que surgiram na minha ausência. Deixo meu muito obrigada ao João, Célia, Lya e Socorro.

Por fim, agradeço a Deus por ter permitido que eu chegasse até aqui.

# Sumário

	Página
<b>LISTA DE FIGURAS</b>	<b>x</b>
<b>LISTA DE TABELAS</b>	<b>xi</b>
<b>RESUMO</b>	<b>xv</b>
<b>SUMMARY</b>	<b>xvii</b>
<b>1 APRESENTAÇÃO</b>	<b>1</b>
<b>2 INTRODUÇÃO</b>	<b>3</b>
2.1 Objetivos . . . . .	4
2.2 Justificativa . . . . .	5
2.3 Organização da tese . . . . .	9
<b>3 DEPRESSÃO - ASPECTOS HISTÓRICOS E ECONÔMICOS</b>	<b>10</b>
3.1 Breve histórico sobre a depressão . . . . .	10
3.2 Depressão e economia . . . . .	11
<b>4 PESQUISA NACIONAL DE SAÚDE 2013</b>	<b>15</b>
4.1 Definições e etapas da pesquisa . . . . .	15
4.2 Esquema de Amostragem da PNS . . . . .	17
4.2.1 Peso das Unidades Primárias de Amostragem - UPAs . . . . .	20
4.2.2 Peso dos domicílios e de todos os moradores . . . . .	21
4.2.3 Peso do morador selecionado . . . . .	22



4.3	Características sociodemográficas e percepção sobre o estado de saúde mental . . . . .	24
<b>5</b>	<b>FUNDAMENTAÇÃO TEÓRICA E METODOLOGIA</b>	<b>27</b>
5.1	Análise de Classes Latentes . . . . .	27
5.1.1	O modelo de classes latentes . . . . .	30
5.1.2	Método de estimação . . . . .	32
5.1.3	Algoritmo EM . . . . .	33
5.1.4	Erros-padrão para $\hat{\theta}$ . . . . .	36
5.1.5	O modelo de classes latentes com covariáveis . . . . .	36
5.2	LCA e depressão . . . . .	37
5.3	Material e Métodos . . . . .	38
5.3.1	Suporte computacional . . . . .	40
<b>6</b>	<b>RESULTADOS</b>	<b>41</b>
6.1	Análise Descritiva . . . . .	41
6.1.1	Características Sociodemográficas . . . . .	41
6.1.2	Itens relacionados ao PHQ-9 . . . . .	43
6.2	Análise de Classes Latentes . . . . .	45
6.2.1	Escolha do número de classes . . . . .	45
6.2.2	Modelos separados para as UF's . . . . .	47
6.2.3	Ajuste de um modelo global . . . . .	66
6.2.4	Estimativas dos $\gamma$ 's (prevalências) . . . . .	66
6.2.5	Estimativas dos $\rho$ 's (probabilidades condicionais de respostas) . . . . .	69
6.2.6	Estimativas dos $\alpha$ 's (efeitos das covariáveis) . . . . .	70
6.3	Exploração da relação entre a classificação por classes latentes e o uso de pontos de corte . . . . .	75
<b>7</b>	<b>DISCUSSÃO E CONCLUSÕES</b>	<b>78</b>

	ix
<b>ANEXOS</b>	<b>81</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>83</b>

## Lista de Figuras

	Página
5.1 Diagrama de uma estrutura latente . . . . .	29
6.1 Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9, no modelo global, PNS-2013, Brasil. . . . .	73

## Lista de Tabelas

	Página
6.1 Características sociodemográficas para adultos brasileiros - PNS-2013, Brasil (total populacional estimado = 146.308.458, $n = 60.202$ ) . . . . .	43
6.2 Proporção amostral das respostas em cada item do PHQ-9, PNS-2013, Brasil . . . . .	45
6.3 Resumo das estatísticas de ajuste para selecionar o número de classes latentes, PNS-2013, Brasil . . . . .	46
6.4 Estimativas das prevalências e respectivos erros-padrão (E.P.) segundo unidades da federação, PNS-2013, Brasil . . . . .	49
6.5 Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 para o estado do Pará, PNS-2013, Brasil . .	51
6.6 Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 para o estado da Paraíba, PNS-2013, Brasil .	52
6.7 Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 para o estado do Espírito Santo, PNS-2013, Brasil . . . . .	53
6.8 Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 para o estado de Santa Catarina, PNS-2013, Brasil . . . . .	54
6.9 Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 para o estado do Mato Grosso, PNS-2013, Brasil	55
6.10 Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 para o estado de São Paulo, PNS-2013, Brasil	56

6.11 Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 para o estado do Rio Grande do Sul, PNS-2013, Brasil . . . . .	57
6.12 Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis no estado do Pará . .	59
6.13 Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis no estado da Paraíba	60
6.14 Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis no estado do Espírito Santo . . . . .	61
6.15 Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis no estado de Santa Catarina . . . . .	62
6.16 Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis no estado do Mato Grosso . . . . .	63
6.17 Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis no estado de São Paulo	64
6.18 Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis no estado do Rio Grande do Sul . . . . .	65
6.19 Estimativas das prevalências e respectivos erros-padrão (E.P.) segundo unidades da federação no modelo global, PNS-2013, Brasil . . . . .	68
6.20 Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 no modelo global, PNS-2013, Brasil . . . . .	72
6.21 Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis, PNS-2013, Brasil . .	74
6.22 Proporções condicionais, dada a classe latente, avaliadas nos itens relacionados ao PHQ-9, PNS-2013, Brasil . . . . .	76

6.23 Proporções condicionais, avaliado nos itens relacionados ao PHQ-9, PNS- 2013, Brasil . . . . .	77
--	----

## Lista de Siglas

**APA** Associação Americana de Psicologia

**CID** Classificação Internacional de Doenças

**CONEP** Comissão Nacional de Ética em Pesquisa

**DALY** *Disability-Adjusted Life Year*

**DCNT** Doenças Crônicas Não Transmissíveis

**DSM-5** Manual de Diagnóstico e Estatístico de Transtornos Mentais (5<sup>a</sup> Edição)

**IBGE** Instituto Brasileiro de Geografia e Estatística

**LCA** *Latent Class Analysis*

**OMS** Organização Mundial de Saúde

**PHQ-9** *Patient Health Questionnaire-9*

**PNS** Pesquisa Nacional de Saúde

**WHO** *World Health Organization*

**YLL** *Years of Life Lost*

**ESTIMAÇÃO DA PREVALÊNCIA DA DEPRESSÃO VIA O  
MODELO DE CLASSES LATENTES UTILIZANDO A AMOSTRA DA  
PESQUISA NACIONAL DE SAÚDE, 2013**

Autora: RITA DE CÁSSIA DE LIMA IDALINO

Orientadora: Profa. Dra. LUZIA APARECIDA TRINCA

**RESUMO**

A depressão é apontada como a doença crônica não transmissível que terá maior abrangência até 2030, atingindo direta ou indiretamente vários setores nos quais a população está inserida. No Brasil, a situação da depressão é alarmante, respondendo pela maior taxa do continente latino-americano. A Pesquisa Nacional de Saúde (PNS) é um inquérito de base domiciliar de abrangência nacional. Trata-se de uma iniciativa do Ministério da Saúde em parceria com o Instituto Brasileiro de Geografia e Estatística (IBGE) e tem como objetivo caracterizar a situação de saúde e os estilos de vida da população brasileira, e assim conhecer como acontece a atenção à saúde em diversos grupos da população. Neste trabalho a depressão será o objeto de estudo por ser caracterizada como uma doença complexa de difícil mensuração e observação devido a suas causas multifatoriais. O levantamento de dados da PNS



utilizou um planejamento amostral complexo, o que demanda uma atenção especial em relação à análise das informações coletadas. Considerando a magnitude da pesquisa e levando em consideração as diversidades regionais, foram ajustados modelos com base na teoria de classes latentes. Essa abordagem identifica grupos baseados nos padrões de respostas observadas nas variáveis categóricas utilizando um modelo probabilístico. Assim é possível classificar cada indivíduo como pertencente a um grupo, estimar a prevalência e identificar características decisivas para o surgimento dos grupos. A partir de itens que tratam sobre a saúde mental, foi possível propor classes com o perfil de respostas associadas com questões sociodemográficas de adultos brasileiros. Os resultados demonstraram a influência de respostas quanto a gravidade de forma mais acentuada para a idade, sexo e escolaridade. Foi possível investigar sob a ótica de uma metodologia promissora o comportamento em termos populacionais da depressão e apesar da complexidade, identificar possíveis subtipos.

**Palavras-Chave:** depressão, classes latentes, amostragem complexa, saúde mental.

**ESTIMATION OF THE PREVALENCE OF DEPRESSION BY  
LATENT CLASS MODEL USING THE NATIONAL HEALTH  
SURVEY SAMPLE, 2013**

Author: RITA DE CÁSSIA DE LIMA IDALINO

Adviser: Prof. Dr. LUZIA APARECIDA TRINCA

**SUMMARY**

Depression is indicated as the non-transmissible chronic disease that will be more widespread until 2030, reaching directly or indirectly several sectors in which the population is inserted. In Brazil, the situation of depression is alarming, accounting for the highest rate in the Latin American continent. The National Health Survey (PNS) is a nationwide household survey. It is an initiative of the Ministry of Health in partnership with the Brazilian Institute of Geography and Statistics (IBGE) and aims to characterize the health situation and the lifestyles of the Brazilian population, and thus to know how health care happens in different groups of the population. In this study, depression will be the object of study because it is characterized as a complex disease of difficult measurement and observation due to its multifactorial causes. The PNS data collection used a complex sampling plan, which

demands special attention in relation to the analysis of the information collected. Considering the magnitude of the research and taking into account the regional diversities, models were adjusted based on latent class theory. This approach identifies groups based on the patterns of responses observed in the categorical variables using a probabilistic model. Thus it is possible to classify each individual as belonging to a group, to estimate the prevalence and identify decisive characteristics for the emergence of groups. Based on items dealing with mental health, it was possible to propose classes with the profile of responses associated with sociodemographic issues of Brazilian adults. The results showed the influence of severity responses more markedly for age, sex and schooling. It was possible to investigate the behavior, in population terms, of depression and in spite of the complexity, to identify possible subtypes.

**Keywords:** depression, latent classes, complex sampling, mental health.

# 1 APRESENTAÇÃO

Nascida em Natal-RN, conclui a graduação em Estatística na Universidade Federal do Rio Grande do Norte - UFRN em 2008, no ano seguinte ingressei no mestrado em Biometria e Estatística Aplicada na Universidade Federal Rural de Pernambuco - UFRPE e lá desenvolvi minha dissertação na linha de pesquisa Estatística Aplicada e Experimental que abordou a análise das homologias em genes relacionados à resistência à mastite. Ainda como aluna de mestrado iniciei a carreira docente como professora substituta no Departamento de Estatística e Informática - DEINFO da mesma instituição.

Em 2010 prestei concurso para trabalhar na Universidade Federal do Piauí - UFPI assumindo a função de professora efetiva no ano de 2011 e permaneci lecionando disciplinas ligadas ao curso de bacharelado em Estatística bem como aos cursos ligados a outros departamento por três anos. Em 2014 prestei seleção junto ao Programa de Pós-Graduação em Biometria, na Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP para realizar o curso de doutorado, o qual foi iniciando no ano de 2015.

Desde a graduação tive uma relação mais próxima com a aplicação da estatística em cenários clínicos, passando por projetos de iniciação científica relacionados ao diagnóstico das disfunções sexuais em mulheres usuárias do SUS, comportamento sexual e o risco de infecção pelo HIV entre mulheres esterilizadas e não esterilizadas, Dengue BR: uso de previsões climáticas em um modelo de risco de reprodução da doença e análise comparativa dos determinantes da fecundidade

do nordeste e no estado de São Paulo. Na UFPI coordenei um projeto de pesquisa sobre a sobrevivência de pacientes com neoplasias tratados em um hospital de referência da cidade de Teresina - PI e realizei parcerias de pesquisas junto ao Programa de Pós-Graduação Ciências e Saúde gerando uma publicação na *Special Care in Dentistry* sobre saúde bucal em crianças e adolescentes com hemofilia.

Seguindo a temática da saúde, o estudo da teoria de classes latentes, tema associado à pesquisa de doutorado, se mostrou atrativo e com grande potencial de desenvolvimento. O campo das doenças psíquicas, especificamente a depressão se adequou bem à metodologia devido a subjetividade envolvida no processo de identificação do perfil populacional dada a complexidade em torno das etapas envolvidas desde o diagnóstico até o tratamento.

## 2 INTRODUÇÃO

Certos fenômenos nas ciências sociais, comportamentais e da saúde podem ser representados por um modelo que considera a existência de subtipos ou categorias distintas numa população de indivíduos, embora a categoria específica a qual cada indivíduo pertence não pode ser observada diretamente ou objetivamente. Na área da saúde em particular, muitas doenças são diagnosticadas ou suas gravidades acessadas por meio de variáveis manifestas ou indicadoras da doença, variáveis essas passíveis de mensuração.

Para auxiliar na pesquisa do diagnóstico, por exemplo, de um determinado transtorno mental, é comum o uso de algum questionário envolvendo diversos itens, cada um com múltiplas categorias de resposta, que recebem um escore numérico de acordo com a resposta dada pelo indivíduo. É também comum usar algum ponto de corte para a pontuação nos itens ou em combinações de conjuntos de itens como critério para identificar e/ou classificar a gravidade da doença.

Alternativamente, do ponto de vista de conhecimento populacional, é interessante explorar as relações entre os itens e associações de padrões de respostas aos possíveis subtipos subjacentes que formam a população. O uso de um modelo que postula a existência de subtipos caracterizando uma variável de classes latentes e a relaciona às variáveis manifestas é de grande potencial. Tal modelo, conhecido como modelo de classes latentes, introduzido por Lazarsfeld & Henry (1968) tem se mostrado com ampla aplicação nas áreas sociais e também da saúde. Do ponto de vista metodológico, trata-se de uma área em pleno desenvolvimento, tanto sob

a ótica de generalização de tipos de variáveis possíveis e métodos de estimação, quanto sob a perspectiva de tratamento de dados omissos, extensão para planos amostrais complexos, observações longitudinais e aspectos computacionais (Schafer & Kang, 2013; Collins & Lanza, 2010).

A partir de uma revisão sistemática, Ulbricht et al. (2018) discutem as inconsistências e complexidades observadas na identificação de subtipos de depressão. As características multifatoriais envolvidas no processo de adoecimento exigem que estudos possam ser conduzidos para melhor compreender a heterogeneidade em torno de todo o processo da depressão. Nesse sentido, vale salientar que pesquisas populacionais trazem uma imensa variabilidade em torno das questões investigadas, o que contribui para justificar a aplicação de um método de análise que considera a subjetividade para acessar determinadas informações quanto a formação de grupos.

## 2.1 Objetivos

Com o destaque dado, pelas organizações de saúde, aos problemas relacionados à depressão, cujo diagnóstico não obedece uma característica objetivamente mensurável, aliado à disponibilidade de uma base de dados nacional e pública, a Pesquisa Nacional de Saúde, PNS (2013), este trabalho se propõe a construir um modelo de classes latentes com o objetivo de contribuir para o conhecimento dos perfis e potenciais subgrupos quanto aos transtornos mentais, em meio à abundância de heterogeneidades que compõem a população brasileira avaliada sob a ótica da PNS.

Com esse objetivo geral em mente, especificamente, explora-se a aplicação do modelo de classes latentes a um conjunto de dados obtido por esquema de amostragem complexo. Tal modelagem permite a estimação das prevalências de cada classe, estratificadas em grupos de interesse ou não, e explora as relações entre os itens manifestos, além de permitir a estimação dos efeitos de covariáveis sobre a

probabilidade do indivíduo pertencer a determinada classe.

## 2.2 Justificativa

Entender a dinâmica de associação de uma doença que atinge uma parcela expressiva da população, provocando danos de forma direta e indireta em diversos seguimentos sociais, como é o caso da depressão, pode ajudar a evitar e controlar o seu surgimento, trazendo um ganho imensurável na qualidade de vida das pessoas, além de contribuir para o melhor funcionamento dos mais variados setores da sociedade que são afetados. Isso porque a depressão atinge tanto pacientes quanto seus círculos de convivência familiar, profissional e social de maneira debilitadora, pois reduz a disposição para o exercício de tarefas e dificulta as relações interpessoais. Não raro, sua ocorrência é simultânea a outras doenças crônicas, como diabetes, câncer, insuficiência renal, o que também dificulta a identificação da doença e por sua vez o tratamento apropriado, intensificando assim a incidência da mesma. O termo depressão é empregado para designar tanto um estado afetivo normal (a tristeza), quanto um sintoma, uma síndrome e uma (ou várias) doença(s) (Del Porto, 1999).

Nesse cenário, o quadro depressivo traz custos não desprezíveis para toda a sociedade. A Organização Mundial de Saúde (OMS), descreve a depressão como uma situação epidêmica, acometendo mais de 350 milhões de pessoas ao redor do mundo em 2015, o que representava à época 5% da população mundial, aproximadamente. Este número antecipou a previsão feita pela própria entidade no *World Health Report 2001*, que projetava que, até 2020, a depressão seria a segunda doença mais onerosa à sociedade, ficando atrás apenas do infarto do miocárdio (Organização Mundial da Saúde, 2001).

A prevalência da depressão no mundo apresenta diferentes aspectos culturais, geográficos e socioeconômicos. Países europeus de clima mais frio e países



mais pobres apresentam maiores prevalências da doença. Regiões do norte da África e do Oriente Médio apresentam altas prevalências do transtorno, reforçando a ideia de que condições socioeconômicas e culturais têm uma influência importante no desenvolvimento da doença (Ferrari et al., 2013).

O conceito moderno da depressão, aceito pela maioria dos psiquiatras está definido nas duas classificações oficiais de saúde, a Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde (CID-10) e o Manual Diagnóstico e Estatístico de Distúrbio Mental em sua 4ª edição (DSM-IV) (Paykel, 2008). Sendo assim, a depressão é definida como uma doença composta por um conjunto de sintomas psicológicos e físicos, associada a altos índices de comorbidades médicas, incapacitação e mortalidade prematura. A depressão está entre as Doenças Crônicas Não Transmissíveis (DCNT) que constituem o maior problema de saúde global, respondendo por 60% de todas as mortes no mundo além de gerar perda da qualidade de vida, com alto grau de limitação e incapacidade, além de serem responsáveis por impactos econômicos para famílias e comunidades e para a sociedade em geral (Malta et al., 2014).

De acordo com a Associação Americana de Psiquiatria (APA), na edição mais atualizada do *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), “a depressão é uma doença que faz com que indivíduos apresentem sentimento de tristeza, vazio e irritabilidade, acompanhado de alterações somáticas e cognitivas que afetam sua capacidade de funcionar. Em casos mais graves, a depressão pode levar ao suicídio” (American Psychiatric Association, 2013). Similarmente, a OMS definiu em 2016 a depressão como “uma doença caracterizada por tristeza, perda de interesse ou prazer, sentimento de culpa ou baixa autoestima, sono ou apetite perturbados, sensação de cansaço e dificuldade de concentração”. Ainda sobre as características, a OMS aponta que “a depressão pode ser perene ou recorrente, prejudicando sensivelmente a capacidade de o indivíduo funcionar no

trabalho, na escola ou lidar com a vida cotidiana”.

Atualmente, os critérios utilizados para diagnosticar e classificar os estados depressivos se encontram no Manual Diagnóstico e Estatístico dos Transtornos Mentais, sendo esses os indicativos de um estado depressivo: sentir-se deprimido a maior parte do tempo; anedonia (interesse diminuído ou perda de prazer para realizar as atividades de rotina); sensação de inutilidade ou culpa excessiva; dificuldade de concentração: habilidade frequentemente diminuída para pensar e se concentrar; fadiga ou perda de energia; distúrbios do sono (insônia ou hipersônia praticamente diárias); problemas psicomotores (agitação ou retardo psicomotor); perda ou ganho significativo de peso, na ausência de regime alimentar; ideias recorrentes de morte ou suicídio (American Psychiatric Association, 2014). Do ponto de vista de diagnóstico de um paciente, de acordo com o número de itens respondidos afirmativamente, o estado depressivo pode ser classificado em três grupos;

1. Depressão menor: 2 a 4 sintomas por duas ou mais semanas, incluindo estado deprimido ou anedônia;
2. Distímia: 3 ou 4 sintomas, incluindo estado deprimido, durante dois anos, no mínimo;
3. Depressão maior: 5 ou mais sintomas por duas semanas ou mais, incluindo estado deprimido ou anedônia.

Barros et al. (2017) relatam que os transtornos depressivos podem interagir, agravar ou mesmo constituir-se em fator de risco independente para doenças crônicas, além de influenciar de forma importante a adoção e manutenção de vários comportamentos relacionados à saúde.

Os sintomas depressivos se manifestam numa parte significativa da população, afetando, atualmente, 4,4% da população mundial. Em se tratando do

Brasil, esse dado chega a 5,8% sendo a maior taxa do continente latino-americano, e as projeções sugerem um crescimento alarmante dessa estatística. Segundo estudiosos da área, até 2030, a depressão poderá ser a doença mais comum no mundo (World Health Organization, 2017).

Desta forma, o estudo dos padrões populacionais do país quanto ao perfil de percepção dos sintomas depressivos e associações com características regionais, oferece uma grande contribuição principalmente pelo uso de dados públicos que já estão coletados e validados por organizações de renome e conseqüentemente não requerendo recursos financeiros extras.

## 2.3 Organização da tese

Esta tese encontra-se dividida em sete capítulos. No terceiro capítulo é apresentado um breve histórico da depressão e em seguida é realizada uma descrição dos impactos da doença na economia.

O quarto capítulo descreve, de maneira resumida, o processo de amostragem adotado na PNS seguido das etapas em torno da construção dos pesos amostrais. Uma apresentação das características sociodemográficas e percepção sobre o estado de saúde mental da amostra também está descrito no capítulo.

O capítulo 5 apresenta uma introdução sobre a Análise de Classes Latentes (LCA), a formulação da modelagem e a descrição do método de estimação. No decorrer do capítulo é também apresentado o modelo de classes latentes na presença de covariáveis. A metodologia trás as etapas em torno das análises realizadas sob o contexto da LCA e por fim são descritos aspectos em torno do suporte computacional utilizados para o desenvolvimento dessa tese.

O sexto capítulo contém os resultados para a modelagem de classes latentes aplicada aos dados de um questionário constituído de nove perguntas que avalia a presença de sintomas para o episódio de depressão maior adaptado na PNS, 2013. As discussões e conclusões estão presentes no sétimo capítulo.

## 3 DEPRESSÃO - ASPECTOS HISTÓRICOS E ECONÔMICOS

Neste capítulo são apresentados dois tópicos em torno do tema de depressão. O primeiro versa sobre um breve histórico sobre a doença, desde os primeiros relatos até a consolidação do termo na área médica e em seguida é apresentada uma descrição dos impactos causados na economia.

### 3.1 Breve histórico sobre a depressão

O termo “depressão” é consideravelmente novo em toda a história, tendo surgido em meados de 1680, designando um estado de desânimo ou perda de interesse. O estabelecimento de tal conceito emergiu com a paulatina queda das crenças sobrenaturais que fundamentavam a compreensão dos transtornos mentais até então. O conceito de depressão como se concebe na atualidade teve seu início no século XVII em que sua origem pode ser inferida a partir das menções de alterações de humor, sobretudo nas referências ao estado conhecido como melancolia (Rousseau, 2000). As primeiras definições trazem a depressão como loucura, mania, fúria divina, possessão, bruxaria, tristeza, demência, psicose (Nemade et al., 2007).

Segundo Cordas (2002), a depressão é uma enfermidade conhecida desde os primórdios da humanidade. Há registros relacionados a ela em todas as culturas e regiões do mundo desde muitos séculos, com denominações distintas, porém, descrições semelhantes. Apesar da sua longa história e dos muitos avanços modernos na sua terapêutica, não se pode dizer que se trata de um quadro plenamente esclarecido: há controvérsias sobre o diagnóstico diferencial, as fronteiras entre tristeza e apatia consideradas normais e aquelas patológicas, os fatores que predisõem e aqueles que são determinantes, os neurotransmissores envolvidos, aspectos culturais e psicodinâmicos, tais como sentimentos de culpa. Sob o enfoque

estatístico, segue a proposição de que os diversos construtos<sup>1</sup> (termo utilizado na psicologia para se referir à variáveis latentes) são multivariados.

Atribui-se a Hipócrates a diferenciação entre medicina e filosofia, bem como a transição da explicação da doença centrada em termos sobrenaturais para termos científicos. O desenvolvimento da compreensão da doença apoiou-se no conceito dos quatro fluidos – bile, fleuma, sangue e bile negra como manifestação de um desequilíbrio entre esses fluidos e também num modelo de quatro humores: colérico, fleumático, sanguíneo e melancólico. Segundo ele, esses humores explicariam a regulação das emoções e a formação do caráter dos indivíduos. Fundamentado nesse modelo, Hipócrates formulou a primeira classificação médica dos transtornos mentais registrada na história, nomeou e descreveu a melancolia, a mania e a paranoia. Assim, por exemplo, a etimologia do termo melancolia – *melan* significa negro e *cholis*, bile – revela a visão de Hipócrates de que o quadro clínico da melancolia, em que se observava aversão à comida, falta de ânimo, inquietação, irritabilidade, medo ou tristeza que perduravam por longo período, seria resultante de uma intoxicação do cérebro pela bile negra (Cordas, 2002).

Com a consolidação da psiquiatria como especialidade médica, que ocorreu no século XIX, foi imposto à comunidade científica o desafio de mudança do paradigma diagnóstico dos transtornos mentais. O avanço da neurociência nas últimas décadas lançou as bases para um paradigma fisiopatológico do diagnóstico psiquiátrico.

### 3.2 Depressão e economia

A prevalência crescente da depressão no mundo vem preocupando os órgãos internacionais de saúde não só pelo problema em si, como também pelo alto

---

<sup>1</sup>Ideia ou teoria construída a partir de elementos conceituais ou subjetivos, não baseados em evidências empíricas.

custo que o transtorno acaba impondo à sociedade. Tais custos trazem ônus aos sistemas de saúde públicos e privados pelo aumento de gastos diretos e indiretos; e à sociedade, pela queda da produtividade do trabalhador. Nesse cenário, Berto et al. (2000) conduziram uma revisão sistemática da literatura a fim de descrever e comparar o custo da depressão em diferentes países. Foram 11 estudos revisados, todos destacando um custo não desprezível da doença para a sociedade. Greenberg et al. (2003) estimaram que custos diretos e indiretos da depressão totalizavam, em 1990, nos Estados Unidos, cerca de US\$ 94 bilhão. Já Kind & Sorenson (1993) estimaram o custo da depressão no Reino Unido em cerca de US\$ 1,32 bilhão.

Em se tratando dos impactos, estes vão além da vida pessoal do indivíduo. Dada a diversidade dos fatores que provocam o sofrimento psíquico, os mesmos podem ter várias origens. Muitas vezes, os indivíduos com depressão podem se sentir incapazes de participar de trabalhos e atividades comuns do dia-a-dia, não conseguindo desempenhar suas responsabilidades dentro do contexto profissional e podendo vir a sofrer discriminação por parte da sociedade (Santos & Kassouf, 2007).

Em termos econômicos, são diversos os impactos provocados pelos transtornos mentais, ainda que apenas parte desse ônus possa ser mensurável, como o aumento da necessidade de serviços sociais e de saúde, a perda de emprego, a redução de produtividade, o aumento do nível de criminalidade e da mortalidade prematura. Entre os imensuráveis, destacam-se os custos de oportunidade dos indivíduos e famílias. No entanto, na maior parte do mundo, a saúde mental dos indivíduos não é vista com a mesma importância que a da saúde física ().

Ferrari et al. (2013) estudaram a sobrecarga dos transtornos depressivos em diversas regiões do mundo através de dados do *Global Burden of Disease*, da OMS, e constataram que a depressão era a terceira doença mais onerosa para a sociedade no ano de 2010, representando uma das principais causas de YLD

(*years lost due to disability*), ou seja, anos perdidos pela deficiência. Embora a doença não seja tão letal, refletido no YLL (*years of life lost*), ou seja, anos de vida perdidos por motivo de morte, os autores ponderam que o transtorno depressivo tem um peso importante no DALY (*disability adjusted life year*), que representa o custo de uma doença em termos de anos perdidos de vida saudável. Assim, como citado pela OMS, entende-se que “a soma dos DALY de cada indivíduo de uma população representa o hiato entre o atual *status* de saúde e um *status* ideal em que toda a população vive até uma idade avançada, livre de doenças e deficiências”. Isso destaca a importância de quantificar o custo social de doenças não fatais (ou menos fatais). Os autores ainda enfatizam que os custos sociais da depressão estão associados ao sexo, à idade e à região de residência.

No Brasil, o custo social da depressão foi estimado por Leite et al. (2015) que apontaram o transtorno como o mais custoso entre mulheres no ano de 2008 (25,2 DALY/1000 hab.) e a sétima entre os homens (7,1 DALY/1000 hab.). Isso significa que, no referido ano, a soma do custo da doença representou 25,1 anos de vida saudável para cada 1000 mulheres e 7,1 anos de vida saudável para cada 1.000 homens, evidenciando como a doença impõe uma carga pesada à sociedade.

Os custos não se restringem somente às vidas perdidas, mas também existem os associados ao tratamento dos indivíduos com transtornos mentais e comportamentais (Santos & Kassouf, 2007). Nesse cenário, considerando somente as coberturas de saúde feitas pelo Sistema Único de Saúde (SUS) em 2002, o percentual de internações por transtornos mentais e comportamentais representou 2,9% do total de todos os grupos de causas. A psiquiatria foi responsável em 2003, por aproximadamente 2,9% do total de internações de todas as especialidades, somando R\$ 466,6 milhões e representando quase 8% do total pago por todas as especialidades.

Em suma, o cenário multifatorial em torno da depressão envolve diver-



os seguimentos de pesquisa que ainda são considerados escassos no que diz respeito a compreensão desse fenômeno, o que reitera a importância em se produzir pesquisas nessa área.

## **4 PESQUISA NACIONAL DE SAÚDE 2013**

Neste capítulo será descrito o processo de desenvolvimento da Pesquisa Nacional de Saúde (PNS) seguido das etapas em torno da construção dos pesos amostrais no contexto de um plano amostral complexo, seguido de uma apresentação das características sociodemográficas e percepção sobre o estado de saúde mental dos indivíduos da amostra.

### **4.1 Definições e etapas da pesquisa**

Em termos de pesquisa científica, é de fundamental importância que haja múltiplos esforços para a criação e a manutenção das estatísticas de saúde e a contínua reformulação das bases de dados, de modo que seja possível responder questões fundamentais sobre a oferta e o monitoramento das condições de saúde da população, bem como o desempenho do sistema de saúde que é oferecido. Com esse propósito, a PNS foi elaborada por meio de uma iniciativa do Ministério da Saúde em parceria com o Instituto Brasileiro de Geografia e Estatística - IBGE e, hoje, constitui o mais amplo inquérito sobre a saúde já realizado no Brasil. A PNS começou a ser elaborada no início de 2009, mas somente em julho de 2013 o estudo foi aprovado pela Comissão Nacional de Ética em Pesquisa - CONEP, sendo em agosto daquele ano iniciado o trabalho de campo, que se estendeu até fevereiro de 2014. A PNS contou com mais de 1.000 técnicos do IBGE coletando dados em 1.600 municípios brasileiros (Malta & Szwarcwald, 2017).

A PNS é definida como um inquérito de base domiciliar de âmbito

nacional, que surgiu da necessidade do aprimoramento e ampliação dos Suplementos de Saúde da Pesquisa Nacional por Amostra de Domicílios - PNAD. Os inquéritos são essenciais para se estabelecer avaliações sob as diversas óticas às quais estejam associados. Em relação ao sistema de saúde, um inquérito visa, sobretudo, conhecer e avaliar o serviço prestado a uma população sob a ótica de quem oferece, bem como daquele que é usuário. Os inquéritos são instrumentos eficientes para o entendimento das necessidades reais da população e essenciais para a elaboração e melhoria de programas e políticas públicas (Malta & Szwarcwald, 2015). A PNS surge como uma ferramenta essencial para fornecer informações que possam orientar os serviços sociais e de saúde para melhor compreender as necessidades e expectativas da população seja na promoção e prevenção da saúde ou nas demais áreas sociais que afetam a longevidade e a qualidade de vida do indivíduo.

A PNS visa a caracterização da situação de saúde e dos estilos de vida da população brasileira, conhecendo, assim, como acontece a atenção à saúde sob critérios sociais, geográficos, educacionais, entre outros. O inquérito da PNS foi composto por três questionários: a) o domiciliar, que se refere às características do domicílio, nos moldes do censo demográfico e da PNAD; b) o relativo a todos os moradores do domicílio, que dá continuidade ao Suplemento Saúde da Pesquisa Nacional por Amostras de Domicílios (PNAD); c) o individual, a ser respondido por um morador do domicílio com 18 anos ou mais, selecionado com equiprobabilidade entre todos os residentes elegíveis, que dá enfoque às principais doenças crônicas não transmissíveis, aos estilos de vida e ao acesso ao atendimento médico (PNS, 2013).

A população que compôs a PNS, constituiu-se por moradores de domicílios particulares permanentes pertencentes à área de abrangência geográfica da pesquisa. Define-se como abrangência geográfica todo o território nacional, dividido nos setores censitários da Base Operacional Geográfica de 2010, excluídas as áreas com características especiais e com pouca população, que são classificadas pelo IBGE,

na base de setores, como aldeias indígenas, quartéis, bases militares, alojamentos, acampamentos, embarcações, penitenciárias, colônias penais, presídios, cadeias, asilos, orfanatos, conventos e hospitais. Também foram excluídos os setores censitários localizados em terras indígenas. O levantamento de dados da PNS utilizou um planejamento amostral complexo, necessitando, portanto, de uma atenção especial em relação à análise das informações coletadas. Um plano de amostragem complexa é aquele que considera a combinação de vários métodos de amostragem (múltiplos estágios) no delineamento de um estudo para seleção de uma amostra representativa da população (Lohr, 2009)

## 4.2 Esquema de Amostragem da PNS

A PNS é uma pesquisa domiciliar e o plano amostral empregado foi amostragem conglomerada em três estágios, com estratificação das unidades primárias de amostragem (UPA). Os setores censitários ou conjunto de setores formam as unidades primárias de amostragem (UPAs), os domicílios são as unidades de segundo estágio e os moradores com 18 anos ou mais de idade definem as unidades de terceiro estágio (Souza-Junior et al., 2015).

Como parte do Sistema Integrado de Pesquisas Domiciliares (SIPD) do IBGE, as UPAs da pesquisa foram obtidas da Amostra Mestra, que é um conjunto de unidades de áreas selecionadas para atender a diversas pesquisas. Tais unidades são consideradas unidades primárias de amostragem (UPA), dentro do planejamento amostral de cada uma das pesquisas que utilizam a Amostra Mestra – incluída a PNS.

No segundo estágio foi selecionado, por amostragem aleatória simples, um número fixo de domicílios particulares permanentes em cada UPA selecionada no primeiro estágio. A seleção foi feita do Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE). Nestes domicílios selecionados foram investigadas características do domicílio, características de todos os moradores, inclusive dos

idosos. Dentro de cada domicílio, um morador com 18 anos ou mais de idade foi selecionado para responder ao questionário específico, também por amostragem aleatória simples através de uma lista construída no momento da entrevista.

O tamanho da amostra foi definido considerando o nível de precisão desejado para as estimativas de alguns indicadores de interesse, que são basicamente proporções de pessoas em determinadas categorias. Como os indicadores referem-se a grupos populacionais diferentes, também foi preciso avaliar a proporção de domicílios que possuíam pessoas nestes grupos, com base nos dados do Censo 2010, para que fosse possível definir o tamanho de amostra necessário para estimar com a precisão os grupos de interesse. Considerou-se os valores do efeito de plano amostral (EPA), medida que indica o quanto o plano amostral por conglomerados é menos eficiente (maior variância) que a amostragem aleatória simples, por isso os tamanhos da amostra obtidos para AAS foram aumentados para alcançar a mesma precisão. Os dados utilizados nos cálculos amostrais foram obtidos no Censo Demográfico 2010 e na PNAD 2008.

Os cálculos foram realizados com base nas fórmulas das expressões (1) e (2), as quais foram utilizadas para definir o tamanho amostral dentro de cada domínio.

Seja  $n_{AAS}$  o tamanho de amostra da indivíduos de uma população sob o esquema de amostragem aleatória simples dado por

$$n_{AAS} = \frac{N}{N-1} \times P \times Q \times \frac{1}{CV^2 \times P^2 + \frac{P \times Q}{N-1}} \quad (1)$$

Define-se, então,  $n_{AC}$  como o tamanho da amostra sob amostragem conglomerada, dada por

$$n_{AC} = n_{AAS} \times EPA, \quad (2)$$

em que:

$N$  é o número total de pessoas na população;

$P$  é a proporção de pessoas numa categoria de interesse, por exemplo sexo, obtida do censo;

$Q = 1 - P$ ;

$CV$  é o coeficiente de variação desejado da estimativa da proporção.

O  $EPA$  é o efeito do plano amostral, calculado para corrigir o efeito que a conglomeração das unidades primárias de amostragem causa nas estimativas. O  $EPA$  indicará o quanto a amostragem por conglomerados, que é o caso da PNS, é menos eficiente em relação a AAS.

Os pesos amostrais são necessários para corrigir as imperfeições carregadas na amostra que podem levar a presença de viés, que ocorrem devido a inclusão das unidades amostrais com probabilidades desiguais, falta de cobertura populacional e não-resposta.

Em específico, os propósitos da ponderação são:

- compensar probabilidades desiguais de seleção;
- compensar a falta de resposta (não-resposta);
- ajustar a distribuição da amostra ponderada para as principais variáveis de interesse, como por exemplo, idade, raça e sexo.

Uma vez coletado um conjunto de observações provenientes de uma amostra, é necessário o uso da ponderação para garantir que a amostra seja representativa da população na qual foi coletada. No desenvolvimento dos pesos apresentados nas seções 4.2.1 a 4.2.3, os índices  $i$ ,  $h$ ,  $j$  e  $k$  são usados para identificar a UPA, estrato, domicílio e morador, respectivamente.

Ajustar os dados por ponderação reduz o impacto de cada respondente e reduz determinados vieses. Se ao resumir as características de uma população,

os pesos forem ignorados, os resultados das estimativas serão tendenciosos (Schulz et al., 1995). Considerando a ponderação adequada, é possível tornar a amostra representativa à distribuição encontrada na população. Uma técnica de ajuste de pesos só pode ser realizada quando variáveis auxiliares adequadas estiverem disponíveis. As variáveis auxiliares devem ter sido medidas no levantamento e a distribuição da população deve ser conhecida. Variáveis auxiliares são normalmente o sexo, idade, estado civil e região do país e tais informações, geralmente, são obtidas através de institutos nacionais de estatística. Ao comparar a distribuição de frequência observada de uma variável com a sua distribuição populacional, é possível estabelecer se a resposta da pesquisa é representativa em relação a essa variável. Se houver diferença substancial entre a distribuição da resposta e a distribuição da população, conclui-se, então, que há uma falta de representatividade com relação a essa variável.

A ponderação é um dos principais componentes da amostragem por inquérito. Para um determinado levantamento amostral, a cada unidade da amostra selecionada é associado um peso que é utilizado para obter estimativas de parâmetros populacionais de interesse. Tal processo é frequentemente considerado como restrito à amostragem por inquérito e à produção de estatísticas relacionadas com populações finitas. Na PNS foi considerado o cálculo de pesos amostrais para as UPAs, para os domicílios, todos os seus moradores e para o morador selecionado.

#### 4.2.1 Peso das Unidades Primárias de Amostragem - UPAs

Os pesos das UPAs foram calculados considerando a probabilidade de seleção da UPA para a Amostra Mestra e a probabilidade de seleção para a amostra da pesquisa. O peso básico da UPA  $i$  do estrato  $h$  na PNS é  $w_{hi}$ , dado por

$$w_{hi} = \times \frac{N_h}{N_{hi}} \times \frac{1}{m_h^{PNS}}, \quad (3)$$

em que:  $m_h$  é o número de UPAs selecionadas no estrato  $h$  para a Amostra Mestra;

$N_{hi}$  é o número de domicílios particulares permanentes ocupados, ocupados sem entrevistas realizadas (equivalentes aos domicílios fechados) e vagos na UPA  $i$  do estrato  $h$ , dados atualizados do Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE) no momento da seleção das UPAs para a Amostra Mestra;

$N_h$  é o número de domicílios particulares permanentes ocupados, ocupados sem entrevistas realizadas (equivalentes aos domicílios fechados) e vagos no estrato  $h$ , dados atualizados do CNEFE no momento da seleção das UPAs para a Amostra Mestra;

$m_h^{PNS}$  é o número de UPAs selecionadas no estrato  $h$  para a PNS.

#### 4.2.2 Peso dos domicílios e de todos os moradores

Quanto aos domicílios e moradores, os pesos foram utilizados para estimação das características investigadas, levando-se em consideração o peso da UPA correspondente e ajustes para correção de não respostas e também para calibrar as estimativas com totais populacionais conhecidos de outras fontes de registros nacionais. Os domicílios foram selecionados com probabilidade igual em cada UPA, assim o peso de seleção dos domicílios dentro da UPA, denotado por  $w_{j|hi}$ , é dado por:

$$w_{j|hi} = \frac{N_{hi}^*}{n_{hi}}, \quad (4)$$

em que:  $N_{hi}^*$  é o número de domicílios particulares permanentes ocupados e fechados na UPA  $i$  do estrato  $h$ , dados da última atualização do CNEFE no momento da seleção dos domicílios;

$n_{hi}$  é o número de domicílios selecionados na UPA  $i$  do estrato  $h$ .

Assim, o peso básico do domicílio é obtido pela seguinte expressão:

$$w_{hij} = w_{hi} \times w_{j|hi} = \frac{1}{m_h} \times \frac{N_h}{N_{hi}} \times \frac{m_h}{m_h^{PNS}} \times \frac{N_{hi}^*}{n_{hi}} \quad (5)$$

Após a determinação do peso básico, este foi ajustado para compensar as perdas de entrevistas por não resposta, ou seja, entrevistas não realizadas por recusa do informante, por não contato com o morador ou por outro motivo para



perda em domicílios ocupados, utilizando

$$w_{hij}^* = w_{hij} \times \frac{n_{hi}^*}{n_{hi}^{**}}, \quad (6)$$

na qual:

$n_{hi}^*$  é o número de domicílios selecionados com morador na UPA  $i$  do estrato  $h$ ;

$n_{hi}^{**}$  é o número de domicílios selecionados com entrevista realizada na UPA  $i$  do estrato  $h$ .

Outro ajuste feito nos pesos foi para calibrar os resultados provenientes da pesquisa a totais conhecidos de outras fontes. Os pesos foram ajustados para que ao se estimar o total populacional de certos níveis geográficos, a estimativa obtida coincida com a estimativa populacional produzida pela Coordenação de População e Indicadores Sociais (COPIS) da Diretoria de Pesquisas. A calibração foi feita para a população do dia 27 de julho de 2013. Assim, o peso final do domicílio é

$$w_{hij}^{**} = w_{hij}^* \times \frac{P_a^{tri}}{\hat{P}_a^{tri}}, \quad (7)$$

em que:

$P_a^{tri}$  é a estimativa da proporção populacional produzida pela COPIS para o nível geográfico  $a$  para do dia 27 de julho de 2013;

$\hat{P}_a^{tri}$  é a estimativa da proporção populacional obtida com os dados da pesquisa para o nível geográfico  $a$ .

#### 4.2.3 Peso do morador selecionado

Para definir o peso do morador selecionado foi preciso, primeiramente, definir a probabilidade de seleção do morador dentro do domicílio. A seleção foi feita por amostragem aleatória simples. Assim, o peso básico é dado por:

$$w_{hijk} = \frac{1}{m_h} \times \frac{N_h}{N_{hi}} \times \frac{m_h}{m_h^{PNS}} \times \frac{N_{hi}^*}{n_{hi}} \times O_{hij}, \quad (8)$$

no qual  $O_{hij}$  é o número de moradores com 18 anos ou mais de idade no domicílio  $j$  na UPA  $i$  do estrato  $h$ . O peso com ajuste de não resposta do domicílio inteiro é

dado por:

$$w_{hijk} \times \frac{n_{hi}^*}{n_{hi}^{**}}. \quad (9)$$

Quando ocorre a não resposta do domicílio inteiro, o peso com ajuste de não resposta é dado por:

$$w_{hijk}^M = w'_{hijk} \times \frac{\sum_j^{n_{hijk}^{**}} w'_{hijk} \times \alpha_{hijk}^M}{\sum_i^{n_{hijk}^{**}} w'_{hijk} \times \alpha_{hijk}^M} \quad (10)$$

$$w_{hijk}^F = w'_{hijk} \times \frac{\sum_j^{n_{hijk}^{**}} w'_{hijk} \times \alpha_{hijk}^F}{\sum_i^{n_{hijk}^{**}} w'_{hijk} \times \alpha_{hijk}^F} \quad (11)$$

em que:

$\alpha_{hijk}^M$  indica se o morador selecionado no domicílio  $j$  da UPA  $i$  do estrato  $h$  é do sexo masculino;

$\alpha_{hijk}^F$  indica se o morador selecionado no domicílio  $j$  da UPA  $i$  do estrato  $h$  é do sexo feminino.

É esperado que após a seleção do morador ocorra perda de entrevista e, portanto, é necessário que haja a correção dos pesos para compensar a não resposta. Na PNS, após a análise das estimativas de pessoas por sexo e das características dos moradores não respondentes, decidiu-se fazer o ajuste por sexo, porque percebeu-se que a perda foi maior entre os homens do que entre as mulheres. Essa calibração é feita da seguinte forma:

$$w_{hijk}^{M*} = w_{hijk}^M \times \frac{\hat{P}_{a,M,c}^M}{\hat{P}_{a,M,c}^S}, \quad (12)$$

para o sexo masculino e

$$w_{hijk}^{F*} = w_{hijk}^F \times \frac{\hat{P}_{a,F,c}^F}{\hat{P}_{a,F,c}^S}, \quad (13)$$

para o sexo feminino.

nas quais:

$\hat{P}_{a,M,c}^M$  é a estimativa populacional obtida com os dados dos moradores dos domicílios da pesquisa para o nível geográfico  $a$  do sexo masculino e classe de idade  $c$ ;

$\hat{P}_{a,F,c}^S$  é a estimativa populacional obtida com os dados dos moradores dos domicílios da pesquisa para o nível geográfico  $a$  do sexo feminino e classe de idade  $c$ .

Por se tratar de dados provenientes de um plano amostral complexo que gera estimativas a partir de dados coletados para o domicílio e todos os seus moradores (segundo estágio de seleção da amostra), assim como para os moradores de 18 anos ou mais de idade selecionados aleatoriamente (terceiro estágio), incluído um processo de calibração dos pesos em ambos os estágios, tal característica precisa ser levada em consideração em todo o processamento e análise dos dados. Toda amostra oriunda de pesquisas por múltiplos estágios deve ser analisada em programas estatísticos com algoritmos para a análise de dados complexos, capazes de corrigir o efeito que a conglomeração das unidades primárias de amostragem causa nas estimativas, o efeito do plano amostral (EPA).

### **4.3 Características sociodemográficas e percepção sobre o estado de saúde mental**

A PNS conta com diversos módulos que englobam questões associadas com informações do domicílio, visitas da equipe de saúde da família e agentes de endemias, pessoas com deficiência, cobertura de plano de saúde, utilização de serviços de saúde, saúde dos indivíduos com 60 anos ou mais, crianças com menos de dois anos, percepção do estado de saúde, acidentes, violências e agressões, estilo de vida (uso de álcool, tabagismo, consumo de alimentos, atividade física e hábito de assistir televisão), doenças crônicas transmissíveis e não transmissíveis, cobertura

de mamografia entre mulheres com 50 anos ou mais, saúde da mulher, atendimento pré-natal e saúde bucal.

Em estudos populacionais, os resultados preliminares geralmente são descritos em termos de características sociodemográficas (idade, sexo, raça, estado civil, entre outras) e tais informações dão base para conhecer o comportamento da amostra sob os mais variados aspectos. A partir das informações obtidas na PNS foi possível conhecer tais aspectos da população.

Além das questões sobre a situação de saúde, estão presentes na PNS, nove perguntas que avaliam a presença de sintomas depressivos. Avaliando essas perguntas foi possível associá-las ao *Patient Health Questionnaire-9*, abreviado pela sigla PHQ-9 (Kroenke et al., 2001). Uma reprodução exata das questões inseridas na PNS são apresentadas no anexo desse texto. A frequência de cada sintoma nas últimas duas semanas é, frequentemente, avaliada por meio de uma escala *Likert* de 0 a 3 correspondendo às alternativas “nenhum dia”, “menos da metade dos dias”, “mais da metade dos dias” e “quase todos os dias”, respectivamente. Os sintomas considerados são desinteresse, desânimo, sono (excesso ou falta), cansaço ou falta de energia, apetite (excesso ou falta), insatisfação pessoal, falta de concentração, lentidão e/ou agitação e pensamento suicida.

Usualmente, a gravidade da depressão pode ser avaliada por meio de questionários e calculada através de pontuações que são atribuídas. Entre os muitos instrumentos usados para diagnosticar a depressão, o PHQ-9 é considerado um instrumento de fácil aplicação devido ao pouco número de perguntas, característica que o coloca em vantagem quando comparado a outros instrumentos utilizados em estudos epidemiológicos. O PHQ-9 é derivado do PRIME-MD (Avaliação da atenção primária de transtornos mentais) que foi originalmente desenvolvido e validado no início dos anos 90 para diagnosticar cinco transtornos mentais comuns em atenção

primária à saúde, sendo eles a depressão, a ansiedade, o uso abusivo de álcool, os transtornos somatiformes (presença de sintomas físicos que sugerem uma condição médica geral) e os transtornos da alimentação (Spitzer et al., 1994).

Em termos de validação, Santos et al. (2013) concluíram que o PHQ-9 é um teste apropriado para rastreamento de depressão entre adultos da população geral vivendo na área urbana de cidades de porte médio semelhantes à Pelotas, RS. O PHQ-9 possui eficácia na detecção da depressão e se mostra compatível quando comparado com outros instrumentos, se tornando uma ferramenta propícia para o rastreamento dos transtornos depressivos nos mais diversos contextos sociais (Kroenke et al., 2001; de Lima et al., 2009; Monahan et al., 2009; Woldetensay et al., 2018).

## 5 FUNDAMENTAÇÃO TEÓRICA E METODOLOGIA

O objetivo deste capítulo é introduzir a teoria sobre Análise de Classes Latentes, apresentar a formulação da modelagem, descrever como são os métodos de estimação dos parâmetros e especificamente o método sanduíche, uma alternativa para estimar os erros-padrão dos estimadores dos parâmetros. No decorrer do capítulo será discutido o modelo de classes latentes básico, na presença de agrupamentos, e na presença de covariáveis. Por fim na seção 5.2 apresenta-se a metodologia específica utilizada nesta tese para alcançar os objetivos estabelecidos, na qual descreve-se as etapas de análise dos dados da PNS. Por fim são descritos aspectos em torno do suporte computacional.

### 5.1 Análise de Classes Latentes

Uma variável latente é definida como aquela que não é observável ou passível de mensuração, mas sim avaliada indiretamente através de um conjunto de duas ou mais variáveis que são possíveis de se observar. As variáveis observáveis estão sujeitas a erros ou variação aleatória e são chamadas de variáveis manifestas ou indicadoras porque o modelo de variáveis latentes postula que estas são manifestações da característica latente. Elas podem também ser referidas como itens, principalmente quando são questões inseridas em um questionário. Nesse contexto, a variável latente influencia as respostas que serão dadas às perguntas, de forma que os múltiplos indicadores e/ou escalas se destinam a informar sobre a característica não mensurada diretamente.

Existem diversos tipos de análise de variáveis latentes, cada uma recebendo um nome específico de acordo com os tipos de variáveis envolvidas (categórica ou contínua). Quando tanto as latentes quanto as indicadoras são contínuas tem-se a popular análise fatorial; quando as latentes são contínuas e as indicadoras categóricas tem-se a análise de traço latente, também conhecida por teoria de resposta ao item. Quando as variáveis latentes são categóricas e as indicadoras contínuas tem-se a análise de perfil latente e quando tanto as latentes quanto as indicadoras são categóricas tem-se a análise de classes latentes (Collins & Lanza, 2010).

É importante ressaltar que todas essas análises seguem o mesmo princípio, tal que, as variáveis indicadoras são causadas pelas latentes (não observadas) e não o contrário. É usual representar um modelo de variáveis latentes por uma figura na qual as variáveis latentes são representadas por círculos (ou elipses) e as indicadoras por quadrados (ou retângulos). Setas no sentido das latentes às indicadoras mostram a relação de causa. Na figura 5.1 a variável latente é representado por uma elipse. As variáveis indicadoras observadas medindo a latência por quadrados rotulados por  $X_1, X_2, \dots, X_9$ . Os círculos contendo as letras  $e_1, e_2, \dots, e_9$  representam os componentes de erro associados a  $X_1, X_2, \dots, X_9$ , respectivamente.

A proposta deste trabalho é identificar subgrupos relacionados à gravidade da depressão através dos nove itens presentes na PNS associados ao PHQ-9, cada um com quatro categorias de resposta. Portanto, a modelagem se insere na análise de classes cujos principais fundamentos metodológicos são apresentados.

A análise de classes latentes ou LCA, do inglês *Latent Class Analysis*, é uma técnica estatística multivariada que tem como objetivo identificar subgrupos,

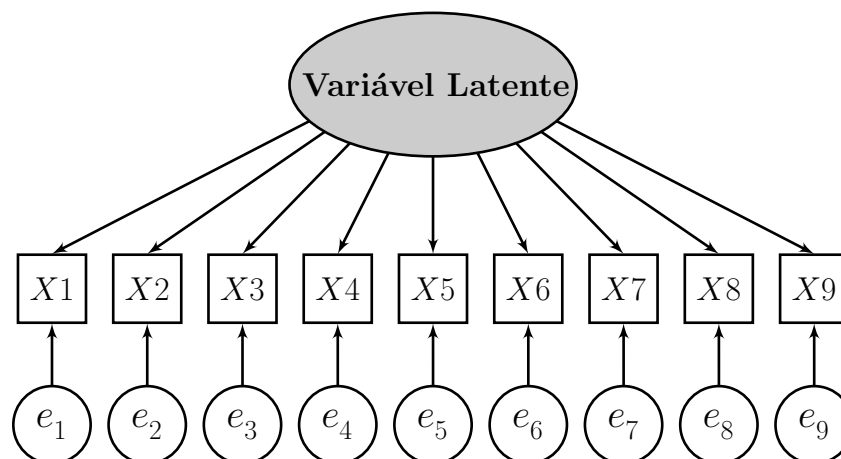


Figura 5.1: Diagrama de uma estrutura latente

tipos ou categorias de indivíduos de uma população com base nos padrões de características qualitativas (Collins & Lanza, 2010).

A LCA foi originalmente proposta por Lazarsfeld & Henry (1968), no entanto foi Goodman (1974) que descreveu os métodos de estimação por máxima verossimilhança, definindo a técnica como uma mistura de modelos que postula a existência de uma variável categórica subjacente não observada que divide uma população em classes latentes mutuamente exclusivas e exaustivas. A ideia intuitiva da LCA é descobrir subgrupos ocultos caracterizados pela interseção de múltiplos aspectos em uma população. A técnica produz a separação de uma população cuja heterogeneidade não é objetivamente ou diretamente observada.

A LCA tem sido utilizada de forma crescente em epidemiologia psiquiátrica como um método específico para dados categóricos. O método possui a vantagem de não se basear nos pressupostos dos métodos clássicos para variáveis contínuas (como análise fatorial, análise discriminante, entre outros) como o relacionamento linear, distribuição normal e homogeneidade, o que minimiza os vieses em amostras que não seguem estas suposições (Tófoli, 2008).



### 5.1.1 O modelo de classes latentes

O ponto inicial para efetuar a análise de classes latentes em um conjunto de dados seria a construção de tabelas de contingência envolvendo os itens de interesse na pesquisa. Por exemplo, para  $m$  itens cada um com  $R$  categorias, a tabela que relaciona todos os  $m$  itens contém  $R^m$  caselas. Se  $m = 9$  e  $R = 4$  tem-se 262.144 caselas. Cada casela corresponde a um padrão de resposta ao item. O modelo de classes latentes com os parâmetros, prevalências de classes e probabilidades de resposta aos itens, permite a obtenção das frequências esperadas em cada uma das caselas (Collins & Lanza, 2010).

Para compreensão do modelo é necessário estabelecer a notação matemática e definição dos parâmetros envolvidos. O desenvolvimento que se segue acompanha o trabalho de Schafer & Kang (2013), autores do pacote *lcca* presente no *software* R (R Core Team, 2018).

Seja  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iM})$  o vetor de respostas dos  $M$  itens categóricos associado a cada indivíduo  $i$ . Considera-se que  $Y_{im}$  assume um dos possíveis valores no conjunto  $1, \dots, r_m$ , correspondendo à resposta do item  $m$ . O valor observado para  $\mathbf{y}_i$  é denotado por  $\mathbf{y}_i = (y_{i1}, \dots, y_{iM})$ . No modelo usual com C-classes, assume-se que  $y_{i1}, \dots, y_{iM}$  são condicionalmente independentes dada uma variável latente categórica  $C_i$ , que assume valores (classes) de  $1, \dots, C$ . A indicação de que  $Y_{im}$  assume uma dada categoria  $r$  é denotada pela função indicadora tal que  $I(y_{im} = r) = 1$  se o indivíduo  $i$  escolhe a categoria  $r$  com  $r \in (1, \dots, r_m)$  do item  $m$  e 0 caso contrário.

No modelo LCA é possível lidar com populações multigrupos de forma que cada indivíduo é classificado, a priori, em um de  $G$  grupos ( $g = 1, \dots, G$ ). Os valores dos parâmetros e/ou a probabilidade de adesão a cada uma das classes pode variar ou não entre os grupos. Seja  $g_i$  o grupo ao qual o indivíduo  $i$  pertence, os

parâmetros de mensuração, são definidos por:

$$\rho_{m,r|c,g} = Pr(Y_{im} = r \mid L_i = c, g_i = g), \quad (14)$$

ou seja, a probabilidade de um indivíduo  $i$  qualquer, pertencente à classe  $c$  e ao grupo  $g$ , escolher a resposta  $r$  do item  $m$ . Como são probabilidades, esses parâmetros satisfazem a restrição  $\sum_{r=1}^{r_m} \rho_{m,r|c,g} = 1$  para cada combinação de  $m = 1, \dots, M, c = 1, \dots, C$  e  $g = 1, \dots, G$ . Ao assumir que os  $\rho$ 's são iguais entre os grupos, uma condição conhecida como invariância de medição, resulta no número de parâmetros livres de mensuração dado por  $C \sum_{m=1}^M (r_m - 1)$ . Caso contrário, o número de parâmetros de mensuração livres é  $GC \sum_{m=1}^M (r_m - 1)$ .

A prevalência ou probabilidade de um indivíduo  $i$  do grupo  $g$  pertencer a uma determinada classe  $c$  é dada por:

$$\gamma_{c|g} = Pr(L_i = c \mid g_i = g), \quad (15)$$

sujeito às restrições  $\sum_{c=1}^C \gamma_{c|g} = 1$  para  $g = 1, 2, \dots, G$ . Se as prevalências forem considerados iguais entre os grupos, então o número de parâmetros livres desse tipo é  $(C - 1)$ . Para  $\gamma$ 's heterogêneos entre grupos, o número de parâmetros livres é  $G(C - 1)$ .

Aplicando os conceitos de probabilidade condicional e independência, obtém-se:

$$Pr(\mathbf{Y}_i = \mathbf{y}_i, L_i = c \mid g_i = g) = \gamma_{c|g} \prod_{m=1}^M \prod_{r=1}^{r_m} \rho_{mr|cg}^{I(y_{im}=r)}, \quad (16)$$

para a classe  $c$  e

$$Pr(\mathbf{Y}_i = \mathbf{y}_i, L_i \mid g_i = g) = \prod_{c=1}^C \gamma_{c|g}^{I(L_i=c)} \prod_{c=1}^C \prod_{m=1}^M \prod_{r=1}^{r_m} \rho_{mr|cg}^{I(L_i=c) I(y_{im}=r)}, \quad (17)$$

considerando a distribuição conjunta para todas as classes, na qual  $I(L_i = c) = 1$  se o indivíduo  $i$  pertencer à classe  $c$  e 0 caso contrário.

Tomando-se a marginal de  $\mathbf{Y}_i$  sobre a variável latente, tem-se:

$$Pr(\mathbf{Y}_i = \mathbf{y}_i \mid g_i = g) = \sum_{c=1}^C \gamma_{c|g} \prod_{m=1}^M \prod_{r=1}^{r_m} \rho_{mr|cg}^{I(y_{im}=r)}. \quad (18)$$

Essa descrição caracteriza o modelo com agrupamento a priori dos indivíduos, por exemplo, por estado ou região, mas sem covariáveis. Com posse de dados amostrais, o ajuste do modelo envolve estimação de dois conjuntos de parâmetros: mensuração e prevalência. Na seção 5.1.2 o método de estimação é considerado.

### 5.1.2 Método de estimação

Fazendo o produto da expressão (17) para os  $n$  indivíduos constrói-se a função de verossimilhança cuja maximização leva às estimativas dos parâmetros  $\gamma$ 's e  $\rho$ 's. No entanto para modelos de classes latentes a maximização é complexa e não existe solução analítica. É necessário o uso de um método iterativo que se inicia quando um conjunto de valores iniciais são dados aos parâmetros que vão sendo atualizados de acordo com um determinado algoritmo.

Dados de amostragem complexa são acompanhados por pesos. O peso do indivíduo  $i$ , denotado por  $w_i$  pode ser interpretado como sendo o número de indivíduos na população representados pelo indivíduo  $i$  na amostra. No caso da PNS, a obtenção dos pesos está descrita na seção 4.2.1 e maiores detalhes podem ser consultados em Souza-Junior et al. (2015). O método de estimação dos parâmetros é similar ao caso da amostragem aleatória simples, porém agora é como se o indivíduo  $i$  fosse clonado  $w_i$  vezes, caracterizando o método de máxima pseudo-verossimilhança.

A implementação de Schafer & Kang (2013) permite dados na forma agregada na qual  $f_i$  representa o número de indivíduos com o mesmo padrão de resposta  $\mathbf{y}_i$ . Assim, nas fórmulas que se seguem, no caso de dados com pesos  $w_i$ , estes substituem  $f_i$ . Seja  $\boldsymbol{\theta}$  o vetor com todos os parâmetros ( $\rho$ 's e  $\gamma$ 's) livres que serão estimados. Seja  $I_i(\boldsymbol{\theta})$  a contribuição do indivíduo  $i$  do grupo  $g$  no logaritmo

da verossimilhança, tal que o log da verossimilhança de todos os dados observados é dada por:

$$\begin{aligned}
 I(\boldsymbol{\theta}) &= \sum_{i=1}^n I_i(\boldsymbol{\theta}) \\
 &= \sum_{i=1}^n \log Pr(\mathbf{Y}_{i,obs} = \mathbf{y}_{i,obs} \mid g_i = g)^{f_i} \\
 &= \sum_{g=1}^G \sum_{i \in g} f_i \log \left( \sum_{c=1}^C \gamma_{c|g} \prod_{m \in obs_i} \prod_{r=1}^{r_m} \rho_{mr|cg}^{I(y_{im}=r)} \right), \tag{19}
 \end{aligned}$$

lembrando que,  $f_i$  pode representar a frequência de indivíduos em dados agregados ou pesos em amostragem complexa. Para dados não agregados dentro de um plano de amostragem simples  $f_i = 1$ . Analiticamente, é muito complicado maximizar a função  $I(\boldsymbol{\theta})$ . Nesse caso, é necessário recorrer a um método iterativo que computacionalmente gera uma sequência de soluções aproximadas que vão melhorando conforme iterações são executadas.

### 5.1.3 Algoritmo EM

O algoritmo EM (*expectation-maximization*) proposto por Dempster et al. (1977) é uma ferramenta computacional utilizada para o cálculo do estimador de máxima verossimilhança (EMV) de forma iterativa. Em cada iteração do algoritmo EM, maximiza-se o valor esperado da função de log- verossimilhança que aumenta

os dados observados com a variável latente  $L_i$ , dada da seguinte forma:

$$\begin{aligned}
I^*(\boldsymbol{\theta}) &= \sum_{i=1}^n I_i^*(\boldsymbol{\theta}) \\
&= \sum_{i=1}^n \log Pr(\mathbf{Y}_{i,obs} = \mathbf{y}_{i,obs}, L_i | g_i = g)^{f_i} \\
&= \sum_{g=1}^G \sum_{i \in g} f_i \log \left( \prod_{c=1}^C \gamma_{c|g}^{I(L_i=c)} \prod_{c=1}^C \prod_{m \in obs_i} \prod_{r=1}^{r_m} \rho_{mr|cg}^{I(L_i=c)I(y_{im}=r)} \right) \quad (20) \\
&= \sum_{g=1}^G \sum_{i \in g} \sum_{c=1}^C f_i I(L_i = c) \log \gamma_{ic} \\
&+ \sum_{g=1}^G \sum_{i \in g} \sum_{c=1}^C \sum_{m \in obs_i} \sum_{r=1}^{r_m} f_i I(L_i = c) I(y_{im} = r) \log \rho_{mr|cg}.
\end{aligned}$$

**Etapa E:** Nesta etapa, calcula-se a esperança de (20) com relação à distribuição dos dados ausentes, dado os dados observados, fixando-se os parâmetros desconhecidos nos valores das estimativas atuais. Na referida equação, tem-se uma função linear dos indicadores de  $I(L_i = c)$ . Substituindo-se tais indicadores pelas suas esperanças, que são as probabilidades posteriores de associação à classe  $c$  dados os itens observados. As probabilidades posteriores serão

$$\eta_{ic} = \Pr(L_i = c | \mathbf{Y}_{i,obs} = \mathbf{y}_{i,obs}, g_i = g) \quad (21)$$

$$= \frac{\gamma_{c|g} \prod_{m \in obs_i} \prod_{r=1}^{r_m} \rho_{mr|cg}^{I(y_{im}=r)}}{\sum_{c'=1}^C \gamma_{c'|g} \prod_{m \in obs_i} \prod_{r=1}^{r_m} \rho_{mr|c'g}^{I(y_{im}=r)}}. \quad (22)$$

A etapa E consiste em calcular o vetor de probabilidades posteriores  $(\eta_{i1}, \dots, \eta_{iC})$  para  $i = 1, \dots, n$  com base nas estimativas atualizadas para  $\gamma's$  e  $\rho's$ .

**Etapa M:** Para realizar esta etapa, nota-se que a esperança em (20) é a soma de dois termos. Um depende dos  $\rho's$  e o outro depende dos  $\gamma's$ . O máximo global com relação a  $\boldsymbol{\theta}$  é obtido maximizando os dois termos separadamente. Quando não assume-se igualdade dos  $\rho's$  entre os grupos, a maximização em relação aos  $\rho's$

é dada por

$$\begin{aligned}\hat{\rho}_{mr|cg} &= \frac{\sum_{i \in \text{obs}_{(g,m)}} f_i \eta_{ic} I(y_{im} = r)}{\sum_{r'=1}^{r_m} \sum_{i \in \text{obs}_{(g,m)}} f_i \eta_{ic} I(y_{im} = r')} \\ &= \frac{\sum_{i \in \text{obs}_{(g,m)}} f_i \eta_{ic} I(y_{im} = r)}{\sum_{i \in \text{obs}_{(g,m)}} f_i \eta_{ic}}, \quad r = 1, \dots, r_m\end{aligned}\quad (23)$$

para cada combinação do grupo  $g$ , classe  $c$  e item  $m$ , em que  $\text{obs}_{(g,m)}$  denota o subconjunto de casos dentro do grupo  $g$  para o qual  $Y_{im}$  não está ausente (*non-missing*). Se a igualdade de  $\rho$ 's para todos os grupos for assumida, ou seja

$$\rho_{mr|c1} = \rho_{mr|c2} = \dots, \rho_{mr|cG} = \rho_{mr|c}, \quad (24)$$

então, o máximo ocorre para  $\rho$ 's dados por

$$\begin{aligned}\hat{\rho}_{mr|c} &= \frac{\sum_{i \in \text{obs}_{(m)}} f_i \eta_{ic} I(y_{im} = r)}{\sum_{r'=1}^{r_m} \sum_{i \in \text{obs}_{(m)}} f_i \eta_{ic} I(y_{im} = r')} \\ &= \frac{\sum_{i \in \text{obs}_{(m)}} f_i \eta_{ic} I(y_{im} = r)}{\sum_{i \in \text{obs}_{(m)}} f_i \eta_{ic}}, \quad r = 1, \dots, r_m\end{aligned}\quad (25)$$

para cada combinação da classe  $c$  e item  $m$ , onde  $\text{obs}_{(m)}$  denota o subconjunto de casos para os quais  $Y_{im}$  não está ausente (*non-missing*). Se a igualdade dos  $\gamma$ 's nos grupos não for assumida, então o máximo ocorre para  $\gamma$ 's dadas por

$$\hat{\gamma}_{c|g} = \frac{\sum_{i \in g} f_i \eta_{ic}}{\sum_{c'=1}^C \sum_{i \in g} f_i \eta_{ic'}} \quad (26)$$

para  $c = 1, \dots, C$  e  $g = 1, \dots, G$ . Se os  $\gamma$ 's são restritos a serem iguais entre os grupos, então, o máximo ocorre quando

$$\hat{\gamma}_{c|g} = \hat{\gamma}_c = \frac{\sum_{i=1}^n f_i \eta_{ic}}{\sum_{c'=1}^C \sum_{i=1}^n f_i \eta_{ic'}}. \quad (27)$$

Para iniciar o algoritmo é necessário fornecer valores iniciais para os parâmetros, que serão então atualizados pelo método descrito com o objetivo de encontrar os melhores valores, ou seja, aqueles que maximizam o log da função de verossimilhança dada em (20). Quando  $f_i$  é substituído por  $w_i$ , ou seja, o peso do indivíduo  $i$  devido ao esquema da amostragem complexa a função se torna de pseudo-verossimilhança.

### 5.1.4 Erros-padrão para $\hat{\boldsymbol{\theta}}$

Seja  $\psi_i(\boldsymbol{\theta}) = \frac{\partial I_i(\boldsymbol{\theta})}{\partial(\boldsymbol{\theta})}$ , conhecida como função escore e  $\boldsymbol{\psi}_i(\boldsymbol{\theta}) = \sum_{i=1}^n \psi_i(\boldsymbol{\theta})$ . Uma vez estimados os parâmetros do modelo, é necessário obter os erros-padrão das estimativas. Uma das alternativas é aproximá-los pelo método conhecido como sanduíche (Carroll et al., 1998). O método aproxima a variância de  $\hat{\boldsymbol{\theta}}$  através de

$$\hat{V}(\hat{\boldsymbol{\theta}}) \approx \left( - \sum_{i=1}^n \hat{\boldsymbol{\psi}}_i' \right)^{-1} \left( \sum_{i=1}^n f_i^{-1} \hat{\boldsymbol{\psi}}_i \hat{\boldsymbol{\psi}}_i^T \right) \left( - \sum_{i=1}^n \hat{\boldsymbol{\psi}}_i' \right)^{-1}, \quad (28)$$

na qual

$$\boldsymbol{\psi}_i'(\boldsymbol{\theta}) = \frac{\partial^2 I_i(\boldsymbol{\theta})}{\partial(\boldsymbol{\theta})\partial(\boldsymbol{\theta})} \quad e \quad \hat{\boldsymbol{\psi}}_i' = \boldsymbol{\psi}_i'(\hat{\boldsymbol{\theta}}). \quad (29)$$

No entanto, para levar em conta o esquema de amostragem complexo com  $H$  estratos cada um com  $U_h$  conglomerados, o estimador sanduíche obedece a modificação dada por

$$V(\hat{\boldsymbol{\theta}}) \approx \left( - \sum_{i=1}^n \boldsymbol{\psi}_i' \right)^{-1} \left( \sum_{h=1}^H \sum_{u=1}^{U_h} (\boldsymbol{\psi}_{uh} - \bar{\boldsymbol{\psi}}_h)(\boldsymbol{\psi}_{uh} - \bar{\boldsymbol{\psi}}_h)^T \right) \left[ \left( - \sum_{i=1}^n \boldsymbol{\psi}_i' \right)^{-1} \right]^T, \quad (30)$$

em que  $\boldsymbol{\psi}_{uh} = \sum_{i=1}^{n_{uh}} \boldsymbol{\psi}_{iuh}$  é o escore total dentro do cluster  $u$  no estrato  $h$ ,  $\bar{\boldsymbol{\psi}}_h = U_h^{-1} \sum_{u=1}^{U_h} \boldsymbol{\psi}_{uh}$  é a média dos totais dos conglomerados dentro do estrato  $h$  e

$$\boldsymbol{\psi}_i' = \frac{\partial}{\partial \boldsymbol{\theta}^T} \boldsymbol{\psi}_i(\boldsymbol{\theta}).$$

Para obter  $\hat{V}(\hat{\boldsymbol{\theta}})$  substitui-se as funções dos parâmetros avaliados em  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ .

### 5.1.5 O modelo de classes latentes com covariáveis

Assim como acontece em um modelo de regressão convencional, o propósito de introduzir covariáveis em um modelo de classes latentes é identificar características que predizem a associação nas várias classes latentes que possam

representar um dado fenômeno.

Seja  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$  um vetor de dimensão  $p \times 1$  de covariáveis associado ao indivíduo  $i$  que pode influenciar a probabilidade de pertencer a classe  $L_i = c$  para  $c = 1, \dots, C$ . É suposto que  $\mathbf{x}_i$  influencia os itens  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iM})^T$  somente através do classificador latente  $L_i$ .

As covariáveis podem ser relacionadas às probabilidade de classificação através de um modelo de regressão logística categórico com umas das classes como referência conforme apresentado em Agresti (2003). A probabilidade que um indivíduo  $i$  dentro do grupo  $g$  pertencer a classe  $c$  é modelada por:

$$\gamma_{ic|g} = Pr(L_i = c \mid g_i = g) = \frac{\exp(\mathbf{x}_i^T \boldsymbol{\alpha}_{c|g})}{\sum_{c'=1}^C \exp(\mathbf{x}_i^T \boldsymbol{\alpha}_{c'|g})}, \quad (31)$$

em que  $\boldsymbol{\alpha}_{c|g} = (\alpha_{1c|g}, \alpha_{2c|g}, \dots, \alpha_{pc|g})^T$  é um vetor  $p \times 1$  de coeficientes a serem estimados. Para identificabilidade dos parâmetros deve-se escolher uma das classes para servir de referência para as demais. Assim, se a classe  $d$  for a referência, define-se  $\boldsymbol{\alpha}_{d|g} = (0, 0, \dots, 0)^T$  para essa classe. Considerando tal restrição, tem-se:

$$\log \left( \frac{\gamma_{ic|g}}{\gamma_{id|g}} \right) = \mathbf{x}_i^T \boldsymbol{\alpha}_{c|g}, \quad (32)$$

de forma que os elementos de  $\boldsymbol{\alpha}_{ci/g}$  tornam-se o *log* da razão de chances para distinguir a classe  $c$  da classe referência  $d$ . Para estimação dos parâmetros o método é similar ao descrito na seção 5.1.2, porém agora  $\gamma_{ic|g}$  é função das covariáveis e dos parâmetros  $\alpha$ 's.

## 5.2 LCA e depressão

A configuração dos transtornos mentais, em especial a depressão, contempla uma diversidade de áreas de conhecimentos. No entanto, dada a complexidade envolvida no processo de diagnóstico e tratamento é difícil prever um perfil sociodemográfico e comportamental e em se tratando de grandes populações,



o fator de complexidade aumenta substancialmente considerando o objetivo de delinear possíveis grupos de risco e elucidar subtipos de depressão. Abelha (2014) descreve a depressão como um problema relevante e urgente de saúde pública.

Ulbricht et al. (2018) apresenta uma revisão sistemática sobre o uso da LCA na identificação de subtipos de depressão, técnica que vem sendo utilizada no processo de identificação de padrões em relação aos subtipos de depressão. Esses subtipos, geralmente se baseiam na etiologia, sintomas, tempo de diagnóstico, local de residência do indivíduo, e por meio da LCA é possível criar subgrupos a partir de uma grande quantidade de informações e assim buscar alternativas quanto aos muitos critérios de classificação adotados frequentemente, tais como a soma de escores e pontos de cortes.

O uso da LCA como método de análise conduz a característica em torno da gravidade da depressão atrelando probabilidades às respostas contidas em um determinado instrumento de investigação, tal com o PHQ-9. O problema em agrupar variáveis ou rotular uma determinada classe que pode ser definida pelo número de sintomas e não pela severidade gera muitos desencontros com relação a eficiência da LCA na identificação de padrões. Ao considerar uma pesquisa de base populacional é esperado que haja heterogeneidade entre e intra aos diversos seguimentos que compõem o espaço amostral.

### **5.3 Material e Métodos**

No contexto de uma pesquisa, especificamente a PNS que abrange todo o território nacional, a premissa da LCA é empregada para investigar a formação de classes a partir do agrupamento dos indivíduos com base em cada um dos itens da pesquisa e os fatores que podem exercer maior ou menor força na formação dos mesmos, sob os diversos cenários que compõem a PNS.

Os dados da PNS são disponibilizados através de microdados que se constituem no menor nível de desagregação de dados recolhidos na pesquisa possibilitando a construção de tabelas, índices e demais medidas de informações que sejam necessárias. O processo de importação da base de dados para o *software* R foi realizado com auxílio do pacote *lodown*, proposto por Damico (2017). O pacote prepara o ambiente adequado para realizar o *downloads* de microdados públicos.

Os inquéritos populacionais abrangem um grande número de informações em que as mesmas possuem variadas estruturas de respostas. Indivíduos e domicílios condensam dados demográficos, educacionais, rendimentos e especialmente, no caso da PNS, é possível ter acesso a questões a cerca das condições e do acesso aos serviços de saúde. Ao todo 64.348 entrevistas domiciliares e 60.202 entrevistas individuais com o morador selecionado no domicílio formam o banco de dados da PNS (Damacena et al., 2015). As primeiras investigações realizadas nos dados da PNS foram no âmbito sociodemográfico para compreender características básicas da população.

Com o intuito de examinar os padrões de respostas de indivíduos que responderam à PNS quanto à sua própria percepção dos sintomas depressivos, modelos de classes latentes serão ajustados e os grupos distintos (classes latentes) serão identificados baseados nos padrões de respostas observadas em variáveis categóricas. Tudo com base em um modelo probabilístico, com a capacidade de identificar características que indicam bem os grupos, estimar a prevalência de cada grupo e classificar cada indivíduo dentro dos grupos.

Dada a complexidade e a dimensão da PNS e levando em consideração as diversidades regionais quanto a economia, o desenvolvimento, o índice de desenvolvimento humano, a oferta e o acesso aos serviços de saúde entre tantos outros seguimentos, faz-se necessário considerar possíveis diferenças que podem

existir entre as regiões quanto ao problema aqui abordado. Nesse sentido, foram realizadas análises em separado para cada unidade da federação e assim, um modelo de classes latentes com covariáveis foi ajustado para cada uma delas. No entanto, para apresentação foram considerados na primeira etapa os resultados para cinco estados, um em cada região do país e dois como referência.

Quanto ao número de classes definidas no modelo, não existe um teste estatístico formal para esse parâmetro, mas sim medidas descritivas (critérios de informação) e aspectos conceituais do problema, o que pode ser em grande parte elucidado por especialista da área da aplicação. Considerando que os indivíduos da amostra podem ser classificados de acordo com a gravidade do transtorno, vislumbrou-se a separação em quatro classes visto as possibilidades de respostas quanto a frequência de sintomas inseridos no questionário. Algumas variáveis como raça e renda apresentaram problemas no ajuste do modelo devido as baixas frequências em determinadas categorias.

### 5.3.1 Suporte computacional

No presente trabalho, as análises foram conduzidas no *software* estatístico R Core Team (2018), versão 3.5.1. A importação da base de dados da PNS foi realizada através do pacote *lodown* destinado ao *download* de microdados, proposto por Damico (2017). A manipulação e elaboração das análises estatísticas foram conduzidas com auxílio do pacote *survey*, criado para a análise de pesquisas cujo desenho amostral é complexo, Lumley (2004). O processo de estimação das classes latentes foi conduzido por meio do pacote *lcca* (*Latent Class Causal Analysis*), Schafer & Kang (2013). O texto foi redigido na linguagem  $\text{\LaTeX}$  (Lamport, 1994).

## 6 RESULTADOS

Neste capítulo são apresentados os resultados referentes a modelagem de classes latentes para estudar os perfis de subtipos dos brasileiros no que concerne a depressão utilizando os dados relacionados ao PHQ-9 da PNS, 2013, incorporando covariáveis tradicionalmente relacionadas ao tema. Na seção 6.1 apresenta-se estatísticas descritivas dessas covariáveis e dos itens do PHQ-9. A seção 6.2 se refere aos resultados da LCA, iniciando com uma exploração do número de classes latentes apropriado, análises separadas por UF seguido do ajuste de um modelo global.

### 6.1 Análise Descritiva

De posse do conjunto de dados da Pesquisa Nacional de Saúde, composto por uma amostra de tamanho  $n = 60.202$ , foram realizadas análises descritivas preliminares com o objetivo de conhecer características gerais da população bem como aquelas que podem estar associadas ao problema específico da pesquisa, ou seja, o perfil de resposta quanto a questões associadas a depressão e seus respectivos agravos.

#### 6.1.1 Características Sociodemográficas

Inicialmente, na Tabela 6.1 é apresentada uma descrição das características sociodemográficas em termos populacionais e amostrais. No geral a amostra apresenta certo balanceamento quanto a relação percentual entre os sexos, sendo que a categoria feminina corresponde a 56,95% da amostra. No que diz respeito a idade, as categorias mais expressivas estão relacionadas às classes iniciais,

em que, aproximadamente 66% da amostra corresponde ao grupo de idade entre 18 e 50 anos. Vale ressaltar que inicialmente a variável idade foi distribuída em 6 classes ([18;30); [30;40); [40;50); [60;70); [70+)) porém, avaliando os resultados da modelagem preliminar verificou-se a possibilidade do agrupamentos em 4 classes que são apresentadas na Tabela 6.1.

A variável escolaridade foi subdividida em 4 classes: SI/FI (Sem Instrução ou Fundamental Incompleto), F/MI (Fundamental Completo ou Médio Incompleto), M/SI (Médio Completo ou Superior Incompleto) e S (Superior). A frequência mais elevada corresponde à categoria SI/FI (40,00%), seguida da M/SI (31,81%).

Mais da metade dos respondentes (57,34%) declararam ter companheiro(a) enquanto que apenas 14,02% responderam viver sozinho(a).

As raças mais predominantes na amostra correspondem à parda e branca que equivalem, respectivamente, a 49,03% e 40,04%.

A renda foi mensurada em número de salários mínimos a considerar o valor referente ao ano de 2013, sendo este igual a R\$ 678,00. A maior proporção ficou associada ao grupo que declarou receber até 2 salários mínimos, (67,84%). Não foi possível obter estimativas populacionais devido a alta frequência de não respostas.

As covariáveis raça e renda não foram inseridas no ajuste do modelo de classes latentes devido a problemas de estimação causados pelas baixas frequências em algumas categorias.

Tabela 6.1. Características sociodemográficas para adultos brasileiros - PNS-2013, Brasil (total populacional estimado = 146.308.458,  $n = 60.202$ )

Variáveis	Categoria	População		Amostra	
		N	%	$n$	%
Sexo	Masculino	68.916.470	47,10	25.920	43,05
	Feminino	77.391.988	52,90	34.282	56,95
Idade	[18; 30)	38.157.850	26,08	14.321	23,79
	[30; 40)	31.643.091	21,63	14.269	23,70
	[40; 60)	50.099.686	34,24	20.435	33,94
	[60+)	26.407.831	18,05	11.177	18,57
Escolaridade	SI/FI	56.960.795	38,93	24.083	40,00
	F/MI	22.715.539	15,53	9.215	15,31
	M/SI	47.995.480	32,80	19.149	31,81
	S	18.636.644	12,74	7.755	12,88
Tem companheiro(a)	Sim	895.373.28	61,20	34.522	57,34
	Não	567.711.30	38,80	25.680	42,66
Vive sozinho(a)	Sim	100.751.21	6,88	8.438	14,02
	Não	136.233.338	93,12	51.764	85,98

### 6.1.2 Itens relacionados ao PHQ-9

Os itens relacionados aos sintomas responsáveis por identificar o perfil do indivíduo quanto a um possível quadro depressivo foram mensurados em termos dos escores considerados e o percentual das respostas estão apresentados na Tabela 6.2. Os escores de 0 a 3 correspondem às frequências dos sintomas observados em 0 (nenhum dia), 1 (menos da metade dos dias), 2 (mais da metade dos dias) e 3 (quase todos os dias), respectivamente, nas duas últimas semanas de referência do estudo.

Quanto ao quesito desinteresse, observa-se que aproximadamente 76% dos indivíduos não apresentaram o sintoma em nenhum dia durante o período considerado. Em menos da metade dos dias (escore 1), 15% não relataram o sintoma. O item associado ao desânimo obteve cerca de 78% das respostas relativas à ausência de sintomas no período, ao passo que 19% relataram a presença de cansaço em menos da metade dos dias considerados.

Variações no sono e falta de concentração apresentaram proporções equivalentes quanto à ausência do sintoma (escore 0) com frequências de respostas próximas de 75%. O sintoma se apresentou em menos da metade dos dias para aproximadamente 15% daqueles que o relataram. Quanto aos itens que tratam sobre o cansaço ou falta de energia e alterações no apetite, em torno de 83% dos indivíduos não apresentaram os sintomas em nenhum momento. Cerca de 10% notaram os itens em menos da metade dos dias no período considerado, enquanto que discretos 8% observaram maior presença dos itens (escore 2 e 3).

Insatisfação pessoal e lentidão e/ou agitação foram sintomas que não se manifestaram para aproximadamente 87% dos indivíduos, enquanto 8,4% notaram a presença dos sintomas em menos da metade dos dias (escore 1). O pensamento suicida esteve presente em menos da metade dos dias para 2,6% dos indivíduos, enquanto que 1,4% dos indivíduos declararam a presença do sintoma em mais da metade do dias (escore 2) ou em quase todos os dias (escore 3).

Tabela 6.2: Proporção amostral das respostas em cada item do PHQ-9, PNS-2013, Brasil

Itens	Escore			
	0	1	2	3
Desinteresse	0,765	0,155	0,038	0,042
Desânimo	0,776	0,142	0,041	0,041
Sono (excesso ou falta)	0,703	0,145	0,056	0,096
Cansaço ou falta de energia	0,688	0,189	0,055	0,068
Apetite (excesso ou falta)	0,817	0,103	0,040	0,041
Insatisfação pessoal	0,871	0,083	0,023	0,023
Falta de concentração	0,825	0,113	0,030	0,032
Lentidão e/ou agitação	0,861	0,085	0,026	0,029
Pensamento suicida	0,960	0,026	0,007	0,007

## 6.2 Análise de Classes Latentes

Para ajustar um modelo de classes latentes o pesquisador encontra alguns grandes desafios como a escolha do número de classes latentes, a decisão sobre as restrições que devem ser impostas aos parâmetros  $\rho$ 's assim como aos parâmetros  $\gamma$ 's, no caso de haver agrupamento natural na população de interesse, além da seleção de covariáveis.

### 6.2.1 Escolha do número de classes

Primeiramente é necessário decidir sobre o número de classes latentes ( $C$ ) que devem ser postuladas. Como em toda modelagem estatística, tais decisões devem ser norteadas por parcimônia e interpretabilidade dos resultados, além de critérios estatísticos. Como modelos com diferentes valores de  $C$  não são encaixados, o teste estatístico de razão de verossimilhança, usualmente empregado para comparar modelos, não pode ser aplicado, restando as alternativas do uso



de medidas descritivas, como os critérios de informação (AIC, BIC, por exemplo), agregando-se também aspectos conceituais do problema, o que pode ser em grande parte elucidado por especialista da área de aplicação. Collins & Lanza (2010) sugerem a seleção de  $C$ , para amostras aleatórias simples, via valores de AIC, BIC e logaritmo da função de verossimilhança ( $\log(MV)$ ). No entanto, para o caso de amostragem em sistema complexo, AIC e BIC não estão consensualmente definidos, conforme apontam (Schafer & Kang, 2013). Assim, utilizou-se a sugestão desses autores de, inicialmente, ignorar a estrutura complexa de amostragem para se ter uma ideia sobre o número de classes.

Dada a complexidade e dimensão da Pesquisa Nacional de Saúde e o interesse em explorar as heterogeneidades regionais, neste trabalho considerou-se importante estimar as prevalências das classes latentes em cada unidade da federação (UF) e assim, essa variável foi considerada como agrupamento a priori na modelagem, apresentando 27 grupos. Para decidir sobre o número de classes ajustou-se então quatro modelos ( $C = 2, 3, 4, 5$ ), sem covariáveis e com a restrição de  $\rho$ 's homogêneos entre as UF's. Esses ajustes foram também refeitos considerando o delineamento amostral dos quais obteve-se os valores do logaritmo do máximo da função pseudo-verossimilhança ( $\log(MPV)$ ). Algumas estatísticas de ajuste são apresentadas na Tabela 6.3.

Tabela 6.3: Resumo das estatísticas de ajuste para selecionar o número de classes latentes, PNS-2013, Brasil

$C$	$p$	G.L.	AIC	BIC	$\log(MV)$	Dif*	$\log(MPV)$
2	81	7.077.806	599.089,86	599.819,30	-299.463,93	-	-291.422,67
3	135	7.077.752	578.530,21	579.745,94	-289.130,10	10.333,83	-281.335,43
4	189	7.077.698	571.465,09	573.167,12	-285.543,54	3.586,56	-277.698,32
5	243	7.077.644	566.932,38	569.120,70	-283.223,18	2.320,36	-275.319,28

Dif\*: diferença entre  $\log(MV)$  de dois modelos ( $\log(MV_C) - \log(MV_{C-1})$ ), um com  $C$  e o outro com  $C - 1$  classes ( $C = 1, 2, 3, 4, 5$ ),  $p$ : número de parâmetros.

Lembrando que na comparação de modelos, aqueles com  $AIC$  ou  $BIC$  menores e  $\log(MV)$  maiores são preferíveis, as estatísticas da Tabela 6.3 indicam ganho nos ajustes com números de classe maiores, embora, relativamente, o impacto do aumento de 4 para 5 classes seja bem menor do que de 2 para 4, por exemplo. Assim, optou-se por 4 classes já que o uso de muitas classes, além de não resultar em modelo parcimonioso, dificulta a interpretação. A ideia é que quatro classes permitam separar a população em grupos dos tipos “sem sintomas”, “gravidade leve”, “gravidade moderada” e “gravidade alta”, embora tal rotulação dependa da interpretação dos parâmetros estimados após o ajuste.

Então, fixando  $C = 4$ , ajustou-se modelos separados para cada UF como forma de se ganhar conhecimento sobre as heterogeneidades regionais, considerando como covariáveis as características relacionadas ao perfil social dos indivíduos, descritas na Tabela 6.1.

### 6.2.2 Modelos separados para as UF's

Ajustar modelos separados foi a alternativa viável encontrada ao problema devido ao número expressivo de parâmetros para se estimar no modelo global (número de parâmetros  $p = 3.807$ ). Já para cada UF o número de parâmetros no modelo é 141. Essas análises separadas tem objetivo apenas exploratório.

Na apresentação dos resultados dessas análises preliminares e exploratórias não haverá preocupação com interpretações. O interesse é explorar o comportamento geral nas UF's, ou seja, investigar padrões de homogeneidade ou não nas estimativas dos parâmetros.

Inicia-se com a investigação das prevalências das classes latentes nas 27 UF's, apresentadas na Tabela 6.4. As classes estão ordenadas no sentido da gravidade ou maior risco, sendo a classe 4 a mais grave.

As prevalências estimadas na classe 1 variam de 53,3% (RS) a 70,0% (RJ) com certo destaque de valores mais baixos para as regiões Sul e Nordeste e valores mais altos para a região Sudeste. Em relação às classes de gravidade mais altas já aparece maior heterogeneidade das prevalências dentro de cada região. A UF com maior valor na classe 3 foi o RS (14,3%) seguido por MT (11,9%), RR (10,4%) e PA (10,1%). Demais UF's apresentaram valores inferiores a 10% nessa classe. Em relação à classe 4 a UF com maior prevalência foi AP (11,5%) seguida de PB (9,1%). As UF's com menores valores para a classe 3 foram MG (3,5%), BA (4,0%) e PR (4,4%) e para a classe 4 foram PA (3,5%), PI (4,1%) e RS (4,2%). Tais resultados mostram considerável heterogeneidade entre as UF's no que diz respeito aos parâmetros  $\gamma$ 's (prevalências das classes). Estimativas para o estado do Acre não puderam ser obtidas devido a não adequação do ajuste do modelo ao número de classes consideradas. Observou-se também um número elevado de dados faltantes para o referido estado.

Tabela 6.4: Estimativas das prevalências e respectivos erros-padrão (E.P.) segundo unidades da federação, PNS-2013, Brasil

UF	$n$	%	Classes latentes							
			1	(E.P.)	2	(E.P.)	3	(E.P.)	4	(E.P.)
RO	1.694	2,81	0,657	0,025	0,232	0,024	0,065	0,014	0,046	0,009
AC	1.814	3,01	-	-	-	-	-	-	-	-
AM	2.586	4,30	0,602	0,020	0,265	0,016	0,074	0,018	0,060	0,009
RR	1.591	2,64	0,639	0,028	0,168	0,035	0,104	0,025	0,089	0,010
PA	2.004	3,33	0,619	0,036	0,245	0,039	0,101	0,019	0,035	0,006
AP	1.332	2,21	0,565	0,024	0,240	0,170	0,080	0,016	0,115	0,184
TO	1.515	2,52	0,565	0,031	0,309	0,031	0,061	0,021	0,065	0,017
MA	1.774	2,95	0,690	0,022	0,170	0,023	0,064	0,014	0,075	0,012
PI	1.804	3,00	0,539	0,062	0,330	0,045	0,090	0,032	0,041	0,007
CE	2.560	4,25	0,609	0,029	0,249	0,025	0,093	0,015	0,048	0,009
RN	1.691	2,81	0,538	0,105	0,325	0,053	0,074	0,140	0,064	0,015
PB	1.943	3,23	0,586	0,026	0,255	0,024	0,068	0,017	0,091	0,019
PE	2.591	4,30	0,658	0,022	0,222	0,024	0,058	0,008	0,062	0,009
AL	1.748	2,90	0,572	0,027	0,265	0,024	0,092	0,016	0,070	0,019
SE	1.553	2,58	0,603	0,040	0,287	0,027	0,060	0,026	0,050	0,048
BA	2.641	4,39	0,677	0,023	0,226	0,027	0,040	0,006	0,057	0,014
MG	3.779	6,28	0,661	0,030	0,237	0,027	0,035	0,006	0,067	0,012
ES	1.724	2,86	0,683	0,030	0,198	0,028	0,048	0,009	0,070	0,012
RJ	3.486	5,79	0,700	0,018	0,201	0,017	0,049	0,009	0,050	0,011
SP	5.305	8,81	0,671	0,021	0,222	0,024	0,056	0,012	0,051	0,007
PR	3.012	5,00	0,570	0,029	0,320	0,023	0,044	0,008	0,066	0,015
SC	1.623	2,70	0,631	0,025	0,248	0,026	0,059	0,021	0,062	0,015
RS	2.913	4,84	0,533	0,029	0,281	0,031	0,143	0,025	0,042	0,010
MS	1.809	3,00	0,592	0,037	0,288	0,035	0,050	0,024	0,071	0,022
MT	1.476	2,45	0,602	0,041	0,211	0,079	0,119	0,050	0,068	0,010
GO	2.423	4,02	0,608	0,021	0,283	0,025	0,053	0,012	0,056	0,013
DF	1.811	3,01	0,573	0,048	0,297	0,030	0,076	0,023	0,053	0,010

Para cada UF tem-se 108 estimativas de parâmetros tipo  $\rho$  que são as probabilidades de resposta aos itens dada a classe latente. Devido a grande quanti-

dade de estimativas geradas, decidiu-se apresentar em detalhes aqueles relativos a apenas uma UF pertencente a cada uma das cinco regiões do país. As UF's foram selecionadas por sorteio tal que o estado do Pará representa a região norte, a Paraíba representa o nordeste. O Espírito Santo representa o sudeste. Santa Catarina e Mato Grosso representam as regiões sul e centro-oeste, respectivamente. A título de comparação, também serão discutidos os resultados para os estados de São Paulo e Rio Grande do Sul, já que estes são os estados nos quais se encontra o maior número de estudos publicados sobre a problemática da depressão (Cross et al., 2018).

Para os sete estados selecionados, observa-se alguns padrões gerais (Tabelas 6.5 a 6.11) de relações entre as classes latentes e as respostas dos itens. Todos os itens se mostram relacionados com a variável latente, dada a heterogeneidade dos valores em cada linha das tabelas. O item de relação mais fraca parece ser o relativo a pensamento suicida, embora tenha algum destaque nos estados de SC, SP e RS. A grosso modo pode-se dizer que as estimativas dos  $\rho$ 's nas UF's obedecem padrão similar e portanto, visando uma simplificação da modelagem, esses parâmetros serão considerados homogêneos nas UF's.

Tabela 6.5: Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 para o estado do Pará, PNS-2013, Brasil

Item	escore	Classes Latentes			
		1	2	3	4
Desinteresse	0	<b>0,980</b>	<b>0,659</b>	<b>0,280</b>	0,249
	1	0,014	<b>0,301</b>	<b>0,536</b>	0,146
	2	0,001	0,036	0,172	0,083
	3	0,005	0,005	0,011	<b>0,522</b>
Desânimo	0	<b>0,962</b>	<b>0,778</b>	<b>0,295</b>	<b>0,259</b>
	1	0,034	0,215	<b>0,534</b>	0,176
	2	0,004	0,001	0,161	0,196
	3	0,001	0,007	0,010	<b>0,369</b>
Sono ( <i>excesso ou falta</i> )	0	<b>0,918</b>	<b>0,473</b>	<b>0,350</b>	<b>0,264</b>
	1	0,062	<b>0,359</b>	<b>0,358</b>	0,091
	2	0,013	0,060	0,221	0,104
	3	0,008	0,107	0,072	<b>0,541</b>
Cansaço ou falta de energia	0	<b>0,966</b>	<b>0,407</b>	<b>0,318</b>	0,078
	1	0,026	<b>0,485</b>	<b>0,458</b>	0,150
	2	0,005	0,072	0,189	0,139
	3	0,004	0,036	0,035	<b>0,634</b>
Apetite ( <i>excesso ou falta</i> )	0	<b>0,956</b>	<b>0,660</b>	<b>0,550</b>	<b>0,268</b>
	1	0,034	0,248	<b>0,258</b>	0,095
	2	0,006	0,050	0,174	0,196
	3	0,005	0,042	0,018	<b>0,440</b>
Insatisfação pessoal	0	<b>0,979</b>	<b>0,956</b>	<b>0,378</b>	<b>0,465</b>
	1	0,020	0,042	<b>0,485</b>	0,138
	2	0,001	0,001	0,103	0,136
	3	0,000	0,001	0,035	<b>0,261</b>
Falta de concentração	0	<b>0,989</b>	<b>0,758</b>	<b>0,436</b>	0,155
	1	0,005	0,240	<b>0,453</b>	0,200
	2	0,003	0,001	0,110	0,122
	3	0,004	0,001	0,002	<b>0,523</b>
Lentidão e/ou agitação	0	<b>0,989</b>	<b>0,865</b>	<b>0,491</b>	<b>0,325</b>
	1	0,007	0,121	<b>0,436</b>	0,222
	2	0,004	0,012	0,072	0,108
	3	0,000	0,001	0,002	<b>0,345</b>
Pensamento suicida	0	<b>0,998</b>	<b>0,996</b>	<b>0,834</b>	<b>0,849</b>
	1	0,002	0,003	0,149	0,021
	2	0,000	0,001	0,014	0,024
	3	0,000	0,001	0,004	0,106

Tabela 6.6: Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 para o estado da Paraíba, PNS-2013, Brasil

Item	score	Classes Latentes			
		1	2	3	4
Desinteresse	0	<b>0,988</b>	<b>0,576</b>	0,116	<b>0,273</b>
	1	0,006	<b>0,361</b>	<b>0,758</b>	0,216
	2	0,006	0,044	0,112	0,245
	3	0,001	0,019	0,014	<b>0,266</b>
Desânimo	0	<b>0,974</b>	<b>0,692</b>	0,086	0,243
	1	0,021	<b>0,276</b>	<b>0,753</b>	0,216
	2	0,004	0,026	0,137	0,216
	3	0,001	0,006	0,024	<b>0,326</b>
Sono ( <i>excesso ou falta</i> )	0	<b>0,860</b>	<b>0,501</b>	0,247	0,216
	1	0,076	<b>0,348</b>	<b>0,438</b>	0,152
	2	0,017	0,080	0,155	0,095
	3	0,048	0,070	0,160	<b>0,537</b>
Cansaço ou falta de energia	0	<b>0,918</b>	<b>0,435</b>	0,131	0,138
	1	0,062	<b>0,435</b>	<b>0,733</b>	0,123
	2	0,002	0,087	0,087	<b>0,302</b>
	3	0,019	0,044	0,050	<b>0,437</b>
Apetite ( <i>excesso ou falta</i> )	0	<b>0,948</b>	<b>0,714</b>	0,124	<b>0,403</b>
	1	0,041	<b>0,257</b>	<b>0,684</b>	0,123
	2	0,006	0,013	0,154	0,139
	3	0,006	0,016	0,038	<b>0,335</b>
Insatisfação pessoal	0	<b>0,999</b>	<b>0,840</b>	0,247	<b>0,478</b>
	1	0,000	0,151	<b>0,679</b>	0,150
	2	0,000	0,005	0,069	0,129
	3	0,001	0,005	0,005	0,242
Falta de concentração	0	<b>0,990</b>	<b>0,723</b>	0,213	<b>0,393</b>
	1	0,003	<b>0,241</b>	<b>0,598</b>	0,149
	2	0,000	0,029	0,167	0,136
	3	0,007	0,007	0,022	<b>0,322</b>
Lentidão e/ou agitação	0	<b>0,988</b>	<b>0,816</b>	<b>0,299</b>	<b>0,595</b>
	1	0,006	0,170	<b>0,636</b>	0,108
	2	0,004	0,008	0,053	0,109
	3	0,002	0,006	0,012	0,189
Pensamento suicida	0	<b>0,999</b>	<b>0,991</b>	<b>0,895</b>	<b>0,686</b>
	1	0,000	0,008	0,101	0,157
	2	0,000	0,001	0,002	0,061
	3	0,000	0,001	0,002	0,097

Tabela 6.7: Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 para o estado do Espírito Santo, PNS-2013, Brasil

Item	score	Classes Latentes			
		1	2	3	4
Desinteresse	0	<b>0,971</b>	<b>0,591</b>	<b>0,340</b>	0,243
	1	0,023	<b>0,307</b>	<b>0,304</b>	0,221
	2	0,004	0,070	<b>0,351</b>	0,109
	3	0,002	0,032	0,004	<b>0,428</b>
Desânimo	0	<b>0,971</b>	<b>0,551</b>	0,129	0,169
	1	0,027	<b>0,399</b>	<b>0,259</b>	0,107
	2	0,002	0,015	<b>0,594</b>	0,168
	3	0,001	0,036	0,018	<b>0,556</b>
Sono ( <i>excesso ou falta</i> )	0	<b>0,934</b>	<b>0,348</b>	0,177	0,220
	1	0,035	<b>0,318</b>	0,216	0,234
	2	0,018	0,028	<b>0,580</b>	0,009
	3	0,013	<b>0,307</b>	0,028	<b>0,537</b>
Cansaço ou falta de energia	0	<b>0,925</b>	<b>0,327</b>	<b>0,277</b>	0,230
	1	0,064	<b>0,401</b>	<b>0,419</b>	0,118
	2	0,005	0,096	<b>0,273</b>	0,202
	3	0,007	0,177	0,032	<b>0,450</b>
Apetite ( <i>excesso ou falta</i> )	0	<b>0,959</b>	<b>0,670</b>	<b>0,499</b>	<b>0,389</b>
	1	0,033	0,152	<b>0,284</b>	0,139
	2	0,006	0,062	0,212	0,210
	3	0,003	0,115	0,005	0,262
Insatisfação pessoal	0	<b>0,994</b>	<b>0,817</b>	<b>0,440</b>	<b>0,382</b>
	1	0,004	0,181	<b>0,252</b>	0,098
	2	0,001	0,001	<b>0,282</b>	0,170
	3	0,001	0,001	0,027	<b>0,350</b>
Falta de concentração	0	<b>0,976</b>	<b>0,707</b>	0,200	<b>0,466</b>
	1	0,020	<b>0,272</b>	<b>0,342</b>	0,123
	2	0,004	0,016	<b>0,406</b>	0,216
	3	0,001	0,006	0,053	0,196
Lentidão e/ou agitação	0	<b>0,987</b>	<b>0,850</b>	<b>0,380</b>	<b>0,660</b>
	1	0,009	0,138	0,114	0,080
	2	0,002	0,008	<b>0,448</b>	0,040
	3	0,002	0,005	0,057	0,221
Pensamento suicida	0	<b>0,999</b>	<b>0,985</b>	<b>0,674</b>	<b>0,861</b>
	1	0,001	0,014	0,205	0,098
	2	0,000	0,001	0,118	0,005
	3	0,000	0,001	0,003	0,036



Tabela 6.8: Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 para o estado de Santa Catarina, PNS-2013, Brasil

Item	escore	Classes Latentes			
		1	2	3	4
Desinteresse	0	<b>0,980</b>	<b>0,497</b>	0,160	0,075
	1	0,018	<b>0,435</b>	0,185	<b>0,258</b>
	2	0,002	0,053	0,246	0,116
	3	0,001	0,014	<b>0,409</b>	<b>0,551</b>
Desânimo	0	<b>0,931</b>	<b>0,426</b>	0,005	<b>0,443</b>
	1	0,061	<b>0,494</b>	0,098	0,198
	2	0,004	0,062	<b>0,312</b>	0,091
	3	0,004	0,018	<b>0,585</b>	<b>0,269</b>
Sono ( <i>excesso ou falta</i> )	0	<b>0,876</b>	<b>0,459</b>	<b>0,309</b>	0,233
	1	0,077	<b>0,342</b>	0,069	0,052
	2	0,019	0,089	0,132	0,027
	3	0,029	0,111	<b>0,490</b>	<b>0,689</b>
Cansaço ou falta de energia	0	<b>0,902</b>	<b>0,338</b>	0,105	0,091
	1	0,078	<b>0,493</b>	0,135	0,020
	2	0,007	0,132	0,169	0,155
	3	0,013	0,037	<b>0,592</b>	<b>0,734</b>
Apetite ( <i>excesso ou falta</i> )	0	<b>0,961</b>	<b>0,743</b>	<b>0,492</b>	<b>0,573</b>
	1	0,010	0,196	0,133	0,078
	2	0,005	0,031	0,233	0,031
	3	0,024	0,030	0,142	<b>0,318</b>
Insatisfação pessoal	0	<b>0,985</b>	<b>0,676</b>	0,174	<b>0,823</b>
	1	0,005	<b>0,283</b>	<b>0,251</b>	0,056
	2	0,003	0,040	0,189	0,003
	3	0,008	0,001	<b>0,387</b>	0,118
Falta de concentração	0	<b>0,968</b>	<b>0,599</b>	0,199	<b>0,525</b>
	1	0,013	0,333	<b>0,414</b>	0,081
	2	0,003	0,047	0,133	0,008
	3	0,015	0,021	<b>0,255</b>	<b>0,386</b>
Lentidão e/ou agitação	0	<b>0,975</b>	<b>0,777</b>	<b>0,402</b>	<b>0,511</b>
	1	0,022	0,183	0,211	0,128
	2	0,001	0,038	0,195	0,044
	3	0,002	0,001	0,192	<b>0,318</b>
Pensamento suicida	0	<b>0,998</b>	<b>0,961</b>	<b>0,518</b>	<b>0,938</b>
	1	0,001	0,033	<b>0,283</b>	0,022
	2	0,000	0,006	0,139	0,003
	3	0,000	0,001	0,060	0,036

Tabela 6.9: Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 para o estado do Mato Grosso, PNS-2013, Brasil

Item	escore	Classes Latentes			
		1	2	3	4
Desinteresse	0	<b>0,991</b>	<b>0,722</b>	<b>0,395</b>	0,141
	1	0,005	0,228	<b>0,556</b>	0,240
	2	0,003	0,023	0,048	0,199
	3	0,000	0,028	0,001	<b>0,420</b>
Desânimo	0	<b>0,966</b>	<b>0,792</b>	0,236	0,074
	1	0,033	0,124	<b>0,704</b>	<b>0,259</b>
	2	0,001	0,045	0,048	0,246
	3	0,002	0,038	0,012	<b>0,422</b>
Sono ( <i>excesso ou falta</i> )	0	<b>0,930</b>	<b>0,434</b>	<b>0,410</b>	0,160
	1	0,046	0,241	<b>0,503</b>	0,162
	2	0,018	0,073	0,066	<b>0,272</b>
	3	0,006	<b>0,252</b>	0,021	<b>0,406</b>
Cansaço ou falta de energia	0	<b>0,965</b>	<b>0,396</b>	<b>0,379</b>	0,114
	1	0,031	<b>0,328</b>	<b>0,588</b>	0,159
	2	0,001	0,118	0,029	0,240
	3	0,003	0,158	0,004	<b>0,488</b>
Apetite ( <i>excesso ou falta</i> )	0	<b>0,965</b>	<b>0,736</b>	<b>0,581</b>	0,218
	1	0,031	0,146	<b>0,387</b>	0,133
	2	0,003	0,041	0,026	<b>0,251</b>
	3	0,001	0,077	0,007	<b>0,398</b>
Insatisfação pessoal	0	<b>0,985</b>	<b>0,955</b>	<b>0,523</b>	<b>0,281</b>
	1	0,012	0,043	<b>0,463</b>	<b>0,256</b>
	2	0,003	0,001	0,007	0,238
	3	0,000	0,001	0,007	0,226
Falta de concentração	0	<b>0,990</b>	<b>0,784</b>	<b>0,409</b>	<b>0,280</b>
	1	0,010	0,154	<b>0,543</b>	0,148
	2	0,000	0,042	0,044	0,233
	3	0,000	0,020	0,004	<b>0,339</b>
Lentidão e/ou agitação	0	<b>0,981</b>	<b>0,721</b>	<b>0,616</b>	<b>0,269</b>
	1	0,011	0,122	<b>0,316</b>	0,243
	2	0,008	0,079	0,053	0,155
	3	0,001	0,078	0,016	<b>0,333</b>
Pensamento suicida	0	<b>0,995</b>	<b>0,992</b>	<b>0,901</b>	<b>0,768</b>
	1	0,005	0,006	0,062	0,117
	2	0,000	0,001	0,029	0,054
	3	0,000	0,001	0,008	0,060

Tabela 6.10: Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 para o estado de São Paulo, PNS-2013, Brasil

Item	score	Classes Latentes			
		1	2	3	4
Desinteresse	0	<b>0,983</b>	<b>0,553</b>	0,165	0,209
	1	0,011	<b>0,403</b>	<b>0,345</b>	0,089
	2	0,002	0,016	<b>0,412</b>	0,063
	3	0,003	0,029	0,078	<b>0,640</b>
Desânimo	0	<b>0,981</b>	<b>0,535</b>	0,212	0,184
	1	0,014	<b>0,434</b>	<b>0,287</b>	0,098
	2	0,003	0,022	<b>0,431</b>	0,092
	3	0,001	0,010	0,071	<b>0,626</b>
Sono ( <i>excesso ou falta</i> )	0	<b>0,890</b>	<b>0,493</b>	<b>0,331</b>	<b>0,253</b>
	1	0,056	<b>0,298</b>	0,177	0,100
	3	0,023	0,079	<b>0,280</b>	0,097
	4	0,032	0,129	0,212	<b>0,551</b>
Cansaço ou falta de energia	0	<b>0,926</b>	<b>0,384</b>	<b>0,282</b>	0,187
	1	0,053	<b>0,485</b>	0,192	0,058
	3	0,008	0,051	<b>0,427</b>	0,067
	4	0,014	0,080	0,099	<b>0,688</b>
Apetite ( <i>excesso ou falta</i> )	0	<b>0,970</b>	<b>0,706</b>	<b>0,449</b>	<b>0,411</b>
	1	0,019	0,212	0,171	0,153
	2	0,005	0,042	<b>0,299</b>	0,065
	3	0,006	0,040	0,081	<b>0,371</b>
Insatisfação pessoal	0	<b>0,994</b>	<b>0,756</b>	<b>0,451</b>	<b>0,363</b>
	1	0,005	0,235	<b>0,289</b>	0,135
	2	0,001	0,008	0,216	0,109
	3	0,000	0,001	0,045	<b>0,394</b>
Falta de concentração	0	<b>0,988</b>	<b>0,686</b>	<b>0,475</b>	<b>0,404</b>
	1	0,007	<b>0,276</b>	<b>0,274</b>	0,050
	2	0,001	0,013	0,247	0,096
	3	0,004	0,025	0,004	<b>0,450</b>
Lentidão e/ou agitação	0	<b>0,992</b>	<b>0,791</b>	<b>0,527</b>	<b>0,395</b>
	1	0,005	0,175	0,204	0,071
	2	0,001	0,016	0,243	0,092
	3	0,001	0,019	0,027	<b>0,443</b>
Pensamento suicida	0	<b>0,999</b>	<b>0,958</b>	<b>0,761</b>	<b>0,743</b>
	1	0,001	0,040	0,208	0,106
	2	0,000	0,001	0,029	0,052
	3	0,000	0,000	0,002	0,099

Tabela 6.11: Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 para o estado do Rio Grande do Sul, PNS-2013, Brasil

Item	score	Classes Latentes			
		1	2	3	4
Desinteresse	0	<b>0,996</b>	<b>0,598</b>	0,190	0,090
	1	0,003	<b>0,374</b>	<b>0,548</b>	0,007
	2	0,000	0,008	0,200	<b>0,250</b>
	3	0,000	0,019	0,062	<b>0,652</b>
Desânimo	0	<b>0,981</b>	<b>0,648</b>	0,182	0,129
	1	0,017	<b>0,335</b>	<b>0,496</b>	0,069
	2	0,001	0,012	0,214	0,169
	3	0,001	0,004	0,109	<b>0,633</b>
Sono ( <i>excesso ou falta</i> )	0	<b>0,887</b>	<b>0,549</b>	<b>0,306</b>	0,099
	1	0,082	<b>0,309</b>	<b>0,257</b>	0,079
	2	0,009	0,083	0,187	0,061
	3	0,022	0,059	<b>0,250</b>	<b>0,760</b>
Cansaço ou falta de energia	0	<b>0,927</b>	<b>0,426</b>	0,172	0,097
	1	0,064	<b>0,474</b>	<b>0,413</b>	0,121
	2	0,001	0,082	0,195	0,056
	3	0,008	0,019	0,220	<b>0,726</b>
Apetite ( <i>excesso ou falta</i> )	0	<b>0,969</b>	<b>0,732</b>	<b>0,433</b>	<b>0,285</b>
	1	0,022	0,185	<b>0,334</b>	0,069
	2	0,006	0,065	0,147	0,178
	3	0,003	0,018	0,086	<b>0,468</b>
Insatisfação pessoal	0	<b>0,996</b>	<b>0,870</b>	<b>0,430</b>	<b>0,299</b>
	1	0,003	0,128	<b>0,413</b>	0,109
	2	0,001	0,002	0,105	0,184
	3	0,001	0,001	0,052	<b>0,408</b>
Falta de concentração	0	<b>0,986</b>	<b>0,754</b>	<b>0,333</b>	0,225
	1	0,010	0,237	<b>0,459</b>	0,018
	2	0,002	0,002	0,178	0,177
	3	0,003	0,008	0,031	<b>0,580</b>
Lentidão e/ou agitação	0	<b>0,993</b>	<b>0,803</b>	<b>0,393</b>	0,240
	1	0,004	0,190	<b>0,404</b>	0,105
	2	0,002	0,006	0,150	0,180
	3	0,002	0,001	0,054	<b>0,476</b>
Pensamento suicida	0	<b>0,998</b>	<b>0,994</b>	<b>0,851</b>	<b>0,501</b>
	1	0,002	0,005	0,132	0,112
	3	0,000	0,000	0,009	0,203
	4	0,000	0,001	0,008	0,184

Para a interpretação dos efeitos das covariáveis sobre as prevalências das classes latentes pode-se utilizar as estimativas das razões de chances, semelhante à análise de regressão logística multinomial. Estimativas das razões de chances e intervalos de confiança a 95% para o efeito de cada covariável incluída no modelo são apresentadas nas Tabelas 6.12 a 6.18. A classe 1 é a referência em todas as tabelas.

No geral, conclui-se que a covariável sexo está fortemente associada à variável latente sinalizando que mulheres tem maior chance de serem classificadas nas classes mais “graves”, sobretudo na classe 4.

Das demais covariáveis, o nível educacional e a idade se mostraram com alguma associação à variável latente, indicando que quanto mais baixo o nível educacional, e mais avançada a idade maiores as chances de classificação nas classes 3 ou 4. Esses efeitos não foram significativos em todas as UF’s, porém, neste momento não se tem interesse em teste formais, mesmo porque, uma vez tratando-se cada UF separadamente perde-se a estrutura dos pesos da amostragem complexa. Por essa razão e por saber que as amostras nas UF’s são de tamanhos distintos, o fato de não se encontrar significância dessas covariáveis em alguns estados e encontrar para outros era esperado.

As demais covariáveis “Tem companheiro(a)” e “Vive sozinho(a)” não se mostraram importantes nas análises separadas por UF’s.

Destas explorações, tirou-se que assumir os efeitos das covariáveis como sendo homogêneos nas UF’s vai de encontro com o princípio da parcimônia e esse princípio será utilizado no modelo global apresentado na seção 6.2.3.

Tabela 6.12: Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis no estado do Pará

Covariáveis	Classes Latentes		
	2/1	3/1	4/1
<b>Intercepto</b>	0,077 (0,015; 0,39)	0,185 (0,028; 1,211)	0,019 (0,003; 0,113)
<b>Sexo*</b>			
Feminino	1,157 (0,727; 1,84)	1,852 (0,959; 3,576)	5,879 (2,209; 15,643)
<b>Educação</b>			
Sem Educ. Formal	-	-	-
Ensino Fundamental	1,692 (0,431; 6,646)	0,787 (0,237; 2,61)	0,527 (0,168; 1,654)
Ensino Médio	1,417 (0,367; 5,467)	0,555 (0,177; 1,737)	0,294 (0,079; 1,093)
Ensino Superior	1,332 (0,303; 5,86)	0,471 (0,098; 2,256)	0,261 (0,036; 1,902)
<b>Idade (anos)</b>			
[18; 30)	-	-	-
[30; 40)	1,459 (0,842; 2,531)	0,977 (0,488; 1,958)	0,774 (0,313; 1,918)
[40; 50)	1,187 (0,641; 2,199)	1,265 (0,52; 3,079)	0,997 (0,363; 2,741)
[50; 60)	1,441 (0,656; 3,166)	1,010 (0,368; 2,776)	0,926 (0,288; 2,978)
[60; 70)	0,910 (0,318; 2,6)	0,972 (0,296; 3,194)	0,163 (0,037; 0,714)
≥ 70	2,274 (0,712; 7,262)	1,579 (0,426; 5,849)	0,800 (0,206; 3,099)
<b>Tem Companheiro</b>			
Não	1,159 (0,742; 1,809)	1,424 (0,763; 2,659)	1,410 (0,686; 2,896)
<b>Vive sozinho</b>			
Não	2,513 (0,958; 6,595)	0,721 (0,261; 1,992)	2,177 (0,617; 7,679)

★: Efeito significativo, valor-p  $\leq 0,05$ .

Tabela 6.13: Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis no estado da Paraíba

Covariáveis	Classes Latentes		
	2/1	3/1	4/1
<b>Intercepto</b>	0,054 (0,014; 0,207)	0,006 (0,001; 0,056)	0,090 (0,024; 0,344)
<b>Sexo*</b>			
Feminino	2,039 (1,369; 3,037)	2,836 (1,451; 5,546)	2,337 (1,298; 4,207)
<b>Educação</b>			
Sem Educ. Formal	-	-	-
Ensino Fund.	2,136 (0,814; 5,609)	1,172 (0,406; 3,379)	0,906 (0,43; 1,906)
Ensino Médio	2,215 (0,817; 6,008)	1,496 (0,428; 5,232)	0,974 (0,41; 2,31)
Ensino Superior	2,440 (0,817; 6,615)	1,033 (0,247; 4,318)	0,942 (0,319; 2,777)
<b>Idade (anos)*</b>			
[18; 30)	-	-	-
[30; 40)	1,183 (0,685; 2,041)	5,182 (1,372; 19,575)	1,062 (0,408; 2,762)
[40; 50)	1,152 (0,640; 2,075)	5,026 (1,397; 18,084)	1,344 (0,563; 3,207)
[50; 60)	1,522 (0,841; 2,753)	9,461 (2,438; 36,720)	1,680 (0,670; 4,210)
[60; 70)	1,045 (0,46; 2,376)	6,983 (1,642; 29,698)	0,705 (0,239; 2,084)
≥ 70	1,849 (0,734; 4,655)	8,341 (1,69; 41,17)	3,214 (1,326; 7,789)
<b>Tem Companheiro</b>			
Não	1,254 (0,780; 2,017)	1,631 (0,762; 3,491)	0,877 (0,485; 1,583)
<b>Vive sozinho</b>			
Não	2,138 (1,070; 4,273)	1,413 (0,495; 4,037)	0,913 (0,422; 1,973)

★: Efeito significativo, valor-p  $\leq 0,05$ .

Tabela 6.14: Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis no estado do Espírito Santo

Covariáveis	Classes Latentes		
	2/1	3/1	4/1
<b>Intercepto</b>	0,655 (0,173; 2,475)	0,577 (0,049; 6,742)	0,114 (0,015; 0,881)
<b>Sexo*</b>			
Feminino	1,965 (1,212; 3,186)	1,051 (0,451; 2,451)	3,081 (1,285; 7,390)
<b>Educação*</b>			
Sem Educ. Formal	-	-	-
Ensino Fund.	0,480 (0,213; 1,080)	0,152 (0,033; 0,709)	0,321 (0,099; 1,037)
Ensino Médio	0,475 (0,192; 1,174)	0,190 (0,031; 1,147)	0,129 (0,036; 0,468)
Ensino Superior	0,545 (0,207; 1,438)	0,079 (0,010; 0,636)	0,191 (0,043; 0,852)
<b>Idade (anos)*</b>			
[18; 30)	-	-	-
[30; 40)	0,500 (0,254; 0,982)	1,318 (0,420; 4,138)	1,332 (0,494; 3,589)
[40; 50)	0,860 (0,439; 1,684)	0,797 (0,212; 2,998)	0,938 (0,313; 2,812)
[50; 60)	1,203 (0,649; 2,230)	0,715 (0,116; 4,423)	0,895 (0,290; 2,763)
[60; 70)	1,0578 (0,482; 2,322)	0,588 (0,036; 9,558)	1,036 (0,281; 3,818)
≥ 70	0,277 (0,104; 0,740)	0,357 (0,067; 1,902)	0,162 (0,033; 0,795)
<b>Tem Companheiro*</b>			
Não	0,605 (0,333; 1,096)	3,248 (1,218; 8,662)	1,721 (0,755; 3,921)
<b>Vive sozinho</b>			
Não	0,830 (0,347; 1,983)	0,292 (0,082; 1,040)	1,439 (0,561; 3,687)

\*: Efeito significativo, valor-p  $\leq 0,05$ .



Tabela 6.15: Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis no estado de Santa Catarina

Covariáveis	Classes Latentes		
	2/1	3/1	4/1
<b>Intercepto</b>	0,366 (0,098; 1,372)	0,012 (0,000; 6,231)	0,032 (0,000; 3,361)
<b>Sexo*</b>			
Feminino	1,710 (1,148; 2,549)	30,636 (0,924; 1016,2)	1,883 (0,474; 7,491)
<b>Educação</b>			
Sem Educ. Formal	-	-	-
Ensino Fund.	1,233 (0,494; 3,08)	1,077 (0,027; 43,33)	0,554 (0,107; 2,857)
Ensino Médio	1,313 (0,497; 3,466)	0,517 (0,014; 18,716)	0,387 (0,054; 2,786)
Ensino Superior	1,019 (0,367; 2,823)	0,418 (0,018; 9,973)	0,857 (0,161; 4,555)
<b>Idade (anos)</b>			
[18; 30)	-	-	-
[30; 40)	1,488 (0,838; 2,639)	3,923 (0,49; 31,422)	1,234 (0,054; 28,089)
[40; 50)	0,813 (0,435; 1,518)	2,570 (0,504; 13,108)	2,705 (0,224; 32,682)
[50; 60)	1,885 (0,989; 3,594)	2,038 (0,361; 11,513)	4,927 (0,369; 65,771)
[60; 70)	1,908 (0,925; 3,934)	2,482 (0,462; 13,347)	5,025 (0,312; 80,884)
≥ 70	0,717 (0,177; 2,901)	0,010 (0; 1000)	11,544 (0,465; 286,4)
<b>Tem Companheiro</b>			
Não	0,818 (0,494; 1,353)	0,679 (0,213; 2,169)	0,875 (0,17; 4,502)
<b>Vive sozinho</b>			
Não	0,551 (0,265; 1,147)	0,347 (0,092; 1,302)	1,380 (0,137; 13,877)

★: Efeito significativo, valor-p  $\leq 0,05$ .

Tabela 6.16: Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis no estado do Mato Grosso

Covariáveis	Classes Latentes		
	2/1	3/1	4/1
<b>Intercepto</b>	0,079 (0,006; 0,982)	0,542 (0,140; 2,102)	0,031 (0,005; 0,186)
<b>Sexo*</b>			
Feminino	2,223 (1,306; 3,783)	2,007 (1,128; 3,573)	4,089 (1,981; 8,443)
<b>Educação*</b>			
Sem Educ. Formal	-	-	-
Ensino Fund.	1,415 (0,449; 4,461)	0,651 (0,298; 1,425)	2,024 (0,655; 6,255)
Ensino Médio	1,499 (0,369; 6,0871)	0,616 (0,248; 1,534)	1,337 (0,390; 4,586)
Ensino Superior	3,503 (1,013; 12,116)	0,363 (0,031; 4,288)	2,530 (0,694; 9,222)
<b>Idade (anos)</b>			
[18; 30)	-	-	-
[30; 40)	1,450 (0,620; 3,395)	0,694 (0,297; 1,619)	0,904 (0,373; 2,190)
[40; 50)	0,992 (0,415; 2,375)	0,922 (0,421; 2,022)	1,841 (0,770; 4,397)
[50; 60)	2,475 (1,090; 5,623)	0,663 (0,192; 2,289)	0,894 (0,308; 2,598)
[60; 70)	2,191 (0,781; 6,144)	1,006 (0,335; 3,020)	1,334 (0,385; 4,623)
≥ 70	2,550 (0,663; 9,814)	1,337 (0,474; 3,768)	1,867 (0,467; 7,467)
<b>Tem Companheiro</b>			
Não	1,032 (0,560; 1,901)	0,807 (0,363; 1,792)	0,671 (0,311; 1,450)
<b>Vive sozinho</b>			
Não	1,175 (0,479; 2,883)	0,460 (0,226; 0,939)	0,818 (0,297; 2,249)

★: Efeito significativo, valor-p  $\leq 0,05$ .

Tabela 6.17: Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis no estado de São Paulo

Covariáveis	Classes Latentes		
	2/1	3/1	4/1
<b>Intercepto</b>	0,255 (0,119; 0,544)	0,150 (0,045; 0,502)	0,071 (0,024; 0,216)
<b>Sexo*</b>			
Feminino	2,111 (1,646; 2,708)	3,462 (2,122; 5,647)	1,903 (1,202; 3,0111)
<b>Educação*</b>			
Sem Educ. Formal	-	-	-
Ensino Fund.	0,824 (0,475; 1,429)	0,349 (0,150; 0,815)	0,715 (0,362; 1,414)
Ensino Médio	0,637 (0,360; 1,128)	0,208 (0,089; 0,483)	0,408 (0,196; 0,850)
Ensino Superior	0,686 (0,388; 1,213)	0,199 (0,080; 0,494)	0,213 (0,085; 0,538)
<b>Idade (anos)*</b>			
[18; 30)	-	-	-
[30; 40)	0,892 (0,623; 1,278)	0,870 (0,437; 1,731)	1,260 (0,596; 2,664)
[40; 50)	1,151 (0,801; 1,654)	1,037 (0,494; 2,177)	1,214 (0,569; 2,589)
[50; 60)	0,850 (0,565; 1,277)	0,878 (0,420; 1,835)	1,826 (0,888; 3,754)
[60; 70)	0,968 (0,632; 1,482)	0,682 (0,315; 1,478)	1,800 (0,837; 3,872)
≥ 70	0,938 (0,583; 1,508)	0,161 (0,061; 0,427)	1,452 (0,727; 2,898)
<b>Tem Companheiro</b>			
Não	1,200 (0,922; 1,562)	1,512 (0,928; 2,465)	1,055 (0,647; 1,721)
<b>Vive sozinho</b>			
Não	1,134 (0,795; 1,618)	0,945 (0,514; 1,735)	1,063 (0,607; 1,862))

★: Efeito significativo, valor-p  $\leq 0,05$ .

Tabela 6.18: Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis no estado do Rio Grande do Sul

Covariáveis	Classes Latentes		
	2/1	3/1	4/1
<b>Intercepto</b>	0,828 (0,278; 2,469)	0,558 (0,171; 1,817)	0,046 (0,008; 0,272)
<b>Sexo*</b>			
Feminino	1,764 (1,274; 2,443)	3,021 (1,99; 4,586)	5,972 (2,901; 12,293)
<b>Educação*</b>			
Sem Educ. Formal	-	-	-
Ensino Fund.	0,664 (0,272; 1,623)	0,445 (0,205; 0,962)	0,479 (0,16; 1,434)
Ensino Médio	0,951 (0,379; 2,384)	0,311 (0,128; 0,757)	0,137 (0,026; 0,712)
Ensino Superior	0,816 (0,326; 2,043)	0,246 (0,096; 0,629)	0,145 (0,037; 0,571)
<b>Idade (anos)</b>			
[18; 30)	-	-	-
[30; 40)	0,702 (0,419; 1,175)	0,933 (0,462; 1,884)	1,277 (0,382; 4,273)
[40; 50)	0,800 (0,481; 1,33)	1,305 (0,683; 2,495)	1,128 (0,39; 3,262)
[50; 60)	0,729 (0,42; 1,266)	1,426 (0,758; 2,682)	1,902 (0,686; 5,273)
[60; 70)	0,645 (0,352; 1,18)	1,026 (0,49; 2,147)	1,235 (0,323; 4,717)
≥ 70	0,886 (0,447; 1,758)	0,587 (0,259; 1,327)	0,848 (0,23; 3,125)
<b>Tem Companheiro</b>			
Não	1,000 (0,672; 1,488)	0,968 (0,613; 1,527)	0,805 (0,399; 1,626)
<b>Vive sozinho</b>			
Não	0,750 (0,453; 1,243)	0,588 (0,333; 1,036)	1,376 (0,488; 3,881)

★: Efeito significativo, valor-p  $\leq 0,05$ .

### 6.2.3 Ajuste de um modelo global

Embora os resultados dos ajustes separados tenham indicado algumas heterogeneidades entre as UF's, os relacionamentos entre as classes e os itens e entre as classes e as covariáveis seguiram, razoavelmente um padrão comum, ou seja, tem-se o mesmo tipo de relacionamento, embora particularidades em relação à significância ou não dos efeitos tenham ocorrido, principalmente devido aos diferentes tamanhos amostrais das UF's, inclusive com a impossibilidade do ajuste do modelo para o estado do Acre.

Também ressalta-se que a significância nessas análises separadas não devem ser levadas muito em conta devido à quebra da estrutura complexa. Assim, postulou-se um modelo de classes latentes, com agrupamento dado pelas UF's, porém incorporando as restrições de que os parâmetros  $\rho$ 's e  $\alpha$ 's não variam com as UF's.

Nesse modelo, a covariável idade foi reclassificada com o objetivo de simplificar a interpretação, já que nas análises preliminares os resultados indicam similaridade entre as classes [40; 50) e [50; 60). Também observou-se frequências muito baixas na classe  $\geq 70$  e esta foi então agrupada na classe  $\geq 60$ .

A covariável nível educacional também teve suas classes levemente redefinidas visto ser esta última classificação mais usual nas pesquisas sociodemográficas e da saúde.

### 6.2.4 Estimativas dos $\gamma$ 's (prevalências)

Primeiramente serão apresentados os resultados que descrevem as prevalências estimadas para cada UF, com o intuito de explorar a heterogeneidade regional entre as unidades da federação. A Tabela 6.19 apresenta o tamanho amostral em valor absoluto e percentual de cada UF seguidos pelas estimativas

das prevalências estimadas das classes latentes e seus respectivos erros-padrão. É possível interpretar a classe 1 como representante do grupo de indivíduos sem risco de experienciar qualquer sintoma do PHQ-9 enquanto que a classe 4 representa o grupo de indivíduos sob constante risco de experienciar até oito dos sintomas do PHQ-9, sugerindo a ordem 1, 2, 3 e 4 das classes em termos de gravidade numa escala de depressão. Essa interpretação tem como base a exploração das estimativas dos  $\rho$ 's apresentadas na Seção 6.2.5 que serão discutidas posteriormente. As estimativas das prevalências nesse modelo global, em geral, seguem, em termos relativos, aquelas das análises separadas (Tabela 6.4) mas com algumas inversões que, talvez, possam ser justificadas pelos altos erros-padrão das análises separadas.

As prevalências estimadas na classe 1 variam de 53,4% (MA) a 71,0% (RJ) com certo destaque de valores mais baixos para as regiões Sul e Nordeste e valores mais altos para a região Sudeste. Em relação às classes de gravidade mais altas já aparece maior heterogeneidade das prevalências dentro de cada região. A UF com maior valor na classe 3 foi SC (8,9%) seguida por PR (7,9%), DF (7,8%), PB (7,7%), BA (7,4%) e AP (7,3%).

Demais UF's apresentaram valores inferiores a 7% nessa classe. Em relação à classe 4, a UF com maior prevalência foi também SC (8,8%) seguida de RN (7,7%) e RR (7,1%). As UF' com menores valores para a classe 4 foram PA (3,5%) e ES (4,7%). Tais resultados mostram considerável heterogeneidade de prevalências nas classes entre as UF's, mesmo entre aquelas pertencentes à mesma região do país.

Tabela 6.19: Estimativas das prevalências e respectivos erros-padrão (E.P.) segundo unidades da federação no modelo global, PNS-2013, Brasil

UF	$n$	%	Classes latentes							
			1	E.P.	2	E.P.	3	E.P.	4	E.P.
RR	1.694	2,81	0,633	0,037	0,253	0,021	0,063	0,017	0,071	0,008
AC	1.814	3,01	0,614	0,037	0,268	0,022	0,060	0,016	0,058	0,007
TO	2.586	4,30	0,643	0,035	0,245	0,021	0,061	0,016	0,056	0,009
RO	1.591	2,64	0,575	0,039	0,300	0,022	0,054	0,019	0,051	0,007
AM	2.004	3,33	0,646	0,038	0,274	0,023	0,049	0,017	0,050	0,006
AP	1.332	2,21	0,565	0,040	0,312	0,022	0,073	0,021	0,050	0,007
PA	1.515	2,52	0,568	0,039	0,311	0,023	0,066	0,020	0,031	0,005
RN	1.774	2,95	0,597	0,042	0,284	0,024	0,057	0,021	0,077	0,008
PE	1.804	3,00	0,601	0,043	0,301	0,025	0,056	0,020	0,070	0,006
AL	2.560	4,25	0,592	0,040	0,285	0,017	0,066	0,027	0,069	0,006
MA	1.691	2,81	0,534	0,038	0,320	0,021	0,069	0,021	0,062	0,009
CE	1.943	3,23	0,599	0,037	0,279	0,014	0,072	0,030	0,057	0,008
BA	2.591	4,30	0,556	0,040	0,301	0,020	0,074	0,024	0,051	0,006
PB	1.748	2,90	0,563	0,036	0,291	0,017	0,077	0,024	0,050	0,006
PI	1.553	2,58	0,595	0,039	0,292	0,022	0,070	0,018	0,042	0,005
SE	2.641	4,39	0,637	0,038	0,265	0,027	0,047	0,014	0,042	0,005
MG	3.779	6,28	0,635	0,037	0,249	0,029	0,061	0,009	0,055	0,009
SP	1.724	2,86	0,663	0,038	0,210	0,028	0,066	0,012	0,061	0,006
RJ	3.486	5,79	0,710	0,030	0,184	0,021	0,052	0,009	0,058	0,005
ES	5.305	8,81	0,661	0,033	0,232	0,022	0,049	0,012	0,054	0,005
SC	3.012	5,00	0,556	0,035	0,310	0,018	0,089	0,020	0,088	0,011
RS	1.623	2,70	0,590	0,038	0,270	0,025	0,052	0,015	0,058	0,006
PR	2.913	4,84	0,567	0,036	0,296	0,016	0,079	0,024	0,045	0,005
MT	1.809	3,00	0,604	0,038	0,280	0,024	0,064	0,015	0,055	0,007
GO	1.476	2,45	0,628	0,037	0,261	0,023	0,056	0,017	0,054	0,007
DF	2.423	4,02	0,580	0,038	0,287	0,027	0,078	0,014	0,053	0,007
MS	1.811	3,01	0,577	0,039	0,300	0,021	0,070	0,021	0,052	0,007

### 6.2.5 Estimativas dos $\rho$ 's (probabilidades condicionais de respostas)

As estimativas das probabilidades condicionais de resposta ( $\rho$ 's), dado cada classe latente, para cada um dos itens são apresentadas na Tabela 6.20 e na Figura 4.1. Antes de proceder às interpretações das estimativas vale a pena apontar padrões que as guiam:

1. Como dado uma classe latente, para cada item há 4 possíveis alternativas, então, caso os indivíduos respondam ao acaso, a probabilidade esperada em cada alternativa é igual a 0,25. Assim, para facilitar, as estimativas maiores que 0,25 foram destacadas na tabela.
2. Relação fraca entre o item e a variável latente é indicada quando as probabilidades de uma resposta se mantêm constante em relação a todas as classes. Estudo cuidadoso desses padrões formam a base para a rotulação das classes latentes.

Primeiramente observa-se que todos os itens são fortemente relacionados com a variável latente categórica visto a heterogeneidade das estimativas em cada linha da tabela. O item de relação mais fraca parece ser o relativo a pensamento suicida, embora tenha algum destaque com probabilidades de resposta nas categorias “mais da metade dos dias” ou “quase todos os dias” dado as classes 3 ou 4. Avaliando o comportamento das estimativas em termos do “distanciamento” do valor 0,25 nota-se que as classes não são muito bem separadas mas é possível estabelecer os seguintes perfis de indivíduos:

Classe 1: Grupo sem risco ou risco muito baixo de experienciar qualquer sintoma;

Classe 2: Grupo sob risco esporádico de experienciar sintomas como desinteresse, desânimo, cansaço e problemas de sono;

Classe 3: Grupo sob risco habitual de experienciar sintomas como desinteresse, desânimo, cansaço, e problemas de sono, de apetite e de concentração, lentidão/agitação, insatisfação pessoal;



Classe 4: Grupo sob risco constante de experienciar sintomas como desinteresse, desânimo, cansaço, problemas de sono, de apetite, concentração e lentidão/agitação.

A interpretação das estimativas dos  $\rho$ 's fica mais imediata através da Figura 6.1 que mostra, numa escala de tons de cinza, como as probabilidades aumentam com os escores dos itens conforme a ordem da classe latente também aumenta. Dada a classe 1, tons mais fortes aparecem, em bem pequenas probabilidades, apenas nos quesitos “sono” e “cansaço”, indicando que pessoas “saudáveis” também estão sujeitas a reportar esses problemas. Pela figura fica clara a subdivisão das classes e a ordenação que foi sugerida. Nota-se também que, embora o quesito “pensamento suicida” não seja muito reportado pelos brasileiros, nota-se uma probabilidade pequena mas de destaque dada a classe 4.

Vale ressaltar que, embora os resultados da modelagem forneçam interpretações que pareçam estar de acordo com o esperado nesse tipo de problema, alguns detalhes parecem deficientes de explanação, como é o caso de indivíduos da classe 4 apresentarem probabilidades altas de reportarem “nenhum dia” ou “quase todos os dias” nos quesitos apetite, insatisfação pessoal, falta de concentração, lentidão e/ou agitação.

### 6.2.6 Estimativas dos $\alpha$ 's (efeitos das covariáveis)

Estimativas das razões de chances e intervalos de confiança a 95% para o efeito de cada covariável incluída no modelo são apresentadas na Tabela 6.21. A classe 1 “sem risco a experienciar sintomas depressivos” é a referência em todas as tabelas.

Com exceção da variável “Viver sozinho”, todas as covariáveis apresentaram efeitos significativos sobre a prevalência das classes. A variável “Sexo”

sinalizou que mulheres apresentam maiores chances de pertencerem às classes que apontam para experiências de sintomas maiores do que os homens, sobretudo nas classes 3 e 4 “risco frequente ou constante”. A variável Nível educacional evidenciou maiores chances de indivíduos sem educação formal ou fundamental incompleto, em contraste aos de níveis educacionais mais avançados, pertencerem às classes 3 e 4. A variável Idade evidenciou maiores chances de indivíduos pertencentes aos grupos mais elevados, acima de 30 anos, em contraste aos de 18 a 29 anos, pertencerem à classe 4. O fato de não ter companheiro também contribui para aumento da chance do indivíduo ser classificado na classe 4.

Tabela 6.20: Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9 no modelo global, PNS-2013, Brasil

Item	Escore	Classes Latentes			
		1	2	3	4
Desinteresse	1	<b>0,982</b>	<b>0,551</b>	0,156	0,235
	2	0,014	<b>0,392</b>	<b>0,441</b>	0,142
	3	0,002	0,035	<b>0,357</b>	0,079
	4	0,002	0,022	0,046	<b>0,544</b>
Desânimo	1	<b>0,975</b>	<b>0,612</b>	0,116	0,192
	2	0,021	<b>0,350</b>	<b>0,414</b>	0,154
	3	0,002	0,029	<b>0,387</b>	0,110
	4	0,001	0,009	0,083	<b>0,545</b>
Sono ( <i>excesso ou falta</i> )	1	<b>0,895</b>	<b>0,480</b>	0,244	0,242
	2	0,060	<b>0,316</b>	0,254	0,090
	3	0,018	0,084	<b>0,314</b>	0,061
	4	0,028	0,121	0,189	<b>0,607</b>
Cansaço ou falta de energia	1	<b>0,929</b>	<b>0,375</b>	0,159	0,197
	2	0,057	<b>0,471</b>	<b>0,338</b>	0,102
	3	0,006	0,083	<b>0,389</b>	0,071
	4	0,009	0,071	0,115	<b>0,630</b>
Apetite ( <i>excesso ou falta</i> )	1	<b>0,960</b>	<b>0,695</b>	<b>0,380</b>	<b>0,435</b>
	2	0,028	0,216	<b>0,283</b>	0,126
	3	0,006	0,050	<b>0,278</b>	0,073
	4	0,006	0,039	0,059	<b>0,367</b>
Insatisfação pessoal	1	<b>0,994</b>	<b>0,818</b>	<b>0,348</b>	<b>0,402</b>
	2	0,005	0,171	<b>0,388</b>	0,159
	3	0,001	0,008	0,218	0,107
	4	0,001	0,003	0,046	<b>0,333</b>
Falta de concentração	1	<b>0,985</b>	<b>0,693</b>	<b>0,283</b>	<b>0,396</b>
	2	0,010	<b>0,272</b>	<b>0,388</b>	0,105
	3	0,002	0,021	<b>0,296</b>	0,089
	4	0,003	0,015	0,033	<b>0,410</b>
Lentidão e/ou agitação	1	<b>0,987</b>	<b>0,776</b>	<b>0,394</b>	<b>0,403</b>
	2	0,009	0,186	<b>0,349</b>	0,131
	3	0,002	0,021	0,225	0,083
	4	0,002	0,018	0,032	<b>0,383</b>
Pensamento suicida	1	<b>0,999</b>	<b>0,974</b>	<b>0,752</b>	<b>0,732</b>
	2	0,001	0,025	0,178	0,117
	3	0,000	0,001	0,062	0,052
	4	0,000	0,000	0,009	0,100

Nos gráficos a seguir é possível visualizar de forma dinâmica a relação condicional da proporção de cada uma das classes nos 9 itens avaliados.

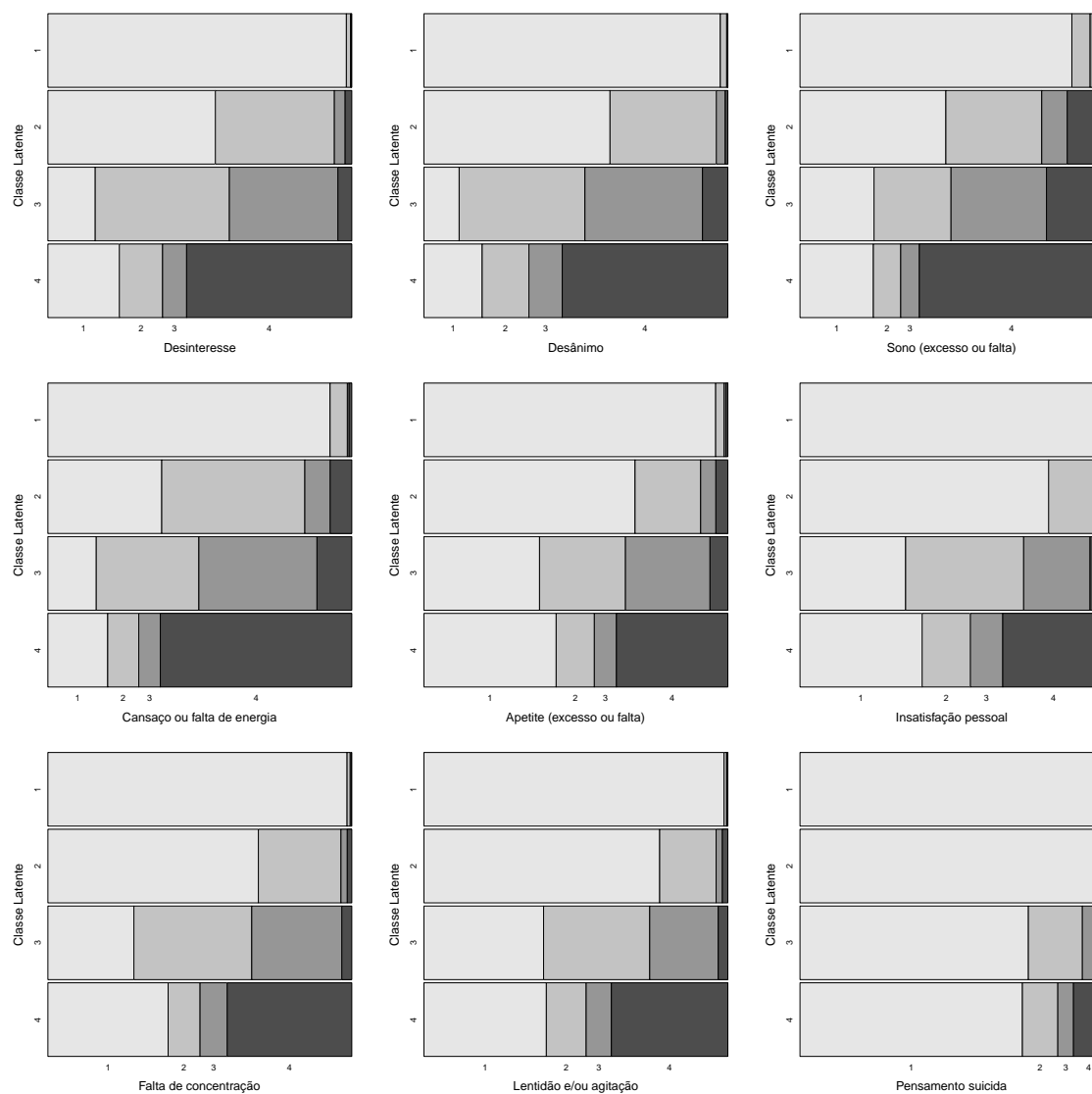


Figura 6.1: Estimativas das probabilidades condicionais (dada a classe) de resposta a cada item ( $\rho$ 's) do PHQ-9, no modelo global, PNS-2013, Brasil.

Tabela 6.21: Estimativas da razão ou da razão de chances e respectivos intervalos de confiança (95%) para os parâmetros das covariáveis, PNS-2013, Brasil

Covariáveis	Classes Latentes		
	2/1	3/1	4/1
<b>Intercepto</b>	0,307 (0,240; 0,392)	0,069 (0,038; 0,125)	0,040 (0,029; 0,055)
<b>Sexo*</b>			
Feminino	1,854 (1,596; 2,154)	2,914 (2,280; 3,723)	2,900 (2,502; 3,361)
<b>Educação*</b>			
SI/FI	-	-	-
F/MI	0,884 (0,782; 0,998)	0,644 (0,523; 0,793)	0,841 (0,696; 1,017)
M/SI	0,869 (0,756; 1,000)	0,521 (0,431; 0,629)	0,527 (0,445; 0,623)
S	0,874 (0,762; 1,003)	0,495 (0,387; 0,633)	0,371 (0,285; 0,482)
<b>Idade *</b>			
[18; 30)	-	-	-
[30; 40)	0,963 (0,870; 1,067)	1,185 (0,979; 1,433)	1,353 (1,101; 1,662)
[40; 60)	1,021 (0,930; 1,121)	1,405 (1,183; 1,667)	1,582 (1,312; 1,908)
≥ 60	0,947 (0,828; 1,084)	0,922 (0,738; 1,151)	1,488 (1,183; 1,872)
<b>Tem Companheiro*</b>			
Não	1,031 (0,945; 1,126)	1,201 (0,996; 1,449)	1,263 (1,081; 1,475)
<b>Vive sozinho</b>			
Não	1,059 (0,915; 1,226)	0,832 (0,674; 1,027)	1,112 (0,920; 1,344)

★: Efeito significativo, valor-p  $\leq 0,05$ .

### **6.3 Exploração da relação entre a classificação por classes latentes e o uso de pontos de corte**

O sucesso de um tratamento clínico está condicionado ao fato do diagnóstico de uma determinada doença ter sido feito com precisão. No caso específico da depressão, existem diversas formas para identificá-la. Considerando as causas multifatoriais associadas à doença, o diagnóstico pode ser muito complexo podendo variar desde questões fisiológicas até estressores psicossociais e até mesmo uma mistura de ambos. Em alguns casos, o paciente pode ser convidado a preencher um ou mais questionários formais padronizados, que são úteis para fins de triagem e como instrumentos de pesquisa, mas não são a única fonte de informação usada para fazer um diagnóstico. Critérios de comparação quanto ao grau em que os sintomas interferem na vida do indivíduo e o período de duração dos mesmos podem ser utilizados como termômetro para diagnosticar a doença. Seja adotando critérios pré-definidos, pontos de corte com base na pontuação em um questionário ou mesmo avaliação clínica, quando trata-se de fatores psicológicos, é inevitável sair da subjetividade.

Num estudo sobre heterogeneidades regionais quanto ao acesso ao tratamento da depressão, Lopes et al. (2016), consideraram um ponto de corte para classificar indivíduos com sintomas depressivos clinicamente relevantes, utilizando a amostra da PNS-2013. No referido estudo, adotou-se pontuações totais no PHQ-9 maiores ou iguais que 10, e estudou-se associações entre relato de diagnóstico, serviço médico, tratamentos e outras características.

Conforme enfatizado ao longo desta tese, o objetivo do modelo de classes latentes não é o de diagnosticar indivíduos mas sim o de investigar subtipos na população com padrões definidos quanto aos sintomas sem utilizar pontos de corte arbitrários.

No entanto considerou-se válido explorar como o uso de classificações usuais via ponto de corte se relaciona com a classificação sinalizada pelo ajuste da LCA. A LCA não se propôs a classificar indivíduos que fazem parte da amostra utilizada para o ajuste do modelo, mas para cada caso amostral, tem-se a probabilidade de adesão à cada classe. Usando essa probabilidade obteve-se uma possível classificação de todos os indivíduos da amostra. Tal classificação foi então associada à classificação sugerida em Lopes et al. (2016) em que os pontos de corte ( $< 10$  e  $\geq 10$ ) são levados em consideração. Os resultados são apresentados na Tabela 6.22.

Os resultados mostram basicamente, que 45,7% dos indivíduos atribuídos à classe 3 “risco frequente” em algum dos sintomas, não se classificam no grupo “depressivos” pelo ponto de corte, e mesmo entre os indivíduos da classe 4 “risco constante”, 15,8% não se classificariam via o uso deste ponto de corte.

Tabela 6.22: Proporções condicionais, dada a classe latente, avaliadas nos itens relacionados ao PHQ-9, PNS-2013, Brasil

Classe Latente	Classificação via ponto de corte			
	$< 10$	$n$	$\geq 10$	$n$
1	1,000	37.285	0,000	0
2	0,991	15.688	0,009	162
3	0,457	1.628	0,543	1.987
4	0,158	550	0,842	2.902

Num segundo momento levou-se em consideração a gravidade da depressão a partir de cinco critérios adotados a partir da pontuação contabilizada no PHQ-9. Os pontos de cortes adotados obedeceram a seguinte ordem:  $[0; 5)$  “Sem sintomas”,  $[5; 10)$  “Leve”,  $[10; 15)$  “Moderada”,  $[15; 20)$  “Mod. Severa”,  $\geq 20$  “Severa”. Na Tabela 6.23 são apresentados os resultados da amostra quanto a severidade da depressão em termos de probabilidade acompanhada do tamanho amostral em cada

classe latente.

Tabela 6.23: Proporções condicionais, avaliado nos itens relacionados ao PHQ-9, PNS-2013, Brasil

CL	Classificação via ponto de corte									
	Sem sintomas	<i>n</i>	Leve	<i>n</i>	Moderada	<i>n</i>	Mod. Severa	<i>n</i>	Severa	<i>n</i>
1	0,997	37.162	0,003	123	0,000	0	0,000	0	0,000	0
2	0,618	9.626	0,373	6.062	0,009	162	0,000	0	0,000	0
3	0,000	0	0,457	1.628	0,449	1.635	0,092	345	0,002	7
4	0,001	2	0,157	548	0,369	1.286	0,284	995	0,189	621

Considerando dois critérios distintos, o padrão de formação das classes ocorre de forma semelhante quando considera-se a presença ou ausência de sintomas. Em termos de separação via classes latentes, mesmo utilizando uma classificação mais refinada, a separação dos subtipos ocorre de maneira distinta à obtida pelo modelo de classes latentes. Em particular, tem-se 15,7% e 36,9% dos indivíduos sendo classificados como leve e moderado quanto a gravidade, sendo que estes, pelo modelo de classes latentes, aparecem na classe 4.



## 7 DISCUSSÃO E CONCLUSÕES

A LCA fornece uma maneira de identificar subgrupos subjacentes caracterizados por múltiplas dimensões que, por sua vez, podem ser usadas para examinar os efeitos de outras covariáveis.

Sendo a depressão uma doença complexa e de difícil mensuração, devido a essa configuração, a LCA representa uma metodologia promissora na discussão desse fenômeno cada vez mais presente na população nos mais variados contextos (Máximo, 2010). Com os dados da PNS, a LCA permitiu a formação de quatro classes com base em nove itens delineados para avaliar sintomas, possibilitando traçar perfis para a população brasileira evitando a utilização de pontos de corte nos escores adotados usualmente.

Exploração de ajustes separados para cada UF levou a formulação de um modelo global que levou em conta diferentes prevalências nas quatro classes latentes, porém similaridades quanto ao perfil de respostas aos itens e efeitos das covariáveis. Nessa análise global foi possível identificar com mais evidências as covariáveis associadas às classes.

Os intervalos de confiança (95%) para as razões de chances para a variável sexo mostraram valores superiores a 1 sinalizando que mulheres tem chance maior de serem classificadas em alguma das classes que apontam risco de experienciar sintomas depressivos, quando contrastadas com os homens. Esse resultado concorda com estudos que mostram que mulheres são mais propensas à

depressão (Angst et al., 2002; Justo & Calil, 2006; Santos & Kassouf, 2007; Brown & Harris, 2012).

A variável nível educacional também se mostrou com efeito significativo, em que indivíduos sem educação formal ou com ensino fundamental incompleto, em contraste aos níveis educacionais mais avançados, apresentam maiores chances de classificação nas classes de maior risco.

A variável idade sinaliza que indivíduos com mais de 30 anos tem maior chance de serem classificados na classe 4. Aqueles com mais de 60 anos apresentam chances de serem classificados nas classes 3 ou 4. Leão et al. (2017) argumentam em sua aplicação de LCA, que a sintomatologia da depressão em idosos pode estar relacionada às condições socioeconômicas, culturais e aos aspectos biológicos. Vale ressaltar que no referido trabalho apenas uma parte da população foi considerada, o que sugere um padrão em termos populacionais, visto que nos resultados encontrados no presente trabalho também encontrou-se resultados semelhantes quanto a classificação de grupos de idades mais elevados nas classes ditas mais graves. Das covariáveis investigadas apenas a binária “viver sozinho” não foi significativa.

Diante das estatísticas e previsões fornecidas pela Organização Mundial de Saúde quanto ao número de pessoas que vivem e que poderão desenvolver algum tipo de transtorno mental nos próximos 10 anos, o tema aqui abordado surge como uma contribuição no meio científico para explorar os possíveis fatores relacionados a depressão por meio de métodos estatísticos que lidam com as características de uma população que não podem ser medidas diretamente.

Os resultados demonstraram a influência de respostas em que grupos específicos de idade, sexo, escolaridade apresentam probabilidade de escolhas

específicas quanto a itens característicos aos sintomas em questão.

A depressão e o comportamento de saúde em adultos brasileiros observados na PNS-2013 foram estudados e os resultados sinalizaram a importância da avaliação da presença de depressão e da frequência e gravidade dos sintomas. Não há uma definição em termos de fatores que carreguem maior ou menor peso na manutenção da doença, mas sugestões em torno da implementação de ações para a promoção de comportamentos saudáveis são destacadas por Barros et al. (2017). Esse apontamento contempla também o processo de acesso aos serviços de saúde, questão atrelada ao fato de que a distribuição da saúde, em termos de serviço e acesso, não constitui um fenômeno aleatório na sociedade. Portanto, é preciso considerar as estruturas sociais e econômicas em uma dada região e as condições de saúde de quem lá habita (Barros, 1986). Vale ressaltar que no presente trabalho, devido à complexidade envolvida na coleta de informações sobre a renda, não foi possível acessar a relevância da mesma, possivelmente importante no problema.

Os resultados encontrados corroboram com a discussão de Ulbricht et al. (2018) em relação a complexidade em torno de delinear um perfil de resposta vista a heterogeneidade em torno do problema biológico e dos componentes populacionais que norteiam essa pesquisa.

O Ministério da Saúde e o IBGE iniciaram o planejamento da próxima PNS, prevista para ter a coleta de dados iniciada no segundo semestre de 2019, trazendo assim perspectivas com relação ao conhecimento sobre alterações que possam ter ocorrido na formação de grupos em termos de respostas para itens associados à saúde mental.

## Anexos

A PNS contém entre os seus módulos um que trata sobre percepção do estado de saúde. Neste módulo são apresentadas as questões relacionadas ao PHQ-9. O entrevistado é convidado a falar sobre problemas que podem ter incomodado nas duas últimas semanas, sendo possível optar entre quatro possibilidades de respostas, entre elas estão “nenhum dia”, “menos da metade dos dias”, “mais da metade dos dias”, “quase todos os dias”. As questões são descritas a seguir:

N10 - 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?

N11 - Nas duas últimas semanas, com que frequência o(a) Sr(a) teve problemas por não se sentir descansado(a) e disposto(a) durante o dia, sentindo-se cansado(a), sem ter energia?

N12 - Nas duas últimas semanas, com que frequência o(a) Sr(a) se sentiu incomodado por ter pouco interesse ou não sentiu prazer em fazer as coisas?

N13 - Nas duas últimas semanas, com que frequência o(a) Sr(a) teve problemas para se concentrar nas suas atividades habituais?

N14 - Nas duas últimas semanas, com que frequência o(a) Sr(a) teve problemas na alimentação, como ter falta de apetite ou comer muito mais do que de costume?

N15 - Nas duas últimas semanas, com que frequência o(a) sr(a) teve lentidão para se movimentar ou falar, ou ao contrário, ficou muito agitado(a) ou inquieto(a)?

N16 - Nas duas últimas semanas, com que frequência o(a) Sr(a) se sentiu deprimido(a), “pra baixo” ou sem perspectiva?

N17 - Nas duas últimas semanas, com que frequência o(a) Sr(a) se sentiu mal consigo mesmo, se achando um fracasso ou achando que decepcionou sua família?

N18 - Nas duas últimas semanas, com que frequência o(a) Sr(a) pensou em se ferir de alguma maneira ou achou que seria melhor estar morto?

## REFERÊNCIAS BIBLIOGRÁFICAS

ABELHA, L. Depression, a matter of public health. **Cadernos Saúde Coletiva**, v.22, n.3, p.223–223, 2014.

AGRESTI, A. **Categorical data analysis**. John Wiley & Sons, 2003. 482v.

AMERICAN PSYCHIATRIC ASSOCIATION. **Diagnostic and Statistical Manual of Mental Disorders (DSM-5)**. American Psychiatric Association, 2013.

AMERICAN PSYCHIATRIC ASSOCIATION. **DSM-5: Manual diagnóstico e estatístico de transtornos mentais**. Artmed Editora, 2014.

ANGST, J.; GAMMA, A.; GASTPAR, M.; LÉPINE, J.-P.; MENDLEWICZ, J.; TYLEE, A. Gender differences in depression. **European Archives of Psychiatry and Clinical Neuroscience**, v.252, n.5, p.201–209, 2002.

ASSOCIATION, A. P.; ET AL. **Diagnostic and statistical manual of mental disorders (DSM-5®)**. American Psychiatric Pub, 2013.

BARROS, M. B. A. A utilização do conceito de classe social nos estudos dos perfis epidemiológicos: uma proposta. **Revista de Saúde Pública**, v.20, p.269–273, 1986.

BARROS, M. B. D. A.; LIMA, M. G.; AZEVEDO, R. C. S. D.; MEDINA, L. B. D. P.; S, C. D. S.; MENEZES, P. R.; MALTA, D. C. Depressão e comportamentos de saúde em adultos brasileiros–PNS 2013. **Rev. Saúde Pública**, v.51, n.suppl 1, 2017.

BERTO, P.; D'ILARIO, D.; RUFFO, P.; VIRGILIO, R. D.; RIZZO, F. Depression: cost-of-illness studies in the international literature, a review. **The Journal of Mental Health Policy and Economics**, v.3, n.1, p.3–10, 2000.

BROWN, G. W.; HARRIS, T. **Social origins of depression: A study of psychiatric disorder in women**. Routledge, 2012.

CARROLL, R. J.; WANG, S.; SIMPSON, D. G.; STROMBERG, A. J.; RUPPERT, D. The sandwich (robust covariance matrix) estimator. **Unpublished manuscript**, 1998.

COLLINS, L. M.; LANZA, S. T. **Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences**. , 2010. 330p.

CORDAS, T. A. Depressão: da bile negra aos neurotransmissores, uma introdução histórica. Editora Lemos, 2002.

CROSS, D.; THOMSON, S.; SINCLAIR, A. **Research in Brazil - A report for CAPES by Clarivate Analytics**. São Paulo: Clarivate Analytics, 2018.

DAMACENA, G. N.; SZWARCOWALD, C. L.; MALTA, D. C.; ET AL. O processo de desenvolvimento da Pesquisa Nacional de Saúde no Brasil, 2013. **Epidemiologia e Serviços de Saúde**, v.24, p.197–206, 2015.

DAMICO, A. J. **lodown: locally download and prepare publiclyavailable microdata**. R package version 0.1.0, 2017.

DEL PORTO, J. A. Conceito e diagnóstico. **Revista Brasileira de Psiquiatria**, v.21, p.6–11, 1999.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society. Series B**, v.39, p.1–38, 1977.

FERRARI, A. J.; CHARLSON, F. J.; NORMAN, R. E.; PATTEN, S. B.; FREEDMAN, G.; MURRAY, C. J.; VOS, T.; WHITEFORD, H. A. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. **PLoS Medicine**, v.10, n.11, p.e1001547, 2013.

GOODMAN, L. A. Exploratory latent structure analysis using both identifiable and unidentifiable models. **Biometrika**, v.61, n.2, p.215–231, 1974.

GREENBERG, P. E.; KESSLER, R. C.; BIRNBAUM, H. G.; LEONG, S. A.; LOWE, S. W.; BERGLUND, P. A.; COREY-LISLE, P. K. The economic burden of depression in the United States: how did it change between 1990 and 2000? **Journal of Clinical Psychiatry**, v.64, n.12, p.1465–1475, 2003.

JUSTO, L. P.; CALIL, H. M. Depressão: o mesmo acometimento para homens e mulheres? **Archives of Clinical Psychiatry**, v.33, p.74–79, 2006.

KIND, P.; SORENSON, J. The costs of depression. **International Clinical Psychopharmacology**, v.7, p.191–195, 1993.

KROENKE, K.; SPITZER, R. L.; WILLIAMS, J. B. The PHQ-9: validity of a brief depression severity measure. **Journal of General Internal Medicine**, v.16, n.9, p.606–613, 2001.

LAMPORT, L. **LATEX: a document preparation system: user's guide and reference manual**. Addison-wesley, 1994.

LAZARSELD, P. F.; HENRY, N. W. **Latent Structure Analysis**. Houghton Mifflin Company, 1968. 294p.

LEÃO, R. C. H.; SILVA, V. L.; MOREIRA, R. S. Latent Class Analysis: a new vision of the phenomenon of depression in elderly men in the Brazilian Northeast. **Revista Brasileira de Geriatria e Gerontologia**, v.20, n.6, p.814–825, 2017.

LEITE, I. D. C.; VALENTE, J. G.; SCHRAMM, J. M. A.; DAUMAS, R. P.; RODRIGUES, R. D. N.; SANTOS, M. F.; OLIVEIRA, A. F. D.; SILVA, R. S. D.; CAMPOS, M. R.; MOTA, J. C. Carga de doença no Brasil e suas regiões, 2008. **Cadernos de Saúde Pública**, v.31, p.1551–1564, 2015.

DE LIMA, O. F.; M, V.; CRIPPA, J. A.; LOUREIRO, S. R. Study of the discriminative validity of the PHQ-9 and PHQ-2 in a sample of Brazilian women in the context



of primary health care. **Perspectives in Psychiatric Care**, v.45, n.3, p.216–227, 2009.

LOHR, S. **Sampling: design and analysis**. Nelson Education, 2009.

LOPES, C. S.; HELFWIG, N.; SILVA, G. A.; MENEZES, P. R. Inequities in access to depression treatment: results of the Brazilian National Health Survey–PNS. **International Journal for Equity in Health**, v.15, n.1, p.154, 2016.

LUMLEY, T. Analysis of complex survey samples. **Journal of Statistical Software**, v.9, n.1, p.1–19, 2004, r package version 2.2.

MALTA, D. C.; MOURA, L. D.; PRADO, R. R. D.; ESCALANTE, J. C.; SCHMIDT, M. I.; DUNCAN, B. B. Mortalidade por doenças crônicas não transmissíveis no Brasil e suas regiões, 2000 a 2011. **Epidemiologia e Serviços de Saúde**, v.23, p.599–608, 2014.

MALTA, D. C.; SZWARCOWALD, C. L. Pesquisa Nacional de Saúde e a Saúde Pública Brasileira. **Revista Brasileira de Epidemiologia**, v.18, n.supl. 2, p.1–2, 2015.

MALTA, D. C.; SZWARCOWALD, C. L. Pesquisas de base populacional e o monitoramento das doenças crônicas não transmissíveis. **Revista de Saúde Pública**, v.51, 2017.

MONAHAN, P. O.; SHACHAM, E.; REECE, M.; KROENKE, K.; ONG’OR, W. O.; OMOLLO, O.; YEBEI, V. N.; OJWANG, C. Validity/reliability of PHQ-9 and PHQ-2 depression scales among adults living with HIV/AIDS in western Kenya. **Journal of General Internal Medicine**, v.24, n.2, p.189, 2009.

MÁXMO, G. C. Aspectos sociodemográficos da depressão e utilização de serviços de saúde no Brasil. Belo Horizonte, 2010. 325p. Thesis(Ph.D.) - Belo Horizonte: Universidade Federal de Minas Gerais.

NEMADE, R.; REISS, N. S.; DOMBECK, M. Historical understandings of depression. **Mental Help. net**, 2007, disponível em <<https://www.mentalhelp.net/articles/historical-understandings-of-depression/1>>.

Acesso em Janeiro de 2018.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. **Relatório sobre a saúde no mundo 2001: Saúde Mental: Nova concepção, Nova esperança**. World Health Organization, 2001.

PAYKEL, E. S. Basic concepts of depression. **Dialogues in clinical neuroscience**, v.10, n.3, p.279, 2008.

PNS. **Pesquisa Nacional de Saúde 2013: percepção do estado de saúde, estilos de vida e doenças crônicas – Brasil, Grandes Regiões e Unidades da Federação**. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2013. 181p.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria, 2018.

ROUSSEAU, G. Depression's forgotten genealogy: notes towards a history of depression. **History of Psychiatry**, v.11, n.41, p.71–106, 2000.

SANTOS, I. S.; TAVARES, B. F.; MUNHOZ, T. N.; ALMEIDA, L. S. P.; SILVA, N. T. B.; TAMS, B. D.; PATELLA, A. M.; MATIJASEVICH, A. Sensibilidade e especificidade do Patient Health Questionnaire-9 (PHQ-9) entre adultos da população geral. **Cadernos de Saúde Pública**, v.29, p.1533–1543, 2013.

SANTOS, M. J. D.; KASSOUF, A. L. Uma investigação dos determinantes socioeconômicos da depressão mental no Brasil com ênfase nos efeitos da educação. **Economia Aplicada**, v.11, n.1, p.5–26, 2007.

SCHAFER, J. L.; KANG, J. LCCA package for R users' guide (Version 1.1. 0). **University Park: The Methodology Center, Penn State. Available from methodology.psu.edu**, 2013.

SCHULZ, K. F.; CHALMERS, I.; HAYES, R. J.; ALTMAN, D. G. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. **Jama**, v.273, n.5, p.408–412, 1995.

SOUZA-JUNIOR, P. R. B.; FREITAS, M. P. S.; ANTONACI, G. A.; SZWARCOWALD, C. L. Desenho da amostra da Pesquisa Nacional de Saúde 2013. **Epidemiologia e Serviços de Saúde**, v.24, p.207 – 216, 2015.

SPITZER, R. L.; WILLIAMS, J. W.; K, K.; ET AL. Utility of a new procedure for diagnosing mental disorders in primary care: The prime-md 1000 study. **JAMA**, v.272, n.22, p.1749–1756, 1994.

TÓFOLI, L. F. Investigação categorial e dimensional sobre sintomas físicos e síndromes somatoformes na população geral, 2008. Tese (Doutorado) - .

ULBRICHT, C. M.; CHRYSANTHOPOULOU, S. A.; LEVIN, L.; LAPANE, K. L. The use of latent class analysis for identifying subtypes of depression: A systematic review. **Psychiatry research**, v.266, p.228–246, 2018.

WOLDETENSAY, Y. K.; BELACHEW, T.; TESFAYE, M.; SPIELMAN, K.; BLESALSKI, H. K.; KANTELHARDT, E. J.; SCHERBAUM, V. Validation of the Patient Health Questionnaire (PHQ-9) as a screening tool for depression in pregnant women: Afaan Oromo version. **PloS one**, v.13, n.2, p.e0191782, 2018.

WORLD HEALTH ORGANIZATION. **Depression and other Common Mental Disorders: Global Health Estimates**. World Health Organization, 2017.