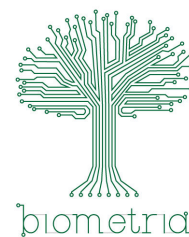




Universidade Estadual Paulista “Júlio de Mesquita Filho”
Instituto de Biociências – Câmpus de Botucatu
Programa de Pós-graduação em Biometria



Modelos de classes latentes: uma aplicação na saúde mental

Daisy Santana Ferreira

Botucatu
2023

Daisy Santana Ferreira

Modelos de classes latentes: uma aplicação na saúde mental

Dissertação de Mestrado apresentada ao Curso de Programa de Pós-graduação em Biometria da Universidade Estadual Paulista “Júlio de Mesquita Filho” como parte dos requisitos necessários para a obtenção do título de Mestre em Biometria.

Orientador: Prof(a). Dr(a). Luzia Aparecida Trinca

Botucatu
2023

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: MARIA CAROLINA A. CRUZ E SANTOS-CRB 8/10188

Ferreira, Daisy Santana.

Modelos de classes latentes : uma aplicação na saúde mental / Daisy Santana Ferreira. - Botucatu, 2023

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Luzia Aparecida Trinca

Capes: 10203001

1. Depressão mental. 2. Análise de Classes Latentes. 3. Agenda de Pesquisa em Saúde. 4. Questionário de Saúde do Paciente. 5. Saúde mental.

Palavras-chave: Depressão; Modelos de classes latentes; PNS-2019; Saúde mental; Variável latente.

Dedico este trabalho a todos aqueles que enfrentam batalhas internas invisíveis. Àqueles que lutam corajosamente contra a angústia e a dor dos pensamentos perturbadores. Eu desejo sinceramente que a cura e o alívio cheguem o mais rápido possível. Cada indivíduo merece uma vida plena de saúde mental e bem-estar.

Agradecimentos

À minha orientadora, professora Luzia Aparecida Trinca, por ter me recebido sob sua orientação e por ter me ajudado além das questões previstas durante o curso de pós-graduação. Sou extremamente grata pela tão grande paciência e generosidade.

À CAPES pelo apoio financeiro que foi fundamental para o desenvolvimento do meu trabalho.

À Universidade Estadual Paulista "Júlio de Mesquita Filho - UNESP" e ao Programa de Pós-Graduação em Biometria pela oportunidade de adquirir conhecimento e por sua contribuição para o meu desenvolvimento pessoal e profissional.

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

Aos meus pais e familiares pelo amor que dedicaram a mim ao longo de minha jornada.

Aos meus colegas e amigos pela disposição em ajudar e pelos preciosos momentos de descontração que compartilhamos juntos.

A todos os funcionários e alunos que, ao desempenharem suas funções, contribuíram para o meu crescimento e aprendizado.

À Deus por Sua graça, misericórdia e Seu amor infinito que me sustentam todos os dias.

Viver sem problemas é impossível. O sofrimento nos constrói ou nos destrói.

Augusto Cury

Resumo

O objetivo do estudo foi identificar grupos de pessoas relacionados ao perfil de saúde mental da população de residentes nas capitais brasileiras, com idade mínima de 15 anos, com base em suas respostas aos itens do Questionário de Saúde do Paciente (PHQ-9) presentes no banco de dados da Pesquisa Nacional de Saúde de 2019 (PNS-2019). Para isso, utilizamos um modelo de variáveis latentes, especificamente o de classes latentes, o qual trata o estado de saúde mental como uma variável latente categórica que se manifesta por outras variáveis também categóricas, porém observáveis. Com isso, verificamos a heterogeneidade regional da variável latente no Brasil, assim como as relações entre variáveis sociodemográficas e as prevalências das classes latentes. O questionário tem a finalidade de verificar a presença e a gravidade dos sintomas de depressão, de acordo com os critérios de diagnóstico para a doença do Manual Diagnóstico e Estatístico de Transtornos Mentais (DSM-5). Portanto, a análise de classes latentes (LCA) foi aplicada para apontar padrões de respostas que caracterizam as classes latentes mutuamente exclusivas e exaustivas por meio de ferramentas computacionais. Utilizamos o pacote *lcmm* do programa R, que permite ajustes de modelos estratificados (múltiplos grupos) e com covariáveis. Além disso, o *lcmm* possibilita a inclusão, no modelo, do esquema de amostragem complexa e dos pesos de calibração dos moradores selecionados constantes na base de dados divulgada pelo IBGE. Por meio do modelo de classes latentes, o estado de saúde mental da população das capitais do Brasil foi caracterizado em quatro classes: saudáveis, início de síndrome de Burnout, transtorno depressivo único e transtorno depressivo recorrente, com base nas probabilidades item-respostas do PHQ-9 e dos conceitos relativos à área da psicologia. Houve heterogeneidade regional das prevalências, mas em todas as capitais, a classe mais prevalente foi das pessoas saudáveis, abrangendo de 39,7% a 71,2% das populações nas 27 capitais do país. A categoria início da síndrome de Burnout foi a segunda classe mais prevalente cujos valores variaram de 18,1% a 42,7%. Em termos das classificações apontando para depressão, as distribuições das prevalências nas últimas duas classes foram similares. Em suma, para transtorno depressivo recorrente os valores variaram de 3,8% a 11,5%. Desta análise a região Nordeste se sobressai com prevalências maiores do que nas outras regiões na classe mais grave. Além disso, neste trabalho também foi ajustado um modelo de classes latentes, estratificado por capital, considerando covariáveis categóricas como sexo, cor/raça, estado civil, escolaridade e renda per capita, bem como covariáveis quantitativas como idade e o número de pessoas vivendo no domicílio. Para interpretação, usando a classe "Saudáveis" como referência, foram estimadas as razões de chances e respectivos intervalos de confiança (95%). A covariável sexo e estado civil se mostraram fortemente associados à variável latente, indicando que as mulheres e os divorciados têm uma maior chance de pertencimento às classes mais "graves" quando comparadas aos homens e aos casados, respectivamente. A variável idade também se revela associada à variável latente, embora com razões de chances próximas de um para todas as classes, em relação à classe "Saudável". Este estudo visou contribuir com o conhecimento da dimensão dos problemas relacionados à saúde mental da população brasileira,

trazendo subsídios úteis para melhor planejamento de serviços e tratamentos especializados para refrear o agravamento do cenário.

Palavras-chave: PNS-2019, Modelos de Classes Latentes, Saúde Mental, Depressão, Variável Latente.

Abstract

The objective of the study was to identify groups of people related to the mental health profile of the population of residents in Brazilian capitals, aged at least 15 years, based on their responses to the items of the Patient Health Questionnaire (PHQ-9) present in the database of the 2019 National Health Survey (PNS-2019). To do this, we use a model of latent variables, specifically that of latent classes, which treats the state of mental health as a categorical latent variable that is manifested by other variables that are also categorical, but observable. With this, we verified the regional heterogeneity of the latent variable in Brazil, as well as the relationships between sociodemographic variables and the prevalence of latent classes. The questionnaire aims to verify the presence and severity of depression symptoms, according to the diagnostic criteria for the disease in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Therefore, latent class analysis (LCA) was applied to identify response patterns that characterize mutually exclusive and exhaustive latent classes using computational tools. We used the *lcca* package in the R program, which allows adjustments to stratified models and with covariates. Furthermore, the *lcca* makes it possible to include, in the model, the complex sampling scheme and the calibration weights of the selected residents contained in the database published by IBGE. Using the latent class model, the mental health status of the population of the capitals of Brazil was characterized into four classes: Healthy, Indicative of Burnout syndrome, Single depressive disorder and Recurrent depressive disorder, based on the item-response probabilities of the PHQ-9 and concepts related to the area of psychology. There was regional heterogeneity in prevalence, but in all capitals, the most prevalent class was healthy people, covering from 39.7% to 71.2% of the populations in the country's 27 capitals. The category Indications of Burnout syndrome was the second most prevalent class, with values ranging from 18.1% to 42.7%. In terms of classifications pointing to depression, the prevalence distributions in the last two classes were similar. In short, for Recurrent depressive disorder the values ranged from 3.8% to 11.5%. From this analysis, the Northeast region stands out with higher prevalence rates than in other regions in the most severe class. Furthermore, this work also adjusted a latent class model, stratified by capital, considering categorical covariates such as sex, color/race, marital status, education and per capita income, as well as quantitative covariates such as age and the number of people living in the residence. For interpretation, using the "Healthy" class as a reference, odds ratios and respective confidence intervals (95%) were estimated. The covariates sex and marital status were strongly associated with the latent variable, indicating that women and divorced people have a greater chance of belonging to the more "serious" classes when compared to men and married people, respectively. The age variable is also associated with the latent variable, although with odds ratios close to one for all classes, in relation to the "Healthy" class. This study aimed to contribute to knowledge of the extent of problems related to the mental health of the Brazilian population, providing useful information for better planning of services and specialized treatments to curb the worsening of the situation.

Keywords: PNS-2019, Latent Class Models, Mental Health, Depression, Latent Variable.

Lista de figuras

Figura 1 – Modelo de variável latente formativa.	5
Figura 2 – Modelo de variável latente reflexiva.	5

Lista de tabelas

Tabela 1 – Distribuição das características sociodemográficas (pessoas com 15 anos ou mais) segundo a amostra nacional ou residentes nas capitais, PNS-2019, considerando o delineamento amostral complexo.	18
Tabela 2 – Proporção das respostas para a amostra nacional e das capitais em cada questão do PHQ-9 da PNS- 2019, levando em consideração o delineamento amostral complexo.	20
Tabela 3 – Resumo das estatísticas de ajuste para selecionar o número de classes latentes da PNS-2019.	21
Tabela 4 – Estimativas das probabilidades item-resposta ($\rho_{j,r_j c}$) do PHQ-9, PNS-2019.	23
Tabela 5 – Estimativas das prevalências segundo as capitais das unidades da federação, PNS-2019.	24
Tabela 6 – Estimativas da razão de chances (RC) e respectivos intervalos de confiança (95%) para o modelo com covariáveis, usando a classe "Saudáveis" como referência.	27

Lista de abreviaturas e siglas

6A70	Transtorno Depressivo Único
6A71	Transtorno Depressivo Recorrente
CID-11	Classificação Internacional das Doenças - 11ª revisão
DSM-5	Manual Diagnóstico e Estatístico de Transtornos Mentais
IBGE	Instituto Brasileiro de Geografia e Estatística
L	Variável Latente
LCA	Análise de Classes Latentes
MS	Ministério da Saúde
PHQ-9	Questionário de Saúde do Paciente
PNS	Pesquisa Nacional de Saúde

Sumário

1	INTRODUÇÃO	1
2	MODELOS DE CLASSES LATENTES	4
3	MATERIAL E MÉTODOS	13
4	RESULTADOS	17
4.1	Análise descritiva	17
4.2	Análise de classes latentes	20
4.2.1	Seleção do número de classes latentes	20
4.2.2	Modelo de classes latentes estratificado por capital	21
4.2.3	Modelo de classes latentes, estratificado por capital, com covariáveis	25
5	CONCLUSÃO	28
5.1	Trabalhos futuros	28
	Referências	29
	Apêndices	32
	APÊNDICE A – DICIONÁRIO DE VARIÁVEIS - PNS 2019	33

1 Introdução

A saúde mental é um tema cada vez mais relevante e importante na sociedade atual, pois ela é essencial para o bem-estar e a qualidade de vida, tanto a nível individual como coletivo. A pandemia da COVID-19, por exemplo, destacou ainda mais a importância de se estar saudável mentalmente para enfrentar os desafios e pressões relacionados à saúde, finanças, trabalho e relacionamentos de forma positiva, contribuindo assim para a sociedade em geral. Aqueles que enfrentam problemas de saúde mental, frequentemente, têm mais dificuldades para encontrar soluções diante de situações cotidianas, como a gestão de relações interpessoais e o desempenho no trabalho. Vale ressaltar que a saúde mental é uma característica latente, por isso não pode ser observada diretamente. Logo, é comum utilizar instrumentos como questionários, cujas respostas podem indicar o estado de saúde mental. No entanto, esses questionários não são suficientes para um diagnóstico e tratamento adequados para doenças psicológicas.

Há diversos problemas relacionados à saúde mental, um deles é a depressão: um transtorno que afeta o humor, os pensamentos e o comportamento do indivíduo. Ela é caracterizada por sentimentos persistentes de tristeza, desesperança, perda de interesse e prazer em atividades que antes eram agradáveis, além de alterações no sono, apetite, concentração e energia. Com a finalidade de auxiliar na identificação da presença e a gravidade dos sintomas de depressão foram criados vários instrumentos ou escalas, um deles é o questionário com nove perguntas chamado PHQ-9 (Patient Health Questionnaire), que se baseia nos critérios de diagnóstico para a depressão do DSM-5 (Manual Diagnóstico e Estatístico de Transtornos Mentais - 5ª edição) (APA - American Psychiatric Association, 2014).

A depressão afeta mais de 300 milhões de pessoas de todas as idades em todo o mundo. É a principal causa de incapacidade em escala global e contribui significativamente para a carga de doenças. Existe uma relação entre a depressão e a saúde física, em que doenças cardiovasculares, por exemplo, podem levar ao desenvolvimento da depressão e vice-versa. Pessoas que passaram por eventos adversos na vida, como desemprego, luto ou trauma psicológico, são mais suscetíveis a desenvolver essa patologia. Mulheres são mais propensas que homens, e em casos extremos, o transtorno pode levar ao suicídio. Anualmente, aproximadamente 800 mil pessoas perdem a vida devido ao suicídio, tornando-o a quarta principal causa de morte entre indivíduos com idade entre 15 e 29 anos (WHO - World Health Organization, 2023).

Esses números alarmantes destacam a necessidade urgente de prevenir e fornecer apoio adequado às pessoas que lutam com problemas de saúde mental que podem levar a essa triste e evitável perda de vida. De acordo com [Pan American Health Organization \(2023\)](#), apesar de haver tratamentos eficazes para a depressão, menos da metade das pessoas afetadas em todo o mundo recebem esses tratamentos, e isso pode chegar a menos de 10% em muitos países. Dificuldades no acesso aos recursos, falta de profissionais capacitados e estigmas sociais associados aos transtornos mentais são algumas das barreiras ao tratamento efetivo. Além disso, a avaliação imprecisa também representa um obstáculo para o atendimento adequado. Em países de todos os níveis de renda, pessoas com depressão, frequentemente, não são diagnosticadas corretamente, enquanto outras que não têm o transtorno são diagnosticadas de forma inadequada, recebendo intervenções desnecessárias.

Em relação aos episódios de depressão, esses podem ser classificados como transtorno depressivo único (6A70) e recorrente (6A71), de acordo com a intensidade dos sintomas apresentados, segundo a CID-11 (Classificação Internacional das Doenças - 11ª revisão). Em um episódio de transtorno depressivo único, o indivíduo vivencia apenas um episódio depressivo, ou seja, é registrado um episódio quase diário de humor deprimido prolongado, ou diminuição acentuada do interesse pelas atividades, que dura pelo menos duas semanas. Já em um episódio de transtorno depressivo recorrente, a pessoa teve pelo menos dois episódios depressivos com alguns meses de intervalo, sem perturbação significativa do humor ([World Health Organization, 2018](#)). Alguns artigos destacam a importância da detecção precoce da depressão e explicam que a identificação antecipada pode levar a melhores resultados a longo prazo, além disso, ambos recomendam o tratamento imediato da doença ([Pan American Health Organization, 2022](#); [BARROSO; MELO; GUIMARÃES, 2014](#); [PROGRESS IN MIND - BRAZIL PSYCHIATRY & NEUROLOGY RESOURCE CENTER, 2022](#)).

Outro problema relacionado à saúde mental é a síndrome de Burnout, também chamada de síndrome do esgotamento profissional. Assim como a depressão, é reconhecida pela Organização Mundial da Saúde (OMS) como uma doença ocupacional e que tem ligação ao contexto de trabalho. Entre os sintomas mais comuns estão o cansaço excessivo, a sensação de não ter descansado após dormir, dores de cabeça frequentes, queda no desempenho profissional, entre outros. A síndrome de Burnout é mais prevalente entre aqueles que trabalham em áreas de alto estresse ou que enfrentam uma jornada dupla de trabalho, como profissionais da saúde, policiais e professores ([COSTA et al., 2022](#); [J.PUC-SP, 2022](#)).

Dada a dimensão do problema e a disponibilidade de dados brasileiros ao nível de nação, através dos projetos de Pesquisa Nacional de Saúde, que incluem instrumentos de saúde mental, o emprego de análises estatísticas que desvendam a característica latente é de extrema relevância. Logo, nesta dissertação, o objetivo principal é identificar perfis relacionados à saúde mental com base nas respostas aos itens do PHQ-9, presentes no banco de dados da Pesquisa Nacional de Saúde de 2019 (PNS-2019). Para isso, utilizaremos um modelo de variáveis latentes, especificamente

o de classes latentes que trata o estado de saúde mental como uma variável latente categórica que se manifesta por outras variáveis também categóricas, porém observáveis, conhecidas como itens ou manifestas. Outros objetivos são estudar a heterogeneidade regional da variável latente no Brasil e associá-la às características sócio-demográficas.

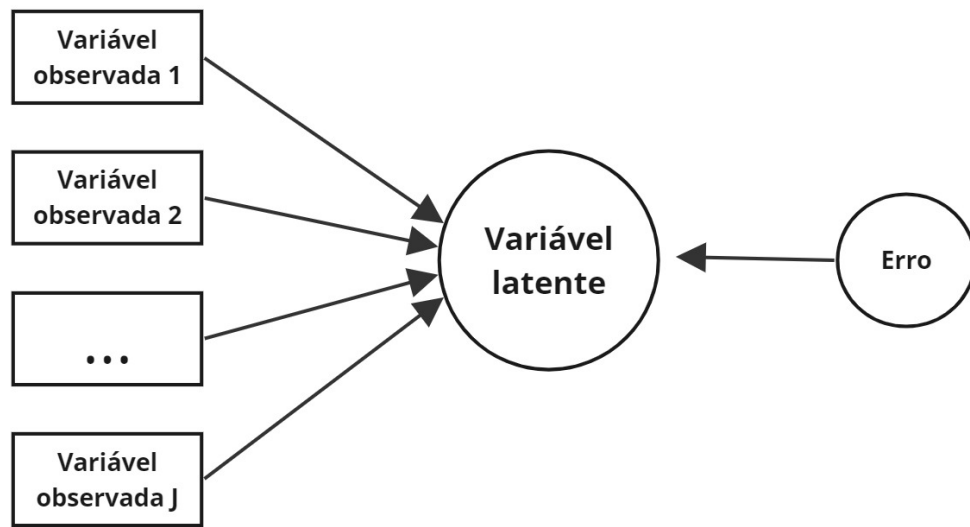
Do ponto de vista de modelagem estatística, para cumprir com os objetivos acima, os modelos deverão incorporar o esquema de amostragem utilizado na coleta de dados, a saber, em múltiplos estágios, caracterizando um esquema complexo, além do uso de pesos de calibração, comuns nos planos de amostragem a nível nacional. A modelagem de variáveis latentes sob esquemas de amostragem complexos ainda é escassa na bibliografia, assim como, os recursos computacionais disponíveis, justificando a contribuição deste trabalho. No capítulo 2, apresentamos uma introdução ao tema variável latente e os conceitos e definições envolvidos no modelo de classes latentes com e sem covariáveis. Já no capítulo 3, descrevemos o banco de dados da Pesquisa Nacional de Saúde (PNS-2019) que está sendo utilizado nesta dissertação e o roteiro de aplicação da modelagem, cujos resultados são apresentados no capítulo 4. Por fim, no capítulo 5, resumimos as conclusões.

2 Modelos de classes latentes

Vários são os modelos de variáveis latentes passíveis de aplicação nas áreas médicas e da saúde, principalmente na Psicologia, que lida com características não diretamente observáveis, mas que se manifestam na forma de sintomas ou respostas. Por exemplo, comer demais ou dificuldade para dormir podem ser algumas das manifestações de um estado de saúde mental desequilibrado. A variável latente e as manifestações podem variar quanto aos seus tipos e podem ser qualitativas ou quantitativas. Nesse sentido, [Lanza, Flaherty e Collins \(2003\)](#) explicam que os tipos da variável latente e manifestas irão determinar a metodologia adequada, tais como: 1. Análise fatorial: quando tanto a variável latente quanto as manifestas são contínuas; 2. Análise de perfil latente: quando a variável latente é categórica e as manifestas são contínuas; 3. Análise de traço latente ou teoria de resposta ao item: quando a variável latente é contínua e as manifestas categóricas; e 4. Análise de classes latentes: quando ambos os tipos são variáveis categóricas.

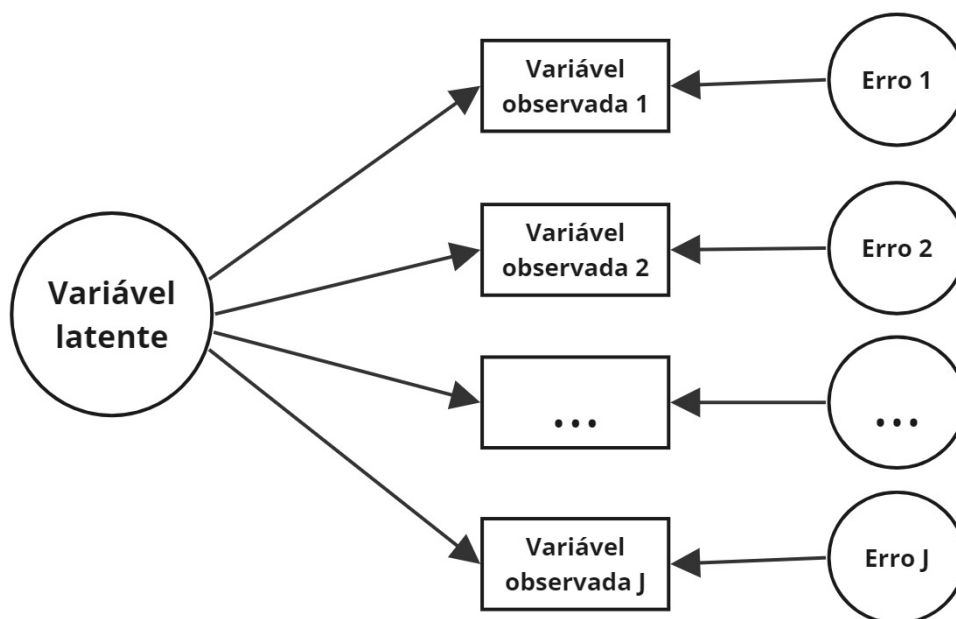
Além dos diferentes modelos, existem os tipos de relação entre as variáveis que podem ser reflexivas ou formativas. As Figuras 2 e 1, adaptadas de [Beaujean \(2014\)](#), ilustram as duas formas sendo que, na Figura 2, a variável latente exerce influência nas variáveis observadas. Já na Figura 1, são as variáveis observadas que influenciam a variável latente. Por convenção, no esquema, a variável latente e o componente de erro são representados por formas circulares, já as variáveis manifestas são representadas por formas retangulares e o sentido das setas mostra a relação de causa entre elas.

Com a finalidade de ilustrar a diferença entre ambas, podemos citar algumas variáveis observadas que indicam ou formam o estado depressivo das pessoas tendo como variável latente o estado de saúde mental. As perguntas para a variável latente reflexiva podem ser relacionadas a frequência em que se sentiu mal consigo mesmo, se achando um fracasso, também a ocorrência em que teve problemas no sono, como dificuldade para adormecer, acordar frequentemente à noite ou dormir mais do que de costume; e a constância em que teve pensamentos suicidas. Já para a variável latente formativa, em que a variável observada contribui para a formação do traço latente, logo podemos pensar no grau de relação com a família e amigos, no nível de produtividade no trabalho e no impacto do trauma em relação ao abuso sexual, físico ou emocional na infância. Entre elas, a forma reflexiva parece ser mais frequente nas áreas de estudo de comportamento, saúde mental ou qualidade de vida.



miro

Figura 1 – Modelo de variável latente formativa.



miro

Figura 2 – Modelo de variável latente reflexiva.

A análise de classes latentes (LCA) é uma técnica estatística multivariada que caracteriza subgrupos de indivíduos que compõem a população, de acordo com os perfis da variável latente sendo estudada. Ela se baseia nos padrões de respostas observados às várias variáveis categóricas, manifestações da variável latente de interesse (MASTELLA, 2015). Sendo assim, a LCA modela os padrões de resposta por meio de um modelo de probabilidades que caracterizam as classes latentes mutuamente exclusivas e exaustivas (LANZA; FLAHERTY; COLLINS, 2003). Isso significa que o modelo postula que cada indivíduo pertence a apenas uma das classes identificadas,

e todos os indivíduos se enquadram à alguma das classes. Ferramentas computacionais viabilizam o ajuste do modelo, produzindo as estimativas de seus parâmetros.

As variáveis categóricas são manifestações ou respostas de uma condição latente, e estão, portanto, sujeitas a erros ou variações aleatórias. A variável latente não é diretamente observável e por isso usamos as manifestações a fim de inferir sobre o estado latente. Para ilustração podemos considerar o comportamento de uma pessoa diagnosticada com a doença depressão. Não existe um método de investigação objetivo da doença, mas ela afeta o emocional da pessoa e provoca algumas dificuldades na execução das atividades do cotidiano. Um questionário com as perguntas relevantes pode ser formulado com escolhas de alternativas para respostas, que indicam o grau de dificuldade ou a severidade do sentimento. Por exemplo, com relação às duas últimas semanas, com que frequência teve problemas no sono, como dificuldade para adormecer, acordar frequentemente à noite ou dormir mais do que de costume, as seguintes alternativas são oferecidas para que a pessoa escolha a que melhor lhe aplica: nenhum dia (1), menos da metade dos dias (2), mais da metade dos dias (3) e quase todos dias (4). A análise do padrão de respostas, nas diversas perguntas, pode auxiliar na identificação dos indivíduos que sofrem com o distúrbio mental.

A ideia de se utilizar as variáveis indicadoras para descrever a população em classes latentes mutuamente exclusivas e exaustivas foi lançada por [Lazarsfeld e Henry \(1968\)](#), porém o método de estimação por máxima verossimilhança foi descrito por [Goodman \(1974\)](#). O desenvolvimento das definições matemáticas que está de acordo com o livro de [Lanza, Flaherty e Collins \(2003\)](#).

A concepção da LCA tem como base a tabela de contingência formada pelo cruzamento das respostas de todas as questões em estudo, ou seja, das variáveis categóricas observadas. Para J variáveis indicadoras, definimos $Y_j = r_j$ tal que $r_j = 1, \dots, R_j$ (o número de categorias possíveis de Y_j) e $j = 1, \dots, J$. Assim, o cálculo para obter o número de caselas para a tabela é dado por $W = \prod_{j=1}^J R_j$. Logo, se houver apenas duas variáveis indicadoras em que cada uma segue com duas alternativas para resposta, a tabela resultante é composta por $W = \prod_{j=1}^2 R_j = 2 \cdot 2 = 4$ caselas, levando ao agrupamento em no máximo quatro classes, cada uma caracterizando um perfil de resposta. No entanto, tal raciocínio é difícil se J ou o número de alternativas for grande. Para quatro alternativas, por exemplo, o número de combinações já sobe para dezesseis e, para $J = 9$ itens, o número de caselas resulta em 262 144 ($W = \prod_{j=1}^9 R_j = 4^9$), o que é um valor consideravelmente grande, sendo necessário algum método capaz de capturar informação relevante de uma maneira mais resumida e interpretável. O modelo de classes latentes, além de cumprir com esse objetivo, permite também a estimação da prevalência em cada classe.

Para o padrão de resposta w ($w = 1, \dots, W$), definimos o vetor de variáveis categóricas descrito por $\mathbf{Y}_w = (Y_{w1}, \dots, Y_{wJ})$. Cada padrão de resposta $\mathbf{Y}_w$ está associado com uma probabilidade $P(\mathbf{Y}_w = \mathbf{y}_w)$ tal que $\sum_{w=1}^W P(\mathbf{Y}_w = \mathbf{y}_w) = 1$.

Seja uma variável latente categórica representada por L com $c = 1, \dots, C$ classes

latentes mutuamente exclusivas e exaustivas. A probabilidade do padrão de resposta é modelada por dois tipos de parâmetros que relacionam as variáveis categóricas e as classes latentes. Ambos tipos de parâmetros também são probabilidades conforme definidos a seguir:

- Prevalência na Classe Latente (γ_c): é a probabilidade relacionada à classe, ou seja, se trata da proporção populacional da variável latente L de categoria c . Neste trabalho essas categorias descrevem os subgrupos que caracterizam o estado de saúde mental na população brasileira;
- Probabilidade do item-resposta ($\rho_{j,r_j|c}$): é a probabilidade de resposta à alternativa r_j da questão j condicionada à classe c . No contexto deste trabalho, ρ é a probabilidade do indivíduo responder determinada alternativa dado o estado de saúde mental.

Dadas as definições, seguem as restrições probabilísticas: $\sum_{c=1}^C \gamma_c = 1$ e $\sum_{r_j=1}^{R_j} \rho_{j,r_j|c} = 1$ para todo j . Usando a função indicadora $I(y_j = r_j)$ que assume o valor um quando a resposta para o item j é igual a r_j , e zero caso contrário, a probabilidade da variável $\mathbf{Y}_i$ assumir um padrão de resposta particular $\mathbf{y}_i$, para um indivíduo i qualquer, é dada por:

$$P(\mathbf{Y}_i = \mathbf{y}_i) = \sum_{c=1}^C \gamma_c \prod_{j=1}^J \prod_{r_j=1}^{R_j} \rho_{j,r_j|c}^{I(y_{ij}=r_j)}. \quad (2.1)$$

A suposição fundamental desse modelo é a independência local. Esse requisito quer dizer que, dada a classe latente, as variáveis observáveis são independentes.

O número de parâmetros do modelo em (2.1) é dado por $P = (C - 1) + C \cdot \sum_{j=1}^J (R_j - 1)$, sendo que o primeiro termo se refere ao número de prevalências de classes e o segundo ao número de probabilidades item-resposta. O método de estimação usual é o de máxima verossimilhança. Devido a complexidade do modelo, não existe solução fechada para os estimadores e métodos iterativos são utilizados. Uma possibilidade é o algoritmo EM (expectation-maximization) proposto por [Dempster \(1977\)](#) que pode ser combinado com o conhecido algoritmo Newton-Raphson. Uma vez encontrada uma solução, a estatística da razão de verossimilhança é utilizada para avaliar a qualidade global do ajuste. Nesse caso, ela reflete o quão bem o modelo de classes latentes ajusta-se aos dados observados.

Sejam f_w e $\hat{f}_w$ as frequências observadas e estimadas (sob o modelo de classes latentes), respectivamente, na casela w . Então, a estatística de razão de verossimilhanças é dada por:

$$G^2 = 2 \cdot \sum_{w=1}^W f_w \cdot \log \left(\frac{f_w}{\hat{f}_w} \right), \quad (2.2)$$

em que o grau de liberdade é $W - P - 1$, tal que valores altos de G^2 indicam falta de ajuste do modelo. O princípio da estatística da razão de verossimilhança também pode ser aplicado para a comparação de dois modelos encaixados fornecendo metodologia para a seleção de modelos. No entanto, vale ressaltar que dois modelos com número de classes latentes diferentes não são

encaixados, e portanto o princípio não pode ser aplicado. Então, para escolher o número de classes, utilizamos os critérios de informação AIC (Critério de Informação de Akaike) e BIC (Critério Bayesiano de Schwarz), que são dados pela estatística G^2 com penalizações conforme as expressões:

$$AIC = G^2 + 2 \cdot P \quad (2.3)$$

$$BIC = G^2 + [\log(n)] \cdot P, \quad (2.4)$$

em que n é o tamanho da amostra, e valores menores dos critérios indicam modelos mais parcimoniosos. Além disso, é importante ressaltar que a interpretabilidade das classes é um quesito de decisão junto aos critérios de seleção para auxiliar a escolha do número de classes latentes no modelo. Como apontado em [Lanza et al. \(2007\)](#), duas são as propriedades desejadas nas classes geradas pelo modelo: a separação e a homogeneidade. A separação entre as classes ocorre quando o perfil de respostas dos indivíduos não se misturam entre as classes, em outras palavras, a distribuição das probabilidades item-resposta são bem distintas entre as classes. Já a homogeneidade da classe ocorre quando as probabilidades de resposta em uma ou algumas categorias são destacadamente altas para todos os itens, indicando padrão de resposta bem definido para os indivíduos que pertencem à mesma classe.

Quando o modelo apresenta classes bem homogêneas e separadas espera-se que a probabilidade a posteriori de atribuir um indivíduo a uma das classes seja alta, dado o seu padrão de resposta. A probabilidade a posteriori do indivíduo i ser classificado na classe c , p_{ic} , é dada por:

$$p_{ic} = P(L_i = c | \mathbf{Y}_i = \mathbf{y}_i) = \frac{(\prod_{j=1}^J \prod_{r_j=1}^{R_j} \rho_{j,r_j|c}^{I(y_{ij}=r_j)}) \cdot \gamma_c}{\sum_{c=1}^C \gamma_c \prod_{j=1}^J \prod_{r_j=1}^{R_j} \rho_{j,r_j|c}^{I(y_{ij}=r_j)}}. \quad (2.5)$$

Uma medida global de incerteza na classificação é baseada na entropia que nesse caso é definida como:

$$E = 1 - \frac{\sum_{i=1}^n \sum_{c=1}^C -p_{ic} \cdot \log p_{ic}}{n \cdot \log C}. \quad (2.6)$$

Essa estatística varia de 0 a 1, e valores altos indicam melhor separação entre as classes. Como os valores de E tendem a diminuir com o aumento do número de classes, esse é um critério que deve ser usado com cuidado, sempre levando o contexto em consideração.

Quando a amostragem é complexa, envolvendo estratificação, conglomerados e pesos de seleção para cada observação, a construção da função de verossimilhança é muito difícil, sendo necessário o uso da função de pseudo-verossimilhança. Segundo [Patterson et al. \(2002\)](#), na pseudo-verossimilhança, o peso de cada indivíduo é incorporado na função e alguma forma de ajuste pelo efeito do delineamento amostral é aplicado. Assim, todas as estatísticas G^2 , AIC e BIC são baseados na pseudo-verossimilhança e não na verossimilhança de fato.

O modelo em (2.1) é o mais básico, apropriado para uma população não estratificada. A metodologia pode ser estendida para populações estratificadas, por exemplo, estratificação regional ou por variáveis sócio-demográficas, de acordo com o contexto da aplicação.

Seja V uma variável de estratificação com Q grupos. Podemos desenvolver o mesmo raciocínio de construção de tabelas de contingências como previamente exposto, porém, agora seguindo a estratificação, tal que o número de caselas se torna $W = Q \cdot \prod_{j=1}^J R_j$. Cada padrão de resposta $\mathbf{y}_w$ no grupo q correspondente está associado com a probabilidade de ocorrência $P(\mathbf{Y}_w = \mathbf{y}_w | V = q)$ tal que $\sum_{w=1}^W P(\mathbf{Y}_w = \mathbf{y}_w | V = q) = 1$ dentro de cada grupo q ($q = 1, \dots, Q$). Assumindo que a variável latente L é composta pelo mesmo número de classes em todos os grupos, a equação referente à probabilidade em (2.1), particularizada para o indivíduo i , é uma probabilidade condicional do padrão de resposta dado o grupo, e segue como:

$$P(\mathbf{Y}_i = \mathbf{y}_i | V_i = q) = \sum_{c=1}^C \gamma_{c|q} \prod_{j=1}^J \prod_{r_j=1}^{R_j} \rho_{j,r_j|c,q}^{I(y_{ij}=r_j)}, \quad (2.7)$$

em que o $\gamma_{c|q}$ é a probabilidade de um indivíduo pertencente ao grupo q ser vinculado à classe c ; e o $\rho_{j,r_j|c,q}$ é a probabilidade de um indivíduo, do grupo q e classe c , responder r_j ao item j . Resumindo, tanto as prevalências das classes latentes e as probabilidades item-resposta são condicionadas ao grupo.

Considerando o vetor $\boldsymbol{\theta}$ que contém todos os parâmetros livres a serem estimados (ρ 's e γ 's) e o termo $I_i(\boldsymbol{\theta})$ que representa a contribuição do indivíduo i do grupo q no logaritmo da verossimilhança, podemos expressar o logaritmo da verossimilhança de todos os dados observados da seguinte forma:

$$\begin{aligned} I(\boldsymbol{\theta}) &= \sum_{i=1}^n I_i(\boldsymbol{\theta}) \\ &= \sum_{i=1}^n \log P(\mathbf{Y}_i = \mathbf{y}_i | V_i = q)^{f_i} \\ &= \sum_{q=1}^Q \sum_{i \in q} f_i \log \left(\sum_{c=1}^C \gamma_{c,q} \prod_{j=1}^J \prod_{r_j=1}^{R_j} \rho_{j,r_j|c,q}^{I(y_{ij}=r_j)} \right), \end{aligned} \quad (2.8)$$

em que f_i é a frequência de indivíduos em dados agregados, com mesmo perfil de resposta, ou pesos em amostragem complexa, no grupo q . No caso de dados agregados, n se refere ao número de padrões de respostas e no caso de dados não agregados, n se refere ao tamanho da amostra. No caso em que cada indivíduo da amostra está associado a um peso de seleção f_i , a função é a verossimilhança de uma pseudo-população, na qual cada indivíduo é visto como sendo clonado f_i vezes. Assim, as estimativas dos parâmetros são obtidas considerando esses pesos. Para obtenção dos erros padrões das estimativas, informação do delineamento amostral como estratificação e conglomerado, devem ser incorporadas, utilizando o estimador tipo sanduíche (SCHAFER; KANG, 2013; PATTERSON et al., 2002).

A principal mudança em relação ao modelo sem estratificação é que, na presença de grupos, o número de parâmetros a serem estimados pode ser muito maior. No caso em que não se incorpora nenhuma restrição nos parâmetros temos: $P = Q \cdot (C - 1) + Q \cdot C \cdot \sum_{j=1}^J (R_j - 1)$, o que pode trazer dificuldades de estimação e também de interpretação. É comum formular modelos

mais parcimoniosos introduzindo a condição de equivalência dos parâmetros nos grupos. Nesse sentido, o modelo sem restrição e o modelo com restrição são encaixados e o critério da razão de verossimilhança pode ser usado na comparação dos dois modelos. Por exemplo, supondo que os valores ρ 's são iguais em todos os grupos, o que chamamos de invariância de medição, resulta em um número de parâmetros de $P = Q \cdot (C - 1) + C \cdot \sum_{j=1}^J (R_j - 1)$.

Outras extensões existem, como o modelo que inclui covariáveis, ou seja, características específicas dos indivíduos, que podem estar associadas às prevalências das classes latentes. O modelo usual é similar ao da regressão logística. Como nos modelos de regressão tradicionais, o objetivo de adicionar covariáveis aos modelos de classes latentes é identificar características dos indivíduos possivelmente associadas à variável latente. Ao introduzir covariáveis, buscamos encontrar fatores que possam explicar a relação entre diferentes categorias latentes para um determinado fenômeno.

Denote o vetor $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$, com dimensão $p \times 1$, o conjunto de covariáveis associadas ao indivíduo i que podem influenciar a probabilidade de pertencer a uma das classes $L_i = c$, para $c = 1, \dots, C$. O modelo supõe que o vetor $\mathbf{x}_i$ influencia os itens $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iM})^T$ somente por meio do classificador latente L_i .

Por meio de um modelo de regressão logística politômica ou multinomial, as covariáveis podem estar relacionadas às probabilidades de pertencimento a cada uma das classes, com uma das classes escolhida como referência, conforme descrito por [Agresti \(2012\)](#). A probabilidade de um indivíduo i pertencer à classe c , dentro do grupo q , é modelada por:

$$\gamma_{ic|q} = P(L_i = c | V_i = q) = \frac{\exp(\mathbf{x}_i^T \cdot \boldsymbol{\alpha}_{c|q})}{\sum_{c'=1}^C \exp(\mathbf{x}_i^T \cdot \boldsymbol{\alpha}_{c'|q})}, \quad (2.9)$$

em que $\boldsymbol{\alpha}_{c|q} = (\alpha_{1c|q}, \alpha_{2c|q}, \dots, \alpha_{pc|q})^T$ representa um vetor de coeficientes a serem estimados, com dimensão $p \times 1$. Para garantir a identificabilidade dos parâmetros, é necessário escolher uma das classes como referência para as demais. Portanto, se a classe d for escolhida como referência, definimos $\boldsymbol{\alpha}_{d|q} = (0, 0, \dots, 0)^T$ para a classe específica. Com essa restrição, podemos expressar o logaritmo da razão das prevalências por:

$$\log\left(\frac{\gamma_{ic|q}}{\gamma_{id|q}}\right) = \mathbf{x}_i^T \cdot \boldsymbol{\alpha}_{c|q}, \quad (2.10)$$

de forma que os elementos de $\boldsymbol{\alpha}_{c|q}$ representam o logaritmo da razão de chances que distingue ou compara a classe c e a classe de referência d . Em geral $x_{i1} = 1$ para todos os indivíduos, de forma que $\alpha_{1c|q}$ representa o intercepto, como na regressão logística multinomial. Aplicando a função exponencial a cada elemento $\boldsymbol{\alpha}_{c|q}$ temos a razão de chances referente ao efeito de uma covariável específica. Na prática, restrições podem ser impostas aos parâmetros $\boldsymbol{\alpha}_{c|q}$, por exemplo, homogeneidade do vetor nos diferentes grupos, o que pode resultar em modelos mais parcimoniosos.

Similarmente à expressão em (2.8), considerando a contribuição do indivíduo ou perfil

de resposta i , o logaritmo da função de verossimilhança é:

$$I(\theta) = \sum_{q=1}^Q \sum_{i=1}^n f_i \log \left(\sum_{c=1}^C \left\{ \frac{\exp(\mathbf{x}_i^T \cdot \boldsymbol{\alpha}_{c|q})}{\sum_{c'=1}^C \exp(\mathbf{x}_i^T \cdot \boldsymbol{\alpha}_{c'|q})} \right\} \prod_{j=1}^J \prod_{r_j=1}^{R_j} \rho_{j,r_j|c,q}^{I(y_j=r_j)} \right). \quad (2.11)$$

Novamente, aplicando o algoritmo EM, considerando a variável latente como omissa, é possível estimar os diversos tipos de parâmetros, inclusive os efeitos das covariáveis. A maximização de (2.11) estimando todos os parâmetros simultaneamente, em um passo, foi proposto por [Dayton e Macready \(1988\)](#) e [Bandein-Roche et al. \(1997\)](#). Outros autores propuseram a estimação em três passos: primeiro estima os γ 's e ρ 's, segundo, atribui cada indivíduo à uma classe (seguindo algum procedimento, por exemplo, modal ou aleatório, baseados nas probabilidades a posteriori) e, por fim, ajusta o modelo logístico, dado a classificação obtida no segundo passo (ver [Vermunt \(2010\)](#) para detalhes). Neste trabalho, usamos a primeira alternativa devido a nossa limitação de software capaz de lidar com o planejamento amostral complexo.

O modelo de classes latentes tem sido regularmente aplicado ao tema saúde mental. Para finalizar esse capítulo fazemos um breve relato de algumas publicações. [Idalino \(2019\)](#), na sua tese de doutorado, empregou a modelagem de classes latentes para investigar os diferentes perfis de estado de saúde mental dos brasileiros. Esta análise se baseou nos itens do PHQ-9, disponíveis no banco de dados da Pesquisa Nacional de Saúde (PNS) de 2013. Já [Santos et al. \(2015\)](#) realizaram um estudo que utilizou a LCA para identificar subgrupos de indivíduos com diferentes níveis de apoio social percebido, fundamentado nas respostas aos itens de uma escala de apoio social, com um dos objetivos sendo o de examinar as associações entre esses subgrupos e a prevalência de Transtornos Mentais Comuns (TMC). [Carmo et al. \(2017\)](#), interessados no estudo da prevalência de transtornos mentais comuns em mulheres brasileiras, assim como nos padrões subjacentes dos subtipos, aplicaram duas alternativas para análise, a CFA (Confirmatory Factor Analysis) conjunta e a LCA. Para as variáveis manifestas foi utilizado a escala SRQ-20 (Self-Reporting Questionnaire-20) que é um instrumento de avaliação de saúde mental utilizado para identificar sintomas de transtornos mentais comuns, como depressão e ansiedade, em contextos de saúde pública e pesquisa. É um questionário de autorrelato com vinte perguntas que abordam vários aspectos do bem-estar psicológico e emocional. Os autores também compararam os resultados destas análises multivariadas à alternativa usual, que utiliza um de ponto de corte pré definido, baseado na soma da pontuação obtida na escala. Por sua vez, [Silva Junior \(2012\)](#) realizou um estudo sobre a epidemiologia dos subtipos de depressão, aplicando análise de classes latentes dos sintomas depressivos em uma amostra populacional da região Metropolitana de São Paulo. Para mencionar alguns estudos em outros países, destacamos [Chen et al. \(2019\)](#) e [Essau e Torre-Luque \(2019\)](#), ambos com dados dos Estados Unidos da América. O primeiro trata de uma pesquisa sobre a prevalência, os fatores de risco e a análise de classes latentes multi-grupo, considerando transtornos de ansiedade ao longo da vida. Diferenças de gênero foram identificadas nos subtipos depressivos entre pacientes com ansiedade, destacando a necessidade de abordagens de prevenção eficientes e específicas para diferentes grupos populacionais. Por fim, [Essau e Torre-Luque \(2019\)](#)

explicaram o perfil de comorbidades de transtornos mentais em adolescentes com essa mesma metodologia.

Em relação aos recursos computacionais que implementam ajustes de modelos na Análise de Classes Latentes, Amorim et al. (2015) destaca os softwares Mplus (MUTHÉN, 1998; ASPAROUHOV; MUTHÉN, 2012), WinLTA (COLLINS et al., 2002) e as bibliotecas do R (R Core Team, 2023), como "e1071" (MEYER et al., 2014) e "poLCA" (Polytomous Variable Latent Class Analysis) (LINZER; LEWIS, 2011). A biblioteca "e1071" lida apenas com itens ou variáveis manifestas binárias, enquanto "poLCA" admite manifestas politômicas e ajusta efeitos de covariáveis. Pesquisadores da universidade estadual da Pensilvânia têm desenvolvido, e disponibilizado livremente em página web, vários recursos computacionais, por exemplo, o WinLTA, o "lcca" (SCHAFER; KANG, 2013; The Methodology Center, 2013) e PROC LCA (LANZA et al., 2015), uma macro para uso no SAS. No entanto, dentre essas opções, exceto "lcca" e PROC LCA, uma amostragem aleatória simples é assumida. Este último, ainda em versões preliminares, permite a incorporação de conglomeração na amostragem, apenas. Já o "lcca" lida com os planejamentos usuais de amostragens de pesquisas a nível nacional. Sua desvantagem é que o pacote não tem sido atualizado para versões recentes do R, sendo necessário o uso de uma versão antiga. Por sua vez, o Mplus, um software comercial, parece ser o estado da arte em flexibilidade de implementações de modelos de variáveis latentes, mas de difícil acesso, no nosso contexto, devido à limitação de recursos financeiros para aquisição de licenças.

3 Material e métodos

A Pesquisa Nacional de Saúde (PNS) é uma pesquisa com âmbito nacional, realizada, em parceria, pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e o Ministério da Saúde, Governo Federal, cujos objetivos são extrair informações sobre o funcionamento do sistema nacional de saúde e verificar a situação da população em relação a doenças e fatores de risco, além de fornecer informações para um melhor planejamento em saúde e alocação de recursos, sendo de fundamental importância para embasar as políticas públicas do Sistema Único de Saúde (SUS) (IBGE - Instituto Brasileiro de Geografia e Estatística, 2020b; FIOCRUZ - Fundação Instituto Oswaldo Cruz, 2020; SZWARCOWALD et al., 2014). Desse modo, a pesquisa, além de tratar questões diretamente relacionadas à saúde da população, também traz a investigação sobre os estilos de vida dos brasileiros sendo, portanto, uma pesquisa de base domiciliar.

Até o momento, duas edições da PNS foram realizadas, a primeira em 2013 e a segunda em 2019. A pesquisa é composta por seis partes, sendo que as duas últimas são referentes às variáveis de amostragens e derivadas, em que cada parte contém módulos. A percepção do estado de saúde pertence ao módulo N (Percepção do Estado de Saúde) da parte 4, que trata do questionário do morador selecionado, incluindo um grupo de nove questões que abordam a saúde mental nos últimos quinze dias. Embora não identificado na pesquisa, essas questões se referem aos itens do Patient Health Questionnaire (PHQ-9) (KROENKE; SPITZER; WILLIAMS, 2001), os quais são descritos na PNS-2019 como:

1. 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?
2. 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?
3. Nas duas últimas semanas, com que frequência o(a) Sr(a) teve pouco interesse ou não sentiu prazer em fazer as coisas?
4. Nas duas últimas semanas, com que frequência o(a) Sr(a) teve problemas para se concentrar nas suas atividades habituais?

5. 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?
6. 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)?
7. Nas duas últimas semanas, com que frequência o(a) Sr(a) se sentiu deprimido(a), "para baixo" ou sem perspectiva?
8. 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?
9. 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?

Essas nove questões têm as mesmas categorias de resposta, que são: nenhum dia (1); menos da metade dos dias (2); mais da metade dos dias (3) e quase todos os dias (4).

A população alvo da PNS-2019 é definida pelos indivíduos de 15 anos ou mais de idade, residentes em domicílios particulares permanentes. Exclui-se domicílios localizados em setores censitários especiais ou de população muito pequena como agrupamento indígenas, quartéis, bases militares, alojamento, acampamento, penitenciárias, colônias penais, presídios, cadeias, asilos, centro de atendimento integrado a crianças e ao adolescentes, conventos, hospitais, assentamentos agrícolas e quilombos (STOPA, 2020).

Para a amostragem, a pesquisa segue a Amostra Mestra que é a estrutura do Sistema Integrado de Pesquisas Domiciliares (IBGE - Instituto Brasileiro de Geografia e Estatística, 2020a). O plano contempla três estágio, no primeiro estágio as unidades primárias de amostragem (setor censitários ou um conjunto de setores censitários) foram sorteadas de acordo com suas estratificações, sendo que a probabilidade de sorteio de cada uma foi proporcional ao seu tamanho (número de domicílios particulares permanentes); no segundo, os domicílios foram sorteados por amostragem aleatória simples; e no terceiro, o morador, com 15 anos ou mais, do domicílio foi sorteado também por meio da amostragem aleatória simples.

O tamanho da amostra foi determinado com base nas análises dos dados da PNS-2013, do efeito do plano amostral e dos critérios específicos do IBGE relativos à precisão de estimativas de indicadores de interesse. Para a formulação específica e tamanhos amostrais planejados e selecionados em cada unidade da federação são disponibilizada às páginas 14 e 15 da publicação que divulga os resultados da Pesquisa Nacional de Saúde - PNS 2019 (IBGE - Instituto Brasileiro de Geografia e Estatística, 2020a). No total, a amostra foi de 8 036 UPAs e 108 525 domicílios selecionados.

Devido ao esquema amostral complexo e com probabilidade distinta de seleção é necessária uma calibração, tanto para domicílios quanto para os moradores selecionados. A calibração

é feita por pesos definidos pelo produto do inverso das expressões das probabilidades de seleção de cada estágio do plano amostral, incluindo correção de não resposta e ajuste dos totais populacionais. Esses pesos já fazem parte da base de dados divulgada pelo IBGE e está disponível publicamente em <<https://www.pns.icict.fiocruz.br/bases-de-dados/>> e <<https://www.ibge.gov.br/estatisticas/downloads-estatisticas.html?caminho=PNS/2019/Microdados/Dados>>.

Para o estudo, realizamos um recorte dos dados, satisfazendo a condição de residência em capital de estados ou Distrito Federal, ter 15 anos ou mais e a pessoa entrevistada ser o próprio indivíduo sendo pesquisado. Vale ressaltar que ao utilizar uma subamostra da amostra é necessário recalcular os pesos de calibração levando em conta o novo tamanho amostral. Tal recalibração foi realizada previamente às análises, seguindo as diretrizes constantes no material de divulgação da PNS, organizado pela Fundação Oswaldo Cruz (<<https://www.pns.icict.fiocruz.br/programacao-em-r/>>). Todas as variáveis da base de dados estão detalhadamente descritas no "dicionário das variáveis da PNS 2019", anexo ao site <<https://www.pns.icict.fiocruz.br/bases-de-dados/>>. Após inspecionar esse dicionário de variáveis, selecionamos aquelas referentes ao estado de saúde mental, do módulo N, e algumas características gerais dos moradores como: idade, escolaridade, número de pessoas no domicílio, renda per capita, estado civil, cor/raça e sexo. Além disso, utilizamos apenas os registros nos quais a própria pessoa respondeu. Isso foi importante porque o respondente não necessariamente era o indivíduo sendo pesquisado. Essa informação foi revelada pela presença de uma variável, em cada módulo, que se referia ao informante da pesquisa, cujas alternativas eram: "A própria pessoa", "Outro morador", "Pessoa não moradora" e "Ignorado". No nosso estudo, só consideramos as respostas em que a opção selecionada foi "A própria pessoa", o que resultou em uma amostra de 22 681 pessoas.

Para a descrição da amostra completa e da subamostra, calculamos as porcentagens de respostas em cada categoria. Mesmo as variáveis quantitativas foram categorizadas de acordo com as instruções dadas no site <<https://www.pns.icict.fiocruz.br/calculo-dos-indicadores-modulo-n-2019/>>. Devido à heterogeneidade nos fatores sócio-demográficos e condições na capacidade do sistema de saúde e também do acesso aos serviços de saúde no país, visamos uma análise estratificada por capitais. Porém, o primeiro passo é a decisão do número de classes que representam a heterogeneidade do estado depressivo na população vivendo nas capitais. Como visto no Capítulo 2, modelos com número de classes distintos não são encaixados e não temos um critério claro para a decisão, que é, então, realizada com base nos valores dos critérios globais de qualidade de ajuste (AIC e BIC), e também levando em conta a parcimônia na interpretação dos perfis das classes obtidas. Para isso, ajustamos vários modelos com número de classes de 2 a 5 e usamos os valores dos critérios aproximados pela pseudo-verossimilhança, assim como as características envolvendo a interpretação, separação e homogeneidade das classes. Uma vez definido o número de classes, passamos à interpretação do ajuste estratificado por capital. Por fim, incluímos as covariáveis sócio-demográficas para investigação de suas associações com a variável latente.

Para os ajustes utilizamos o pacote *lcca* (The Methodology Center, 2013) no programa R

([R Core Team, 2023](#)), por ser disponibilizado livremente e ser capaz de lidar com dados oriundos de esquemas complexos de amostragem. O uso do peso de calibração do morador selecionado assim como, dos fatores de estratificação e conglomeração do plano, resulta em inferências mais fidedignas à população estudada. O pacote `lcca` tem implementado o método de pseudo-máxima verossimilhança para estimar os parâmetros do modelo, que incorpora as ponderações apropriadas. Existem outros pacotes no R capazes de ajustar os modelos da LCA, porém, não encontramos neles, a possibilidade de incorporar esquemas de amostragens diferentes do aleatório simples. É importante ressaltar que, além desse pacote, utilizamos o pacote `survey` ([LUMLEY, 2023](#)), também do software R por causa da amostragem complexa, pois assim asseguramos a consideração adequada de todos os elementos do esquema de seleção adotado, incluindo estratificação, e ajustes para os efeitos da não resposta. Segundo [Vieira, Claro e Assunção \(2019\)](#), ignorar a incorporação dos pesos amostrais na análise, pode levar a estimativas tendenciosas.

4 Resultados

Os resultados da pesquisa estão separados em duas seções. Na seção 4.1 apresentamos uma exploração descritiva da amostra utilizada e na seção 4.2 apresentamos os resultados da análise de classes latentes, considerando, primeiramente a seleção do número de classes, depois o ajuste do modelo estratificado por capital, interpretando o significado das classes e, finalmente, o modelo com covariáveis.

4.1 Análise descritiva

Foram realizadas análises descritivas preliminares do conjunto de dados da Pesquisa Nacional de Saúde com o intuito de se obter informações sobre as características gerais referentes as variáveis que, com base na literatura, estão associadas aos problemas de saúde mental. Então, os itens do PHQ-9 e aspectos sociodemográficos fazem parte deste estudo.

Primeiramente, apresentamos na Tabela 1 as características sociodemográficas das pessoas de 15 anos ou mais na amostra nacional ($n = 63\,340$) e na subamostra ($n = 22\,681$), considerando nesse último apenas as capitais. Vale salientar que todos os percentuais foram obtidos ajustando-se pelos pesos amostrais, através do pacote survey. Podemos perceber que os dados da subamostra corresponde a aproximadamente 36% da amostra nacional e a representa bem, exceto nas variáveis escolaridade e renda per capita, as quais apresentam maiores porcentagens em categorias mais elevadas na subamostra, um resultado esperado já que cidades maiores oferecem melhores oportunidades nesses quesitos. Logo, vemos que apenas 16% das pessoas na amostra nacional têm o nível de escolaridade superior completo e 13% recebe mais de três salários mínimos, enquanto na amostra referente às capitais esses valores correspondem a 26% e 22%, respectivamente.

Tabela 1 – Distribuição das características sociodemográficas (pessoas com 15 anos ou mais) segundo a amostra nacional ou residentes nas capitais, PNS-2019, considerando o delineamento amostral complexo.

Variáveis	Categoria	Nacional		Capitais	
		<i>n</i>	%	<i>n</i>	%
Sexo	Masculino	26 286	39,69	8 959	40,51
	Feminino	37 054	60,31	13 722	59,49
Faixas Etárias	[15; 18)	616	1,96	183	1,40
	[18; 30)	8 934	15,61	3 244	16,06
	[30; 45)	18 579	30,35	6 766	30,07
	[45; 60)	17 287	27,53	6 235	27,58
	[60; 75)	13 481	18,77	4 770	19,23
	[75+)	4 443	5,78	1 483	5,65
Escolaridade ¹	S/Int. ou FI	26 338	37,06	5 814	13,30
	F ou MI	8 719	14,80	3 002	36,72
	M ou SI	18 390	31,78	8 025	22,43
	S	9 893	16,36	5 840	27,56
Renda per capita ²	Até 1 SM	34 150	51,03	9 323	26,92
	1 até 2 SM	15 803	27,69	5 940	11,18
	2 até 3 SM	5 448	9,07	2 513	39,05
	Mais de 3 SM	7 939	12,21	4 905	22,86
Número de pessoas no domicílio	1	14 094	11,18	5 617	12,98
	2	19 371	29,22	6 867	29,74
	3 a 4	23 911	46,51	8 353	45,16
	5 ou mais	5 964	13,09	1 844	12,12
Cor / Raça	Branca	22 490	41,86	8 494	42,39
	Preta	7 417	11,84	2 803	2,22
	Parda	32 397	44,72	10 899	42,57
	Outros	1 036	1,58	485	12,82
Estado civil	Casado	23 980	44,75	7 666	41,62
	Divorciado	6 256	8,45	2 594	9,05
	Viúvo	6 349	8,33	2 127	7,64
	Solteiro	26 755	38,47	10 294	41,68
Total		63 340	100	22 681	35,80

¹ S/Int. ou FI.: Sem instrução ou fundamental incompleto; F ou MI: Fundamental completo ou médio incompleto; M ou SI: Médio completo ou superior incompleto; S: Superior completo.

² SM: Salário mínimo.

Nas capitais temos que 60% são do sexo feminino; a faixa etária de maior prevalência (58%) está entre 30 e 60 anos e o grupo mais jovem, de 15 a 30 anos, corresponde a 17% da amostra. Com relação à escolaridade, 87% têm pelo menos o ensino fundamental e 28% têm ensino superior completo. Já para a renda per capita, 62% da amostra relataram mais de dois

salários mínimos e apenas 27% disseram ganhar até um salário mínimo. No quesito do número de pessoas no domicílio, observa-se que a maioria das residências (75%) possui entre duas a quatro pessoas, o que sugere a presença de uma família com um ou dois filhos. Além disso, cerca de 13% das pessoas moram sozinhas e 12% abrigam pelo menos cinco pessoas na casa, evidenciando frequência baixa de famílias numerosas nas capitais. Ainda na subamostra, verificou-se que 43% dos indivíduos declararam ser da cor/raça parda e 42% brancos, enquanto que 2% se identificaram como pretos. No que diz respeito ao estado civil, observou-se que os solteiros e casados são predominantes, representando 42% da amostra, os demais, divorciados e viúvos, somam 17%.

A escolha das variáveis descritas na Tabela 1 teve como critério fatores que podem estar associados ao desenvolvimento da depressão segundo Barroso, Melo e Guimarães (2014), juntamente com Gaudenci et al. (2015). Deste modo, acreditamos que a condição econômica, a classificação na sociedade como pessoa intelectual, as exigências relativas à maturidade como idade, preconceitos relacionados a cor/raça e sexo, além da convivência com pessoas, essas que podem ser queridas ou não, são bastantes discutidos por serem pontos que geram questionamentos nos indivíduos perante a comunidade que fazem parte. Questionamentos esses que podem ser positivos ou negativos na construção da própria imagem.

A Tabela 2 apresenta as proporções de respostas para as duas amostras em estudo, nacional e capitais, em relação aos nove itens do PHQ-9, sendo que a numeração presente de 1 a 4 corresponde, respectivamente, às alternativas: nenhum dia, menos da metade dos dias, mais da metade dos dias e quase todos os dias. Desta maneira, podemos notar que todas as questões relativas à categoria "nenhum dia" foi a mais prevalente para ambas amostras, com valores variando de 61% a 95%, dos quais os menores valores se referem aos dois primeiros itens: cansaço e problemas no sono. Consequentemente, esses são os itens que aparecem com relatos nas duas categorias mais graves ("mais da metade dos dias" e "quase todos os dias"). Dentre os demais, destacam-se falta de prazer ou interesse, falta de concentração e deprimido ou sem perspectiva, respectivamente, variando em torno de 30%, 21% e 29% para a amostra nacional e 31%, 22% e 27% para a amostra das capitais, reportando terem sofrido do problema em pelo menos um dos dias. A maior diferença entre as duas amostras está relacionada ao item cansaço ou falta de energia, no qual a amostra das capitais apresenta 1% a mais em pelo menos metade dos dias (categorias três e quatro juntas).

Como o objetivo neste estudo é identificar perfis de saúde mental distintos, então não pontuamos cada questão na escala de 0 a 3 porque, no momento, não há intenção de resumir a informação pela pontuação na escala, como muitos estudos fazem para o PHQ-9.

Tabela 2 – Proporção das respostas para a amostra nacional e das capitais em cada questão do PHQ-9 da PNS- 2019, levando em consideração o delineamento amostral complexo.

Questões	Nacional (%)				Capitais (%)			
	1	2	3	4	1	2	3	4
Sono (excesso ou falta)	62,59	17,53	6,77	13,12	62,57	17,07	6,11	14,24
Cansaço ou falta de energia	61,36	23,11	6,83	8,70	60,74	22,35	7,33	9,57
Falta de prazer ou interesse	69,59	20,42	4,90	5,10	69,21	19,77	5,42	5,60
Falta de concentração	78,51	13,87	4,12	3,50	78,01	14,18	4,15	3,66
Apetite (excesso ou falta)	78,22	12,09	4,79	4,90	78,41	11,28	5,45	4,86
Lentidão e/ou agitação	81,21	11,38	3,91	3,50	82,44	10,42	3,71	3,43
Deprimido ou sem perspectiva	70,93	18,67	5,38	5,01	72,58	17,28	4,73	5,40
Sentimento de fracasso	82,31	11,26	3,27	3,16	83,00	10,29	3,19	3,53
Pensamento suicida	94,94	3,20	1,01	0,84	95,35	2,69	0,97	0,99

Categorias de resposta das nove questões são: nenhum dia (1); menos da metade dos dias (2); mais da metade dos dias (3) e quase todos os dias (4).

4.2 Análise de classes latentes

Ao ajustar um modelo de classes latentes, o pesquisador se depara com diversos desafios a serem superados. Dentre eles podemos destacar a difícil escolha do número de classes latentes, bem como a decisão sobre as restrições a serem impostas aos parâmetros do modelo quando a análise é feita por grupos. Além disso, outro desafio relevante consiste em incluir covariáveis no modelo, o que pode informar sobre as relações dessas com a variável latente. Portanto, é imprescindível que o pesquisador esteja atento a essas questões e busque formas de superar esses desafios para que seja possível obter resultados sensatos e relevantes para o modelo de classes latentes.

4.2.1 Seleção do número de classes latentes

A escolha do número de classes latentes é um passo importante na modelagem, e essa decisão é tomada com base no princípio da parcimônia. De acordo com [Lanza, Flaherty e Collins \(2003\)](#), em amostras aleatórias simples, é recomendado escolher o valor de C por meio dos critérios como AIC (Critério de Informação de Akaike), BIC (Critério de Informação Bayesiano) e o logaritmo da função de verossimilhança, levando também em conta, o conhecimento conceitual do problema para tal escolha. Devido à falta de aninhamento entre modelos que possuem números diferentes de C , o teste estatístico de razão de verossimilhanças, comumente utilizado para comparar modelos, não pode ser empregado. Além do mais, sob o planejamento amostral complexo, o método de estimação utilizado é o de Pseudo-Máxima Verossimilhança e não há, na bibliografia, consenso sobre as definições dos critérios de informação AIC e BIC. Sendo assim, calculamos os valores dos critérios usando o log da Pseudo-Verossimilhança nas expressões

usuais. Os modelos ajustados sendo comparados consideram a estratificação por capitais. Assim, o número de padrão de respostas é $W = Q \cdot \prod_{j=1}^9 R_j = 27 \cdot 4^9 = 7\,077\,888$.

Tabela 3 – Resumo das estatísticas de ajuste para seleccionar o número de classes latentes da PNS-2019.

C	P	log (MPV)	AIC	BIC	Entropia
2	81	−130 785,74	261 733,50	262 383,90	0,874
3	135	−125 442,70	251 115,40	252 239,40	0,855
4	189	−123 572,18	247 522,40	249 039,90	0,848
5	243	−122 364,37	245 214,70	247 165,90	0,801

Na Tabela 3 apresentamos o número de parâmetros, os valores do logaritmo do máximo da função pseudo-verossimilhança, os critérios de informação e a entropia para os ajustes de quatro modelos, considerando de duas até cinco classes latentes. Vemos que aumentando o número de classes, os valores de AIC e BIC diminuem, com grandes diferenças na comparação de duas e três classes, indicando ganho menor para cinco classes. A entropia segue análoga aos critérios de informação AIC e BIC, exceto na comparação das diferenças em que o valor mais alto se encontra entre as classes quatro e cinco. Com isso, avaliando também os aspectos de interpretação e as probabilidades item-resposta, chegamos à decisão de usar quatro classes no modelo, devido a qualidade de ajuste nos dados e melhor separação entre as classes latentes. Em uma investigação preliminar, que não levou em conta a estratificação por capitais, obtivemos resultados similares quanto à avaliação do número de classes.

4.2.2 Modelo de classes latentes estratificado por capital

As estimativas dos parâmetros do modelo estão apresentadas nas Tabelas 4 e 5. Antes de iniciar as interpretações do ajuste, esclarecemos que a análise de diagnóstico desse ajuste ficou prejudicada devido a alguns fatores. Primeiro, não localizamos metodologia que nos auxiliasse no cálculo de resíduos considerando a amostragem complexa e, segundo, a aplicação de fórmulas usuais, por exemplo, resíduo de Pearson, resultou em milhares de valores extremos causados por probabilidades esperadas ínfimas em muitos dos padrões de respostas. Esse problema pode ter sido causado pelo número excessivo de categorias para muitos dos indicadores. No entanto, continuamos a análise desconsiderando a estratificação por capital; portanto, o número de padrões de respostas possíveis para nove itens, cada um com quatro categorias, resultou em $W = 262\,144$. Esse número é consideravelmente maior do que o tamanho amostral das capitais ($n = 22\,681$). Mesmo considerando a amostra nacional ($n = 63\,340$), o número de padrões de respostas possíveis é ainda muito grande. Logo, estamos no caso de dados esparsos o qual podem levar a instabilidades nas estimativas, porém, não detectamos esse problema nos nossos ajustes. Como o método é iterativo e seu ponto de partida é uma distribuição aleatória dos indivíduos às classes,

se repetirmos o processo diversas vezes e obtivermos estimativas distintas, teremos indicação de instabilidade. Cada ajuste foi repetido algumas vezes que produziram as mesmas estimativas. De qualquer forma, as interpretações do modelo devem ser cautelosas.

Para facilitar a interpretação do modelo, iniciaremos pelas estimativas das probabilidades item-resposta, que foram supostas homogêneas entre os grupos. Pode-se observar que a primeira e a última classe latente da Tabela 4 são bem separadas. O grupo de pessoas que se encontram na classe latente um tem a tendência em responder que, em relação às duas últimas semanas, em nenhum dia ocorreram os nove sintomas listados. Já a classe latente quatro têm uma probabilidade alta em reportar que quase todos os dias aconteceram os sinais de uma depressão; no entanto, as questões relacionadas à falta de concentração, apetite (excesso ou falta), sentimento de fracasso e lentidão ou agitação são notadas uma probabilidade considerável de indivíduos que responderam a alternativa “em nenhum dia” no questionário; apesar disso, classe um e quatro não se confundem. Ainda comparando ambas, a classe um tem probabilidade mínima de resposta 0,857 em “nenhum dia” para todos os quesitos, representando assim os indivíduos saudáveis, sendo portanto bem homogênea. Por outro lado, dado a classe quatro, os valores das probabilidades item-respostas não excedem a 0,714, desconsiderando o item pensamento suicida, no qual as pessoas costumam afirmar nunca terem experimentado tais pensamentos. Portanto, a classe quatro não é tão homogênea comparada à classe um.

Na segunda classe latente, observa-se uma tendência em responder às duas primeiras alternativas. Além disso, é visto que as variáveis relacionadas ao sono, cansaço, falta de prazer e sentimento depressivo resultaram em uma probabilidade de item resposta superior a 33% para a alternativa “menos da metade dos dias”, o que a distingue da primeira classe. Portanto, parece indicar pessoas que estão sobrecarregadas devido à sua rotina de trabalho. Por conta disso, há um indicativo de que está ocorrendo uma Síndrome de Burnout. Já a classe latente três tem como característica a resposta de que, em relação às duas últimas semanas, as pessoas respondem duas alternativas intermediárias que são “em menos da metade dos dias” e “mais da metade dos dias” ocorreram os sintomas, exceto para o item de pensamento suicida em que 0,756 relatam não ter acontecido em nenhum momento; por outro lado, há 0,014 dos indivíduos que quase todos os dias teve esse tipo de pensamento. Essa variável possui uma influência significativamente maior em relação às outras, devido a ser uma manifestação proeminente em indivíduos que sofrem de depressão. A questão dos problemas no sono não caracteriza essa classe por seus valores estarem bem parecidos, também os itens: falta de concentração, apetite e lentidão ou agitação incluem probabilidades significativas para a primeira alternativa da questão, o que diferencia das demais variáveis. Logo, para a classe latente três é pertinente pensar em pessoas com transtorno depressivo único e para a classe quatro, transtorno depressivo recorrente, porque essa última classe tem uma probabilidade de 0,129 em quase todos os dias para a variável de pensamento suicida.

Tabela 4 – Estimativas das probabilidades item-resposta ($\rho_{j,r_j|c}$) do PHQ-9, PNS-2019.

Questões	Respostas	Classes Latentes			
		1	2	3	4
Sono (excesso ou falta)	1	0,857	0,411	0,210	0,132
	2	0,076	0,334	0,261	0,134
	3	0,018	0,092	0,267	0,044
	4	0,048	0,163	0,262	0,690
Cansaço ou falta de energia	1	0,891	0,309	0,124	0,123
	2	0,081	0,509	0,290	0,077
	3	0,013	0,099	0,394	0,086
	4	0,014	0,083	0,192	0,714
Falta de prazer ou interesse	1	0,971	0,443	0,107	0,148
	2	0,024	0,490	0,398	0,130
	3	0,001	0,043	0,420	0,116
	4	0,004	0,024	0,075	0,606
Falta de concentração	1	0,984	0,638	0,252	0,323
	2	0,012	0,324	0,386	0,137
	3	0,002	0,023	0,331	0,110
	4	0,002	0,014	0,031	0,429
Apetite (excesso ou falta)	1	0,964	0,663	0,309	0,371
	2	0,025	0,239	0,273	0,106
	3	0,004	0,060	0,348	0,103
	4	0,007	0,037	0,070	0,421
Lentidão e/ou agitação	1	0,980	0,745	0,358	0,421
	2	0,013	0,213	0,332	0,122
	3	0,002	0,029	0,264	0,099
	4	0,005	0,013	0,046	0,358
Deprimido ou sem perspectiva	1	0,971	0,568	0,058	0,164
	2	0,024	0,391	0,439	0,148
	3	0,003	0,028	0,408	0,081
	4	0,003	0,013	0,095	0,606
Sentimento de fracasso	1	0,992	0,786	0,241	0,361
	2	0,005	0,202	0,411	0,121
	3	0,001	0,008	0,295	0,085
	4	0,001	0,004	0,054	0,432
Pensamento suicida	1	0,998	0,974	0,756	0,721
	2	0,001	0,025	0,158	0,095
	3	0,000	0,001	0,072	0,055
	4	0,000	0,000	0,014	0,129

Categorias de resposta das nove questões são: nenhum dia (1); menos da metade dias (2); mais da metade dos dias (3) e quase todos os dias (4).

Na Tabela 5 são exibidas as prevalências de cada capital juntamente com os respectivos rótulos das classes latentes. Estes rótulos estão organizados em ordem crescente de gravidade em relação à saúde mental: "Saudáveis", "Indício de Síndrome de Burnout", "Transtorno depressivo único" e "Transtorno depressivo recorrente". Tais rótulos foram indicados pela interpretação da Tabela 4. Entretanto, existem vários outros tipos de problemas mentais como ansiedade, dependência química (álcool, drogas, etc.) e transtornos do sono (insônia, apneia do sono, narcolepsia, etc.) que podem levar ao estado depressivo, mas não foram considerados por não haver informação no PHQ-9 ou por não se destacarem nas probabilidades da Tabela 4.

Tabela 5 – Estimativas das prevalências segundo as capitais das unidades da federação, PNS-2019.

Capital (UF)	n	%	Classes latentes			
			Saudáveis	Indício de síndrome de Burnout	Transtorno depressivo único	Transtorno depressivo recorrente
Porto Velho (RO)	680	3,00	0,598	0,260	0,060	0,082
Rio Branco (AC)	993	4,38	0,617	0,244	0,071	0,068
Manaus (AM)	1 042	4,59	0,620	0,258	0,083	0,038
Boa Vista (RR)	882	3,89	0,500	0,366	0,080	0,054
Belém (PA)	676	2,98	0,524	0,371	0,062	0,043
Macapá (AP)	555	2,45	0,481	0,411	0,057	0,050
Palmas (TO)	434	1,91	0,596	0,254	0,096	0,054
São Luís (MA)	613	2,70	0,624	0,274	0,045	0,056
Teresina (PI)	713	3,14	0,498	0,341	0,078	0,082
Fortaleza (CE)	612	2,70	0,567	0,305	0,076	0,052
Natal (RN)	643	2,83	0,507	0,351	0,065	0,077
João Pessoa (PB)	767	3,38	0,559	0,279	0,083	0,079
Recife (PE)	650	2,87	0,453	0,363	0,108	0,076
Maceió (AL)	890	3,92	0,397	0,427	0,091	0,085
Aracaju (SE)	632	2,79	0,459	0,365	0,061	0,115
Salvador (BA)	878	3,87	0,407	0,410	0,101	0,082
Belo Horizonte (MG)	1 143	5,04	0,484	0,381	0,064	0,071
Vitória (ES)	710	3,13	0,604	0,285	0,053	0,057
Rio de Janeiro (RJ)	1 336	5,89	0,639	0,227	0,073	0,062
São Paulo (SP)	1 802	7,94	0,575	0,273	0,080	0,073
Curitiba (PR)	755	3,33	0,571	0,276	0,090	0,063
Florianópolis (SC)	570	2,51	0,687	0,236	0,039	0,038
Porto Alegre (RS)	783	3,45	0,516	0,333	0,099	0,052
Campo Grande (MS)	944	4,16	0,538	0,287	0,085	0,090
Cuiabá (MT)	653	2,88	0,712	0,181	0,069	0,039
Goiânia (GO)	767	3,38	0,580	0,247	0,079	0,094
Brasília (DF)	1 558	6,87	0,608	0,249	0,078	0,066

Verificamos que, na Tabela 5, entre as regiões Norte, Nordeste, Sul, Sudeste e Centro-Oeste, as capitais com o maior número de pessoas entrevistadas são Manaus (AM), Maceió (AL), São Paulo (SP), Porto Alegre (RS) e Brasília (DF), respectivamente. As prevalências da classe latente com o rótulo de "Transtorno depressivo recorrente" variam de 3,8% (Manaus - AM e Florianópolis - SC) a 11,5% (Aracaju - SE), havendo bastante heterogeneidade dentro de algumas regiões. Na região Norte, as prevalências vão de 3,8% a 8,2% sendo a capital Porto Velho (RO) a de maior e Manaus (AM) a de menor prevalência; na região Nordeste, a variação é de 5,2% (Fortaleza - CE) a 11,5% (Aracaju - SE). A região Sudeste é a mais homogênea, com variação de 5,7% (Vitória - ES) a 7,3% (São Paulo - SP); na região Sul a variação é de 3,8% (Florianópolis - SC) a 6,3% (Curitiba - PR). Por fim, a região Centro-Oeste a variação é de 3,9% (Cuiabá - MT) a 9,4% (Goiânia - GO).

Em relação à classe de transtorno depressivo único, pode-se dizer de maneira geral que suas prevalências estão seguindo o mesmo padrão da classe transtorno depressivo recorrente, com variações dentro das regiões. Foi observado que as capitais com as maiores prevalências de indícios de síndrome de Burnout foram Macapá (41,1%), Maceió (42,7%), Belo Horizonte (38,1%), Porto Alegre (33,3%) e Campo Grande (28,7%). Em relação às prevalências da classe "Saudáveis", as maiores foram encontradas em Manaus (62,0%), São Luís (62,4%), Rio de Janeiro (63,9%), Florianópolis (68,7%) e Cuiabá (71,2%) em cada uma das suas respectivas regiões.

Vale notar que tanto as estimativas das prevalências das classes latentes, quanto das probabilidades item-respostas, foram similares nos modelos, sem covariáveis e com covariáveis.

4.2.3 Modelo de classes latentes, estratificado por capital, com covariáveis

As covariáveis incluídas no modelo foram sexo, número de pessoas vivendo no domicílio, idade, cor/raça, estado civil, escolaridade e renda per capita. Inicialmente todas as variáveis entraram como categóricas, de acordo com as categorias definidas na Tabela 1, porém os resultados indicaram que tanto idade quanto o número de pessoas poderiam se beneficiar da inclusão como quantitativas já que para ambas, os efeitos se mostraram do tipo linear, ou seja, constantes na escala logarítmica da prevalência. Assim, essas duas variáveis foram centradas nas respectivas médias e incluídas como quantitativas. A centralização auxilia na interpretação dos resultados, pois coloca como referência, a subpopulação que tem idade e vive com número de pessoas iguais às respectivas médias, além de pertencerem às categorias referências das demais variáveis.

Para interpretar os efeitos das covariáveis nas prevalências das classes latentes, podemos utilizar as estimativas das razões de chances, de forma semelhante à análise de regressão logística multinomial. As estimativas das razões de chances e os intervalos de confiança de 95% para o efeito de cada covariável incluída no modelo são apresentados na Tabela 6. É importante ressaltar que a classe "Saudável", ou seja, classe 1, é considerada como referência para todas as razões de chances das outras três classes. A covariável sexo está fortemente associada à variável latente,

indicando que as mulheres têm uma maior chance de serem classificadas nas classes mais "graves". Notamos que as razões de chances aumentam com a gravidade do estado de saúde, variando de 1,870 a 3,073, quando comparado aos homens. A variável idade também se mostra associada a variável latente, embora com razões de chances próximas de um para todas as classes, em relação a classe "Saúdavel". Para a covariável estado civil destacamos as razões de chances superiores a um das categorias divorciado e solteiro em relação aos casados, indicando uma propensão de indivíduos nessas classes apresentarem maior chance de pertencerem às classes mais graves. Em relação à escolaridade, os resultados mostram evidência de menor propensão de indivíduos com nível escolar acima do fundamental serem classificados com algum tipo de problema de saúde mental, quando comparados com os indivíduos com instrução abaixo do grau fundamental. Para a covariável renda per capita percebemos uma interpretação parecida com a covariável escolaridade, ou seja, há evidências de chance menor para os indivíduos com renda maior que um salário mínimo pertencer às classes que mostram problemas de saúde mental do que para aqueles indivíduos que possuem renda inferior a um salário. Os resultados para as covariáveis número de pessoas no domicílio e cor/raça não evidenciaram relação com a variável latente. Vale ressaltar que a inclusão das covariáveis no modelo não alterou o modelo de mensuração de classes latentes anteriormente ajustado.

Tabela 6 – Estimativas da razão de chances (RC) e respectivos intervalos de confiança (95%) para o modelo com covariáveis, usando a classe "Saudáveis" como referência.

Covariáveis	Classe latente 2/1		Classe latente 3/1		Classe latente 4/1	
	RC	$IC_{95\%}(RC)$	RC	$IC_{95\%}(RC)$	RC	$IC_{95\%}(RC)$
Intercepto:	0,256	(0,180 ; 0,363)	0,055	(0,032 ; 0,092)	0,066	(0,044 ; 0,099)
Idade:	0,991	(0,987 ; 0,996)	0,987	(0,979 ; 0,994)	0,992	(0,985 ; 0,999)
Nº pessoas:	0,983	(0,934 ; 1,035)	1,026	(0,944 ; 1,116)	0,994	(0,921 ; 1,073)
Sexo:						
Homem	1		1		1	
Mulher	1,870	(1,649 ; 2,119)	2,957	(2,345 ; 3,729)	3,073	(2,437 ; 3,874)
Cor / Raça:						
Branca	1		1		1	
Parda	0,977	(0,858 ; 1,112)	0,938	(0,754 ; 1,168)	0,964	(0,754 ; 1,233)
Preta	1,170	(0,958 ; 1,429)	1,148	(0,844 ; 1,561)	1,054	(0,771 ; 1,442)
Outro	0,986	(0,681 ; 1,427)	0,694	(0,369 ; 1,306)	0,557	(0,328 ; 0,945)
Estado civil:						
Casado	1		1		1	
Divorciado	1,268	(1,045 ; 1,539)	1,427	(1,063 ; 1,917)	1,732	(1,283 ; 2,338)
Solteiro	1,202	(1,045 ; 1,383)	1,180	(0,909 ; 1,532)	1,266	(1,016 ; 1,579)
Viúvo	1,042	(0,833 ; 1,302)	0,943	(0,656 ; 1,356)	1,194	(0,850 ; 1,678)
Escolaridade ¹:						
S/Int. ou FI	1		1		1	
F ou MI	0,803	(0,662 ; 0,975)	0,610	(0,432 ; 0,861)	0,901	(0,673 ; 1,208)
M ou SI	0,861	(0,738 ; 1,004)	0,641	(0,500 ; 0,820)	0,543	(0,420 ; 0,701)
S	0,956	(0,796 ; 1,147)	0,642	(0,456 ; 0,902)	0,436	(0,300 ; 0,634)
Renda ²:						
Até 1 SM	1		1		1	
1 até 2 SM	0,819	(0,711 ; 0,942)	0,921	(0,728 ; 1,165)	0,842	(0,659 ; 1,075)
2 até 3 SM	0,844	(0,687 ; 1,036)	0,763	(0,544 ; 1,070)	0,985	(0,700 ; 1,385)
Mais de 3 SM	0,906	(0,741 ; 1,108)	0,693	(0,475 ; 1,012)	0,594	(0,359 ; 0,982)

¹ S/Int. ou FI.: Sem instrução ou fundamental incompleto; F ou MI: Fundamental completo ou médio incompleto; M ou SI: Médio completo ou superior incompleto; S: Superior completo.

² SM: Salário mínimo.

5 Conclusão

Por meio do modelo de classes latentes, o estado de saúde mental da população das capitais do Brasil foi caracterizado em quatro classes: saudáveis, início de síndrome de Burnout, transtorno depressivo único e transtorno depressivo recorrente, com base nas probabilidades item-respostas do PHQ-9 e dos conceitos relativos à área da psicologia. Houve heterogeneidade regional das prevalências, mas em todas as capitais, a classe mais prevalente foi das pessoas saudáveis, que ficaram em torno de 39,7% a 71,2%. A categoria início a síndrome de Burnout foi a segunda classe mais prevalente cujos valores variaram de 18,1% a 42,7%. Em termos das classificações apontando para depressão, a distribuição das prevalências nas últimas duas classes foi similar. Em suma, para transtorno depressivo recorrente os valores variaram de 3,8% a 11,5%. Desta análise preliminar a região Nordeste se sobressai com prevalências maiores do que nas outras regiões na classe mais grave.

Em termos de associação de covariáveis com o estado de saúde mental, mais uma vez, nossa análise evidencia associação dos fatores sexo, estado civil e escolaridade com o estado de saúde mental, também encontrados em outros trabalhos como, especificamente, [Idalino \(2019\)](#) e [Munhoz et al. \(2016\)](#) que analisaram os dados da PNS-2013.

5.1 Trabalhos futuros

Vários aspectos da modelagem de classes latentes apresentada merecem investigações e explorações mais aprofundadas e, em particular, destacamos os seguintes pontos:

- Avaliar os benefícios da utilização de agrupamento de categorias similares de itens do PHQ-9, que resultaram em frequências baixas na amostra;
- Avaliar se a suposição de mesmo número de classes latentes para todas as capitais se justifica;
- Relaxar a restrição de probabilidades item-respostas homogêneas nas capitais;
- Investigar o efeito do delineamento amostral complexo nas estimativas dos parâmetros do modelo, comparando os resultados com um ajuste sem levar em conta tal delineamento, ou seja, assumindo uma amostragem aleatória simples.

Referências

- AGRESTI, A. *Categorical data analysis*. [S.l.]: John Wiley & Sons, 2012. v. 792. 10
- AMORIM, L. D. A. F. et al. Análise de classes latentes: um tutorial usando software estatístico. Brasil, 2015. 12
- APA - American Psychiatric Association. *DSM-5: Manual diagnóstico e estatístico de transtornos mentais*. [S.l.]: Artmed Editora, 2014. 1
- ASPAROUHOV, T.; MUTHÉN, B. Using mplus tech11 and tech14 to test the number of latent classes. *Mplus web notes*, v. 14, n. 22, p. 1–17, 2012. 12
- BANDEEN-ROCHE, K. et al. Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, Taylor & Francis, v. 92, n. 440, p. 1375–1386, 1997. 11
- BARROSO, S. M.; MELO, A. P. S.; GUIMARÃES, M. D. C. Depressão em comunidades quilombolas no brasil: triagem e fatores associados. *Revista Panamericana de Salud Pública, SciELO Public Health*, v. 35, p. 256–263, 2014. 2, 19
- BEAUJEAN, A. A. *Latent variable modeling using R: A step-by-step guide*. [S.l.]: Routledge, 2014. 4
- CARMO, M. B. Barreto do et al. Screening for common mental disorders using the srq-20 in brazil: what are the alternative strategies for analysis? *Brazilian Journal of Psychiatry, SciELO Brasil*, v. 40, p. 115–122, 2017. 11
- CHEN, H. et al. Prevalence, risk factors and multi-group latent class analysis of lifetime anxiety disorders comorbid depressive symptoms. *Journal of Affective Disorders, Elsevier*, v. 243, p. 360–365, 2019. 11
- COLLINS, L. M. et al. Winlta user's guide. *State College, PA: Pennsylvania State University, University Park Campus, Methodology Center*, 2002. 12
- COSTA, J. A. et al. Síndrome de Burnout: uma análise da saúde mental dos residentes médicos de um Hospital Escola. *Revista Brasileira de Educação Médica*, v. 46, n. 1, p. 1–10, 2022. ISSN 0100-5502. 2
- DAYTON, C. M.; MACREADY, G. B. Concomitant-variable latent-class models. *Journal of the american statistical association*, Taylor & Francis, v. 83, n. 401, p. 173–178, 1988. 11
- DEMPSTER, A. P. Maximum likelihood from incomplete data via em algorithm. *J. Royal Stat. Soc. B.*, v. 39, n. 1, p. 1–38, 1977. 7
- ESSAU, C. A.; TORRE-LUQUE, A. de la. Comorbidity profile of mental disorders among adolescents: A latent class analysis. *Psychiatry research, Elsevier*, v. 278, p. 228–234, 2019. 11
- FIOCRUZ - Fundação Instituto Oswaldo Cruz. *PNS – Pesquisa Nacional de Saúde*. 2020. Disponível em: <<https://www.pns.icict.fiocruz.br/>>. 13

- GAUDENCI, E. M. et al. Qualidade de vida, sintomas depressivos e incapacidade física de pacientes com hanseníase. *Hansenologia Internationalis: hanseníase e outras doenças infecciosas*, v. 40, n. 2, p. 48–58, 2015. 19
- GOODMAN, L. A. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, Oxford University Press, v. 61, n. 2, p. 215–231, 1974. 6
- IBGE - Instituto Brasileiro de Geografia e Estatística. *Pesquisa Nacional de Saúde 2019: Atenção primária à saúde e informações antropométricas*. 2020. Disponível em: <<https://abeso.org.br/wp-content/uploads/2021/07/Pesquisa-Nacional-de-Saude-2019.pdf>>. 14
- IBGE - Instituto Brasileiro de Geografia e Estatística. *PNS - Pesquisa Nacional de Saúde*. 2020. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/saude/9160-pesquisa-nacional-de-saude.html?=&t=o-que-e>>. 13
- IDALINO, R. C. L. Estimacão da Prevalência da Depressão via o Modelo de Classes Latentes Utilizando a Amostra da Pesquisa Nacional de Saúde, 2013. 2019. 11, 28
- J.PUC-SP. *Síndrome de Burnout já é classificada como doença ocupacional*. 2022. <<https://j.pucsp.br/noticia/sindrome-de-burnout-ja-e-classificada-como-doenca-ocupacional>>. Acesso em 15 de março de 2023. 2
- KROENKE, K.; SPITZER, R. L.; WILLIAMS, J. B. The phq-9: validity of a brief depression severity measure. *Journal of general internal medicine*, Wiley Online Library, v. 16, n. 9, p. 606–613, 2001. 13
- LANZA, S. T. et al. PROC LCA: A SAS procedure for latent class analysis. *Structural Equation Modeling*, Routledge, v. 14, n. 4, p. 671–694, 2007. ISSN 10705511. 8
- LANZA, S. T. et al. Proc lca & proc lta users' guide (version 1.3. 2). *University Park: The Methodology Center, Penn State*, 2015. 12
- LANZA, S. T.; FLAHERTY, B. P.; COLLINS, L. M. Latent class and latent transition analysis. *Handbook of psychology*, Wiley Online Library, p. 663–685, 2003. 4, 5, 6, 20
- LAZARSFELD, P.; HENRY, N. Latent structure analysis. *Houghton Mifflin Company*, p. 294, 1968. 6
- LINZER, D. A.; LEWIS, J. B. polca: An r package for polytomous variable latent class analysis. *Journal of statistical software*, v. 42, p. 1–29, 2011. 12
- LUMLEY, T. *survey: analysis of complex survey samples*. 2023. R package version 4.2. 16
- MASTELLA, J. O. Análise de classes latentes: da teoria à prática. 2015. 5
- MEYER, D. et al. *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien. R package version 1.6-3*. 2014. 12
- MUNHOZ, T. N. et al. A nationwide population-based study of depression in brazil. *Journal of Affective Disorders*, Elsevier, v. 192, p. 226–233, 2016. 28
- MUTHÉN, L. Mplus user's guide . muthén & muthén. *Los Angeles, CA*, v. 2012, 1998. 12
- Pan American Health Organization. *Saúde mental dos adolescentes*. 2022. <<https://www.paho.org/pt/topicos/saude-mental-dos-adolescentes>>. Acesso em 30 de agosto de 2023. 2

Pan American Health Organization. *Depressão*. 2023. <<https://www.paho.org/pt/topicos/depressao>>. Acesso em 15 de março de 2023. 2

PATTERSON, B. H. et al. Latent Class Analysis of Complex Sample Survey Data: Application to Dietary Data. v. 97, n. 459, p. 721–729, 2002. 8, 9

PROGRESS IN MIND - BRAZIL PSYCHIATRY & NEUROLOGY RESOURCE CENTER. *Otimização de resultados no transtorno depressivo maior*. 2022. <<https://brazil.progress.im/pt-br/content/otimiza%C3%A7%C3%A3o-de-resultados-no-transtorno-depressivo-maior>>. Acesso em 30 de agosto de 2023. 2

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2023. Disponível em: <<https://www.R-project.org/>>. 12, 16

SANTOS, L. M. et al. Measuring the level of social support using latent class analysis. *Social science research*, Elsevier, v. 50, p. 139–146, 2015. 11

SCHAFER, J. L.; KANG, J. *LCCA package for R users' guide (Version 1.1.0)*. The Methodology Center, Penn State. [S.l.], 2013. Disponível em: <<http://stat.psu.edu/research/technical-reports/2010-technical-reports/TR10-05.pdf>>. 9, 12

Silva Junior, C. A. d. *Epidemiologia dos subtipos de depressão: análise de classes latentes dos sintomas depressivos em uma amostra populacional da região Metropolitana de São Paulo*. Tese (Doutorado) — Universidade de São Paulo, 2012. 11

STOPA, S. R. Pesquisa Nacional de Saúde 2019: histórico, métodos e perspectivas. *Epidemiol. Serv. Saude, Brasília*, v. 29, n. 5, p. 2020315, 2020. 14

SZWARCWALD, C. L. et al. Pesquisa nacional de saúde no brasil: concepção e metodologia de aplicação. *Ciência & Saúde Coletiva*, SciELO Public Health, v. 19, p. 333–342, 2014. 13

The Methodology Center. *lcca: Latent-class causal analysis*. [S.l.], 2013. R package version 2.0.0. 12, 15

VERMUNT, J. K. Latent class modeling with covariates: Two improved three-step approaches. *Political analysis*, Cambridge University Press, v. 18, n. 4, p. 450–469, 2010. 11

VIEIRA, M. d. T.; CLARO, R. M.; ASSUNÇÃO, A. Á. Desenho da amostra e participação no estudo educatel. *Cadernos de Saúde Pública*, SciELO Public Health, v. 35, p. e00167217, 2019. 16

WHO - World Health Organization. *Mental health*. 2023. <<https://www.who.int/health-topics/mental-health#>>. Acesso em 15 de março de 2023. 1

World Health Organization. *International classification of diseases for mortality and morbidity statistics (11th Revision)*. 2018. 2

Apêndices

A Dicionário de variáveis - PNS 2019

Aqui se encontra apenas a descrição das variáveis escolhidas para estudo deste trabalho.

Perguntas relacionadas ao PHQ-9

1. 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?
2. 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?
3. Nas duas últimas semanas, com que frequência o(a) Sr(a) teve pouco interesse ou não sentiu prazer em fazer as coisas?
4. Nas duas últimas semanas, com que frequência o(a) Sr(a) teve problemas para se concentrar nas suas atividades habituais?
5. 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?
6. 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)?
7. Nas duas últimas semanas, com que frequência o(a) Sr(a) se sentiu deprimido(a), “pra baixo” ou sem perspectiva?
8. 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?
9. 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?

Alternativas para cada pergunta do questionário do ano de 2019 (segunda versão) em estudo são:

- Nenhum dia;
- Menos da metade dos dias;
- Mais da metade dos dias;
- Quase todos dias.