

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

**Um perfil do comportamento e características dos
indivíduos frente a COVID-19 a partir da PNAD
COVID19**

Samantha Navarro Janine

Trabalho de Conclusão de Curso

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

Um perfil do comportamento e características dos indivíduos
frente a COVID-19 a partir da PNAD COVID19

Samantha Navarro Janine

Orientador(a): Pedro Ferreira Filho

Trabalho de Conclusão de Curso apresentado
como parte dos requisitos para obtenção do
título de Bacharel em Estatística.

São Carlos

Setembro de 2022

FEDERAL UNIVERSITY OF SÃO CARLOS
EXACT AND TECHNOLOGY SCIENCES CENTER
DEPARTMENT OF STATISTICS

The behavioral profile and characteristics of individuals due to
Covid-19 from PNAD COVID19 survey

Samantha Navarro Janine

Advisor: Pedro Ferreira Filho

Bachelors dissertation submitted to the Department of Statistics, Federal University of São Carlos - DEs-UFSCar, in partial fulfillment of the requirements for the degree of Bachelor in Statistics.

São Carlos
September 2022

Samantha Navarro Janine

Um perfil do comportamento e características dos indivíduos frente a Covid-19 a partir da PNAD COVID19

Este exemplar corresponde à redação final do trabalho de conclusão de curso devidamente corrigido e defendido por Samantha Navarro Janine e aprovado pela banca examinadora.

Aprovado em 09 de setembro de 2022

Banca Examinadora:

- Prof. Pedro Ferreira Filho
- Dr. Eduardo Schneider de Oliveira
- Prof. Dr. Márcio Luis Lanfredi Viola

*Aos meus pais, amigos e familiares por todo apoio.
Em solidariedade a todas as famílias vítimas da COVID-19.*

Agradecimentos

Ao meu orientador Professor Pedro Ferreira Filho por todo apoio e auxílio na execução desse trabalho.

Aos membros da banca Dr. Eduardo Schneider de Oliveira e Professor Dr. Márcio Luis Lanfredi Viola, pelas sugestões e correções.

À minha mãe Claudia por ser minha maior incentivadora.

A todos que diretamente ou indiretamente me apoiaram durante o desenvolvimento deste trabalho.

“Não se conhece completamente uma ciência enquanto não se souber da sua história.”

(Auguste Comte)

Resumo

A Pesquisa Nacional por Amostra de Domicílios (PNAD), realizada pelo IBGE, consiste em avaliar características gerais da população brasileira. Com o surgimento da COVID-19 em março de 2020, as condições impostas pela pandemia exigiram do IBGE modificações nos procedimentos de realização da PNAD. Dentre as alterações implementadas, merece destaque a incorporação de questões relacionadas à COVID-19, surgindo, dessa forma, a denominada PNAD COVID19.

É importante lembrar que no período inicial da pandemia poucos estudos foram realizados na perspectiva de avaliar o comportamento dos indivíduos frente a esse contexto. Neste sentido, faz-se necessário destacar [Barbosa e Alves \(2021\)](#) que abordam a importância do conhecimento de estatísticas cívicas, de interesse público, para a formação de cidadãos conscientes da sua realidade.

Diante desse quadro, este trabalho procura realizar um estudo na perspectiva de caracterizar os indivíduos que apresentaram pelo menos um sintoma no contexto da disseminação do coronavírus a partir das informações coletadas pela pesquisa PNAD COVID19 entre maio e novembro de 2020, período inicial da pandemia. Utilizando procedimentos de análise descritiva e exploratória de dados, assim como de análise fatorial de correspondência, foram relacionados os sintomas por COVID-19, dos indivíduos entrevistados, com variáveis a respeito de seus comportamentos diante desse cenário.

Os resultados indicaram um padrão para os três meses estudados (maio, agosto e novembro), tanto na análise descritiva quanto na análise fatorial de correspondência, na qual a maioria dos indivíduos apresentou até três sintomas conjuntos. Além disso, foram identificadas algumas relações entre a ocorrência e o número de determinados sintomas, em particular, a associação entre tosse, dor de garganta, febre e dor no peito, sendo esses alguns dos sintomas conjugados pelo IBGE mais suscetíveis à presença do vírus da COVID-19.

Palavras-chave: *PNAD, COVID-19, IBGE, sintomas, Análise Fatorial de Correspondência.*

Abstract

The National Household Sample Survey (PNAD), which is conducted by IBGE, consists of general characteristics of the population, education, labor, income and housing. The COVID-19 pandemic, started in March 2020, imposed some conditions that made IBGE modify the procedures for carrying out the PNAD. Among the changes implemented, it is worth mentioning the incorporation of questions related to COVID-19, thus giving rise to PNAD COVID19.

It is important to remember that in the beginning of the pandemic, few studies were developed to evaluate the behavior of the people in this period. In that regard, it is necessary to stand out [Barbosa e Alves \(2021\)](#), who approach the importance of knowing civic statistics to the formation of citizens aware of their reality.

Due to this fact, this study aims to characterize the individuals who had at least one symptom during the spread of the coronavirus from information collected by PNAD COVID19 between may and november 2020, initial period of the pandemic. Exploratory descriptive analysis and factorial correspondence analysis were conducted to relate the symptoms of COVID-19 from the people interviewed with variables regarding their behavior in this scenario.

Results indicated a pattern between the three months studied (march, august and november), both for exploratory descriptive analysis and factorial correspondence analysis, in which the majority of the individuals had up to three symptoms simultaneously. Furthermore, some relationships between the occurrence and the number of symptoms were identified, in particular, the association between cough, sore throat, fever and chest pain, these being some of the symptoms conjugated by IBGE most susceptible to the presence of the COVID-19 virus.

Keywords: *PNAD, COVID-19, IBGE, symptoms, Factorial Correspondence Analysis.*

Lista de Figuras

3.1	Distribuição da proporção de indivíduos de acordo com a quantidade de sintomas.	40
3.2	Proporção de indivíduos que buscaram um estabelecimento de saúde de acordo com a quantidade de sintomas.	41
3.3	Proporção de indivíduos internados de acordo com a quantidade de sintomas.	45
3.4	Proporção de indivíduos intubados de acordo com a quantidade de sintomas.	46
3.5	Proporção de testes feitos para COVID-19 de acordo com a quantidade de sintomas.	48
3.6	Proporção de resultados positivos de acordo com a quantidade de sintomas.	50
3.7	Primeiro plano fatorial para o mês de maio.	54
3.8	Primeiro plano fatorial para o mês de agosto.	54
3.9	Primeiro plano fatorial para o mês de novembro.	54
3.10	Gráfico das contribuições das categorias do mês de maio para o primeiro plano fatorial.	55
3.11	Gráfico das contribuições das categorias do mês de agosto para o primeiro plano fatorial.	55
3.12	Gráfico das contribuições das categorias do mês de novembro para o primeiro plano fatorial.	56
3.13	Primeiro plano fatorial para o mês de agosto.	58
3.14	Primeiro plano fatorial para o mês de novembro.	58

Lista de Tabelas

2.1	Representação de uma tabela de contingência.	30
3.1	Tamanho amostral e número de indivíduos selecionados.	35
3.2	Proporção de indivíduos por faixa etária nos meses de estudo.	36
3.3	Proporção de indivíduos de acordo com o sexo nos meses de estudo.	36
3.4	Proporção de indivíduos de acordo com a raça nos meses de estudo.	36
3.5	Proporção de indivíduos de acordo com grau de escolaridade nos meses de estudo.	37
3.6	Proporção de indivíduos de acordo com a região nos meses de estudo.	37
3.7	Proporção de indivíduos que apresentaram cada um dos sintomas nos meses de estudo sem utilizar o <i>survey</i>	38
3.8	Proporção de indivíduos que apresentaram cada um dos sintomas nos meses de estudo utilizando <i>survey</i>	38
3.9	Proporção de indivíduos que apresentaram cada possível número de sintomas.	39
3.10	Proporção de indivíduos que procuraram estabelecimentos de saúde segundo número de sintomas.	40
3.11	Proporção de indivíduos que buscaram um estabelecimento de saúde nos meses de estudo.	41
3.12	Proporção de indivíduos que buscaram atendimentos segundo tipo de local.	42
3.13	Proporção de indivíduos (em %) que buscaram um estabelecimento de saúde segundo possuir ou não um plano de saúde privado.	42
3.14	Proporção de tipos de locais (em %) em que se buscou atendimento segundo possuir ou não plano de saúde privado.	43
3.15	Proporção de indivíduos por internação nos meses de estudo.	43
3.16	Proporção de indivíduos por intubação nos meses de estudo.	44
3.17	Proporção dos tipos de providências adotadas nos meses de estudo.	44

3.18	Proporção de indivíduos (em %) que passaram por internação de acordo com a quantidade de sintomas.	44
3.19	Proporção de indivíduos (em %) que passaram por intubação de acordo com a quantidade de sintomas.	46
3.20	Proporção de indivíduos que fizeram teste de COVID-19 de acordo com a quantidade de sintomas.	47
3.21	Proporção de testes de COVID-19 realizados.	48
3.22	Proporção de resultados pelo tipo de teste de COVID-19 realizado.	49
3.23	Proporção de resultados positivos de acordo com a quantidade de sintomas.	49
3.24	Proporção de indivíduos de acordo com a restrição de contatos.	50
3.25	Proporção das restrições de contato para indivíduos com resultados de teste da COVID-19 positivos.	51
3.26	Proporção de indivíduos com até 3 sintomas de acordo com cada sintoma apresentado nos meses de estudo.	51
3.27	Codificação das categorias apresentadas na representação gráfica.	52
3.28	Características de 7 observações do mês de maio com 4 ou mais sintomas.	56
3.29	Codificação das categorias apresentadas na representação gráfica.	57
A.1	Dicionário dos dados utilizados.	65
B.1	Coordenadas das categorias para as dimensões de acordo com os meses.	72
B.2	Contribuições das categorias para as dimensões de acordo com os meses.	73

Sumário

1	Introdução	23
2	Material e Métodos	27
2.1	Planejamento Amostral	28
2.2	Questionário	29
2.3	Análise de Correspondência	30
3	Resultados	35
3.1	Análise Descritiva	35
3.1.1	Questões incorporadas a partir do mês de agosto	46
3.2	Análise de Correspondência	52
3.2.1	Questões incorporadas a partir do mês de agosto	57
4	Conclusão	59
	Referências Bibliográficas	62
A	Dicionário	65
B	Coordenadas e contribuições da Análise de Correspondência	71
C	Códigos utilizados	75

Capítulo 1

Introdução

A Pesquisa Nacional por Amostra de Domicílios (PNAD), realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), consiste em identificar características gerais da população, tais como educação, trabalho, renda e habitação, a qual tem o domicílio como unidade de investigação.

A partir de 2016, a pesquisa foi substituída pela PNAD Contínua, a qual avalia, trimestralmente, as características de interesse com uma cobertura territorial mais abrangente por meio da construção de indicadores referentes aos temas pertinentes para informação do Brasil (IBGE, 2020a). Com o surgimento da COVID-19, um dos assuntos de interesse a ser estudado foram os sintomas relacionados à doença.

De acordo com o Ministério da Saúde (2020), a COVID-19 é uma infecção respiratória aguda causada pelo coronavírus SARS-CoV-2, potencialmente grave, de elevada transmissibilidade e de distribuição global. Os sintomas mais comuns relacionados à doença são febre, tosse, dificuldade para respirar, dores no corpo, perda do olfato e paladar (Lapostolle *et al.*, 2020).

Os primeiros casos foram registrados na cidade de Wuhan, na China, cujas vítimas tinham frequentado um mercado na cidade conhecido pela comercialização de animais silvestres como alimentos. Esses foram relatados pela Organização Mundial da Saúde (OMS) como casos de pneumonia de causa desconhecida. Segundo Gruber (2020), apesar do contato em comum com o mercado, não é possível afirmar que exista uma relação com a origem epidemiológica. Muitos cientistas questionam a origem do vírus SARS-CoV-2 e supõem que esteja relacionada a uma espécie de morcegos, com um hospedeiro intermediário, o qual transmite ao homem.

Logo, a lista de países com casos suspeitos ao redor do mundo se estendeu em poucos

dias. Em fevereiro de 2020 foi confirmado o primeiro caso de um indivíduo infectado por coronavírus no Brasil (Sanar Saúde, 2020).

Com a declaração da pandemia pela OMS em março de 2020 e o aumento do número de casos de COVID-19 no Brasil, o IBGE adaptou os procedimentos de execução da PNAD Contínua. Para dar continuidade à pesquisa, foi desenvolvida a PNAD COVID19, com o objetivo adicional de estimar a quantidade de pessoas com sintomas relacionados à doença, além de avaliar os impactos da pandemia no mercado de trabalho e na renda da população brasileira.

Segundo o [IBGE](#), para a realização da pesquisa, foi utilizada como referência a amostra de domicílios da Pesquisa Nacional por Amostra de Domicílios Contínua – PNAD Contínua do 1º Trimestre de 2019. O plano amostral da PNAD Contínua é dado pela seleção de amostras por conglomerados em dois estágios com estratificação das unidades primárias de amostragem.

Os dados foram coletados entre maio e novembro de 2020, em entrevistas realizadas por telefone, por meio de uma amostra fixa, isto é, os mesmos domicílios foram entrevistados ao longo dos meses da pesquisa, totalizando cerca de 193 mil domicílios. O questionário foi dividido em três partes: a primeira para avaliar as condições de saúde, os sintomas associados à COVID-19 e as respectivas providências relacionadas à situação; a segunda parte é referente às condições de trabalho e a terceira sobre rendimento e possíveis auxílios de outras fontes.

[Barbosa e Alves \(2021\)](#) ressaltam a importância do conhecimento de estatísticas cívicas juntamente ao aprendizado das metodologias estatísticas. Dado o contexto da pandemia da COVID-19 e os constantes estudos apresentados pela mídia com o desenvolvimento de modelos estatísticos, como pesquisas por amostra para estimar a prevalência de infecções, ou estudos clínicos para medir a eficácia das vacinas, os autores reforçam a importância do letramento estatístico para a formação de cidadãos conscientes.

A partir da motivação dada pelo trabalho realizado por [Barbosa e Alves \(2021\)](#), considerando o cenário inicial da pandemia e a busca por informações relacionadas aos respectivos sintomas, deseja-se estudar as informações presentes no questionário da PNAD COVID19. Ainda, pretende-se identificar possíveis características e relações entre as variáveis no período de estudo, com o objetivo de compreender o comportamento, atitudes e procedimentos diante da necessidade de conviver com a perspectiva de contrair ou estar sujeito a contrair o vírus da COVID-19.

Este trabalho está organizado de forma que no Capítulo 2 são apresentados os materiais e métodos a serem utilizados, no Capítulo 3 os resultados das diferentes análises realizadas e, finalmente, no Capítulo 4 são apresentadas as conclusões deste estudo.

Capítulo 2

Material e Métodos

A partir do surgimento da pandemia de COVID-19, houve um grande esforço de pesquisadores, das mais diferentes áreas do conhecimento, em desenvolver estudos para melhor entender e conhecer os seus desdobramentos e consequências do ponto de vista da saúde e bem estar social. Por parte dos profissionais e pesquisadores em Estatística, a maioria dos trabalhos desenvolvidos no início da pandemia buscava construir modelos na perspectiva de prever o crescimento da COVID-19 ao longo do tempo, incluindo o número de casos e de mortes. Por outro lado, poucos estudos foram desenvolvidos com o objetivo de avaliar o comportamento dos indivíduos frente à realidade de contrair ou vir a contrair o vírus da COVID-19.

Nesse sentido, motivados pelo trabalho de [Barbosa e Alves \(2021\)](#), o desafio a ser superado é o de traçar um perfil do comportamento dos indivíduos diante da necessidade de convivência com a COVID-19, a partir dos dados obtidos pela PNAD COVID19, cujos resultados estão disponíveis para livre acesso na página do [IBGE](#). Inicialmente, foi realizada a identificação, nas bases de dados dos meses de estudo (maio, agosto e novembro), das informações que possam contribuir para a caracterização de um perfil do comportamento dos indivíduos que apresentaram sintomas de Covid nesses períodos. Com o uso de métodos descritivos univariados e multivariados, pretende-se identificar um perfil dos indivíduos que apresentaram pelo menos um dos sintomas apontados na pesquisa. Para este conjunto de indivíduos, deseja-se relacionar o número de sintomas apresentados com variáveis que identifiquem o comportamento dos mesmos frente a este diagnóstico, bem como com as demais características presentes na PNAD COVID19 (educação, trabalho, condições de saúde, entre outros). Além disso, verificar se tais relações entre as variáveis se mantêm nos três meses avaliados.

2.1 Planejamento Amostral

Segundo o [IBGE](#), para realizar a pesquisa PNAD COVID19 foi utilizada a amostra de domicílios da PNAD Contínua do 1º Trimestre de 2019. A amostra original da PNAD Contínua foi obtida a partir de um plano amostral conglomerado em dois estágios com estratificação das unidades primárias de amostragem (UPAs). No primeiro estágio foram selecionadas UPAs com probabilidade proporcional ao número de domicílios dentro de cada estrato definido. No segundo estágio foram selecionados 14 domicílios particulares permanentes (que podem estar ou não ocupados) dentro de cada UPA da amostra do primeiro estágio. O sorteio dos domicílios em cada UPA foi feito por amostragem aleatória simples, considerando os endereços listados no Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE) atualizado para cada UPA ([IBGE, 2020c](#)).

Dessa forma, a amostra de domicílios tomada como base consistiu em 211.000 no total. Como a coleta de dados para a PNAD COVID19 foi feita por meio de telefone, foi realizado um processo de pareamento com outras bases de dados para obter o número de telefone de cada domicílio. Ao final desse processo, o número de domicílios, com pelo menos um telefone disponível, foi de 193.662, representando cerca de 92% da amostra base.

Considerando um dos objetivos de produzir estimativas semanais da ocorrência de sintomas associados à COVID-19 e as respectivas providências tomadas em casos de sintomas, a amostra total da pesquisa foi dividida em 4 semanas. Ainda, a amostra foi considerada como fixa, isto é, todos os domicílios entrevistados no mês inicial permaneceram na amostra nos meses posteriores.

Estabelecido o planejamento amostral, [Barbosa e Alves \(2021\)](#) ressaltam a importância de incorporar o desenho amostral nas estimativas calculadas a partir de pacotes estatísticos, em particular, a biblioteca *survey* do software R como uma ferramenta para incorporação do plano amostral.

Ainda de acordo com [Barbosa e Alves \(2021\)](#), o autor da biblioteca, Thomas Lumley, afirma que ela permite aliar as informações dos metadados do desenho amostral à base de dados em um objeto que integra os ajustes necessários para a análise de dados correta. Assim, possibilita a extração de subconjuntos do banco de dados e preserva os metadados por meio de subgrupos do objeto criado, garantindo estimativas apropriadas para as subpopulações, além de fornecer as respectivas estimativas da variância para as estatísticas

calculadas.

De acordo com [Lumley \(2004\)](#), inicialmente, deve-se definir o objeto de estrutura do *survey* contendo os dados e metadados para que seja possível obter as estimativas apropriadas. Como argumento da função *svydesign*, deve ser definido o *ids* que representa a variável de identificação de cada unidade da amostra. Neste caso, são as unidades primárias de amostragem (UPAs). Além disso, é possível definir os estratos os quais cada indivíduo representa, bem como os pesos atribuídos à amostragem.

Após definido o objeto, diferentes funções estatísticas podem ser aplicadas, incorporando o plano amostral. Como, por exemplo, tabelas de frequência, medidas resumo e ajuste de modelos. Entretanto, algumas metodologias não estão implementadas no pacote como, por exemplo, a Análise de Correspondência, a qual é utilizada neste trabalho. Dessa forma, para estes casos, foram considerados os métodos e pacotes usuais para a realização das análises.

2.2 Questionário

O questionário da pesquisa, elaborado pelo [IBGE](#), é dividido em três partes, sendo uma direcionada a questões dos sintomas associados à COVID-19, a segunda, a questões de trabalho e, a última, para questões de rendimento de outras fontes.

Nas questões de saúde, investiga-se a ocorrência de alguns dos principais sintomas da COVID-19 no período de referência, considerando todos os moradores do domicílio. Para aqueles que apresentaram algum sintoma, foi perguntado quais as providências tomadas para alívio dos mesmos; se buscaram por atendimento médico devido a esse fato; e o tipo de estabelecimento de saúde procurado ([IBGE, 2020d](#)). Além disso, outros aspectos foram questionados como, por exemplo, se o indivíduo possui plano de saúde; se foi internado e, conseqüentemente, colocado em respiração artificial. Vale ressaltar que questões a respeito da realização de teste para detectar a presença do vírus, bem como o resultado do mesmo, foram incluídas no questionário aplicado a partir do mês de agosto.

A [Tabela A.1](#), disponível no [Apêndice A](#), apresenta as questões selecionadas que fizeram parte deste estudo, na forma original do banco de dados, bem como o modo como foram organizadas/codificadas para serem utilizadas ao longo deste trabalho.

2.3 Análise de Correspondência

Segundo [Johnson *et al.* \(2014\)](#), a Análise de Correspondência é um procedimento utilizado para representar graficamente associações de uma tabela de contingência. Quando se trata de uma única tabela de contingência (duas variáveis em estudo), utiliza-se a Análise de Correspondência Simples. Nos casos que envolvem múltiplas tabelas, ou análise conjunta de três ou mais variáveis, a Análise Fatorial de Correspondência Múltipla é utilizada.

Para o estudo da associação entre duas variáveis (X e Y), os dados são organizados em uma tabela de contingência $I \times J$, isto é com I linhas e J colunas, retratado na [Tabela 2.1](#), em que n_{ij} representa o número de elementos pertencentes à categoria i da variável X e à categoria j da variável Y ([Mingoti, 2005](#)).

Tabela 2.1: Representação de uma tabela de contingência.

Variável X	Variável Y				Total
	1	2	...	J	
1	n_{11}	n_{12}	...	n_{1J}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2J}	$n_{2.}$
⋮	⋮	⋮	...	⋮	⋮
I	n_{I1}	n_{I2}	...	n_{IJ}	$n_{I.}$
Total	$n_{.1}$	$n_{.2}$...	$n_{.J}$	$n_{..} = N$

A Análise de Correspondência Simples consiste em estudar a existência de associação entre duas variáveis categóricas e suas respectivas categorias. Para verificar se a associação é estatisticamente significativa, utiliza-se o teste qui-quadrado, na qual a estatística χ^2 é dada pela razão entre os resíduos (diferença entre as frequências observadas e esperadas da tabela de contingência) ao quadrado e a frequência esperada.

O desenvolvimento algébrico da análise de correspondência parte, inicialmente, da construção de uma matriz de proporções (ou matriz de correspondência) $\mathbf{P} = \{p_{ij}\}$, tal que $p_{ij} = \frac{n_{ij}}{N}$. Na sequência, são definidas as matrizes de perfil linha (\mathbf{D}_r) e coluna (\mathbf{D}_c), as quais são matrizes diagonais, cujos elementos são dados por $\mathbf{r}' = \left(\frac{n_{.1}}{N}, \frac{n_{.2}}{N}, \dots, \frac{n_{.J}}{N}\right)$ e $\mathbf{c}' = \left(\frac{n_{1.}}{N}, \frac{n_{2.}}{N}, \dots, \frac{n_{I.}}{N}\right)$, respectivamente.

Logo, considerando a matriz $\tilde{\mathbf{P}} = \mathbf{P} - \mathbf{r}\mathbf{c}'$, pode-se decompô-la em autovalores e autovetores, isto é, em valores singulares ([Mingoti, 2005](#)), conforme Expressão (2.1).

$$\tilde{\mathbf{P}}_{I \times J} = \mathbf{A}\mathbf{\Lambda}\mathbf{B}', \quad (2.1)$$

em que $\mathbf{A} = \mathbf{D}_r^{1/2}\mathbf{U}_{I \times k}$ e $\mathbf{B} = \mathbf{D}_c^{1/2}\mathbf{V}_{J \times k}$. \mathbf{U} e \mathbf{V} são matrizes ortogonais, contendo os

autovetores da matriz $\tilde{\mathbf{P}}\tilde{\mathbf{P}}'$ e $\tilde{\mathbf{P}}'\tilde{\mathbf{P}}$, respectivamente. $\mathbf{\Lambda}$ é uma matriz de dimensão $k \times k$, a qual contém os autovalores da matriz $\tilde{\mathbf{P}}$ ordenados de forma decrescente.

As coordenadas principais das linhas e colunas da matriz $\tilde{\mathbf{P}}$ são definidas conforme as Equações (2.2) e (2.3), respectivamente.

$$\mathbf{Y}_{I \times k} = \mathbf{D}_r^{-1} \mathbf{A}_{I \times k} \mathbf{\Lambda}_{k \times k}, \quad (2.2)$$

$$\mathbf{Z}_{J \times k} = \mathbf{D}_c^{-1} \mathbf{B}_{J \times k} \mathbf{\Lambda}_{k \times k}. \quad (2.3)$$

Logo, a matriz $\tilde{\mathbf{P}}$ pode ser escrita em função dos autovalores e das coordenadas principais, conforme Expressão (2.4).

$$\tilde{\mathbf{P}} = \mathbf{P} - \mathbf{r}\mathbf{c}' = \sum_{i=1}^k \hat{\lambda}_i \tilde{a}_i \tilde{b}_i', \quad (2.4)$$

em que \tilde{a}_i e \tilde{b}_i representam a i -ésima coluna da matriz \mathbf{A} e \mathbf{B} , respectivamente, $k = \text{posto}(\tilde{\mathbf{P}}) = \min(I - 1, J - 1)$.

A partir das coordenadas principais obtidas, é possível representá-las em um gráfico denominado mapa perceptual, o qual permite avaliar a associação entre as linhas e colunas da tabela, isto é, identificar possíveis relações entre as categorias das variáveis. Para identificar como determinada linha ou coluna contribui para a construção de cada eixo do mapa perceptual, utilizam-se os componentes da variação total existente no sistema, denominada de inércia total, a qual é expressa na Equação (2.5).

$$\sum_{i=1}^k \lambda_i^2, \quad (2.5)$$

em que λ_i são os autovalores não-nulos da diagonal da matriz $\mathbf{\Lambda}$, $i = 1, 2, \dots, k$. Ainda, a Equação (2.5) pode ser escrita em termos da estatística qui-quadrado, isto é, $\sum_{i=1}^k \lambda_i^2 = \frac{\chi^2}{N}$.

A proporção da explicação da i -ésima coordenada principal em relação à inércia total é apresentada na Equação (2.6).

$$\frac{\lambda_i^2}{\sum_{i=1}^k \lambda_i^2} \quad (2.6)$$

Dado que os autovalores são definidos de forma decrescente, as primeiras coordenadas principais apresentam a maior parte da inércia total definida. Dessa forma, o mapa perceptual das primeiras coordenadas principais identifica, nos casos em que o qui-quadrado é significativo, as categorias das variáveis em estudo que apresentam os maiores desvios da hipótese de não associação entre as variáveis e, conseqüentemente, que ocorrem simultaneamente acima do esperado.

A Análise Fatorial de Correspondência Múltipla é uma extensão da Análise de Correspondência Simples, na qual estuda-se a associação conjunta de mais de duas variáveis categóricas. Nesta técnica não é possível utilizar a estatística χ^2 para verificar a existência da associação de três ou mais categorias simultaneamente, pois tal estatística é calculada com base em uma tabela de contingência bidimensional (Fávero e Belfiore, 2017). Contudo, é possível calcular os autovalores referentes às categorias das variáveis a fim de definir suas coordenadas para serem representadas no mapa perceptual e, conseqüentemente, visualizar como estão distribuídas.

Na impossibilidade de representar todas as categorias em uma tabela bidimensional, é necessário obter uma forma de representar conjuntamente as Q variáveis com suas respectivas categorias. Dessa forma, um primeiro procedimento da análise de correspondência múltipla consiste em construir uma tabela lógica, em que, para cada categoria de cada variável atribui-se 1 se uma determinada observação apresenta a característica e 0, caso contrário. A tabela lógica se caracteriza como uma matriz binária \mathbf{X} , na qual nas linhas estão as unidades de observação e nas colunas as diferentes categorias de cada uma das q variáveis presentes na análise (Ferreira Filho *et al.*, 1998).

Na seqüência, a partir da relação $\mathbf{B} = \mathbf{X}'\mathbf{X}$, obtém-se a matriz de Burt (quadrada e simétrica), na qual as matrizes do bloco diagonal da mesma contêm a distribuição de frequência de cada variável, e as demais matrizes (fora do bloco diagonal) apresentam todas as possíveis tabelas duas a duas, das variáveis presentes no estudo. Mais detalhes da construção da tabela lógica e da obtenção da tabela de Burt podem ser vistos em Fávero e Belfiore (2017) e Ferreira Filho *et al.* (1998).

A análise de correspondência múltipla pode ser desenvolvida tanto a partir da tabela lógica quanto pela tabela de Burt. Por se tratar de duas tabelas, embora com características específicas, procedimentos análogos ao da análise de correspondências simples são utilizados, com o objetivo de obter as coordenadas de cada uma das categorias das variáveis a serem representadas no mapa perceptual. Considerando um estudo com Q

variáveis e que a q -ésima variável possua J_q categorias, a inércia principal total de \mathbf{X} é obtida pela Expressão (2.7).

$$I_T = \frac{\sum_{q=1}^Q (J_q - 1)}{Q} = \frac{J - Q}{Q}, \text{ em que } J = \sum_{q=1}^Q J_q. \quad (2.7)$$

A inércia depende somente do número de variáveis e suas respectivas categorias. Pela relação dada na Equação (2.7), nota-se que quanto menor for a frequência de uma determinada categoria, maior a sua contribuição para a inércia. Desta forma, categorias com baixa frequência acabam se destacando no mapa perceptual, muitas vezes prejudicando a identificação das relações entre as categorias das demais variáveis. Nestes casos, recomenda-se que, se possível, estas categorias sejam agrupadas a outras, ou então excluídas da análise e estudadas separadamente. Eliminados estes casos de baixa frequência, o mapa perceptual, com as coordenadas das demais categorias de todas as variáveis, é interpretado de forma análoga ao caso da análise de correspondência simples. Dessa forma, é esperado que as primeiras coordenadas principais contenham uma boa informação da maior parte do conjunto de categorias presentes no estudo. Portanto, a interpretação das mesmas será suficiente para identificação das principais relações/associações simultâneas entre categorias das diferentes variáveis.

Capítulo 3

Resultados

3.1 Análise Descritiva

Inicialmente, foi realizada uma Análise Descritiva univariada e bivariada, referente às variáveis de cada mês de estudo, isto é, maio, agosto e novembro. A análise foi desenvolvida com o objetivo de investigar as características individuais e em alguns casos, duas a duas das variáveis presentes no estudo, a fim de identificar a frequência das categorias das variáveis e possíveis relações a serem investigadas na Análise de Correspondência.

Considerando o objetivo de estudar os indivíduos que apresentaram pelo menos um dos sintomas do questionário, foram selecionados para o estudo somente os respondentes que atendessem a este critério. Dentre os sintomas abordados no questionário, estão: febre, tosse, dor de garganta, dificuldade para respirar, dor de cabeça, dor no peito, náusea, nariz entupido ou escorrendo, fadiga, dor nos olhos, perda de olfato ou paladar e dor muscular. Vale ressaltar que, para os questionários dos meses de agosto e novembro, o sintoma diarreia foi incluído.

A [Tabela 3.1](#) apresenta o tamanho total da amostra observada em cada um dos meses de estudo, bem como o número de respondentes que atendiam ao requisito de interesse, isto é, terem apresentado pelo menos um dos sintomas no período.

Tabela 3.1: Tamanho amostral e número de indivíduos selecionados.

Meses	Mai	Agosto	Novembro
Tamanho Amostral	349306	386520	381438
Indivíduos com pelo menos um sintoma	37203	21938	13738

Observa-se, na [Tabela 3.1](#), que agosto foi o mês em que foi possível obter o maior

número de respondentes. Além disso, identifica-se que, no mês de maio, apesar de ser o mês com menor tamanho amostral obtido, houve um maior número de indivíduos com pelo menos um sintoma de COVID-19, com uma tendência decrescente nos dois trimestres seguintes.

A partir dos indivíduos selecionados, foi obtida a distribuição univariada, além da frequência percentual das principais características dos respondentes e dos respectivos sintomas apresentados nos meses de maio, agosto e novembro.

Nas Tabelas 3.2 e 3.3 são apresentadas as proporções de indivíduos de acordo com a Faixa etária e Sexo, respectivamente. É possível observar que estas variáveis apresentam proporções balanceadas entre as categorias, ou seja, não há uma categoria com uma frequência muito inferior em relação às demais.

Tabela 3.2: Proporção de indivíduos por faixa etária nos meses de estudo.

Faixa etária	Proporção(%)		
	Maio	Agosto	Novembro
0-13	10,22	13,27	13,82
14-29	24,27	23,00	23,02
30-49	37,75	34,67	32,71
50-59	12,80	12,59	12,85
60 ou mais	14,96	16,47	17,60

Tabela 3.3: Proporção de indivíduos de acordo com o sexo nos meses de estudo.

Sexo	Proporção(%)		
	Maio	Agosto	Novembro
Masculino	43,38	42,80	43,37
Feminino	56,62	57,20	56,63

Tabela 3.4: Proporção de indivíduos de acordo com a raça nos meses de estudo.

Raça	Proporção (%)		
	Maio	Agosto	Novembro
Branca	40,35	43,72	44,28
Preta	9,55	9,72	9,51
Amarela	0,96	0,70	0,72
Parda	48,63	45,47	45,15
Indígena	0,51	0,39	0,34

A distribuição univariada das demais variáveis relacionadas às características dos respondentes estão disponíveis nas Tabelas 3.4, 3.5 e 3.6. Nota-se, pelos resultados apresentados nas Tabelas 3.2, 3.3, 3.4, 3.5 e 3.6 que a distribuição das categorias de cada variável

Tabela 3.5: Proporção de indivíduos de acordo com grau de escolaridade nos meses de estudo.

Escolaridade	Proporção(%)		
	Maio	Agosto	Novembro
Sem instrução	6,75	8,72	9,05
Fundamental incompleto	26,11	28,76	28,28
Fundamental completo	6,82	6,57	6,77
Médio incompleto	9,53	9,49	8,89
Médio completo	27,38	23,99	23,18
Superior incompleto	7,35	7,16	7,30
Superior completo	12,33	11,65	11,85
Pós graduação, Mestrado ou Doutorado	3,73	3,66	4,68

Tabela 3.6: Proporção de indivíduos de acordo com a região nos meses de estudo.

Região	Proporção(%)		
	Maio	Agosto	Novembro
Norte	13,86	8,60	7,90
Nordeste	28,81	27,93	27,56
Centro-Oeste	4,96	8,05	7,05
Sudeste	40,30	39,85	42,01
Sul	12,08	15,57	15,48

segue o mesmo padrão entre os meses. Destaca-se que a maioria dos respondentes está concentrada entre as faixas etárias de 14 a 49 anos, com a proporção maior de mulheres, a maioria de raça branca ou parda, maior concentração de pessoas com ensino fundamental incompleto ou médio completo, sendo, em grande parte, da região Sudeste ou Nordeste. Tais características correspondem com a distribuição da população brasileira, ou seja, o grupo de indivíduos que apresentaram pelo menos um sintoma de COVID-19 se aproxima da caracterização da população brasileira (SIDRA, 2020).

Vale ressaltar que, para o cálculo de tais medidas descritivas, utilizou-se a biblioteca *survey* do software R para incorporar o plano amostral no efeito das estimativas.

Na comparação da diferença entre as estimativas obtidas por métodos usuais com aquelas utilizando o pacote, isto é, considerando uma amostragem aleatória simples contra amostragem aleatória estratificada, são apresentados os resultados sobre a distribuição de sintomas nas Tabelas 3.7 e 3.8, respectivamente. As frequências estão ordenadas de forma decrescente para o mês de maio.

Nota-se, pelos resultados apresentados nas Tabelas 3.7 e 3.8, que as estimativas obtidas considerando os diferentes planos amostrais são muito próximas, em geral, menores para o pacote *survey*. Apesar da semelhança entre os valores, a fim de atender ao objetivo de

Tabela 3.7: Proporção de indivíduos que apresentaram cada um dos sintomas nos meses de estudo sem utilizar o *survey*.

Sintomas	Proporção (%)		
	Maio	Agosto	Novembro
Dor de cabeça	46,51	43,34	41,91
Nariz escorrendo	35,16	38,08	37,80
Tosse	31,59	34,67	36,89
Dor muscular	28,19	24,94	25,43
Dor de garganta	23,43	28,24	31,17
Febre	21,54	20,56	21,21
Perda de olfato ou paladar	16,15	11,34	10,34
Fadiga	15,50	17,07	19,01
Dificuldade para respirar	13,27	11,17	10,73
Dor nos olhos	12,83	9,66	8,21
Dor no peito	10,93	8,18	7,43
Náusea	10,03	8,95	9,06
Diarreia	-	9,68	10,30

Tabela 3.8: Proporção de indivíduos que apresentaram cada um dos sintomas nos meses de estudo utilizando *survey*.

Sintomas	Proporção (%)		
	Maio	Agosto	Novembro
Dor de cabeça	46,16	42,05	40,55
Nariz escorrendo	36,70	39,77	38,83
Tosse	30,60	33,48	36,90
Dor muscular	27,53	23,93	25,03
Dor de garganta	22,63	26,92	30,95
Febre	20,99	18,52	20,44
Perda de olfato ou paladar	16,21	10,68	9,72
Fadiga	15,30	16,40	18,42
Dificuldade para respirar	13,22	10,97	11,06
Dor nos olhos	12,54	9,12	8,23
Dor no peito	10,46	7,45	7,34
Náusea	9,69	8,73	9,11
Diarreia	-	9,56	10,30

incorporar o plano amostral na análise realizada, para o cálculo das proporções em toda a análise descritiva apresentada, foram considerados os resultados obtidos utilizando o pacote *survey*.

A partir dos valores apresentados na [Tabela 3.8](#), observa-se, de modo geral, que a distribuição dos sintomas é próxima entre os meses de maio, agosto e novembro. Ainda, nota-se que os sintomas mais predominantes são dor de cabeça, nariz escorrendo, tosse, dor muscular e dor de garganta todos com porcentagem superior a 20%.

Vale ressaltar que um indivíduo pode apresentar mais do que um sintoma simultaneamente. Logo, avalia-se a distribuição do total de sintomas apresentados pelos entrevistados para cada mês.

Tabela 3.9: Proporção de indivíduos que apresentaram cada possível número de sintomas.

Quantidade de Sintomas	Proporção (%)		
	Maio	Agosto	Novembro
1	45,74	44,65	40,99
2	21,73	22,86	24,60
3	10,65	11,65	12,38
4	6,38	6,73	6,86
5	4,17	4,00	4,63
6	2,92	2,78	2,82
7	2,27	1,88	2,06
8	1,59	1,45	1,54
9	1,26	0,93	1,10
10	0,94	0,77	0,92
11	0,85	0,64	0,74
12	1,49	0,58	0,60
13	-	1,06	0,76
Total ¹	37203	21938	13738

¹ Número de indivíduos que apresentaram pelo menos um sintoma.

Na [Tabela 3.9](#) é possível notar que o número de sintomas mais frequentes entre os indivíduos são entre 1 e 3, os quais representam conjuntamente cerca de 70% da amostra de cada mês.

Para ilustrar a distribuição do número de casos, a representação gráfica é apresentada na [Figura 3.1](#). Como destacado na [Tabela 3.9](#), para os três meses estudados, nota-se que a proporção de indivíduos diminui conforme aumenta o número de sintomas apresentados conjuntamente. Tal comportamento é dado para os três meses de estudo.

Dentre os indivíduos que apresentaram pelo menos um dos sintomas, avalia-se a proporção daqueles que foram ao estabelecimento de saúde, conforme [Tabela 3.10](#). É possível identificar que, independente do mês de estudo, quanto maior o número de sintomas, maior a proporção de pessoas que dirigiram-se a um estabelecimento de saúde.

A distribuição da procura por estabelecimentos de saúde pelo número de sintomas é representada graficamente na [Figura 3.2](#). Identifica-se o crescimento da proporção de indivíduos que buscaram um estabelecimento de saúde conforme a presença de um maior número de sintomas. Além disso, nota-se que, ao longo dos meses, para um mesmo número de sintomas, a proporção da busca por estabelecimentos aumentou. Tal fato pode ser

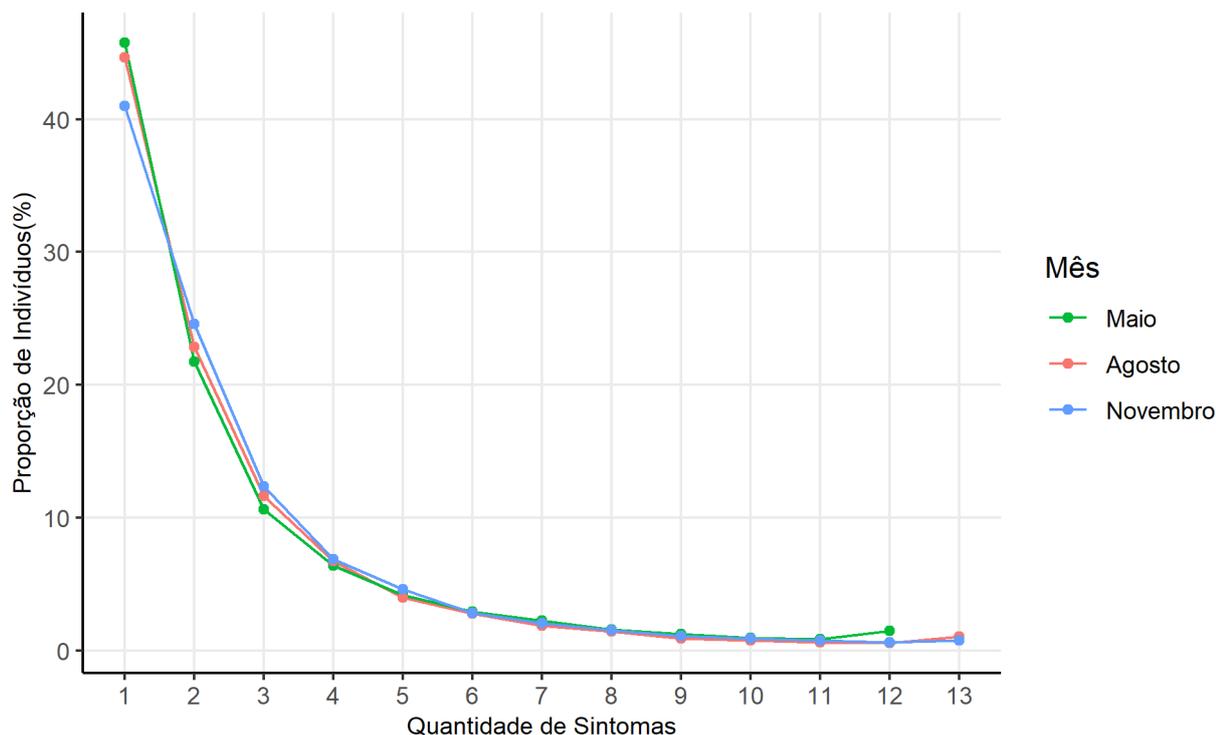


Figura 3.1: Distribuição da proporção de indivíduos de acordo com a quantidade de sintomas.

Tabela 3.10: Proporção de indivíduos que procuraram estabelecimentos de saúde segundo número de sintomas.

Quantidade de sintomas	Proporção (%)					
	Maio		Agosto		Novembro	
	Sim	Não	Sim	Não	Sim	Não
1	8,18	91,82	11,65	88,35	15,32	84,68
2	12,54	87,46	20,52	79,48	22,51	77,49
3	18,64	81,37	26,14	73,86	34,92	65,08
4	23,88	76,13	34,51	65,49	45,95	53,98
5	24,09	75,91	43,70	56,30	47,13	52,87
6	26,99	73,01	50,92	49,08	56,67	43,33
7	35,64	64,36	47,48	52,52	56,58	43,42
8	40,77	59,23	57,68	42,32	67,93	32,07
9	40,96	59,04	59,52	40,48	61,91	38,09
10	47,59	52,41	63,77	36,22	68,68	31,32
11	52,34	47,66	60,21	39,79	80,39	19,61
12	54,89	45,11	66,54	33,46	75,48	24,52
13	-	-	78,21	21,79	78,60	21,40

justificado, possivelmente, pela maior disseminação de informações sobre a COVID-19 ao longo do ano, bem como o aumento de casos e mortes pela doença, resultando em uma maior preocupação da população por um atendimento de saúde.

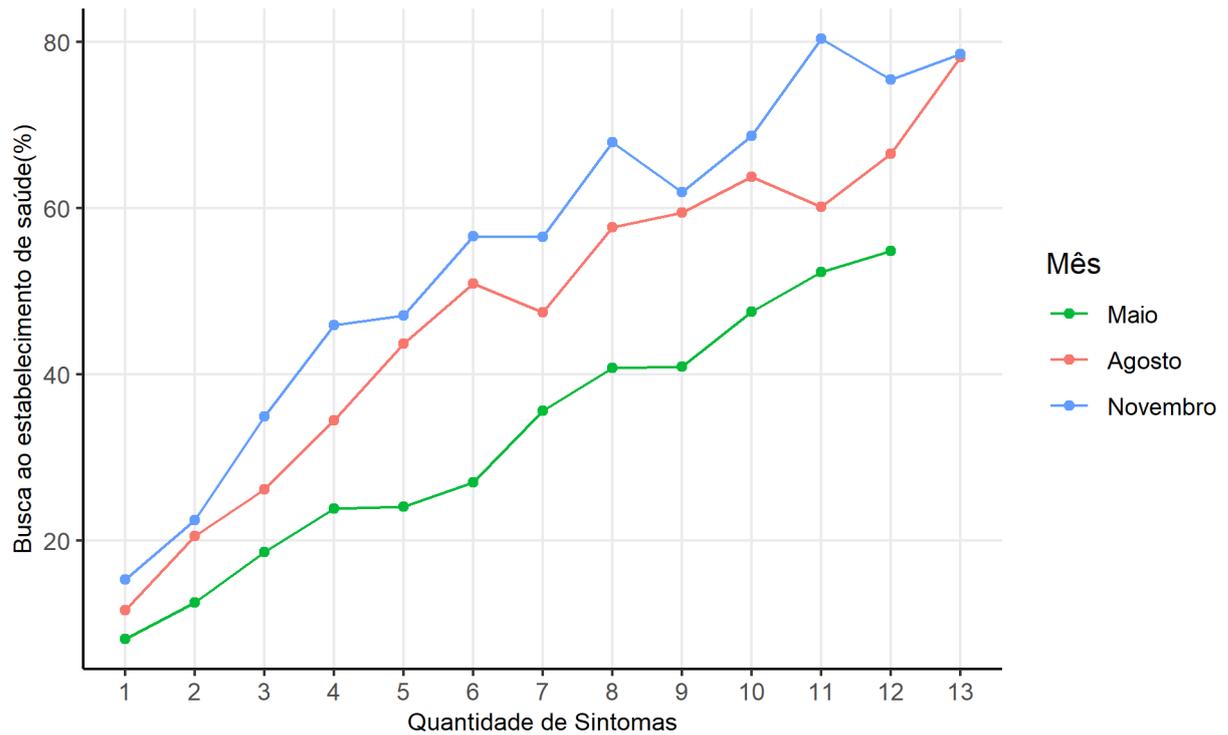


Figura 3.2: Proporção de indivíduos que buscaram um estabelecimento de saúde de acordo com a quantidade de sintomas.

É necessário ressaltar que o comportamento citado é observado dentre aqueles indivíduos que procuraram por algum estabelecimento de saúde. A [Tabela 3.11](#) mostra, entretanto, que a maioria dos indivíduos não procurou por um estabelecimento de saúde nos três meses do estudo.

Tabela 3.11: Proporção de indivíduos que buscaram um estabelecimento de saúde nos meses de estudo.

Foi ao estabelecimento	Proporção (%)		
	Maio	Agosto	Novembro
Sim	15,46	22,71	28,17
Não	84,54	77,29	71,83

Dentre as pessoas que buscaram atendimento em um estabelecimento de saúde, identifica-se a proporção dos lugares visitados, conforme [Tabela 3.12](#). Nota-se, para os três meses, que o local de atendimento mais procurado pelos indivíduos que apresentaram pelo menos um dos sintomas foi o posto de saúde, enquanto que o menos procurado foi o pronto socorro privado.

Vale ressaltar que, no geral, os locais de saúde públicos (Posto de saúde, Pronto Socorro e Hospital do SUS) foram mais procurados em comparação aos atendimentos privados.

Tabela 3.12: Proporção de indivíduos que buscaram atendimentos segundo tipo de local.

Locais	Proporção (%)		
	Maió	Agosto	Novembro
Posto de saúde	44,57	45,51	40,01
Pronto Socorro SUS	23,74	23,07	22,01
Hospital SUS	17,76	17,93	16,31
Ambulatório/Consultório privado	10,18	9,79	10,05
Pronto Socorro privado	4,52	4,24	6,57
Hospital privado	12,16	9,78	13,64

A partir deste fato, é importante avaliar a relação entre a busca por estabelecimentos e o fato do indivíduo possuir plano de saúde privado, conforme [Tabela 3.13](#). É possível notar que a procura por um estabelecimento de saúde não parece ter relação com o indivíduo possuir plano de saúde, uma vez que as proporções dentre os indivíduos que procuraram, ou não, a um estabelecimento de saúde são muito próximas para aqueles que possuem um plano de saúde privado.

Tabela 3.13: Proporção de indivíduos (em %) que buscaram um estabelecimento de saúde segundo possuir ou não um plano de saúde privado.

Foi ao estabelecimento	Possui plano					
	Maio		Agosto		Novembro	
	Sim	Outros	Sim	Outros	Sim	Outros
Sim	30,29	69,71	29,59	70,41	37,02	62,98
Não	29,77	70,23	26,81	73,19	27,34	72,66

Ressalta-se que a maioria dos respondentes (aproximadamente 70% nos três meses) não possui um plano de saúde.

É importante destacar que, em determinadas questões, havia a presença de três ou mais alternativas. Na maior parte destes casos, o interesse foi pelo fato dos indivíduos apresentarem ou não uma determinada característica, tal como na [Tabela 3.13](#) (possuir plano de saúde privado). Para estas questões, foram atribuídas apenas duas alternativas: “Sim” no caso de apresentar a característica e “Outros” para as demais alternativas. As codificações estão disponibilizadas no [Apêndice A](#), tal como comentado na [Seção 2.2](#).

Em complemento, avaliam-se os locais em que se buscou atendimento a partir dos indivíduos que possuem ou não um plano de saúde privado, conforme apresentado na [Tabela 3.14](#). Como esperado, os indivíduos que não possuem plano de saúde privado, independente do mês de estudo, buscaram atendimento em locais de saúde públicos, tais como posto de saúde, pronto socorro e hospital do SUS. Já os que possuem um plano de

saúde privado representam a maior parte dos que buscaram atendimento em ambulatórios, consultórios, pronto socorro ou hospitais privados.

Tabela 3.14: Proporção de tipos de locais (em %) em que se buscou atendimento segundo possuir ou não plano de saúde privado.

Locais	Possui plano					
	Maio		Agosto		Novembro	
	Sim	Não	Sim	Não	Sim	Não
Posto de saúde	14,69	85,31	16,65	83,35	18,14	81,86
Pronto Socorro SUS	14,69	85,31	17,23	82,77	18,00	82,00
Hospital SUS	18,47	81,53	14,79	85,21	19,87	80,13
Ambulatório/Consultório privado	63,76	36,24	68,90	31,10	76,68	23,32
Pronto Socorro privado	85,59	14,41	86,22	13,78	91,10	8,90
Hospital privado	82,95	17,05	85,83	14,17	90,50	9,50

Ainda, considerando os indivíduos que buscaram um estabelecimento de saúde, investiga-se a proporção dos que foram internados, como apresentado na [Tabela 3.15](#).

Tabela 3.15: Proporção de indivíduos por internação nos meses de estudo.

Foi internado	Proporção(%)		
	Maio	Agosto	Novembro
Sim	10,26	8,40	8,36
Não	87,54	90,07	90,18
Não foi atendido	2,20	1,53	1,46

De acordo com a [Tabela 3.15](#), a maioria das pessoas, que buscou atendimento em estabelecimento de saúde, não necessitou ser internada. Destaca-se uma maior proporção de internações no mês de maio, além de uma pequena parcela da população que nem ao menos foi atendida.

Para os que foram internados, avalia-se a proporção dos indivíduos intubados, sedados ou colocados em respiração artificial, conforme [Tabela 3.16](#). Observa-se que a maior proporção dos respondentes que foram internados, não foram intubados, sedados ou colocados em respiração artificial. Vale ressaltar que, para os que passaram por este procedimento, houve uma maior proporção de indivíduos no mês de maio, reduzindo nos meses seguintes.

Com relação às pessoas que não foram até um estabelecimento de saúde, investiga-se quais providências foram tomadas. De acordo com a [Tabela 3.17](#), para os três meses de estudo, a maioria dos indivíduos ficou em casa. A segunda providência mais adotada foi a utilização de medicamentos por conta própria, a qual nota-se que a respectiva proporção aumentou ao longo dos meses.

Tabela 3.16: Proporção de indivíduos por intubação¹ nos meses de estudo.

Foi intubado	Proporção(%)		
	Maio	Agosto	Novembro
Sim	27,64	20,49	16,65
Não	72,36	79,51	83,35

¹ Indica se o indivíduo foi intubado, sedado ou colocado em respiração artificial.

Tabela 3.17: Proporção dos tipos de providências adotadas nos meses de estudo.

Providências	Proporção(%)		
	Maio	Agosto	Novembro
Ficar em casa	82,81	73,09	71,58
Ligar para profissional	5,69	3,85	3,24
Tomar remédio por conta própria	56,76	58,42	61,49
Tomar remédio por orientação médica	14,57	11,88	10,98
Receber visita de profissional do SUS	2,60	2,14	1,93
Receber visita de profissional particular	0,83	0,41	0,45
Outras	2,39	2,65	2,41

Em complemento, realiza-se a análise da distribuição da proporção de internados de acordo com o número de sintomas, conforme [Tabela 3.18](#).

Tabela 3.18: Proporção de indivíduos (em %) que passaram por internação de acordo com a quantidade de sintomas.

Quantidade de sintomas	Internação					
	Maio		Agosto		Novembro	
	Sim	Outros	Sim	Outros	Sim	Outros
1	8,13	91,87	6,95	93,05	9,26	90,75
2	6,14	93,86	6,16	93,84	5,50	94,49
3	6,72	93,28	6,53	93,47	8,52	91,48
4	10,26	89,74	10,96	89,04	7,65	92,35
5	13,62	86,38	9,73	90,27	8,34	91,66
6	4,70	95,30	9,13	90,88	16,13	83,87
7	12,92	87,08	5,23	94,77	5,26	94,74
8	9,46	90,04	11,14	88,86	8,13	91,87
9	17,63	82,37	10,44	89,56	6,50	93,50
10	12,11	87,89	8,22	91,78	4,73	95,27
11	13,88	86,12	8,69	91,31	4,23	95,77
12	23,96	76,03	15,67	84,33	21,23	78,77
13	-	-	19,73	80,27	10,29	89,71

A partir da [Tabela 3.18](#) e [Figura 3.3](#) nota-se, para os três meses, que a proporção de indivíduos internados varia de acordo com a quantidade de sintomas, não demonstrando um padrão. Além disso, observa-se que a proporção de internações parece não aumentar

consideravelmente para uma maior quantidade de sintomas, com exceção dos que apresentaram 12 sintomas, o qual registrou a maior proporção de internados para todos os meses.

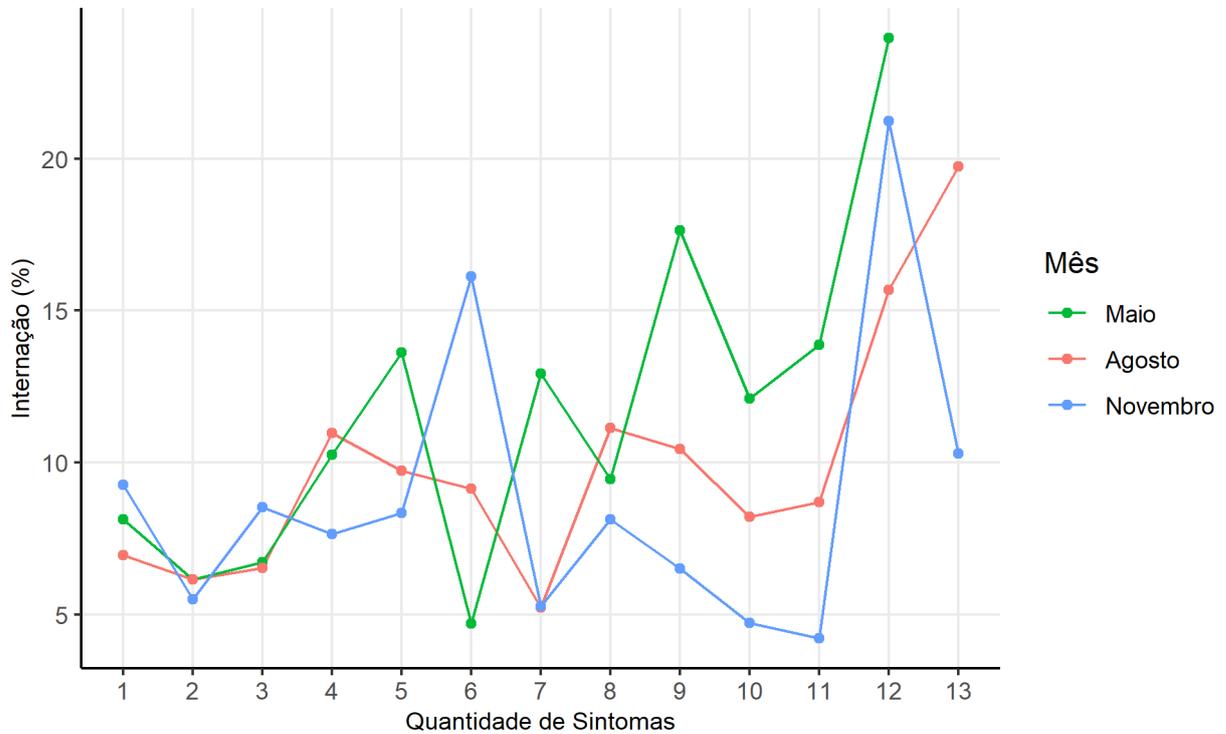


Figura 3.3: Proporção de indivíduos internados de acordo com a quantidade de sintomas.

Ainda, dentre os indivíduos que foram internados, avalia-se a proporção dos que foram intubados, sedados ou colocados em respiração artificial conforme o número de sintomas, por meio da [Tabela 3.19](#) e [Figura 3.4](#). Nota-se que o comportamento da distribuição de intubações varia de acordo com os meses e quantidade de sintomas. Destaca-se que os indivíduos que apresentaram 7 sintomas correspondem à menor proporção de internados no mês de maio e nenhuma internação para os meses de agosto e novembro.

Além disso, não identifica-se um padrão da proporção de intubações conforme o número de sintomas. Por exemplo, indivíduos que apresentaram 11 sintomas correspondem à maior proporção de intubações no mês de agosto (77,44%), enquanto que, para os outros meses, tal proporção não representa a maior quantidade de intubações (26,86% para o mês de maio e 14,80% em novembro).

Tabela 3.19: Proporção de indivíduos (em %) que passaram por intubação de acordo com a quantidade de sintomas.

Quantidade de sintomas	Intubação					
	Maio		Agosto		Novembro	
	Sim	Outros	Sim	Outros	Sim	Outros
1	10,27	89,73	16,48	83,52	24,97	75,03
2	17,02	82,98	19,24	80,76	4,37	95,63
3	17,32	82,68	20,59	79,41	21,12	78,88
4	33,20	66,80	15,92	84,08	17,85	82,15
5	29,87	70,13	9,73	27,30	10,49	89,51
6	32,70	67,30	7,54	92,46	15,73	84,27
7	9,67	90,33	0	100	0	100
8	28,05	71,95	36,82	63,18	16,74	83,26
9	39,72	60,28	13,65	86,35	36,87	65,13
10	52,83	47,17	21,30	78,70	54,90	45,10
11	26,86	73,14	77,44	22,56	14,80	85,20
12	47,77	52,23	37,41	62,59	0	100
13	-	-	13,32	86,68	14,22	85,78

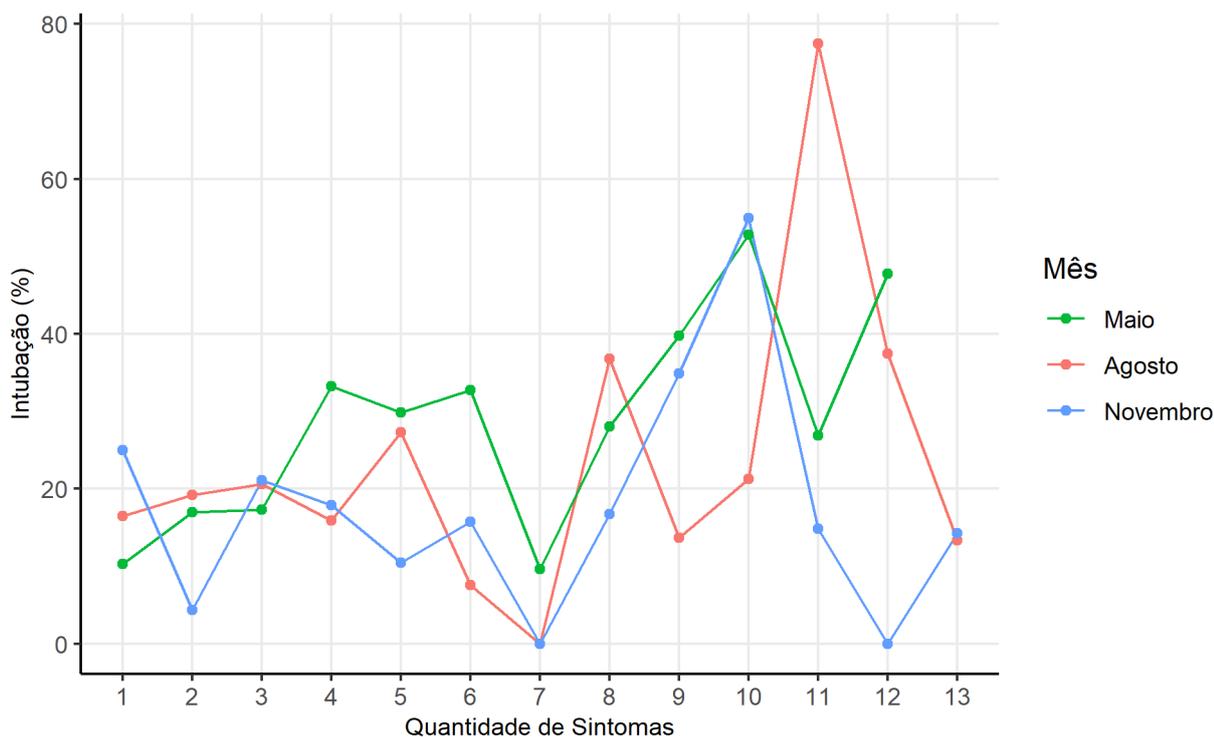


Figura 3.4: Proporção de indivíduos intubados de acordo com a quantidade de sintomas.

3.1.1 Questões incorporadas a partir do mês de agosto

Algumas questões foram incorporadas no questionário a partir da pesquisa PNAD COVID19 realizada no mês de agosto de 2020. Desta forma, estas questões foram avaliadas

para os meses de agosto e novembro.

Inicialmente, avalia-se a proporção de indivíduos que realizaram teste de COVID-19 de acordo com o número de sintomas, conforme apresentado na [Tabela 3.20](#). Nota-se, tanto para o mês de agosto como novembro, que quanto maior o número de sintomas, maior a proporção de pessoas que fizeram o teste para COVID-19. Tal fato pode ser justificado com a associação dos possíveis sintomas relacionados à COVID-19. Dessa forma, quanto maior o número de sintomas possíveis, maior a proporção de pessoas que se preocupam em realizar o teste.

Tabela 3.20: Proporção de indivíduos que fizeram teste de COVID-19 de acordo com a quantidade de sintomas.

Quantidade de sintomas	Proporção(%)			
	Agosto		Novembro	
	Sim	Não	Sim	Não
1	12,08	87,92	19,79	80,21
2	16,46	83,54	25,09	74,91
3	22,94	77,06	33,09	66,91
4	27,49	72,51	38,30	61,70
5	30,44	69,56	42,40	57,60
6	41,35	58,65	49,15	50,85
7	43,39	56,61	46,41	53,59
8	40,54	59,46	56,60	43,40
9	47,06	52,94	59,56	40,44
10	59,03	40,97	56,04	43,96
11	50,33	49,67	62,99	37,01
12	52,83	47,17	58,53	41,47
13	72,18	27,82	62,65	27,46

A [Figura 3.5](#) apresenta o comportamento dos indivíduos que fizeram teste de COVID-19 de acordo com a quantidade de sintomas. Observa-se que a proporção de indivíduos que fizeram teste de COVID-19 é maior com o aumento da quantidade de sintomas. Merece ainda ser destacado que, para a maioria, a proporção de pessoas que fizeram teste foi maior para o mês de novembro, com exceção daqueles que tiveram 10 e 13 sintomas.

Na sequência, investiga-se a proporção de testes realizados segundo o tipo de teste, conforme [Tabela 3.21](#). Dentre as opções estão: teste realizado por cotonete, teste de sangue coletado na veia (representado como “Sangue 1”) e teste de sangue coletado por furo no dedo (representado como “Sangue 2”). Vale destacar que um indivíduo pode ter realizado mais de um tipo de teste. O teste mais frequente foi o coletado por cotonete, cuja proporção aumentou de agosto para novembro.

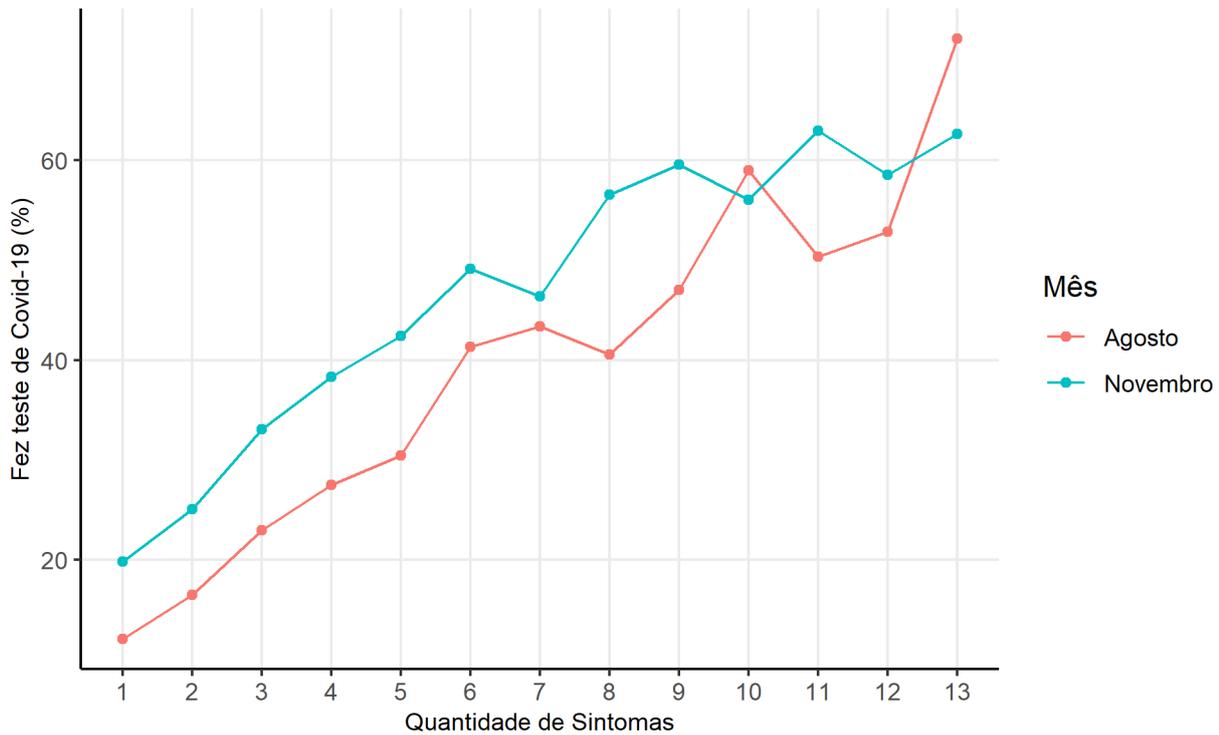


Figura 3.5: Proporção de testes feitos para COVID-19 de acordo com a quantidade de sintomas.

Tabela 3.21: Proporção de testes de COVID-19 realizados.

Tipo de teste	Proporção(%)	
	Agosto	Novembro
Cotonete	50,36	56,82
Sangue 1	37,47	36,35
Sangue 2	29,09	25,97

Considerando os testes realizados, avaliam-se os resultados dos mesmos, conforme [Tabela 3.22](#). Nota-se que, para ambos os meses, independentemente do tipo de teste, a maior proporção dos resultados é negativa. Além disso, é possível observar que, para o tipo de teste coletado por cotonete, a proporção de resultados positivos é superior se comparada à proporção dos demais testes em ambos os meses.

Para os indivíduos que realizaram o teste de COVID-19, estuda-se aqueles que tiveram resultado positivo, independente do tipo de teste. Inicialmente, verifica-se a proporção de resultados positivos de acordo com o número de sintomas, conforme [Tabela 3.23](#). Observa-se que, independente do mês, as maiores proporções de resultados positivos são para até 4 sintomas. Destaca-se que, no mês de agosto, os indivíduos que apresentaram 13 sintomas também representam uma parcela significativa da proporção de resultados

Tabela 3.22: Proporção de resultados pelo tipo de teste de COVID-19 realizado.

Resultado do teste	Proporção(%)					
	Agosto			Novembro		
	Cotonete	Sangue 1	Sangue 2	Cotonete	Sangue 1	Sangue 2
Positivo	38,10	29,43	34,96	31,33	21,34	29,69
Negativo	47,61	67,47	58,57	51,52	76,02	65,76
Inconclusivo	0,36	1,00	1,15	0,58	0,60	1,06
Ainda não recebeu	13,93	2,10	5,32	16,57	2,04	3,49

positivos. Com relação às demais quantidades de sintomas, suas proporções variam, sendo que, para maiores números de sintomas tais como 11 e 12, a proporção de resultados positivos é menor se comparada a uma menor quantidade de sintomas.

Tabela 3.23: Proporção de resultados positivos de acordo com a quantidade de sintomas.

Quantidade de Sintomas	Proporção(%)	
	Agosto	Novembro
1	16,19	22,36
2	14,69	18,14
3	11,30	11,15
4	10,93	8,67
5	7,51	7,80
6	5,77	7,05
7	6,46	4,55
8	4,75	4,23
9	3,47	3,37
10	4,36	3,56
11	3,02	2,08
12	3,23	2,31
13	8,32	4,72

A [Figura 3.6](#) mostra o comportamento da proporção de resultados positivos de acordo com a quantidade de sintomas. Nota-se que, de forma geral, a proporção de resultados positivos varia segundo a quantidade de sintomas, com uma tendência decrescente. Contudo, para os indivíduos com 13 sintomas, a proporção de positivos aumenta.

Outra variável que foi incorporada nos meses de agosto e novembro refere-se às restrições tomadas pelos indivíduos devido à pandemia. Dentre as opções, estavam: o indivíduo não fez restrição; reduziu o contato com as pessoas mas continuou saindo de casa para demais atividades; ficou em casa e saiu somente para necessidades básicas ou ficou rigorosamente em casa. Os resultados são apresentados na [Tabela 3.24](#). É possível notar que a maior parte dos respondentes reduziu o contato com as pessoas ou saiu somente o necessário. Vale destacar que, em ambos os meses, há uma pequena parcela da população

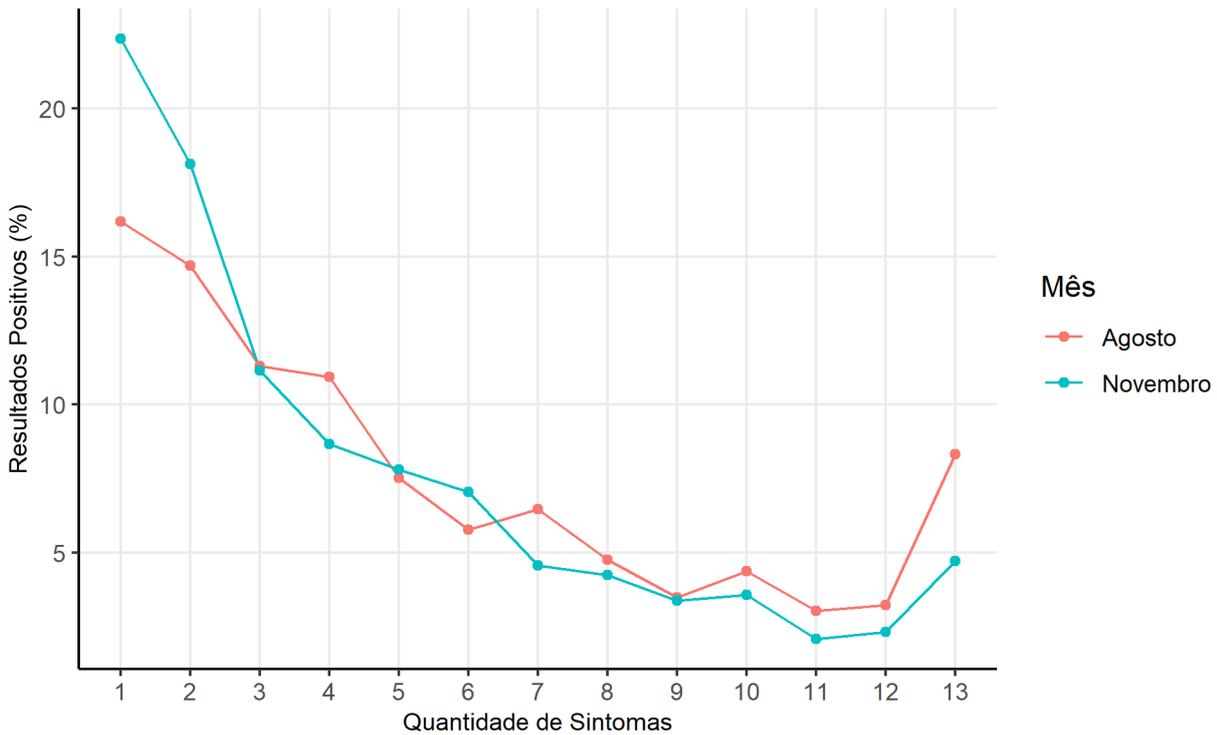


Figura 3.6: Proporção de resultados positivos de acordo com a quantidade de sintomas.

que não aderiu a nenhuma restrição, conduzindo seus hábitos da mesma forma do que antes da pandemia. Outro ponto a ser destacado é a maior flexibilidade de aderir às restrições entre os meses, em que uma menor proporção dos indivíduos ficou rigorosamente em casa no mês de novembro em comparação com agosto.

Tabela 3.24: Proporção de indivíduos de acordo com a restrição de contatos.

Restrição	Proporção(%)	
	Agosto	Novembro
Nenhuma	2,15	4,37
Reduziu contato	34,09	39,37
Saiu somente o necessário	42,29	39,62
Ficou rigorosamente em casa	21,47	16,64

Em complemento, dentre as pessoas que realizaram testes para COVID-19, são analisadas as restrições conduzidas para aqueles que tiveram resultados positivos, apresentada na [Tabela 3.25](#). Nota-se, para ambos os meses, que a maior parcela dos indivíduos, que tiveram resultado positivo, ficou rigorosamente em casa. Por outro lado, existe uma parcela significativa da população que, apesar de ter seus resultados positivos, somente reduziu o contato com outras pessoas, continuando com suas atividades habituais.

Outro fato que é possível identificar, assim como na [Tabela 3.24](#), é o fato de que a

Tabela 3.25: Proporção das restrições de contato para indivíduos com resultados de teste da COVID-19 positivos.

Restrição	Proporção(%)	
	Agosto	Novembro
Nenhuma	1,68	4,13
Reduziu contato	30,02	28,10
Saiu somente o necessário	27,38	32,83
Ficou rigorosamente em casa	40,92	34,93

proporção de indivíduos, que testaram positivo e ficaram rigorosamente em casa no mês de agosto, é maior se comparada com o mês de novembro. Com relação aos que não realizaram nenhuma restrição no mês de agosto, observa-se que a proporção de indivíduos é maior no mês de novembro.

Por fim, considerando alguns resultados obtidos nas Tabelas 3.9 e 3.23, em que as maiores proporções de indivíduos com sintomas e resultados positivos foram de até 3 sintomas, são observados os sintomas mais frequentes para este subgrupo, conforme Tabela 3.26.

Tabela 3.26: Proporção de indivíduos com até 3 sintomas de acordo com cada sintoma apresentado nos meses de estudo.

Sintomas	Proporção(%)		
	Mai	Agosto	Novembro
Dor de cabeça	36,36	32,54	30,30
Nariz escorrendo	30,10	33,57	30,74
Tosse	21,22	23,99	27,02
Dor muscular	16,23	13,76	13,86
Dor de garganta	13,34	17,39	20,50
Febre	10,49	8,79	10,94
Perda de olfato ou paladar	6,24	3,72	3,44
Dificuldade para respirar	5,78	4,18	4,26
Fadiga	4,55	6,55	7,95
Dor nos olhos	4,38	3,10	2,35
Náusea	3,24	3,30	3,53
Dor no peito	3,15	2,06	2,08
Diarreia	-	5,39	6,34

A partir da Tabela 3.26 nota-se que os sintomas mais frequentes para os três meses foram dor de cabeça e nariz escorrendo, com cerca de 30% dos respondentes para os três meses. Outros sintomas que também se destacaram com uma maior frequência foram tosse, dor muscular, dor de garganta e febre.

3.2 Análise de Correspondência

A fim de estudar possíveis relações conjuntas entre os sintomas e respectivas características dos respondentes, foi utilizada a Análise Fatorial de Correspondência Múltipla.

Considerando os três meses de estudo, maio, agosto e novembro, nas Figuras 3.7, 3.8 e 3.9 são apresentados os resultados da Análise de Correspondência. Vale ressaltar que, para uma melhor visualização dos pontos nos gráficos, os nomes das variáveis foram codificados. Também nesse sentido, o número de sintomas foi dividido em quatro categorias, agrupando os indivíduos com quatro ou mais sintomas de forma a evitar categorias com baixa frequência que teriam impacto na Análise de Correspondência. Na Tabela 3.27 são disponibilizadas as siglas e as respectivas categorias das variáveis associadas a elas.

Tabela 3.27: Codificação das categorias apresentadas na representação gráfica.

Categoria	Codificação
Faixa etária 0-13	FE1
Faixa etária 14-29	FE2
Faixa etária 30-49	FE3
Faixa etária 50-59	FE4
Faixa etária 60 ou mais	FE5
Sexo Feminino	F
Sexo Masculino	M
1 Sintoma	NS1
2 Sintomas	NS2
3 Sintomas	NS3
4 ou mais sintomas	NS4
Febre : S - presente / N - ausente	FES/FEN
Tosse : S - presente / N - ausente	TOS/TON
Dor de garganta : S - presente / N - ausente	DGS/DGN
Dificuldade para respirar : S - presente / N - ausente	DRS/DRN
Dor nos olhos : S - presente / N - ausente	DOS/DON
Dor de cabeça : S - presente / N - ausente	DCS/DCN
Náusea : S - presente / N - ausente	NAS/NAN
Dor no peito : S - presente / N - ausente	DPS/DPN
Nariz escorrendo : S - presente / N - ausente	NZS/NZN
Fadiga : S - presente / N - ausente	FAS/FAN
Perda de olfato : S - presente / N - ausente	POS/PON
Dor muscular : S - presente / N - ausente	DMS/DMN
Diarreia : S - presente / N - ausente	DIS/DIN

A partir das Figuras 3.7, 3.8 e 3.9 nota-se que, em geral, a disposição das categorias no gráfico apresentam um mesmo padrão. Para os três meses, a proporção da inércia explicada no primeiro plano fatorial corresponde a cerca de 60%.

Dentre os comportamentos semelhantes, pode-se destacar a faixa etária 1, a qual

representa indivíduos de 0 a 13 anos. Observa-se que este ponto está muito distante dos demais, com uma alta contribuição somente para a segunda dimensão, o que indica que esta categoria não apresenta uma relação com os sintomas, fato que pode ser justificado, possivelmente, pela baixa incidência dos mesmos nessa faixa etária, ou seja, não existe nenhum sintoma na qual a ocorrência se destaque na faixa etária entre 0 e 13 anos.

Outro comportamento que pode ser identificado em comum entre os gráficos é a proximidade da presença de sintomas como dificuldade para respirar, nariz escorrendo, perda de olfato, dor de cabeça, náusea e fadiga. Observa-se que estes pontos estão próximos da categoria dos que apresentaram quatro ou mais sintomas, indicando que tais sintomas são os mais comuns de ocorrerem conjuntamente. Vale ressaltar que, para os meses de agosto e novembro, a presença do sintoma diarreia, também ficou próxima deste grupo.

Nota-se que a ocorrência de tosse, dor de garganta, febre e dor no peito estão próximos tanto na presença, quanto na ausência, ou seja, se um indivíduo apresenta alguns desses sintomas, é esperado que também apresente os demais. Este argumento pode ser complementado pelo fato da ausência desses sintomas estarem próximos do número de sintomas igual a um, isto é, se um indivíduo apresentar somente um sintoma, está mais relacionado a não ser tosse, dor de garganta, febre ou dor no peito, exclusivamente.

Além disso, é possível identificar que há relação entre a ausência de determinados sintomas como, por exemplo, dor muscular, fadiga, náusea e nariz escorrendo.

Um ponto também a ser destacado é a proximidade entre as faixas etárias 3 e 4 nos três meses, o que pode ser justificado por possíveis semelhanças entre indivíduos nessas faixas etárias.

Ainda, para os três gráficos, observa-se que a quantidade de um e dois sintomas estão próximos das categorias da ausência de determinados sintomas. Tal comportamento pode ser explicado pela mínima quantidade de sintomas, isto é, se um indivíduo apresenta somente um ou dois sintomas, conseqüentemente, não tem os demais.

Com relação à variável quantidade de sintomas apresentados, nota-se, para os três meses, que as categorias estão ordenadas no eixo horizontal de acordo com a quantidade de sintomas, definindo que a primeira dimensão é orientada/interpretada por esta variável. Ou seja, caso o interesse seja em atribuir um nome a esta primeira dimensão, em função da participação das variáveis na mesma, o mais apropriado seria o número de sintomas.

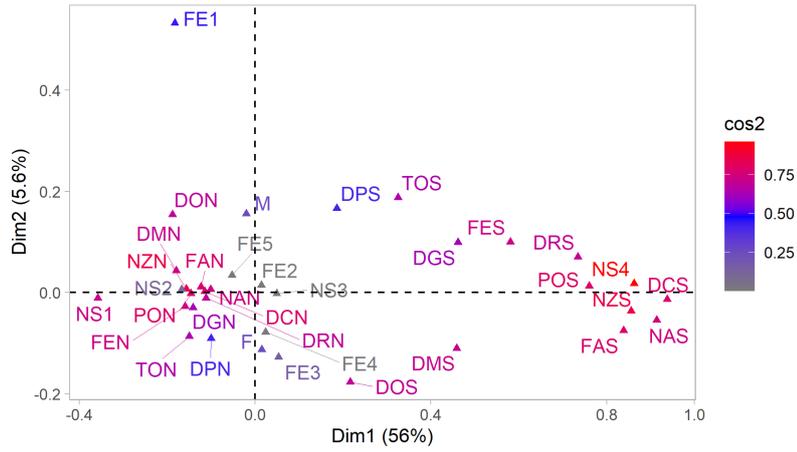


Figura 3.7: Primeiro plano fatorial para o mês de maio.

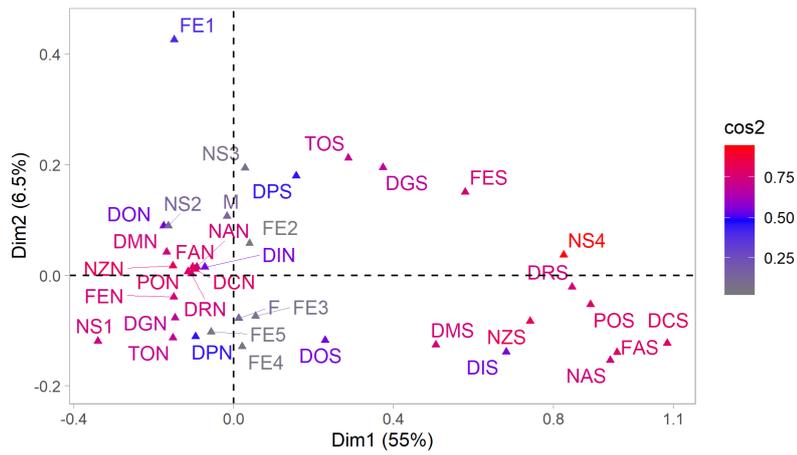


Figura 3.8: Primeiro plano fatorial para o mês de agosto.

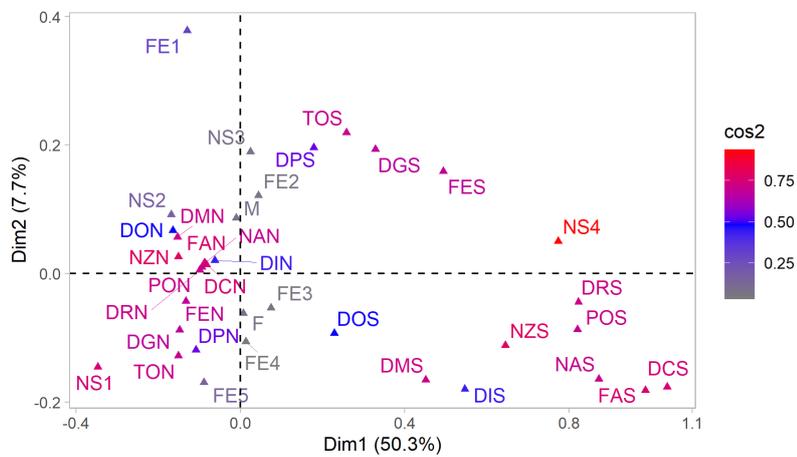


Figura 3.9: Primeiro plano fatorial para o mês de novembro.

Destacam-se as categorias referentes aos sintomas, e suas respectivas quantidades, as quais apresentam altos valores do coeficiente \cos^2 , ou seja, possuem uma alta associação com os eixos representados no plano fatorial.

Para avaliar a contribuição das categorias para o primeiro plano fatorial, para os três meses de estudo, os gráficos são apresentados nas Figuras 3.10, 3.11 e 3.12.

Com relação às variáveis referentes ao sexo e faixa etária, nota-se que não apresentam uma alta contribuição para a explicação da inércia no primeiro plano fatorial. Para as demais variáveis, pelo menos uma categoria apresenta contribuição acima da média.

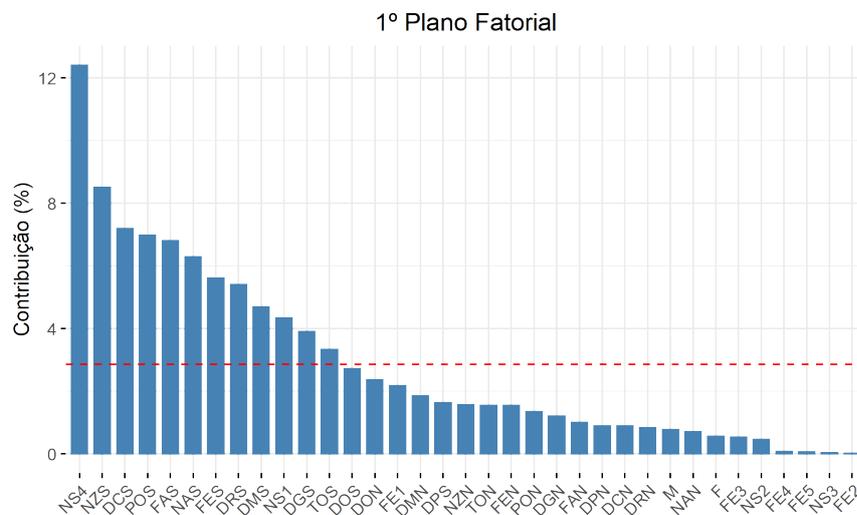


Figura 3.10: Gráfico das contribuições das categorias do mês de maio para o primeiro plano fatorial.

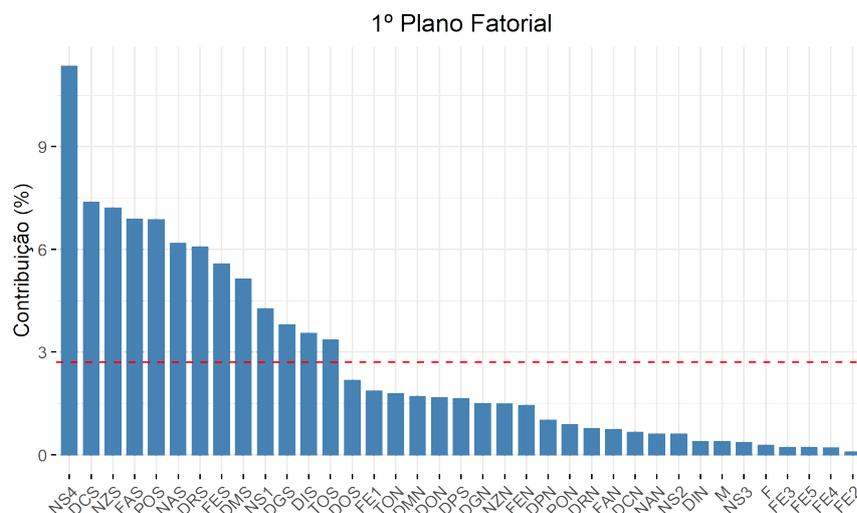


Figura 3.11: Gráfico das contribuições das categorias do mês de agosto para o primeiro plano fatorial.

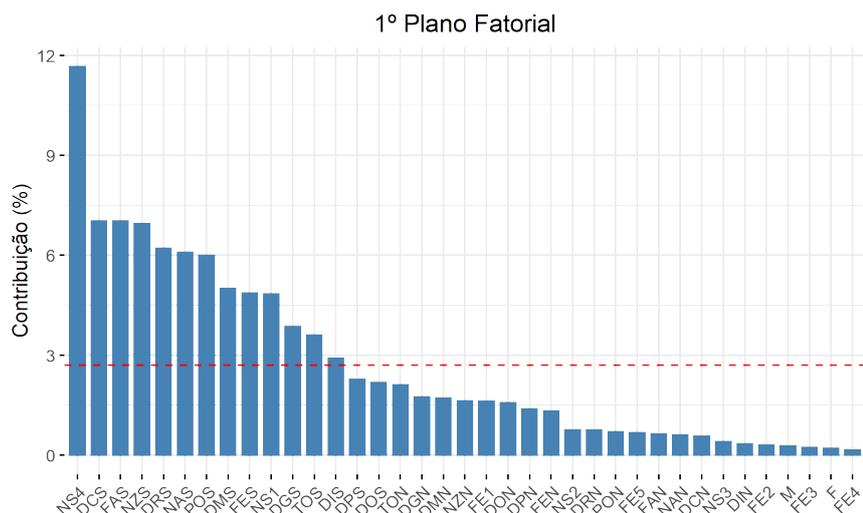


Figura 3.12: Gráfico das contribuições das categorias do mês de novembro para o primeiro plano fatorial.

A fim de identificar as possíveis relações entre a presença de determinados sintomas com indivíduos que apresentam 4 ou mais sintomas, foram selecionadas, aleatoriamente, 7 observações do banco de dados, do mês de maio, com esta característica, as quais são apresentadas na [Tabela 3.28](#). Considerando os indivíduos selecionados, nota-se que, dentre os sintomas identificados nas Figuras 3.7, 3.8 e 3.9 como os mais próximos da categoria de 4 sintomas ou mais (dificuldade para respirar, perda de olfato ou paladar, náusea, nariz escorrendo, dor de cabeça e fadiga), pelo menos três desses ocorrem simultaneamente. Vale destacar a possibilidade de tais sintomas não ocorrerem necessariamente de forma conjunta, sendo que, por exemplo, na terceira observação apresentada na [Tabela 3.28](#), o indivíduo apresentou somente náusea dentre os sintomas apontados.

Tabela 3.28: Características de 7 observações do mês de maio com 4 ou mais sintomas.

Febre	Tosse	Dor gar	Dif resp	Dor olhos	Dor cab	Nausea	Dor pei	Nariz	Fadiga	Perda olf	Dor mus	Faixa etaria	Sexo	Num sintomas
FEN	TOS	DGS	DRS	DOS	DCS	NAS	DPS	NZS	FAN	PON	DMN	FE3	F	NS4
FES	TOS	DGS	DRN	DON	DCN	NAN	DPS	NZN	FAN	PON	DMN	FE1	F	NS4
FES	TON	DGN	DRN	DOS	DCN	NAS	DPS	NZN	FAN	PON	DMS	FE4	F	NS4
FEN	TOS	DGN	DRN	DOS	DCN	NAN	DPS	NZN	FAN	PON	DMS	FE4	F	NS4
FEN	TON	DGN	DRS	DON	DCS	NAN	DPN	NZS	FAN	POS	DMS	FE3	M	NS4
FEN	TOS	DGS	DRS	DOS	DCS	NAN	DPS	NZS	FAN	PON	DMS	FE5	M	NS4
FES	TOS	DGS	DRS	DOS	DCN	NAS	DPS	NZS	FAS	POS	DMS	FE3	M	NS4

Além disso, para aqueles que não apresentaram nenhum dos sintomas do subgrupo destacado (segunda e quarta observação), foi identificado outro conjunto de sintomas relacionados. Estes referem-se a ocorrência de febre, tosse, dor de garganta e dor no peito. A partir da [Tabela 3.28](#), observa-se que pelo menos 2 dos 4 sintomas ocorreram simultaneamente nos indivíduos apresentados.

3.2.1 Questões incorporadas a partir do mês de agosto

Os resultados apresentados da análise de correspondência, correspondem às categorias referentes às questões em comum entre os três meses de estudo. Contudo, conforme mencionado na [Subseção 3.1.1](#), algumas questões foram incorporadas no questionário a partir do mês de agosto, as quais podem agregar informações para a análise de correspondência.

Dessa forma, são apresentados os planos fatoriais dos meses de agosto e novembro, considerando a inclusão das informações sobre se o indivíduo realizou o teste de COVID-19, assim como o resultado do mesmo. As codificações das categorias incluídas estão disponíveis na [Tabela 3.29](#).

Tabela 3.29: Codificação das categorias apresentadas na representação gráfica.

Categoria	Codificação
Fez teste	TES
Não fez teste	TEN
Sem resultado	RE0
Resultado Positivo	RE1
Resultado Negativo	RE2
Resultado Inconclusivo/Não recebeu	RE3

Inicialmente é preciso destacar que, de forma geral, a inclusão destas variáveis não alterou o comportamento apresentado sem as mesmas, seguindo o mesmo padrão dos gráficos apresentados nas Figuras [3.8](#) e [3.9](#). Porém, alguns pontos específicos são destacados na sequência.

Para a questão do resultado do teste, as categorias referentes ao resultado inconclusivo e os que ainda não tinham recebido o resultado foram agrupadas em uma única categoria, uma vez que o interesse é de analisar resultados positivos e negativos.

Conforme apresentado nas Figuras [3.13](#) e [3.14](#), observa-se que as categorias de resultado (Positivo, Negativo e Inconclusivo/Não recebeu) estão muito próximos entre si e dos que realizaram o teste de COVID-19, fato que é esperado, pois os resultados são respectivos aos testes realizados. Por outro lado, indivíduos que não realizaram o teste e, conseqüentemente, não possuem resultados, se encontram no quadrante oposto, próximos às categorias da ausência de sintomas.

Além disso, nota-se que a inclusão destas variáveis no plano fatorial provocou uma concentração dos pontos e uma maior distância dos que fizeram teste para COVID-19 e os respectivos resultados. Tal fato é justificado pela baixa frequência de indivíduos que realizaram o teste (cerca de 20%).

Capítulo 4

Conclusão

A partir do interesse em estudar os indivíduos participantes da PNAD COVID19, que apresentaram pelo menos um sintoma, para os meses de maio, agosto e novembro, foram identificadas características deste grupo.

Independente do tamanho amostral obtido para cada mês, observou-se que a quantidade de indivíduos com pelo menos um sintoma se reduziu ao longo dos meses de estudo. Tais indivíduos são, em maioria, de faixa etária entre 14 e 49 anos, sendo uma maior proporção do sexo feminino, de raça branca ou parda, ensino fundamental incompleto ou médio completo e grande parte da região Sudeste ou Nordeste.

As ocorrências mais frequentes relatadas pelos indivíduos foram dor de cabeça, nariz escorrendo e tosse. Além disso, a maior proporção de indivíduos apresentou até cerca de três sintomas conjuntamente, sendo que a maioria dos respondentes não procurou por um estabelecimento de saúde.

Na análise descritiva, algumas relações foram identificadas como, por exemplo, quanto maior a quantidade de sintomas apresentados, maior a proporção de pessoas que procuraram por um estabelecimento de saúde. De forma semelhante, essa relação com o aumento da quantidade de sintomas também foi observada para indivíduos que realizaram o teste de COVID-19, nos meses de agosto e novembro.

Em complemento, identificou-se que a maior proporção de resultados positivos dos testes, observados nos meses de agosto e novembro, foi referente aos indivíduos que apresentaram até três sintomas conjuntos.

Na análise de correspondência, o primeiro resultado a ser destacado é que a primeira dimensão está ordenada pelo número de sintomas, isto é, a orientação desta dimensão é dada por esta variável. Quanto à segunda dimensão, não há uma variável que se destaque

em todos os meses, observa-se que a faixa etária está completamente ordenada no mês de novembro e parcialmente ordenada nos demais meses. Portanto, com um certo cuidado na segunda dimensão, pode-se considerar que as duas variáveis ordinais presentes no estudo, definem a interpretação das duas primeiras dimensões.

Para além deste primeiro resultado, a análise de correspondência indicou uma associação entre dificuldade para respirar, nariz escorrendo, perda de olfato, dor de cabeça, náusea e fadiga com a categoria de quatro ou mais sintomas. Ou seja, em grande parte, dos indivíduos que apresentaram quatro ou mais sintomas, pelo menos três consistem nos sintomas citados.

Outra relação obtida foi entre a ocorrência de tosse, dor de garganta, febre e dor no peito, os quais demonstraram, de forma geral, estar presentes conjuntamente para pelo menos dois deste conjunto.

Em comparação, nos resultados apresentados pelo IBGE, os sintomas identificados por estarem mais associados à presença do vírus da COVID-19, foram conjugados de acordo com três possíveis categorias: perda de cheiro ou de sabor; tosse e febre e dificuldade para respirar; ou tosse e febre e dor no peito.

O interesse de verificar uma provável associação entre os respectivos sintomas destacados pelo IBGE com os resultados positivos de testes para COVID-19 não foi possível de ser realizado para os meses de agosto e novembro, justificado pela baixa frequência de indivíduos que realizaram o teste.

No caso das manifestações de perda de olfato ou paladar e dificuldade para respirar, estas demonstraram estar relacionadas, porém distantes de febre, tosse, dor de garganta e dor no peito. Logo, as categorias propostas pelo IBGE se assemelham aos resultados obtidos em relação à ocorrência de febre, tosse e dor no peito, os quais indicaram estar presentes de forma conjunta, para a maioria dos indivíduos.

Destaca-se ainda que, dado que a PNAD COVID19 foi realizada somente no ano de 2020, esta pesquisa poderia ter tido continuidade no ano seguinte, em que o número de casos e, conseqüentemente, de sintomas, aumentou de forma significativa. Nesse caso, o IBGE poderia ter realizado a pesquisa, incorporando, por exemplo, novas questões relacionadas à vacinação, uma vez que no ano de 2021 surgiram as primeiras vacinas na prevenção de complicações causadas pelo coronavírus.

Finalmente, é possível afirmar que este trabalho atingiu o seu objetivo. Dificuldades foram encontradas ao longo da análise dado que variáveis de interesse não puderam

ser incorporadas na Análise Fatorial de Correspondência Múltipla devido à presença de categorias de baixa frequência, fato que traria resultados inconclusivos causados pelo impacto das mesmas no mapa fatorial. Contudo, considerando as variáveis analisadas, foram identificados, como demonstram as conclusões apresentadas, tanto na análise descritiva quanto na análise de correspondência, comportamentos e características dos indivíduos frente à COVID-19, particularmente no caso daqueles que apresentaram pelo menos um dos possíveis sintomas da mesma.

Pode ser ainda importante destacar que foram observados aspectos relacionados à COVID-19 que não mereceram destaque nos demais trabalhos desenvolvidos neste tema, a maioria deles direcionados a construir modelos para prever o número de casos e de mortes em função da doença. Por outro lado, como tudo nesta pandemia, que a todos foi tomada de surpresa, importantes aprendizados foram acumulados, inclusive e principalmente, da necessidade de desde o princípio buscar informações que possam melhor entender o comportamento dos indivíduos frente a este contexto e, desta forma, orientar a definição de políticas públicas que possam minimizar os nefastos efeitos que se fizeram presentes no caso da COVID-19, em particular, a perda, até o momento deste trabalho, de aproximadamente 700 mil vidas.

Referências Bibliográficas

- Barbosa, M. T. S. e Alves, D. d. S. B. (2021). As bases de dados oficiais e as bibliotecas do r: Instrumentos para o ensino de uma estatística cívica. *Revista Brasileira de Estatística*, **79**(246), 40–63.
- Fávero, L. P. e Belfiore, P. (2017). *Manual de análise de dados: estatística e modelagem multivariada com Excel®*, *SPSS®* e *Stata®*. Elsevier Brasil.
- Ferreira Filho, P., Bereta, E. M. P. e Ribeiro, F. B. (1998). Tabela de burt. Relatório Técnico 04 - Série C - Notas Didáticas, DEs-UFSCar, São Carlos, SP.
- Gruber, A. (2020). Covid-19: o que se sabe sobre a origem da doença. *Jornal da Universidade de São Paulo*.
- IBGE (2020a). Pnad - pesquisa nacional por amostra de domicílios. Disponível em: < <https://www.ibge.gov.br/estatisticas/sociais/populacao/9127-pesquisa-nacional-por-amostra-de-domicilios.html?=&t=o-que-e>>.
Acessado em: 26 jan 2022.
- IBGE (2020b). Pesquisa nacional por amostra de domicílios - pnad covid19: Downloads. Disponível em: < https://www.ibge.gov.br/estatisticas/investigacoes-experimentais/estatisticas-experimentais/27946-divulgacao-semanal-pnadcovid1?t=downloads&utm_source=covid19&utm_medium=hotsite&utm_campaign=covid_19>. Acessado em: 26 jan 2022.
- IBGE (2020c). Pnad covid19 – plano amostral e ponderação. Disponível em: < <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101726.pdf>>. Acessado em: 13 mar 2022.
- IBGE (2020d). Pesquisa nacional por amostra de domicílios - pnad covid19: maio/2020: resultados mensais / ibge, coordenação de trabalho e rendimento. Disponível

em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv101727.pd>>.

Acessado em: 26 jan 2022.

Johnson, R. A., Wichern, D. W. *et al.* (2014). *Applied multivariate statistical analysis*, volume 6. Pearson London, UK:.

Lapostolle, F., Schneider, E., Vianu, I., Dollet, G., Roche, B., Berdah, J., Michel, J., Goix, L., Chanzy, E., Petrovic, T. *et al.* (2020). Clinical features of 1487 covid-19 patients with outpatient management in the greater paris: the covid-call study. *Internal and emergency medicine*, **15**, 813–817.

Lumley, T. (2004). Analysis of Complex Survey Samples. *Journal of Statistical Software*, **9**(8).

Mingoti, S. A. (2005). *Análise de dados através de métodos estatística multivariada: uma abordagem aplicada*, páginas 257–266. Editora UFMG.

MS (2020). O que é a covid-19? Disponível em: <<https://www.gov.br/saude/pt-br/coronavirus/o-que-e-o-coronavirus>>. Acessado em: 27 jan 2022.

Sanar (2020). Linha do tempo do coronavírus no brasil. Disponível em: <<https://www.sanarmed.com/linha-do-tempo-do-coronavirus-no-brasil>>. Acessado em: 28 jan 2022.

SIDRA (2020). Pesquisa nacional por amostra de domicílios contínua anual. Disponível em: <<https://sidra.ibge.gov.br/pesquisa/pnadca/tabelas>>. Acessado em: 21 set 2022.

Apêndice A

Dicionário

Tabela A.1: Dicionário dos dados utilizados.

Questão	PNAD	Trabalho
Unidade de Federação	UF	Norte, Nordeste, Centro-Oeste, Sudeste e Sul
Idade	em anos	0-13
		14-29
		30-49
		50-59
		60+
Sexo	Homem	M
	Mulher	F
Raça	Branca	Branca
	Preta	Preta
	Amarela	Amarela
	Parda	Parda
	Indígena	Indígena

Continua na próxima página

Tabela A.1 – *Continuação da tabela*

Variável	Categorias	Codificação
Escolaridade	Sem instrução	Sem instrução
	Fundamental incompleto	Fundamental incompleto
	Fundamental completo	Fundamental completo
	Médio incompleto	Médio incompleto
	Médio completo	Médio completo
	Superior incompleto	Superior incompleto
	Superior completo	Superior completo
	Pós graduação, Mestrado ou Doutorado	Pós graduação, Mestrado ou Doutorado
Teve febre	Sim	1
	Não	0
	Não sabe	0
Teve tosse	Sim	1
	Não	0
	Não sabe	0
Teve dor de garganta	Sim	1
	Não	0
	Não sabe	0
Teve dificuldade para respirar	Sim	1
	Não	0
	Não sabe	0
Teve dor de cabeça	Sim	1
	Não	0
	Não sabe	0
Teve dor no peito	Sim	1
	Não	0
	Não sabe	0
Teve náusea	Sim	1
	Não	0
	Não sabe	0

Continua na próxima página

Tabela A.1 – *Continuação da tabela*

Variável	Categorias	Codificação
Teve nariz escorrendo	Sim	1
	Não	0
	Não sabe	0
Teve fadiga	Sim	1
	Não	0
	Não sabe	0
Teve dor nos olhos	Sim	1
	Não	0
	Não sabe	0
Teve perda de olfato ou paladar	Sim	1
	Não	0
	Não sabe	0
Teve dor muscular	Sim	1
	Não	0
	Não sabe	0
Teve diarreia	Sim	1
	Não	0
	Não sabe	0
Quantidade de sintomas	-	1-13
Foi ao estabelecimento de saúde	Sim	Sim
	Não	Não
	Ignorado	
Providência - ficar em casa	Sim	Sim
	Não	Outros
	Ignorado	
Providência - ligar para profissional	Sim	Sim
	Não	Outros
	Ignorado	

Continua na próxima página

Tabela A.1 – *Continuação da tabela*

Variável	Categorias	Codificação
Providência - tomar remédio por conta própria	Sim	Sim
	Não	Outros
	Ignorado	
Providência - tomar remédio por orientação médica	Sim	Sim
	Não	Outros
	Ignorado	
Providência - receber visita de profissional do SUS	Sim	Sim
	Não	Outros
	Ignorado	
Providência - receber visita de profissional particular	Sim	Sim
	Não	Outros
	Ignorado	
Providência - outras	Sim	Sim
	Não	Outros
	Ignorado	
Local - posto de saúde	Sim	Sim
	Não	Não
	Ignorado	
Local - pronto socorro SUS	Sim	Sim
	Não	Não
	Ignorado	
Local - hospital SUS	Sim	Sim
	Não	Não
	Ignorado	
Local de atendimento - ambulatório/ consultório privado	Sim	Sim
	Não	Não
	Ignorado	
Local - pronto socorro privado	Sim	Sim
	Não	Não
	Ignorado	

Continua na próxima página

Tabela A.1 – *Continuação da tabela*

Variável	Categorias	Codificação
Local - hospital privado	Sim	Sim
	Não	Não
	Ignorado	
Foi internado	Sim	
	Não	Sim
	Não foi atendido	Outros
	Ignorado	
Foi entubado	Sim	Sim
	Não	Outros
	Ignorado	
Possui plano de saúde	Sim	Sim
	Não	Outros
	Ignorado	
Fez teste para COVID-19	Sim	Sim
	Não	Não
	Ignorado	
Fez teste de cotonete	Sim	Sim
	Não	Outros
	Ignorado	
Fez teste de sangue 1	Sim	Sim
	Não	Outros
	Ignorado	
Fez teste de sangue 2	Sim	Sim
	Não	Outros
	Ignorado	
Resultado do teste	Positivo	Positivo
	Negativo	Negativo
	Inconclusivo	Inconclusivo
	Ainda não recebeu	Ainda não recebeu

Continua na próxima página

Tabela A.1 – *Continuação da tabela*

Variável	Categorias	Codificação
Resultados positivos	-	1 - Se positivo para alguns dos testes
		0 - caso contrário
Restrição	Nenhuma	Nenhuma
	Reduziu contato	Reduziu contato
	Saiu só necessário	Saiu só necessário
	Ficou em casa	Ficou em casa

Fim da tabela

Apêndice B

Coordenadas e contribuições da Análise de Correspondência

Nas Tabelas [B.1](#) e [B.2](#) algumas categorias apresentaram seus respectivos valores de coordenadas ou contribuições iguais a zero. Tal fato ocorre devido à aproximação utilizada considerando duas casas decimais, nas quais estas categorias apresentaram valores significativos após a segunda casa decimal.

Tabela B.1: Coordenadas das categorias para as dimensões de acordo com os meses.

	Coordenadas					
	Maio		Agosto		Novembro	
	Dim 1	Dim 2	Dim 1	Dim 2	Dim 1	Dim 2
FEN	-0,16	-0,03	-0,15	-0,04	-0,13	0,04
FES	0,58	0,10	0,58	0,15	0,49	-0,16
TON	-0,15	-0,09	-0,15	-0,11	-0,15	0,13
TOS	0,32	0,19	0,29	0,21	0,26	-0,22
DGN	-0,14	-0,03	-0,15	-0,08	-0,15	0,09
DGS	0,46	0,10	0,37	0,19	0,33	-0,19
DRN	-0,11	-0,01	-0,11	0,00	-0,10	-0,01
DRS	0,73	0,07	0,85	-0,02	0,82	0,04
DON	-0,19	0,15	-0,18	0,09	-0,16	-0,07
DOS	0,22	-0,18	0,23	-0,12	0,23	0,09
DCN	-0,12	0,00	-0,10	0,01	-0,08	-0,01
DCS	0,94	-0,01	1,08	-0,12	1,04	0,18
NAN	-0,10	0,01	-0,09	0,02	-0,09	-0,02
NAS	0,91	-0,05	0,94	-0,15	0,87	0,16
DPN	-0,10	-0,09	-0,10	-0,11	-0,11	0,12
DPS	0,19	0,17	0,16	0,18	0,18	-0,20
NZN	-0,16	0,01	-0,15	0,02	-0,15	-0,03
NZS	0,86	-0,04	0,74	-0,08	0,64	0,11
FAN	-0,12	0,01	-0,10	0,01	-0,09	-0,02
FAS	0,84	-0,08	0,96	-0,14	0,98	0,18
PON	-0,15	0,00	-0,11	0,01	-0,09	-0,01
POS	0,76	0,01	0,89	-0,05	0,82	0,09
DMN	-0,18	0,04	-0,17	0,04	-0,15	-0,06
DMS	0,46	-0,11	0,51	-0,13	0,45	0,17
DIN	-	-	-0,07	0,01	-0,06	-0,02
DIS	-	-	0,68	-0,14	0,55	0,18
FE1	-0,18	0,53	-0,15	0,43	-0,13	-0,38
FE2	0,01	0,01	0,04	0,06	0,04	-0,12
FE3	0,05	-0,13	0,05	-0,07	0,07	0,05
FE4	0,02	-0,08	0,02	-0,13	0,01	0,11
FE5	-0,05	0,03	-0,06	-0,10	-0,09	0,17
M	-0,02	0,16	-0,02	0,11	-0,01	-0,09
F	0,02	-0,11	0,01	-0,08	0,01	0,06
NS1	-0,36	-0,01	-0,34	-0,12	-0,35	0,15
NS2	-0,17	0,01	-0,16	0,09	-0,17	-0,09
NS3	0,05	0,00	0,03	0,19	0,02	-0,19
NS4	0,86	0,02	0,83	0,04	0,77	-0,05

Tabela B.2: Contribuições das categorias para as dimensões de acordo com os meses.

	Contribuições					
	Maio		Agosto		Novembro	
	Dim 1	Dim 2	Dim 1	Dim 2	Dim 1	Dim 2
FEN	1,64	0,48	1,51	0,86	1,37	0,92
FES	5,98	1,77	5,83	3,32	5,08	3,44
TON	1,27	4,23	1,28	5,93	1,41	6,62
TOS	2,74	9,15	2,41	11,16	2,42	11,33
DGN	1,26	0,58	1,31	3,02	1,50	3,37
DGS	4,11	1,88	3,33	7,67	3,30	7,45
DRN	0,90	0,08	0,85	0,00	0,86	0,02
DRS	5,89	0,53	6,77	0,03	7,14	0,14
DON	1,56	10,36	1,47	3,25	1,56	1,67
DOS	1,79	11,92	1,92	4,25	2,16	2,31
DCN	0,97	0,00	0,73	0,08	0,63	0,12
DCS	7,90	0,02	8,15	0,88	7,88	1,48
NAN	0,77	0,03	0,66	0,15	0,68	0,16
NAS	6,89	0,25	6,72	1,51	6,78	1,56
DPN	0,54	4,33	0,48	5,44	0,72	5,65
DPS	0,99	7,99	0,79	8,84	1,18	9,30
NZN	1,71	0,03	1,64	0,17	1,83	0,36
NZS	9,33	0,17	7,95	0,84	7,78	1,52
FAN	1,09	0,09	0,81	0,14	0,70	0,15
FAS	7,42	0,59	7,53	1,34	7,84	1,73
PON	1,48	0,00	0,98	0,03	0,79	0,06
POS	7,67	0,02	7,64	0,23	6,84	0,50
DMN	1,91	1,10	1,79	0,94	1,73	1,52
DMS	4,87	2,80	5,39	2,83	5,08	4,47
DIN	-	-	0,41	0,14	0,35	0,24
DIS	-	-	3,81	1,33	3,02	2,13
FE1	0,25	21,32	0,23	15,62	0,19	10,77
FE2	0,00	0,04	0,03	0,50	0,04	1,90
FE3	0,09	4,86	0,08	1,29	0,17	0,57
FE4	0,01	0,73	0,01	1,75	0,00	1,08
FE5	0,04	0,17	0,05	1,46	0,17	3,94
M	0,01	8,34	0,01	3,42	0,004	1,99
F	0,01	6,03	0,01	2,48	0,003	1,44
NS1	4,76	0,05	4,25	4,36	4,74	5,44
NS2	0,50	0,01	0,52	1,30	0,67	1,30
NS3	0,02	0,00	0,01	3,21	0,01	3,00
NS4	13,64	0,06	12,67	0,21	13,40	0,37

Apêndice C

Códigos utilizados

```
library(COVIDIBGE)
```

```
library(survey)
```

```
library(dplyr)
```

```
library(convey)
```

```
library(FactoMineR)
```

```
library(factoextra)
```

```
library(GDAtools)
```

```
library(ggrepel)
```

```
#código base para seleção do conjunto de dados#
```

```
covid_maio<- get_covid(2020,5)
```

```
covid_agosto<- get_covid(2020,8)
```

```
covid_novembro<- get_covid(2020,11)
```

```
sin_maio<-covid_maio$variables %>%
```

```
  mutate(febre = ifelse(B0011=="Sim",1,0),
```

```
         tosse = ifelse(B0012=="Sim",1,0),
```

```
         dor_gar = ifelse(B0013=="Sim",1,0),
```

```
         dif_resp = ifelse(B0014=="Sim",1,0),
```

```
         dor_cab = ifelse(B0015=="Sim",1,0),
```

```
         dor_pei = ifelse(B0016=="Sim",1,0),
```

```
         nausea = ifelse(B0017=="Sim",1,0),
```

```
         nariz = ifelse(B0018=="Sim",1,0),
```

```
         fadiga = ifelse(B0019=="Sim",1,0),
```

```
         dor_olhos = ifelse(B00110=="Sim",1,0),
```

```
         perda_olfpa = ifelse(B00111=="Sim",1,0),
```

```
         dor_mus = ifelse(B00112=="Sim",1,0),
```

```

total_sin = febre+tosse+dor_gar+dif_resp+
  dor_cab+dor_pei+nausea+nariz+fadiga+dor_olhos+
  perda_olfpa+dor_mus)%>%
filter(total_sin>0)

#código base análise descritiva com survey#
tes_maio<- svydesign(ids=~UPA, strata = ~Estrato, weights = ~V1032, data = sin_maio)

prop.table(svytable(~febre, design = tes_maio)*100)

#região
est_sudes<-c("São Paulo","Rio de Janeiro","Minas Gerais","Espírito Santo")

est_sul<-c("Paraná","Rio Grande do Sul","Santa Catarina")

est_norte<-c("Pará","Amazonas","Acre","Rondônia", "Roraima","Amapá","Tocantins")

est_nordeste<-c("Maranhão","Piauí","Ceará","Rio Grande do Norte", "Paraíba",
  "Pernambuco","Alagoas","Sergipe","Bahia")

est_centro<-c("Mato Grosso","Mato Grosso do Sul","Goiás","Distrito Federal")

regioes_maio<- sin_maio)%>%
mutate("Região"=ifelse(is.element(UF, est_centro),"Centro Oeste",
  ifelse(is.element(UF, est_norte),"Norte",
    ifelse(is.element(UF, est_nordeste),"Nordeste",
      ifelse(is.element(UF, est_sul),"Sul",
        ifelse(is.element(UF, est_sudes),
          "Sudeste",""))))))))

tesreg_maio<- svydesign(ids=~UPA, strata = ~Estrato, weights = ~V1032,
data = regioes_maio)

#prop região
reg_maio<-round(prop.table(svytable(~Região, design = tesreg_maio))*100,2)

#separando por faixas etárias#
zero_13<-sin_maio)%>%filter(A002>=0 & A002<=13)%>%
  mutate("Faixa"= "0-13")
treze_29<-sin_maio)%>%
  filter(A002>13 & A002<=29)%>%

```

```

mutate("Faixa"= "14-29")
vinte9_49<-sin_maio%>%
  filter(A002>29 & A002<=49)%>%
  mutate("Faixa"= "30-49")
quarenta9_59<-sin_maio%>%
  filter(A002>49 & A002<=59)%>%
  mutate("Faixa"= "50-59")
sessenta_mais<-sin_maio%>%
  filter(A002>=60)%>%
  mutate("Faixa"= "60+")

maio_etaria<- rbind(zero_13,treze_29,vinte9_49,quarenta9_59,sessenta_mais)
maio_etaria$Faixa<- as.factor(maio_etaria$Faixa)

etaria_maio<- svydesign(ids=~UPA, strata = ~Estrato, weights = ~V1032,
data = maio_etaria)

#exemplo análise bivariada
#plano saúde
plano_sin_maio<-round(prop.table(svytable(~total_sin+B007, design =
tes_maio))*100,2)

#código base para análise de correspondência
sin_maio_rec<- maio_etaria%>%
  mutate(numero_sin = ifelse(total_sin==1,1,
                              ifelse(total_sin==2,2,
                                        ifelse(total_sin==3,3,
                                                "4+"))))

maio_fator<- data.frame(as.factor(sin_maio_rec$febre),
as.factor(sin_maio_rec$tosse),as.factor(sin_maio_rec$dor_gar),
as.factor(sin_maio_rec$dif_resp),as.factor(sin_maio_rec$dor_cab),
as.factor(sin_maio_rec$dor_pei),as.factor(sin_maio_rec$nausea),
as.factor(sin_maio_rec$nariz),as.factor(sin_maio_rec$fadiga),
as.factor(sin_maio_rec$dor_olhos),as.factor(sin_maio_rec$perda_olfpa),
as.factor(sin_maio_rec$dor_mus),as.factor(sin_maio_rec$Faixa),
as.factor(sin_maio_rec$A003),as.factor(sin_maio_rec$numero_sin))

colnames(maio_fator)<-c("Febre","Tosse",
"Dor gar","Dif resp",

```

```

      "Dor olhos", "Dor cab",
      "Nausea", "Dor pei", "Nariz",
      "Fadiga", "Perda olf", "Dor mus",
      "Faixa etaria", "Sexo", "Num sintomas")

#exemplo de codificação de variáveis
levels(maio_fator$Febre)[levels(maio_fator$Febre)%in% "0"] = "FEN"
levels(maio_fator$Febre)[levels(maio_fator$Febre)%in% "1"] = "FES"

options(ggrepel.max.overlaps = Inf)

ac_maio<- MCA(maio_fator, method="Burt")
fviz_mca_var(ac_maio,repel = TRUE ,labelsize=4 ,col.var = "cos2",
gradient.cols=c("grey48","blue","red"))+ labs(title = "")+
  theme_light()+
  theme(panel.grid = element_blank())+
  scale_x_continuous(breaks = c(-0.4,0,0.4,0.8,1.0))

#contribuição das variáveis dim 1 e 2
var_maio<-get_mca_var(ac_maio)

coord_maio<-var_maio$coord[,1:2]#coordenadas
xtable::xtable(coord_maio)

contrib_maio<-var_maio$contrib[,1:2]#contribuição(%)
xtable::xtable(contrib_maio)

#primeiro plano
fviz_contrib(ac_maio, choice = "var",axes = 1:2 )+#top=
  theme(plot.title = element_text(hjust = 0.5))+
  labs(title = "1º Plano Fatorial", y="Contribuição (%)")

#filtrando 7 observações com NS4 para mostrar o comportamento
NS4_maio<- maio_fator%>%
  filter(`Num sintomas`== "NS4")

set.seed(21)
dim(NS4_maio)
split<-sample(1:8293,size=7, replace=FALSE)

xtable::xtable(NS4_maio[split,])

```