

---

Universidade Federal da Paraíba  
Centro de Ciências Exatas e da Natureza  
Departamento de Estatística

Fatores de risco para diagnóstico de COVID-19 no  
Estado da Paraíba considerando o planejamento  
amostral para pesquisa de testagem domiciliar realizada  
no ano de 2020 - Pesquisa Continuar Cuidando

Milleny Fernandes de Sena

Julho/2021

---

Milleny Fernandes de Sena

**Fatores de risco para diagnóstico de COVID-19 no Estado da Paraíba considerando o planejamento amostral para pesquisa de testagem domiciliar realizada no ano de 2020 - Pesquisa Continuar Cuidando**

Trabalho de Conclusão do Curso apresentado no Curso de Bacharelado em Estatística da UFPB, com vistas para a conclusão da disciplina TCC II e obtenção do grau de Bacharela em Estatística da UFPB.

Orientador: Prof. Dr. Hemílio Fernandes Campos Coelho

**João Pessoa  
Julho de 2021**

**Catálogo na publicação**  
**Seção de Catalogação e Classificação**

S474f Sena, Milleny Fernandes de.

Fatores de risco para diagnóstico de COVID-19 no Estado da Paraíba considerando o planejamento amostral para pesquisa de testagem domiciliar realizada no ano de 2020 - Pesquisa Continuar Cuidando / Milleny Fernandes de Sena. - João Pessoa, 2021.

66 f. : il.

Orientação: Hemílio Fernandes Campos Coelho.  
TCC (Graduação/Bacharelado em Estatística) -  
UFPB/CCEN.

1. Regressão. 2. Covid19 - Medidas preventiva. 3. Odds Ratio. I. Coelho, Hemílio Fernandes Campos. II. Título.

UFPB/CCEN

CDU 519.246.8(043.2)

**Fatores de risco para diagnóstico de COVID-19 no Estado da Paraíba  
considerando o planejamento amostral para pesquisa de testagem domiciliar  
realizada no ano de 2020 - Pesquisa Continuar Cuidando**

Milleny Fernandes de Sena

Banca Examinadora

---

Prof. Dr. Hemílio Fernandes Campos Coelho (Orientador)  
Universidade Federal da Paraíba

---

Prof. Dr. Pedro Luis do Nascimento Silva (Membro Externo)  
Escola Nacional de Ciências Estatísticas

---

Profa. Dra. Gilmara Alves Cavalcanti (Membro Interno)  
Universidade Federal da Paraíba

CONCEITO FINAL: \_\_\_\_\_

*Dedico este trabalho a minha família e aos amigos  
que fiz durante minha trajetória acadêmica.*

## AGRADECIMENTOS

Agradeço primeiramente a Deus por todas as oportunidades concedidas, pelos dons compartilhados para a realização de mais uma etapa. Pelo apoio e consolo em situações mais delicadas e forças para superar desafios encontrados durante a trajetória acadêmica.

Agradeço a minha base familiar, mãe (Maria Cilene), pai (Izaias Messias) e irmã (Maria Sophia), por todo o suporte e ajuda. Suas abdições durante esses anos proporcionaram todas as conquistas que venho realizando. Seus conselhos, momentos espontâneos e confiança me incentivaram muito mais que merecido, eu amo vocês.

Aos meus queridos avós, obrigado por acreditarem no meu potencial e seus ensinamentos são ouro na minha vida. Mesmo que não presencialmente, permanecem sempre comigo Dona Lúcia, Dona Rita, seu Zezinho e em especial seu Severino, os senhores são minha perseverança.

Agradeço aos meus familiares, tios, tias e primos, vocês são muitos especiais e agradeço a toda ajuda direta ou indireta durante minha caminhada. Em especial, o exemplo de união familiar.

Professor Hemílio, não tenho palavras para agradecer ao senhor por toda ajuda, apoio, conselhos, incentivos e amizade desde o início. A confiança do senhor em meu trabalho é a base para grande parte da profissional que venho me tornando. Sua compreensão e indicações dentro da instituição tornou viável muitas dessas conquistas, o senhor é fera.

Um agradecimento aos docentes do Departamento de Estatística e amigos por todo conhecimento e conselhos compartilhados. Em especial Tarciana, Everlane, Maria Lídia, Ulisses, Ana Flávia, Pedro, Cláudio, Luiz, Eufrásio, Rodrigo, Telmo e Marcelo, os senhores são exemplos de profissionais. Queria relatar também minha gratidão a professora Gilmara pela participação na banca avaliadora e pelo auxílio durante os processos nessa reta final.

Ao Professor Gualberto, agradeço demais seus ensinamentos e conselhos. O senhor é um grande exemplo como coordenador, profissional e professor. E ao professor Pedro Luiz, pela participação na banca avaliadora, pelos ensinamentos e esclarecimentos recebidos, sua paciência e amplo conhecimento são admiráveis.

A querida secretária do departamento, Dona Ione. Agradeço muito pelas dicas, esclarecimentos, conversas e risadas.

Agradeço aos meus amigos de infância e adolescência, a força do laço que se manteve entre tantas situações me ajudaram e divertiram em muitos momentos. Samara, Layanne, Rebeka, Valentina, Abi, Emanuelli, Fernanda, Fabiana, João Paulo, Natan, José e Osmar.

Aos irmãos do apartamento 306, transformaram minha casa em um local especial e de muito aprendizado, aos caros Samara, João Paulo, Natan e Zé meu agradecimento pelas noites de estudos, receitas atípicas da madrugada, suporte e amizade, vocês são para a vida.

Aos meus amigos conquistados durante a graduação, é difícil esse momento de despedida diária, mas são base e contato para a vida toda, Fernanda, Dani, Natasha, Netto, Mateus, Ulysses, Juliana, Gleize, Franciele e Kleber. Futuros amigos e parceiros de trabalhos, vocês são incríveis.

Ao meu parceiro, modelo de amigo e companheiro, agradeço a todo o suporte nessa reta final. Obrigado por seu cuidado, carinho, compreensão e por estar comigo em todos os momentos, você é essencial na minha vida Borges.

Agradeço aos meus colegas de trabalho da Tely pelos ensinamentos e aprendizados, tem sido uma experiência muito importante para meu desenvolvimento profissional e pessoal. Ao meu gestor Adriano, e meus colegas de trabalho Aline, Felipe Gomes, Felippy Silveira, Wesley, João, Bruno, são exemplos de competência e sabedoria.

Agradeço a instituição Universidade Federal da Paraíba que tornou possível todas essas realizações.

*"Não importa o que aconteça, continue a nadar."  
(WATERS, GRAHAM; PROCURANDO NEMO,  
2003)*



A Organização Mundial da Saúde (OMS) declarou no Brasil, em fevereiro de 2020, que o surto da doença causada pelo novo Corona vírus (SARS-CoV-2, também conhecido como COVID-19) constitui uma emergência de saúde pública de importância internacional, e em março deste mesmo ano a doença foi caracterizada como uma pandemia. Na Paraíba, até o dia 23 de outubro do mesmo ano foram registrados 130.000 casos e 3.038 óbitos confirmados. Os sintomas iniciais da doença incluem febre, tosse, dispneia, fadiga, secreção respiratória e alguns casos podem apresentar sintomas gastrointestinais, como diarreia. A taxa de transmissão da COVID-19 é relativamente alta, a maioria das pessoas contaminadas apresentam sintomas leves, mas as que precisam de atendimento hospitalar, que são casos de moderados a graves, acabam causando uma sobrecarga no sistema de saúde, para assistência e monitoramento. Diante da disseminação do vírus por todo território paraibano, a Pesquisa Continuar Cuidando foi elaborada com o objetivo de analisar a evolução da Pandemia da COVID-19 no Estado da Paraíba e, de modo a conhecer a forma como o vírus afetou e está afetando a vida da população do Estado, e fazer recomendações sobre a sustentação das medidas preventivas para impedir o crescimento do número de casos e de óbitos e manter os avanços já obtidos em todo Estado. As estimativas da pesquisa Continuar Cuidando foram calculadas considerando os pesos básicos de todas as unidades considerando o planejamento amostral realizado. Na pesquisa, não foram abordados estudos sobre identificação de fatores relacionados à infecção pela doença ou para o resultado de um teste específico, o que motiva o presente estudo apresentado nesta monografia. Foram utilizadas fontes primárias da pesquisa, como as informações socioeconômicas, medidas de prevenção consideradas, sintomas recentes e comorbidades dos participantes para buscar evidência de quais destas informações estão relacionadas com a infecção causada pela COVID-19, baseada no resultado do teste de anticorpo IgG. Foi considerado um modelo de regressão logística baseado no planejamento amostral, a fim de avaliar os fatores que possam ser determinantes para este desfecho. Dessa forma, o

presente trabalho permite que seja informado também à população do estado a importância das recomendações básicas no combate a pandemia. Mediante aos dados analisados é possível visualizar que o SARS-CoV-2 é altamente contagioso e com isso as estimativas pontuais evidenciaram o reforço necessário para as medidas preventivas e sobre as conscientizações que a população necessita ter ao enfrentar o cenário pandêmico. Desta forma neste presente trabalho realizado as variáveis analisadas influenciaram no resultado do teste de COVID-19.

**Palavras-chaves:** Covid19, Medidas Preventivas, Odds Ratio, Regressão.

## ABSTRACT

The World Health Organization (OMS) declared in Brazil, in February 2020, that the outbreak of the disease caused by the new coronavirus (SARS-CoV-2, also known as COVID-19) constitutes a public health emergency of international importance, and in March of that same year the disease was characterized as a pandemic. In Paraíba, until the 23rd of October of the same year, 130,000 cases and 3,038 confirmed deaths were registered. The initial symptoms of the disease include fever, cough, dyspnoea, fatigue, respiratory secretions and some cases may present with gastrointestinal symptoms such as diarrhea. The transmission rate of COVID-19 is relatively high, most infected people have mild symptoms, but those who need hospital care, which are moderate to severe cases, end up causing an overload on the health system, for assistance and monitoring. In the light of the spread of the virus throughout paraiban territory, the Continuing Care Survey was developed with the aim of analyzing the evolution of COVID-19 Pandaemia in the State of Parahiba and, in order to know how the virus has affected and is affecting the life of the state population, and make recommendations on the support of preventive measures to prevent the number of cases and deaths from growing and to maintain progress already made across the state. Estimates of the Continuar Cuidando survey were calculated considering the basic weights of all units considering the sampling planning carried out. In the research, studies on the identification of factors related to infection by the disease or the result of a specific test were not addressed, which motivates the present study presented in this monograph. Primary research sources, such as socioeconomic information, considered prevention measures, recent symptoms and comorbidities of the participants, were used to seek evidence of which of these factors might be related to the infection caused by COVID-19, based on the result of the IgG antibody test. A logistic regression model fitted considering the sample design was considered in order to assess the factors that may be determining for this outcome. Thus, the present work also allows the population of the state to be informed of the importance of basic recommendations

in combating the pandemic. Through the analyzed data, it is possible to see that SARS-CoV-2 is highly contagious and with that, the specific estimates showed the necessary reinforcement for preventive measures and on the awareness that the population needs to have when facing the pandemic scenario. Thus, in this present work, the analyzed variables influenced the result of the COVID-19 test.

**Keywords:** Covid19, Preventive Measures, Odds Ratio, Regression.

<b>Lista de Figuras</b>	<b>x</b>
<b>Lista de Tabelas</b>	<b>xi</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Objetivos . . . . .	3
1.1.1 Principal . . . . .	3
1.1.2 Específicos . . . . .	3
<b>2 Pesquisa Continuar Cuidando</b>	<b>4</b>
2.1 Apresentação . . . . .	4
2.2 Planejamento amostral da pesquisa . . . . .	5
2.3 Ponderação da amostra da Pesquisa Continuar Cuidando . . . . .	7
<b>3 Método probabilístico de amostragem</b>	<b>10</b>
3.0.1 Amostragem de Bernoulli (BE) . . . . .	12
3.0.2 Amostragem Proporcional ao Tamanho (PPT) . . . . .	13
3.0.3 Amostragem Estratificada (AE) . . . . .	14
3.0.4 Amostragem Conglomerada (AC) . . . . .	16
<b>4 Modelo de Regressão Logística considerando o planejamento amostral</b>	<b>19</b>
4.1 Modelos Lineares Generalizados . . . . .	19
4.1.1 Componente aleatória . . . . .	20
4.1.2 Componente sistemática e função de ligação . . . . .	21
4.1.3 Função de ligação canônica . . . . .	22
4.1.4 Algoritmo de estimação . . . . .	22
4.2 Regressão Logística . . . . .	26
4.3 Modelo de regressão logística sob o planejamento amostral . . . . .	26

<b>5</b>	<b>Resultados e discussão</b>	<b>29</b>
5.0.1	Regressão logística . . . . .	38
<b>6</b>	<b>Conclusões</b>	<b>43</b>
6.1	Anexo . . . . .	47
6.1.1	Script . . . . .	47

## LISTA DE FIGURAS

2.1	Domínios de estudo considerados para estratificação da amostra . . . . .	6
2.2	Domínios de estudo considerados para estratificação da amostra . . . . .	7

## LISTA DE TABELAS

5.1	Distribuição do resultado do teste IgG segundo as variáveis socioeconômicas	32
5.2	Distribuição do resultado do teste IgG segundo as variáveis sociodemográficas	33
5.3	Distribuição do resultado do teste IgG segundo as variáveis de monitoramento dos participantes da pesquisa . . . . .	34
5.4	Distribuição do resultado do teste IgG segundo as variáveis de sintomas. . .	35
5.5	Distribuição do resultado do teste IgG segundo as variáveis sobre hábitos dos participantes durante a quarentena. . . . .	36
5.6	Distribuição do resultado do teste IgG segundo as variáveis sobre comorbidades. . . . .	37
5.7	Estimativas dos parâmetros do modelo logístico e avaliação dos riscos . . .	39
5.8	Estimativas Razão de Chances e intervalo de confiança . . . . .	41



## LISTA DE ABREVIATURAS

**BE** Amostragem Bernoulli

**AC** Amostragem Conglomerada

**AE** Amostragem Estratificada

**COVID-19** Coronavirus Disease 2019

**IgG** Imunoglobulina G

**IgM** Imunoglobulina M

**MLG** Modelos Lineares Generalizados

**MMV** Método de Máxima Verossimilhança

**OMS** Organização Mundial de Saúde

**OR** Odds Ratio

**PPT** Probabilidade Proporcional ao Tamanho

**UPA** Unidade Primária Amostral

# CAPÍTULO 1

## INTRODUÇÃO

As pandemias são conhecidas como epidemias que se espalham rapidamente por diversos países e afetam uma quantidade relativamente grande de pessoas (DUARTE, M. d. Q. et al.,2020) . De acordo com o tempo em que duram, novas regras e hábitos sociais para a população tem o intuito de conter a proliferação do vírus. Segundo dados da Organização Mundial de Saúde (OMS), o surto da COVID-19 iniciou-se na China em dezembro de 2019 na cidade de Wuhan, localizada na província de Hubei, na China. Esta cidade vivenciou um surto de pneumonia de causa desconhecida. Em janeiro de 2020, pesquisadores chineses identificaram um novo Corona vírus (SARS-CoV-2) como agente etiológico de uma síndrome respiratória aguda grave, denominada doença do Corona vírus 2019, ou simplesmente COVID-19 c.

No Brasil, os primeiros casos foram confirmados no mês de fevereiro de 2020, e diversas ações foram implementadas a fim de conter o avanço da doença. Em 3 de fevereiro de 2020, o país declarou Emergência de Saúde Pública de Importância Nacional (ESPIN), antes mesmo da confirmação do primeiro caso (CAVALCANTE, J. R. et al.,2020). A consolidação dos dados sobre casos e óbitos por COVID-19, coletados e disponibilizados pelas Secretarias Estaduais de Saúde, vem sendo realizada desde o início da pandemia pelo Ministério da Saúde brasileiro. Isso permite o conhecimento da dinâmica da doença no país e, conseqüentemente, o estabelecimento de políticas de contenção.

De acordo com os dados divulgados na OMS, a partir 8 novembro de 2020, um adicional de 11.874.437 casos confirmados de COVID-19 foram relatados globalmente, incluindo 166.688 mortes, elevando o número cumulativo de casos confirmados relatados globalmente para 49.578.590, incluindo 1.245.717 mortes (WASHINGTON, D.,2020). Na paraíba, até a primeira quinzena do mesmo mês, 140.982 pessoas já contraíram a doença, 112.947 já se

recuperaram e 3.232, infelizmente, faleceram. Até então, 444.658 testes para diagnóstico da Covid-19 tinham sido realizados.

As pessoas com COVID-19 podem apresentar sintomas como tosse, dificuldade para respirar, dores de garganta, febre e outros sintomas semelhantes à gripe e outras doenças respiratórias comumente detectadas em período de inverno, no Brasil. Há ainda os portadores do vírus que são assintomáticos, os quais possuem importância epidemiológica, dado que são potenciais transmissores da doença.

Para o diagnóstico do COVID-19, o teste confirmatório é a detecção do material genético do vírus, como o RNA viral, através do teste PCR em tempo real (RT-PCR). Alguns estudos têm mostrado também a importância dos métodos baseados em pesquisa de anticorpos no sangue para o diagnóstico da COVID-19 em indivíduos assintomáticos e sintomáticos. São chamados de testes de ensaio imune enzimático (ELISA), e permitem a detecção de anticorpos específicos. A duração média da detecção de anticorpos IgM (que indicam infecção na fase ativa, pois eles são os primeiros anticorpos a aparecer quando vírus ou bactérias nocivas atacam o nosso corpo) e IgA (proteína encontrada em grandes quantidades nas mucosas) para infecção pelo SARS-CoV-2 é de 5 dias, enquanto o anticorpo IgG (anticorpo que também é uma resposta a vírus e bactérias, porém atuam na fase mais tardia da infecção, indicando que a pessoa está em estado convalescente da doença) pode ser detectada em menos de 14 dias após o início dos sintomas (MAGNO, L. et al, 2020).

Nesta direção, o Governo do Estado da Paraíba, por meio da Secretaria de Estado de Saúde (SES), elaborou uma análise situacional e evolutiva da Pandemia da COVID-19 no Estado e recomendações quanto à necessária sustentação das medidas preventivas para impedir o crescimento do número de casos e de óbitos e manter os avanços já obtidos em todo Estado. A pesquisa foi chamada de **Pesquisa Continuar Cuidando**, e foi fruto de uma parceria entre o Governo do Estado da Paraíba, representado pelas suas secretarias de Saúde e de Educação, Ciência e Tecnologia, e Universidade Federal da Paraíba, representada pelo Observatório de Síndromes Respiratórias da UFPB. A pesquisa foi financiada pela FAPESQ-PB através do termo de cooperação técnica nº 198/2020 SEECT-FAPESQ. Além disso, a pesquisa também possui o Certificado de Apresentação de Apreciação Ética nº 39530620.0.0000.5186 na Plataforma Brasil, e foi devidamente aprovada pelo comitê de ética da Secretaria de Saúde do Estado da Paraíba.

A Pesquisa Continuar Cuidando teve duração de 8 semanas seguidas de coletas domiciliares, entre os meses de novembro e dezembro de 2020, sobre o comportamento da COVID-19 em meio às famílias paraibananas. De acordo com as determinações instruídas pelo Ministério da Saúde, as três medidas que mais protegem a saúde e a vida das pessoas

são: usar máscaras, lavar as mãos e manter o distanciamento social. Sendo decisões e gestos que precisão estar cada vez mais presentes em nosso cotidiano.

Na pesquisa, foram apresentadas diversas estimativas de prevalência para o Estado, e também levando em consideração informações sociodemográficas, comorbidades e informações referentes a cuidados, conforme apresentado por (COELHO H. F. C., 2021). Entretanto, na mesma não foram consideradas análises relacionadas ao uso de modelos estatísticos para identificação de fatores relacionados à infecção ou ao resultado de um teste de anticorpos. Dessa forma, o presente trabalho amplia a análise já feita na Pesquisa Continuar Cuidando, considerando agora o uso de um modelo estatístico levando em conta o planejamento amostral para identificar fatores que possam estar relacionados com o resultado do teste de anticorpo IgG.

## 1.1 Objetivos

### 1.1.1 Principal

Identificar os principais fatores relacionados ao diagnóstico positivo de COVID-19, referente ao resultado do anticorpo IgG;

### 1.1.2 Específicos

- Apresentar as medidas de associação, baseadas no planejamento amostral, entre o diagnóstico positivo de COVID-19 e variáveis demográficas, clínicas e de escalas de sintomas;
- Apresentar as estimativas relacionadas à chance de ter diagnóstico positivo de COVID-19 considerando o planejamento amostral;
- Apresentar os riscos referentes ao diagnóstico positivo de COVID-19 através as medidas como Odds Ratio.

### 2.1 Apresentação

A pesquisa Continuar Cuidando foi uma iniciativa do governo do Estado da Paraíba, por meio de suas secretarias de Saúde (SES) e de Educação, Ciência e Tecnologia (SEECT). O governo, representado por estas secretarias firmou uma parceria com a Universidade Federal da Paraíba, representada pelo Observatório de Síndromes Respiratórias da Universidade Federal da Paraíba, localizado no Departamento de Estatística da UFPB. A pesquisa foi financiada pela Fundação de Apoio à Pesquisa do estado da Paraíba (FAPESQ-PB), conforme termo de cooperação técnica nº 198/2020. A coleta de dados nos domicílios foi realizada pela SCIENCE - Sociedade para o Desenvolvimento da Pesquisa Científica (<https://science.org.br/>). Já a aplicação dos testes (rápido e RT-PCR) nos moradores dos domicílios foi realizada por profissionais de saúde das secretarias municipais de saúde, em acordo com a Secretaria de Saúde do Estado da Paraíba (SES – PB).

A coleta de dados da pesquisa foi realizada no período de 03/11/2020 a 22/12/2020. Todos os domicílios selecionados na pesquisa foram visitados sempre com uma dupla de pesquisadores, sendo um o pesquisador da SCIENCE que registrava as entrevistas em um tablet, e o outro um profissional de saúde, que ficava encarregado da realização do teste rápido e do teste RT-PCR. Todos os pesquisadores de campo foram devidamente treinados para a aplicação do questionário e os participantes da pesquisa precisaram assinar um Termo de Consentimento Livre e Esclarecido (TCLE), concordando com a realização da pesquisa no domicílio. O projeto de pesquisa foi devidamente submetido

ao comitê de ética com Certificado de Apresentação de Apreciação Ética (CAAE) de nº 39530620.0.0000.5186.

Os testes que foram empregados na pesquisa foram os seguintes:

- Teste rápido (Medtest Coronavirus IgG/IgM - sangue). Identifica se indivíduos possuem os anticorpos IgM (indica infecção ativa, ou seja, pessoa está infectada e transmitindo o vírus) e/ou IgG (indica que já ocorreu infecção no passado e que a pessoa já está curada da Covid-19);
- Teste RT-PCR (Swab nasofaringe). Identifica se indivíduos têm o vírus causador da COVID-19 para resultado positivo (detectável) e negativo (não- detectável).

A Pesquisa Continuar Cuidando forneceu estimativas de prevalências baseadas nos dois testes – três marcadores. O teste RT-PCR é considerado mais preciso para detecção de infecção atual pela Covid-19. Teste rápido fornecia resultados imediatos, que eram anotados no tablet e comunicados aos participantes da pesquisa durante abordagem no campo. O teste RT-PCR tinha amostras remetidas ao Laboratório Central de Saúde Pública da Paraíba (LACEN-PB), localizado na cidade de João Pessoa, para análise. Os resultados eram comunicados aos participantes dias depois.

## 2.2 Planejamento amostral da pesquisa

O plano amostral utilizado na Pesquisa Continuar Cuidando foi estratificado e conglomerado em dois estágios, a saber:

- **Estágio 1:** Amostragem de setores censitários em cada estrato, através de plano amostral do tipo PPT;
- **Estágio 2:** Arrolamento simultâneo e seleção de domicílios via Amostragem de Bernoulli.

A descrição dos planos amostrais que foram considerados será apresentada nas seções seguintes. Posteriormente, será apresentada a ponderação da amostra na Pesquisa Continuar Cuidando, considerando a combinação desses planos amostrais.

## Estratificação na Pesquisa Continuar Cuidando

O primeiro nível da estratificação foi definido levando em consideração o interesse em fornecer resultados para quatro domínios de estudo definidos como recortes do território do estado da Paraíba, descritos na Figura 2.1. Estes recortes foram definidos reconhecendo que a epidemia tem se espalhado pelo interior e que a gestão da saúde leva em consideração a divisão regional considerada. Esses quatro domínios de estudo correspondem à estratificação natural da amostra.

Figura 2.1: Domínios de estudo considerados para estratificação da amostra

<b>Estratos naturais</b>	<b>Setores</b>	<b>Domicílios</b>	<b>Pessoas</b>	<b>Setores na amostra por mês</b>	<b>Fração amostral setores (%)</b>	<b>Pessoas na amostra por mês</b>	<b>Fração amostral pessoas (por mil)</b>
João Pessoa	945	213.230	718.744	40	4,2	1.000	1,4
Macro 1 sem João Pessoa	1.566	298.353	1.054.552	44	2,8	1.100	1,0
Macro 2	1.576	309.309	1.060.620	48	3,0	1.200	1,1
Macro 3	1.384	254.318	897.480	60	4,3	1.500	1,7
<b>Total Geral</b>	<b>5.471</b>	<b>1.075.210</b>	<b>3.731.396</b>	<b>192</b>	<b>3,5</b>	<b>4.800</b>	<b>1,3</b>
<b>Domicílios na amostra</b>						<b>1.600</b>	

Além dessa estratificação geográfica mais agregada, foram formados estratos geográficos mais finos, buscando assegurar um espalhamento da amostra no território do estado. A Figura 2.2 revela os estratos geográficos mais finos definidos, com as respectivas alocações da amostra mensal de setores.

A estratificação da amostra foi concluída subdividindo os estratos geográficos da Capital (João Pessoa) em quatro estratos de renda, usando como pontos de corte para delimitar os estratos os quartis da variável “Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com rendimento)” para cada setor, conforme apurados no Censo Demográfico 2010 do IBGE. Nas demais regiões, foram definidos apenas dois estratos de renda, considerando partições dos setores em metades da distribuição da mesma variável.

Com isso, a amostra em cada um dos estratos de seleção, formados pela combinação do estrato natural, com o estrato geográfico e o estrato de renda resultou numa estratificação com 81 estratos, dentro dos quais foram selecionados dois ou quatro setores.

O método para seleção dos setores censitários em cada estrato foi sorteio com probabilidade proporcional ao tamanho (PPT) pelo método de Pareto (ROSÉN, B., 2000). Para o sorteio de domicílios em cada setor censitário foi empregado um método de amostragem Binomial ou de Bernoulli, programado em dispositivos móveis de coleta (SÄRNDAL, C. E.; SWENSSON, B.; WRETMAN, J., 2003).

Figura 2.2: Domínios de estudo considerados para estratificação da amostra

Estrato natural	Região de Saúde	Estrato geográfico	Setores	Domicílios	Pessoas	Setores na amostra por mês	Fração amostral setores (%)	Pessoas na amostra por mês	Fração amostral pessoas (por mil)
João Pessoa	1	João Pessoa.1	166	35.234	123.553	8	4,8	200	1,6
	1	João Pessoa.2	159	39.343	137.610	8	5,0	200	1,5
	1	João Pessoa.3	182	43.464	148.357	8	4,4	200	1,3
	1	João Pessoa.4	155	33.048	112.713	8	5,2	200	1,8
	1	João Pessoa.5	283	62.141	196.511	8	2,8	200	1,0
Macro 1 sem João Pessoa	1	Macro 1 sem João Pessoa.1	647	128.537	456.512	16	2,5	400	0,9
	2	Macro 1 sem João Pessoa.2	471	83.651	294.706	12	2,5	300	1,0
	12	Macro 1 sem João Pessoa.12	229	48.485	168.910	8	3,5	200	1,2
	14	Macro 1 sem João Pessoa.14	219	37.680	134.424	8	3,7	200	1,5
Macro 2	3	Macro 2.3	286	53.125	187.313	8	2,8	200	1,1
	4	Macro 2.4	177	31.175	105.882	8	4,5	200	1,9
	5	Macro 2.5	189	33.978	106.675	8	4,2	200	1,9
	15	Macro 2.15	239	41.415	144.511	8	3,3	200	1,4
	16	Macro 2.16	685	149.616	516.239	16	2,3	400	0,8
Macro 3	6	Macro 3.6	363	63.895	223.416	12	3,3	300	1,3
	7	Macro 3.7	216	40.740	146.150	8	3,7	200	1,4
	8	Macro 3.8	172	30.999	110.811	8	4,7	200	1,8
	9	Macro 3.9	250	48.206	167.424	8	3,2	200	1,2
	10	Macro 3.10	143	31.724	110.381	8	5,6	200	1,8
	11	Macro 3.11	141	22.019	80.866	8	5,7	200	2,5
	13	Macro 3.13	99	16.735	58.432	8	8,1	200	3,4
<b>Total geral</b>			<b>5.471</b>	<b>1.075.210</b>	<b>3.731.396</b>	<b>192</b>	<b>3,5</b>	<b>4.800</b>	<b>1,3</b>
<b>Domicílios na amostra</b>								<b>1.600</b>	

A taxa de amostragem em cada setor foi determinada considerando as informações disponíveis no Censo Demográfico 2010 sobre o tamanho do setor censitário. Esse método permite o sorteio da amostra simultaneamente ao arrolamento dos domicílios nos setores selecionados, essencial nesta pesquisa devido à necessidade de executar com agilidade a coleta das informações em cada setor selecionado.

Cada vez que a unidade arrolada era do tipo domicílio particular permanente, o aplicativo de coleta fazia o sorteio de um número aleatório, e usava esta informação para indicar ao entrevistador se a unidade devia ou não ser incluída na amostra. Para unidades selecionadas para a amostra, o entrevistador então fazia a abordagem do domicílio visando à obtenção da entrevista e aplicação dos testes aos correspondentes moradores. Simultaneamente à regra de amostragem de Bernoulli, foi adotada uma regra de parada do trabalho de coleta em cada setor correspondente à amostragem inversa com probabilidades iguais.

## 2.3 Ponderação da amostra da Pesquisa Continuar Cuidando

Os pesos referentes à Pesquisa Continuando Cuidando, consideraram:

- Cálculo dos pesos básicos para cada domicílio e seus moradores segundo o plano amostral considerado;
- Os pesos básicos foram calibrados segundo informações referentes ao número de



habitantes da população do Estado da Paraíba e seus respectivos estratos conforme a descrição da pesquisa.

O peso básico  $d_{hik}$  de um domicílio  $k$  selecionado no setor  $i$  do estrato  $h$  foi obtido com base na expressão:

$$d_{hik} = \left( m_h \frac{t_{hi}}{T_h} \right)^{-1} \times (\pi_{hi})^{-1} \times \left( \frac{n_{hi}}{m_{hi}} \right)^{-1} \quad (2.1)$$

Em que:

- $m_h$  é o numero de setores selecionados no estrato  $h$ ,
- $t_{hi}$  é a medida de tamanho do setor  $i$  do estrato  $h$ ,
- $T_h$  é a soma das medidas de tamanho dos setores do estrato  $h$ ,
- $\pi_{hi}$  é a probabilidade de seleção de domicílios no setor conforme aplicada no processo de amostragem de Bernoulli em cada setor,
- $m_{hi}$  é o número de domicílios elegíveis entre os domicílios selecionados no setor  $i$  do estrato  $h$ , e
- $n_{hi}$  é o numero de domicílios elegíveis entrevistados no setor  $i$  do estrato  $h$ .

O primeiro termo no lado direito da expressão (2.21) representa o inverso da probabilidade de inclusão do setor na amostra. O segundo termo representa o inverso da probabilidade condicional de seleção de domicílios no setor. O terceiro termo representa uma correção devida ao fato de que domicílios selecionados podem ser inelegíveis (por exemplo, serem de ‘uso ocasional’ ou estarem vagos, situações que só seriam apuradas pelo entrevistador em caso de domicílio selecionado para abordagem pela pesquisa). O produto dessas probabilidades levaria à probabilidade de inclusão na amostra do domicílio em questão, e o peso básico é simplesmente o inverso dessa probabilidade.

Como não houve sorteio de moradores nos domicílios selecionados, os pesos básicos de todos os moradores pesquisados em cada domicílio são iguais ao peso básico do domicílio correspondente.

Dessa forma, uma estimativa da prevalência de COVID-19 para a população do Estado da Paraíba é dada da seguinte forma:

$$\hat{p}_{AEC2PB} = \frac{\sum_{h=1}^H \sum_{i=1}^{n_{hi}} \sum_{k=1}^{m_{hi}} d_{hik} y_{hik}}{N} \quad (2.2)$$

Utilizando a metodologia proposta pelo estimador de ??), temos então que na expressão 2.2 é importante observar que  $y_{hik}$  representa o número de pessoas do domicílio  $k$ , setor censitário  $i$ , estrato  $h$  que apresentaram resultado positivo para a presença do anticorpo IgG. De acordo com as definições apresentadas, o valor de  $N$  pode ser obtido a partir dos pesos básicos, da seguinte forma:

$$\hat{p}_{AEC2PB} = \frac{\sum_{h=1}^H \sum_{i=1}^{n_{hi}} \sum_{k=1}^{m_{hi}} d_{hik} y_{hik}}{\sum_{h=1}^H \sum_{i=1}^{n_{hi}} \sum_{k=1}^{m_{hi}} d_{hik}} \quad (2.3)$$

A partir desta expressão, foi possível também propor estimativas para níveis das macrorregiões de saúde (estratos), além de informações sociodemográficas como sexo, renda, escolaridade, emprego, escalas de sintomas e hábitos de proteção (uso de álcool e uso de máscara, por exemplo). Alguns destes resultados serão apresentados na seção 4.

## CAPÍTULO 3

# MÉTODO PROBABILÍSTICO DE AMOSTRAGEM

A amostragem probabilística tem o papel de atribuir probabilidades de inclusão na amostra a todos os participantes de uma pesquisa, devidamente identificados por meio de um cadastro. Estas probabilidades permitem que o pesquisador apresente estimativas baseadas no processo físico de aleatorização empregado no cadastro. Segundo Silva, P. L. N. (2020) um método probabilístico de amostragem possui as seguintes condições:

1. Todos os elementos definidos no espaço amostral  $S$ , correspondente ao conjunto de todas as amostras  $s$  possíveis, é definido e poderia ser enumerado, ao menos teoricamente;
2. Uma probabilidade  $p(s)$  é associada a cada amostra  $s \in S$ , sendo  $\sum_{s \in S} p(s) = 1$ . A função  $p(s)$  é chamada de plano amostral;
3. Uma única amostra  $s$  ( $s \in S$ ) é selecionada para observação usando um mecanismo de aleatorização com probabilidade  $p(s)$ ;
4. Sendo  $U = \{1, 2, \dots, N\}$  o conjunto de índices que identificam cada um dos elementos da população devidamente identificados no cadastro, cada unidade  $i \in U$  tem uma probabilidade positiva de ser selecionada para a amostra, isto é:  $\pi_i = P(i \in s) = \sum_{i \ni s} p(s) > 0$
5. As probabilidades de inclusão das unidades selecionadas para a amostra e outros aspectos da estrutura do plano amostral são levados em conta ao fazer inferência sobre os parâmetros populacionais.

## Probabilidade de inclusão de primeira e segunda ordem

Considere o interesse de estimar o total populacional  $t = \sum_{i \in U} y_i$ . Sejam  $y_{i1}, \dots, y_{in}$  os dados amostrais para estimar este total populacional.

Um estimador linear  $\hat{t}_w$  do total populacional  $t$  é uma combinação dos valores obtidos da amostra  $y_i$  e seus respectivos pesos amostrais  $w_i$ , como:

$$\hat{t} = \sum_{i \in s} w_i y_i \quad (3.1)$$

A distribuição de probabilidades a partir da função  $p(s)$  definida no conjunto  $S$  obtida com os dados amostrais é chamada de distribuição de aleatorização do estimador. Na amostragem probabilística, as inferências consideram uma única fonte de variação do processo de amostragem utilizando o plano amostral  $p(s)$  que pode resultar em diferentes amostras  $s_1, s_2, \dots \in S$ .

Na maioria dos casos, a definição de todas as amostras  $s$  possível no espaço amostral  $S$  é complicada, devido ao tamanho da população, denotado por  $N$  e o tamanho da amostra, denotado por  $n$ . Dessa forma, de modo a identificar teoricamente a inclusão dos elementos da população na amostra, considere os valores  $\delta_1, \delta_2, \dots, \delta_N$ , em que  $\delta_i$  define a variável indicadora de inclusão na amostra  $s$  do  $i$ -ésimo elemento da população, e é definida como:

$$\delta_i = I(i \in s) = \begin{cases} 1, & i \in s, \forall i \in U \\ 0, & i \ni s \end{cases} \quad (3.2)$$

Dessa forma, a **probabilidade de inclusão de primeira ordem** do  $i$ -ésimo elemento da população é calculada também da seguinte forma:

$$\pi_i = P(i \in s) = \sum_{s \ni i} p(s) = P(\delta_i = 1) = E_p(\delta_i), \forall i \in U \quad (3.3)$$

Adicionalmente, é necessário conhecer também a probabilidade de que dois elementos  $i$  e  $j$  da população estejam incluídos na amostra. Esta probabilidade é chamada de **probabilidade de inclusão de segunda ordem** e é calculada da seguinte forma:

$$\pi_{ij} = P[(i, j) \in s] = \sum_{s \ni (i, j)} p(s) = P(\delta_i \delta_j = 1) = E_p(\delta_i \delta_j), \forall (i, j) \in U \quad (3.4)$$

Dessa forma, é possível agora apresentar as seguintes propriedades da variável indicadora de inclusão:

- $V_p(\delta_i) = \pi_i(1 - \pi_i)$
- $COV_p(\delta_i, \delta_j) = \pi_{ij} - \pi_i \pi_j$
- $\sum_{i \in s} \delta_i = \sum_{i \in U} \delta_i$
- $\sum_{i \in s} y_i = \sum_{i \in U} \delta_i y_i$

Dessa forma, um estimador linear de  $t$  é sempre dado por :

$$\hat{t}_w = \sum_{i \in s} w_i y_i = \sum_{i \in U} \delta_i w_i y_i \quad (3.5)$$

Para que o estimador linear para o total  $t$  seja não viciado os pesos amostrais das unidades selecionadas devem ser iguais ao inverso das respectivas probabilidades de inclusão de primeira ordem,  $w_i = \pi_i^{-1} = 1/\pi_i = d_i, \forall i \in U$ . Os pesos amostrais  $d_i$  são chamados de pesos básicos do plano amostral.

Com os pesos básicos  $d_i$ , foi possível propor um estimador não viciado para o total, que é conhecido como estimador de Horvitz-Thompson ou estimador HT (??):

$$\hat{t}_{HT} = \sum_{i \in s} d_i y_i = \sum_{i \in s} \pi_i^{-1} y_i = \sum_{i \in s} \frac{y_i}{\pi_i} \quad (3.6)$$

Este estimador está definido para qualquer variável de interesse  $y$  e para qualquer plano amostral probabilístico em que  $\pi_i > 0 \forall i \in U$ .

### 3.0.1 Amostragem de Bernoulli (BE)

O plano de amostragem de Bernoulli consiste em uma série de experimentos independentes, um para cada elemento da população. O plano atribui probabilidade igual de

seleção  $\pi$ , e de não seleção,  $1 - \pi$ , a cada elemento listado em cadastro. Consequentemente, o tamanho da amostra, denotado por  $n$  é uma variável aleatória, onde  $n \sim \text{Bin}(N, \pi)$ .

Sob um plano de Bernoulli, tem-se que

$$p(s) = \pi^n (1 - \pi)^{(N-n)} \quad (3.7)$$

Em que a probabilidade de inclusão de primeira é  $\pi_i = \pi$  (constante) e a probabilidade de inclusão de segunda ordem é  $\pi_{ij} = \pi^2, \forall i \neq j$

A partir disso, as etapas de implantação do plano amostral são dadas por:

1. Considere um valor para  $\pi$  ( $0 < \pi < 1$ );
2. Denote por  $\epsilon_1, \epsilon_2, \dots, \epsilon_N$ , uma série de  $N$  realizações de uma distribuição Uniforme(0, 1);
3. Se  $\epsilon_k \leq \pi$ , então, o elemento  $k$  é selecionado para compor a amostra. Caso contrário,  $\epsilon > \pi$ , elemento não selecionado.

Sobre o plano BE, o estimador de HT assume a seguinte forma

$$\hat{t}_{y\pi} = \pi^{-1} \sum_{i \in s} y_i \quad (3.8)$$

### 3.0.2 Amostragem Proporcional ao Tamanho (PPT)

A amostragem com probabilidades proporcionais a uma medida de tamanho que esteja disponível no cadastro agrega métodos de seleção de amostra em que as unidades da população tem probabilidades distintas de seleção, associada a cada unidade uma medida de tamanho, de modo que as probabilidades de inclusão  $\pi$  sejam proporcionais aos valores das variáveis de estudo  $y$ .

Na prática, os valores da população são desconhecidos antes da seleção da amostra, mas em muitos casos, podemos contar com cadastros que contenham valores de uma variável auxiliar  $x$  a qual indica o tamanho das unidades populacionais. Se a variação dos tamanhos das unidades for grande, a informação auxiliar disponível sendo positivamente correlacionada com as variáveis de interesse e contiver informações precisas, pode-se empregar métodos de seleção PPT.

Considerando a estimação de HT definida na sessão 2.2.1, para o plano amostral PPT, a probabilidade de inclusão na amostra é proporcional ao tamanho. Seja  $X$  a variável auxiliar com todos os elementos conhecidos na população  $U$  e listadas em um cadastro. Se  $x > 0, \forall i \in U$  a variável pode ser considerada uma medida de tamanho das unidades populacionais, com isso temos:  $\pi_i = nx_i/t_X$  onde  $t_X = \sum_{i \in U} x_i$ .

Logo, o peso amostral da unidade  $i$  é dado por  $d_i = t_X/(nx_i)$ . O estimador HT do total  $t$  fica dado por:

$$\hat{t}_{HT/PPT} = t_X \times \frac{1}{n} \times \sum_{i \in s} \frac{y_i}{x_i} \quad (3.9)$$

Na Pesquisa Continuar Cuidando, a amostragem PPT foi implementada pela abordagem de Pareto - AP de acordo com ROSÉN, B. (2000). O algoritmo para a seleção de elementos sobre esta abordagem é dado pelas seguintes etapas:

1. Geração de número aleatórios uniformes independentes  $A_i$  para cada unidade  $i$  da população;
2. Cálculo das probabilidades de inclusão da unidade  $i$ :  $\lambda_i = nx_i/t_X$ ;
3. Cálculo do número aleatório modificado  $C_i = \frac{A_i(1 - \lambda_i)}{(1 - A_i)\lambda_i}$ ;
4. Ordenar as unidades crescentemente segundos os valores  $C_i$ ;
5. Seleção para a amostra das  $n$  unidades com os menores valores de  $C_i$ .

Segundo o método Pareto - AP, o estimador do total sob Amostragem PPT é dado por:

$$\hat{t}_{AP} = \sum_{i \in s} \frac{y_i}{\lambda_i} = \frac{1}{n} \sum_{i \in s} \frac{y_i}{p_i} \quad (3.10)$$

### 3.0.3 Amostragem Estratificada (AE)

A Amostragem Estratificada é um plano amostral que consiste na subdivisão da população em grupos disjuntos e homogêneos, com o auxílio de alguma informação auxiliar relevante para o estudo. Cada subgrupo formado é chamado de estrato, sendo  $H$  o total

de grupos e  $n_h$  o tamanho de amostra em cada estrato. As variáveis de estratificação devem contemplar toda a população  $U$  de interesse da pesquisa.

As variáveis de estratificação podem ser geográficas, quando os estratos formam domínios naturais ou de interesse, ou variáveis relacionadas a fatores sociodemográficos dos indivíduos, tais como renda, sexo, idade, habitantes por domicílio. Com a consideração do método podemos tornar a amostra mais representativa, principalmente em populações heterogêneas.

Para pesquisas em que a estimação é requerida para cada estrato definido na população, o processo de amostragem e estimação deve ser replicado separadamente em cada estrato. Para pesquisas onde não são de interesse resultados por estrato, parâmetros dos estratos são estimados separadamente em cada estrato, de acordo com o plano amostral adotado, e estas estimativas são então agregadas para obter estimativas dos parâmetros de interesse referentes ao conjunto da população.

$$U = U_1 \cup \dots \cup U_h \cup \dots \cup U_H = \bigcup_{h=1}^H U_h \quad (3.11)$$

De modo que  $U_h =$  unidades pertencentes ao estrato  $h$ , para  $h = 1, 2, \dots, H$ . Considerando que  $N_h$  é o tamanho de cada estrato, definido como  $U_h$ . O total da população é dado pela soma de todos os tamanhos de estratos. A seleção de uma amostra  $s_h$  em cada estrato é independente conforme um plano amostral  $p_h(s_h)$  de tamanho  $n_h$ , e com isso,  $n = \sum_{h=1}^H n_h$  representa o tamanho total de amostra considerado.

Como diferentes planos amostrais podem ser utilizados em cada estrato, temos possibilidade de implementar um tipo de plano amostral diferente em diversos estratos. A amostragem estratificada simples o método mais utilizado na prática, que corresponde a seleção dos  $n_h$  em todos os estratos a partir de uma amostra aleatória simples sem reposição - AAS.

$$s = s_1 \cup s_2 \cup \dots \cup s_h = \bigcup_{h=1}^H s_h, \quad (3.12)$$

Denote por  $p_h$  o plano amostral implementado pela aleatorização imposta ao estrato  $h$ . Como as amostras  $s_1, s_2, \dots, s_H$  foram geradas independentemente, o plano atribui probabilidade de seleção da amostra  $s$ , dada por



$$p(s) = \prod_{h=1}^H p_h(s_h) \quad (3.13)$$

Um esquema amostral deve ser empregado para implementar cada plano escolhido,  $p_h$ , dentro de cada estrato  $h$ .

### 3.0.4 Amostragem Conglomerada (AC)

A amostragem por conglomerado é definida com um plano amostral de estágios, em que em cada estágio a unidade amostral é composta por grupos (conglomerados ou clusters). A composição de cada conglomerado pode ser por vários critérios, os mais conhecidos são por classificação natural, normalmente pela dificuldade de acesso a unidade amostral diretamente, e planejada pela pesquisa com base nos objetivos da pesquisa.

O método tem como vantagem o fato de que não há a necessidade de uma lista identificando os elementos da população e a redução de custo. A unidade amostral é definida de acordo com o propósito da pesquisa independente do plano amostral determinado.

A população de interesse na pesquisa, representada por  $U$ , é particionada em  $M$  grupos mutuamente exclusivos, denominados conglomerados, pela forma:

$$C_i = C_1 \cup C_2 \cup \dots \cup C_M = \bigcup_{i=1}^M C_i \quad (3.14)$$

Em que  $C_i$  = subgrupos da população conglomerada, para  $i = 1, 2, \dots, M$ . A partir da definição dos conglomerados, é necessário indexar as unidades populacionais em cada grupo considerando dois índices, representados por  $i$  e  $j$ , em que  $i$  se refere ao conglomerado e  $j$  a indicação da unidade em cada conglomerado. Então,

$$C_i = (i, 1), (i, 2), \dots, (i, j), \dots, (i, N_i) \quad (3.15)$$

Em que  $N_i$  é o tamanho populacional de cada conglomerado. O tamanho total da população se dá pela soma de todos os  $N_i$  para todo  $M > 0$ .

### Amostragem conglomerada em um estágio

Amostragem Conglomerada em um estágio - AC1, são incluídas na amostra todas as unidades da população encontradas nos conglomerados selecionados aleatoriamente. A amostra  $s$  de unidades da população  $U$  que serão observadas é dada por:

$$s = C_{i_1} \cup C_{i_2} \cup \dots \cup C_{i_m} = \bigcup_{k=1}^m C_{i_k} \quad (3.16)$$

Seja  $y_{ij}$  a variável de interesse para a unidade  $j$  do conglomerado  $i$ . Os parâmetros populacionais para o total e a média por conglomerado, de forma geral, são dados por

$$t_i = \sum_{j \in C_i} y_{ij} \quad (3.17)$$

e

$$\bar{y}_i = \frac{t_i}{N_i} = \frac{1}{N_i} \sum_{j \in C_i} y_{ij} \quad (3.18)$$

### Amostragem conglomerada em dois estágios

Na abordagem do método de Amostragem Conglomerada em dois estágios a seleção de uma amostra de conglomerados com subamostragem é definida por:

- No primeiro estágio é feita a seleção de uma amostra  $a$  de  $m$  conglomerados.
- No segundo estágio para cada conglomerado é feita a seleção de uma amostra  $s_i$  de  $n_i$  unidades secundárias das  $N_i$  unidades existentes em cada conglomerado.

A amostra completa de unidades é dada por:

$$s = s_{i_1} \cup s_{i_2} \cup \dots \cup s_{i_m} = \bigcup_{i \in a} s_{ik} \quad (3.19)$$

O tamanho final da amostra é dado pela soma de todas as seleções em conglomerados selecionados no primeiro estágio,  $n = \sum_{i \in a} n_i$ .

Isto requer considerar todas as possíveis amostras no primeiro estágio e, todas as possíveis amostras no segundo estágio, dentro de cada amostra de conglomerados do primeiro estágio. O estimador não viciado de HT do total sobre amostragem por conglomerado em dois estágios é dado por:

$$\hat{t}_{AC2/HT} = \sum_{i \in a} \frac{\hat{t}_i}{\pi_i} = \sum_{i \in a} \frac{1}{\pi_i} \sum_{j \in s_i} \frac{y_{ij}}{\pi_{j|i}} = \sum_{i \in a} \sum_{j \in s_i} d_{ij} y_{ij} \quad (3.20)$$

Em que:

- $\pi_i$  é a probabilidade de inclusão da unidade primária amostral  $i$ ;
- $s_i$  é a amostra de unidade selecionadas dentro da unidade primária amostral  $i$ ;
- $\hat{t}_i = \sum_{j \in s_i} y_{ij} / \pi_{j|i}$  é um estimador HT do total  $t_i$  da UPA  $i$ ;
- $\pi_{j|i} = P(j \in s_i | i \in a)$  é a probabilidade de inclusão da unidade  $j$  dado que a UPA  $i$  está na amostra  $a$ ;
- $d_{ij} = \pi_{ij}^{-1} = \pi_i^{-1} \pi_{j|i}^{-1}$  é o peso associado à unidade  $j$  da UPA  $i$ .

Seguindo o mesmo raciocínio, um estimador não viciado HT da média por unidade é dado por:

$$\bar{y}_{AC2/HT} = \frac{\hat{t}_{AC2/HT}}{N} = \frac{1}{N} \times \sum_{i \in a} \frac{\hat{t}_i}{\pi_i} \quad (3.21)$$

## CAPÍTULO 4

# MODELO DE REGRESSÃO LOGÍSTICA CONSIDERANDO O PLANEJAMENTO AMOSTRAL

Neste capítulo, será apresentada a forma como o modelo de regressão logística foi considerado para identificar fatores que podem estar relacionados/associados com o desfecho de interesse (resultado positivo para anticorpo IgG). Inicialmente, faremos a apresentação do modelo de regressão logística, conforme descrição a seguir.

### 4.1 Modelos Lineares Generalizados

Em muitos estudos estatísticos, quer sejam de natureza experimental ou observacional, somos confrontados com problemas em que o objetivo principal é o de estudar a relação entre variáveis, ou mais particularmente, analisar a influência que uma ou mais variáveis (explicativas), medidas em indivíduos ou objetos, tem sobre uma variável de interesse a que damos o nome de variável resposta. O modo como, em geral, o estatístico aborda tal problema é através do estudo de um modelo de regressão que relacione essa variável de interesse com as variáveis ditas explicativas (TURKMAN, M. A. A. et al., 2000).

O modelo linear normal foi predominante na área de modelagem estatística até o século XX, porém muitas variáveis e estudos não cumprem as hipóteses de normalidade e linearidade. Diante da necessidade foram propostos modelos adequados para variáveis não normais. A estrutura de todos os modelos apresentou algumas características semelhantes e a variável resposta segue distribuições de uma família exponencial.

Os modelos lineares generalizados foram propostos por NELDER, J. A. et al (1972) como uma extensão do modelo linear normal. O conceito MLG corresponde à união desses modelos, do ponto de vista teórico, partindo do pressuposto que a distribuição do modelo pertence à família Exponencial. Alguns casos particulares dos MLGs são:

- Modelo normal linear;
- Modelos log-lineares aplicados à análise de tabelas de contingência;
- Modelo de regressão de Poisson;
- Modelo logístico binário;
- Modelo probit para estudo de proporções;
- Modelos Gama.

Na construção de um MLG são definimos a escolha de uma distribuição de probabilidade para a variável resposta, as variáveis preditoras ou explicativas e uma função de ligação. Para a escolha da distribuição da variável dependente do modelo, é recomendado analisar os dados e verificar padrões de acordo com algumas características, como: tipo de variável, variância, simetria, etc.

De uma forma geral, a estrutura de um MLG é formada por três partes; uma componente aleatória, uma componente sistemática e uma função de ligação. Uma importante definição dos MLGs é a suposição de independência entre as observações e escolha de uma única distribuição pertencente à família Exponencial.

#### 4.1.1 Componente aleatória

Considere um vetor de observações  $y = (y_1, \dots, y_n)^T$  referentes às realizações das variáveis aleatórias  $Y = (Y_1, \dots, Y_n)^T$ , independentes e identicamente distribuídas, com médias  $\mu = (\mu_1, \dots, \mu_n)^T$  e pertencentes à família Exponencial de distribuições com função de probabilidade dada por

$$f_Y(y; \theta; \phi) = \exp \phi[y\theta - b(\theta)] + c(y; \phi) \quad (4.1)$$

Em que  $b(\cdot)$  e  $c(\cdot)$  são funções conhecidas para cada observação;  $\phi > 0$  é denominado parâmetro de dispersão e  $\theta$  é denominado parâmetro canônico que caracteriza a distribuição em 3.1.

Escrevendo a log-verossimilhança para uma única observação como

$$l(\theta; \phi; y) = \log f_Y(y; \theta; \phi) \quad (4.2)$$

Temos uma função de  $\theta$  e  $\phi$  para um dado  $y$ . Desse modo, a média e a variância de  $Y$  podem ser calculadas por meio das seguintes relações, respectivamente

$$E\left(\frac{\partial l}{\partial \theta}\right) = 0 \quad (4.3)$$

e

$$E\left(\frac{\partial^2 l}{\partial \theta^2}\right) + E\left(\frac{\partial l}{\partial \theta}\right)^2 = 0 \quad (4.4)$$

Obtendo-se

$$E(Y) = \mu = b''(\theta) \quad (4.5)$$

e

$$Var(Y) = \frac{b''(\theta)}{\phi} \quad (4.6)$$

A variância de  $Y$  dependente do parâmetro canônico  $\theta$  e pode ser escrita em função de  $\mu$ , sendo denominada de função de variância  $V = V(\mu)$ . A função de variância expressa como função de  $\mu$  é reescrita da seguinte forma

$$V(\mu) = b''(\theta) = \frac{d\mu}{d\theta} \quad (4.7)$$

A distribuição escolhida da família Exponencial depende dos dados em questão, que podem ser discretos, contínuos, assimétricos ou proporções.

### 4.1.2 Componente sistemática e função de ligação

A função de ligação relaciona o preditor linear  $\eta$  à média  $\mu$  do vetor de dados  $y$ . Considere a seguinte estrutura linear de um modelo de regressão

$$\eta = X\beta \quad (4.8)$$

Em que  $\eta = (\eta_1, \dots, \eta_n)^T$ ,  $\beta = (\beta_1, \dots, \beta_p)^T$  e  $X$  é uma matriz modelo  $n \times p$  ( $p < n$ ) conhecida de posto  $p$ . Além disso, outra característica da componente sistemática de um MLG é que a média  $\mu$  do vetor  $y$  é expressa por uma função conhecida (monótona e diferenciável) de  $\eta$ ,

$$\mu_i = g^{-1}(\eta_i) \quad (4.9)$$

Em que  $i = 1, \dots, n$  e  $g(\cdot)$  denomina-se função de ligação.

### 4.1.3 Função de ligação canônica

Como a função de ligação conecta o preditor linear  $\eta$  à média  $\mu$  do vetor de dados  $y$ , para cada distribuição tem uma função de ligação específica que está associada ao preditor linear e define uma estatística suficiente com a mesma dimensão de  $\beta$ . Estas ligações são chamadas de canônicas e ocorrem se  $\theta = \eta$ , sendo  $\theta$  é o parâmetro canônico apresentado anteriormente como argumento para a média  $\mu$ . As ligações canônicas para algumas distribuições são :

- Normal  $\eta = \mu$ ;
- Binomial  $\eta = \log(\pi/(1 - \pi))$ ;
- Gama  $\eta = (\mu)^{-1}$ ;
- Poisson  $\eta = \log(\mu)$ .

### 4.1.4 Algoritmo de estimação

Existem inúmeros métodos para estimar os parâmetros  $\beta = (\beta_1, \dots, \beta_p)$ . O mais frequentemente utilizado é o Método de Máxima Verossimilhança (MMV), pois apresenta ótimas propriedades, como, consistência e eficiência assintótica.

O algoritmo de estimação dos parâmetros  $\beta$ 's foi desenvolvido por ??) e baseia-se em um método semelhante ao de Newton Raphson, conhecido como Método Escore de Fisher.

A principal diferença em relação aos modelos de regressão simples é que as equações de máxima verossimilhança são não-lineares.

O método se baseia na resolução do sistema  $U(\beta) = 0$ , onde  $U(\beta)$  é denominado como função escore e  $l(\beta)$  a log-verossimilhança como função de  $\beta$ . Utilizamos a função escore

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta}, \quad (4.10)$$

e a matriz de informação de Fisher

$$K = -E \left( \frac{\partial^2 l(\beta)}{\partial \beta_j \partial \beta_s} \right) = E \left( \frac{\partial U(\beta)}{\partial \beta} \right). \quad (4.11)$$

Expandindo a função escore em série de Taylor até termos a de primeira ordem, obtém-se

$$U(\beta^{(m+1)}) = U(\beta^{(m)}) + \frac{\partial U(\beta^{(m)})}{\partial \beta} [\beta^{(m+1)} - \beta^{(m)}] = 0 \quad (4.12)$$

ou

$$\beta^{(m+1)} = \beta^{(m)} - \left( \frac{\partial U(\beta^{(m)})}{\partial \beta} \right)^{-1} U(\beta^{(m)}) \quad (4.13)$$

Em que o expoente  $m$  significa o valor do termo na  $m$ -ésima iteração. Este é o método de Newton-Raphson para o cálculo da estimativa de máxima verossimilhança.

O método escore de Fisher (1925) é obtido pela substituição de  $-\frac{\partial U(\beta)}{\partial \beta}$  pelo seu valor esperado  $K$ .

Para desenvolver o algoritmo de estimação MLG considera a componente sistemática.

$$\eta_i = g(\mu_i) = \sum_{r=1}^p x_{ir} \beta_r = x_i^T \beta \quad (4.14)$$

,

Em que  $x_i^T$  é a  $i$ -ésima linha de dados  $X$ . E a log-verossimilhança é dada por



$$l(\beta) = \frac{1}{a(\phi)} \sum_{i=1}^n y_i \theta_i - b(\theta_i) + \sum_{i=1}^n c(y_i; \phi). \quad (4.15)$$

Derivando  $l(\beta)$  em relação ao vetor  $\beta$ , tem -se

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \frac{1}{a(\phi)} \sum_{i=1}^n y_i - b'(\theta_i) \frac{\partial \theta_i}{\partial \beta}. \quad (4.16)$$

Calculando

$$\frac{\partial \theta_i}{\partial \beta} = \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta} \quad (4.17)$$

pela regra da cadeia e considerando as equações 3.5, 3.6 e 3.7 obtém-se

$$\mu_i = b'(\theta) \quad (4.18)$$

e

$$V(\mu_i) = b''(\theta_i) = \frac{d\mu_i}{d\theta_i} \quad (4.19)$$

Sendo  $x_i^T$  a  $i$ -ésima linha de  $X$  e  $\eta_i = x_i^T \beta$ , temos

$$\frac{\partial \eta_i}{\partial \beta} = x_i e \frac{\partial \mu_i}{\partial \eta_i} = [g'(\mu_i)]^{-1} \quad (4.20)$$

Em que  $x_i$  é o vetor coluna  $p \times 1$ .

Com isso, a função escore é expressa como

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \frac{1}{a(\phi)} \sum_{i=1}^n y_i - b'(\theta_i) \frac{1}{V(\mu_i) g'(\mu_i)} x_i. \quad (4.21)$$

A matriz de informação para  $\beta$  é dada por

$$K = \frac{1}{a(\phi)} X^T W X, \quad (4.22)$$

Em que  $W$  é uma matriz diagonal de pesos, em que cada elemento da diagonal é dado por

$$w_i = V^{-1} g'(\mu_i)^2. \quad (4.23)$$

Logo, a função escore usando a matriz de pesos é dada por

$$U(\beta) = X^T W z^* \quad (4.24)$$

Em que  $z^*$  é um vetor de dimensão  $n \times 1$  definido como

$$z_i^* = (y_i - \mu_i) \left( \frac{\partial g(\mu_i)}{\partial \mu_i} \right). \quad (4.25)$$

Utilizando os dois últimos resultados, o algoritmo escore de Fisher para calcular a estimativa de máxima verossimilhança de  $\beta$  é expresso por

$$\beta^{(m+1)} = \beta^{(m)} + (X^T W^{(m)} X)^{-1} X^T W^{(m)} z^{(m)}. \quad (4.26)$$

Colocando  $(X^T W^{(m)} X)^{-1}$  em evidência, tem-se

$$\beta^{(m+1)} = (X^T W^{(m)} X)^{-1} X^T W^{(m)} y^{*(m)}, \quad (4.27)$$

onde  $y^{*(m)}$  é a variável resposta do modelo modificada e definida por

$$y^{*(m)} = X \beta^{(m)} + z^{(m)}. \quad (4.28)$$

## 4.2 Regressão Logística

A regressão logística é um caso particular dos modelos lineares generalizados, em que a variável resposta é caracterizada como dicotômica ou *dummy*, ou seja, assume apenas os valores 0 ou 1. Seguindo a distribuição Binomial e considerando a função de ligação 'logit', a qual limita os valores ajustados pela reta ao intervalo da variável resposta.

O modelo permite estimar uma probabilidade associada à ocorrência de determinado evento de interesse com um conjunto de variáveis explicativas ou independentes. O modelo pode ser escrito da seguinte forma:

$$P(Y = 1) = \frac{1}{1 + e^{-g(x)}} \quad (4.29)$$

Em que:

$$g(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$$

A distribuição binomial permite diversas funções de ligação, porém a "logit" permite verificar a influência de cada variável pelo método de razão de chances, uma forma mais simples e interpretativa. Com isso, determina o efeito que os coeficientes exercem sobre a chance de o desfecho de interesse ocorrer  $P(Y = 1)$  com relação ao desfecho não ocorrer  $P(Y = 0)$ .

## 4.3 Modelo de regressão logística sob o planejamento amostral

A metodologia padrão dos modelos lineares generalizados não permite verificar a influência do plano amostral de coleta dos dados. Com isso, modelo de regressão logística padrão é inadequado quando os dados se referem a amostras de projetos de amostragem complexos.

Considere que uma população finita  $U = \{1, \dots, N\}$  é subdividida em  $h = \{1, 2, \dots, H\}$  estratos. Dessa forma, a amostra de cada estrato possui  $n_h$  de unidades amostrais primárias, e cada uma destas possui  $n_{hj}$  unidades amostrais secundárias, e cada uma destas compreendendo  $n_{hji}$  elementos. Supondo que cada elemento selecionado para compor a

amostra consiste de  $m_{hi}$  elementos da amostra secundária selecionados de  $n_{hi}$  dos elementos primários do estrato  $h$ . O número total de observações é dado por  $n = \sum_{h=1}^H \sum_{j=1}^{n_h} \sum_{i=1}^{n_{hj}} n_{hji}$ . Cada unidade amostral tem um peso de amostragem associado dado pelo inverso da probabilidade de inclusão na amostra, denotado por  $w_{hikl} = 1/\pi_{hikl}$ , para cada  $l$  unidade observada, da  $k$ -ésima unidade secundária, da  $i$ -ésima unidade primária do  $h$ -ésimo estrato.

No caso da Pesquisa Continuar Cuidando temos que  $Y_{hikl}$  denota a variável dicotômica que representa se o  $l$ -ésimo indivíduo do domicílio  $k$ , do setor censitário  $i$ , do estrato  $h$  testou positivo para o teste do anticorpo IgG;  $x_{hikl}$  denota a matriz de covariância; e  $\beta$  denota o coeficiente da regressão. Assim, o modelo de regressão logística com base no planejamento amostral é dado por

$$\text{logit}(P(Y_{hikl} = 1|x_{hikl})) = \log\left(\frac{P(Y_{hikl} = 1|x_{hikl})}{(1 - P(Y_{hikl} = 1|x_{hikl}))}\right) = x'_{hikl}\beta. \quad (4.30)$$

Portanto, sob o planejamento de um plano de amostragem complexo, os parâmetros  $\beta$  do modelo de regressão logística são estimados pelo método de máxima pseudo-verossimilhança que incorpora o plano de amostragem e os diferentes pesos na estimação dos coeficientes da regressão. O método tem como objetivo definir uma função que se aproxima da função de verossimilhança da população com uma função de verossimilhança formada pela amostra observada e os pesos de amostragem. Nesse caso, a função de pseudo-log-verossimilhança é dada por

$$l_p(\beta) = \sum_{h=1}^H \sum_{i=1}^{m_h} \sum_{k=1}^{n_{hi}} \sum_{l=1}^{n_{hik}} w_{hikl} y_{hikl} \times \log[P(Y_{hikl} = 1|x_{hikl})] \\ + (1 - y_{hikl}) \times \log[1 - P(Y_{hikl} = 1|x_{hikl})], \quad (4.31)$$

Onde  $w_{hikl}$  é o peso de cada elemento observado. O estimador de máxima pseudo-verossimilhança do estimado para  $\beta$  é obtido derivando a função pseudo-log-verossimilhança sobre  $\beta$  e igualando a zero,

$$\frac{\partial}{\partial \beta} l_p(\beta) = 0 \quad (4.32)$$

Para obter os estimadores de variância por máxima pseudo-verossimilhança considera métodos como a linearização de Taylor, conhecido como método delta. O método resulta no seguinte estimador para a variância de  $\beta$ :

$$\widehat{Var}(\hat{\beta}) = (X'DX)^{-1}S(X'DX)^{-1}, \quad (4.33)$$

Em que  $X$  é a matriz de covariância,  $D = WV$  é a matrix diagonal  $n \times n$  com os elementos  $w_{hikl} \times \hat{P}(Y + hikl = 1|x_{hikl})[1 - \hat{P}(Y + hikl = 1|x_{hikl})]$ , e  $S$  é o estimador de agrupamento para a matriz de covariância em cada estrato, dado por:

$$S = \sum_{h=1}^H (1 - f_h) \frac{n_h}{n_h - 1} \sum_{j=1}^{n_h} (z_{hj} \dots \tilde{z}_{h\dots})(z_{hj} \dots \tilde{z}_{h\dots})', \quad (4.34)$$

Em que  $z_{hikl} = w_{hikl} \times \hat{P}(Y + hikl = 1|x_{hikl})[1 - \hat{P}(Y + hikl = 1|x_{hikl})]$ , sendo a soma de todos os elementos  $n_{hj}$  de unidades primárias  $j$  do estrato  $h$  dado por  $z_{hj} = \sum_{i=1}^{n_{hj}} z_{hikl}$  e a média para cada estrato como  $\tilde{z}_h = \frac{1}{n_h} \sum_{j=1}^{n_h} z_{hj}$ . O fator de correção é dado por  $(1 - f_h)$ , em que  $f_h = \frac{n'_h}{n_h}$  a razão entre o numero de unidades primárias observadas e o total de unidades primárias no estrato  $h$ .

O teste de significância dos coeficientes de regressão e o teste para a qualidade do ajuste do modelo também precisa ser modificada para incorporar o projeto de amostragem e os diferentes pesos das observações. A avaliação de contribuição das covariâncias são ajustadas para o teste Wald, com a estatística de teste dado por:

$$F = \frac{s - p + 1}{sp} \times W \quad (4.35)$$

Em que  $W = \hat{\beta}'[(Var)^{(\hat{\beta})}]^{-1}\hat{\beta}$  e  $s = \sum_{h=1}^H n_h - H$  é o numero total de unidades primarias selecionadas e o numero de estratos. Na equação,  $p$  é o numero de covariâncias. A estatística  $F$  é dado pela distribuição  $F$  com  $p$  e  $(s - p + 1)$  seus graus de liberdade.

## CAPÍTULO 5

# RESULTADOS E DISCUSSÃO

Neste capítulo, serão apresentados os resultados do estudo, do teste de hipóteses e da regressão logística, considerando o plano amostral empregado, com o objetivo de identificar fatores que podem estar relacionados/associados com o desfecho de interesse (resultado positivo para anticorpo IgG).

Os resultados apresentados a seguir foram obtidos com o auxílio do software R, considerando o pacote *survey*. O banco de dados, fornecido pela Pesquisa Continuar Cuidando, composto por 9.763 participantes, dentre os quais foram considerados os indivíduos maiores de 17 anos e apresentaram resultado ao teste rápido para anticorpo IgG:

A partir dos dados da Pesquisa Continuar Cuidando, temos as seguintes estimativas para o Estado da Paraíba segundo os grupos de variáveis: Socioeconômicas (sexo, idade, renda, trabalho, escolaridade, renda), e Hábitos de Higiene/Proteção e Comorbidades (Se saiu de casa, Uso de máscara, Uso de álcool, Diabetes, Hipertensão, Doença no coração, Obesidade, Outra doença crônica).

Considerando a variável sexo, foram observados 1.251.308 homens e 1.539.779 mulheres. De acordo com a idade, 1.831.607 pessoas com faixa etária entre 18 e 49 anos, 413.686 pessoas com faixa etária entre 50 e 59 anos e 545.794 pessoas com faixa etária entre 60 anos ou mais, foram investigadas.

No que se diz respeito a variável faixa de renda constatou-se, 84.560,45 pessoas que possuem renda igual ou inferior a 300 reais, 506.434,42 pessoas com renda entre 301 e 1000 reais, 958.288,81 pessoas com renda entre 1001 e 2000 reais, 910.778,76 pessoas com renda

entre 2001 e 5000 reais, 238.516,54 pessoas com renda superior a 5000 reais e 84.033,51 pessoas que não souberam ou não quiseram responder sobre faixa de renda.

Em se tratando da variável raça ou cor foram verificadas, 903.732,85 brancas, 217.956,53 pretas, 28.179,39 amarela (origem japonesa, chinesa, coreana etc.), 1.597.542,42 parda (mulata, cabocla, cafuza, mameluca ou mestiça), 35.389,76 indígena e 8286,39 pessoas que não souberam ou não quiseram responder sobre raça ou cor.

De acordo com a variável grávida, foram contabilizadas 33.767 gestantes e 1.366.032 mulheres que não estavam grávidas. Dentre aquelas que tiveram filho observa-se, 5.994,69 tiveram filhos e 1.360.037,15 não tiveram. Considerando a variável vive com companheiro, 1.529.436,46 pessoas vivem, 1.252.701,32 não viviam e 8.949,58 não souberam ou não quiseram responder sobre viver com o companheiro.

Considerando a variável situação fora do mercado verifica-se que, 418.296 pessoas que são donas de casa, 99.041 são estudantes, 506.334 são aposentados ou pensionista, 34.300 não procuram por trabalho, 23.047 possuem incapacidade temporária ou em auxílio doença, 18.652 com incapacidade permanente e 10.782 não sabem ou não quis responder.

Conforme a variável sabe ler ou escrever constata-se que, 2.351.006 sabem e 434.616 não sabem e 5.465 não sabem ou não quiseram responder ao item. Considerando a variável frequenta escola, 214.674 frequenta escola, 2.385.028 já frequentou, 182.887 nunca frequentou e 8.498 não sabem ou não quiseram responder.

Considerando a variável acesso à internet, 2.224.990 das pessoas tem acesso a internet, ao passo que, 566.097 delas não tem. Considerando a variável usa o aplicativo monitora do consórcio do Nordeste de sintomas da COVID19, 6.500 acompanham o aplicativo e 2.784.587 não fazem monitoramento. Considerando a variável usa o aplicativo coronavírus do SUS de sintomas da COVID19, 19.755 acompanham o aplicativo e 2.771.332 não fazem monitoramento. Considerando a variável acompanhado por sistema telefônico do consórcio Nordeste sobre sintomas da COVID 19, 2.019 acompanham o aplicativo e 2.789.069 não fazem acompanhamento.

Em se tratando da variável teve sintoma de febre nos últimos 7 dias percebe-se que, 80.690 apresentaram febre, 2.706.968 não tiveram e 3.429 não sabem. De acordo com a variável teve sintoma de tosse nos últimos 7 dias, 252.359 tiveram tosse, 2.536.742 não tiveram e 1.985 não sabem. De acordo com a variável teve sintoma perda de olfato/paladar nos últimos 7 dias, 60.447 pessoas indicaram a perda de olfato e paladar, ao passo que, 2.727.890 não tiveram e 2.750 não sabem. De acordo com a variável teve sintoma dificuldade de respirar nos últimos 7 dias, 103.507 respondentes sentiram dificuldade de

respirar, 2.683.258 não tiveram e 4.323 não sabem.

Considerando a variável Saiu de casa, 1.125.792 pessoas que saem de casa quase todos os dias (6 a 7 dias), 588.695 pessoas que saem de casa entre 3 e 5 dias, 635.131 pessoas que saem de casa entre 1 e 2 dias, 435.994 pessoas que não saem de casa, e 5.474 pessoas que não respondem sobre esta variável.

Em se tratando da variável Uso de máscara, 1.910.700 pessoas afirmaram que sempre usam a máscara, 391.992 respondentes que usam a máscara às vezes e 46.927 pessoas que nunca usam máscara. De acordo com a variável Uso de álcool (gel ou líquido), 1.968.487 pessoas que sempre usam, 334.022 pessoas que usam às vezes, 46.773 pessoas que nunca usam, e 337 que não respondem sobre esta variável;

Considerando a variável Diabetes verifica-se, 255.770 pessoas que têm diabetes, 2.102.165 pessoas que não têm diabetes e 15.884 pessoas que não sabem. De acordo com variável Hipertensão, 678.167 pessoas que têm hipertensão, 2.102.165 pessoas que não têm hipertensão e 10.755 pessoas que não sabem. Ao levar em conta a variável Doença no coração, 146.879 pessoas que têm doença no coração, 2.627.329 pessoas que não têm doença no coração e 16.880 pessoas que não sabem. De acordo com a variável Obesidade, 101.506 pessoas que têm obesidade, 2.680.639 pessoas que não têm obesidade e 13.943 pessoas que não sabem. E de acordo com a variável Outra doença crônica, 218.194 pessoas que têm outra doença crônica, 2.561.254 pessoas que não têm outra doença crônica e 11.638 pessoas que não sabem.

Na análise bivariada, temos as tabelas 5.1, 5.2, 5.3, 5.4, 5.5 as quais apresentam os resultados da tabela cruzada e o teste Qui-Quadrado de Pearson segundo a hipótese nula de independência entre as variáveis, para domínios definidos pelos seguintes grupos de variáveis socioeconômicas, Hábitos de Higiene/Proteção e comorbidades e a variável resposta do teste IgG.



Tabela 5.1: Distribuição do resultado do teste IgG segundo as variáveis socioeconômicas

Variáveis	Resposta IgG		P-valor
	Reagente	Não reagente	
<b>Sexo</b>			
<i>Masculino</i>	96.344	1.154.840	0,000
<i>Feminino</i>	160.658	1.379.120	
<b>Faixa de Renda</b>			
<i>Sem renda</i>	1.248	7.227	0,299
<i>Até 300 reais</i>	5.156	79.404	
<i>Entre 301 e 500 reais</i>	12.579	118.498	
<i>Entre 501 e 1000 reais</i>	25.206	350.152	
<i>Entre 1001 e 2000 reais</i>	84.561	873.604	
<i>Entre 2001 e 5000 reais</i>	97.102	813.677	
<i>Entre 5001 e 8000 reais</i>	15.061	133.111	
<i>Mais de 8000 reais</i>	6.821	83.524	
<i>Não sabe</i>	9.269	74.764	
<b>Acesso internet</b>			
<i>Sim</i>	203.719	2.021.271	-
<i>Não</i>	53.284	512.689	
<i>Não sabe</i>	0	0	
<b>Raca ou cor</b>			
<i>Branca</i>	81.400	822.209	0,657
<i>Preta</i>	24.003	193.953	
<i>Amarela</i>	1.745	26.434	
<i>Parda</i>	145.292	1.452.251	
<i>Indígena</i>	3.407	31.982	
<i>Não sabe</i>	1.155	7.131	
<b>Situacao fora mercado</b>			
<i>Dona-de-casa/do lar</i>	45.808	372.488	0,025
<i>Estudante</i>	6.818	92.223	
<i>Aposentado ou pensionista</i>	51.922	454.288	
<i>Não procura por trabalho</i>	2.482	31.818	
<i>Com incapacidade temporária</i>	1.027	22.020	
<i>Com incapacidade permanente</i>	529	18.123	
<i>Outra</i>	3.053	46.334	
<i>Não sabe</i>	2.602	8.180	

Tabela 5.2: Distribuição do resultado do teste IgG segundo as variáveis sociodemográficas

Variáveis	Resposta IgG		P-valor
	Reagente	Não reagente	
<b>Gravida</b>			
<i>Sim</i>	4.772	28.995	0,289
<i>Não</i>	129.694	1.236.338	
<b>Teve filho</b>			
<i>Sim</i>	1.243	4.751	0,157
<i>Não</i>	128.450	1.231.587	
<b>Vive com companheiro</b>			
<i>Sim</i>	150.872	1.378.440	0,157
<i>Não</i>	106.131	1.146.571	
<i>Não sabe</i>	0	8.950	
<b>Sabe ler escrever</b>			
<i>Sim</i>	210.934	2.139.948	0,187
<i>Não</i>	46.069	388.547	
<i>Não sabe</i>	0	5.465	
<b>Frequenta escolar</b>			
<i>Frequenta</i>	19.978	194.696	0,994
<i>Já frequentou</i>	219.236	2.165.668	
<i>Nunca frequentou</i>	16.876	166.011	
<i>Não sabe</i>	912	7.586	

Tabela 5.3: Distribuição do resultado do teste IgG segundo as variáveis de monitoramento dos participantes da pesquisa

Variáveis	Resposta IgG		P-valor
	Reagente	Não reagente	
<b>Usa aplicativos</b>			
<i>A</i>	1.089	3.357	0,246
<i>AB</i>	0	2.055	
<i>B</i>	2.925	14.774	
<i>C</i>	1.078	2.423	
<i>D</i>	231.564	2.290.703	
<i>E</i>	20.347	220.649	
<b>Monitora</b>			
<i>Sim</i>	1.089	5.412	0,384
<i>Não</i>	255.914	2.528.549	
<b>Coronavirus sus</b>			
<i>Sim</i>	2.925	16.829	0,106
<i>Não</i>	254.077	2.517.131	
<b>Monitora</b>			
<i>Sim</i>	656	1.362	0,179
<i>Não</i>	256.346	2.532.598	
<b>Corona</b>			
<i>Sim</i>	2.236	7.800	0,049
<i>Não</i>	254.767	2.526.161	

Tabela 5.4: Distribuição do resultado do teste IgG segundo as variáveis de sintomas.

Variáveis	Resposta IgG		P-valor
	Reagente	Não reagente	
<b>Febre</b>			
<i>Sim</i>	12.084	68.605	0,020
<i>Não</i>	244.581	2.462.263	
<i>Não sabe</i>	337	3.092	
<b>Tosse</b>			
<i>Sim</i>	35.526	216.832	0,000
<i>Não</i>	221.139	2.315.480	
<i>Não sabe</i>	337	1.648	
<b>Olfato e paladar</b>			
<i>Sim</i>	18.303	42.144	0,000
<i>Não</i>	238.083	2.489.683	
<i>Não sabe</i>	617	2.134	
<b>Dificuldade respirar</b>			
<i>Sim</i>	15.445	88.062	0,004
<i>Não</i>	241.220	2.441.913	
<i>Não sabe</i>	337	3.985	
<b>Estabelecimento médico</b>			
<i>Sim</i>	15.253	58.029	0,056
<i>Não</i>	39.132	249.465	
<i>Não sabe</i>	517	1.322	

Tabela 5.5: Distribuição do resultado do teste IgG segundo as variáveis sobre hábitos dos participantes durante a quarentena.

Variáveis	Resposta IgG		P-valor
	Reagente	Não reagente	
<b>Saiu de casa</b>			
<i>Quase todos os dias</i>	106.904	1.018.889	0,851
<i>Entre 3 a 5 dias</i>	55.073	533.622	
<i>Entre 1 e 2 dias</i>	53.221	581.786	
<i>Não saiu de casa</i>	41.123	394.871	
<i>Não sabe</i>	682	4.792	
<b>Saiu de casa - trabalhar</b>			
<i>Sim</i>	89.810	932.324	0,450
<i>Não</i>	125.388	1.201.973	
<b>Saiu de casa - escola</b>			
<i>Sim</i>	2.298	24.377	0,894
<i>Não</i>	212.900	2.109.920	
<b>Saiu de casa - comprar remédios</b>			
<i>Sim</i>	129.211	1.228.632	0,378
<i>Não</i>	85.987	905.665	
<b>Saiu de casa - encontrar amigos</b>			
<i>Sim</i>	87.404	845.703	0,715
<i>Não</i>	127.794	1.288.594	
<b>Saiu de casa - atendimento saúde</b>			
<i>Sim</i>	33.304	316.265	0,748
<i>Não</i>	181.894	1.818.032	
<b>Saiu de casa - fazer exercícios</b>			
<i>Sim</i>	22.704	230.611	0,896
<i>Não</i>	192.494	1.903.686	
<b>Usou máscara</b>			
<i>Sempre</i>	184.049	1.726.527	-
<i>Às vezes</i>	26.984	365.007	
<i>Nunca</i>	4.165	42.763	
<i>Não sabe</i>	0	0	
<b>Lavou ou álcool</b>			
<i>Sempre</i>	185.072	1.783.291	0,514
<i>Às vezes</i>	27.358	306.664	
<i>Nunca</i>	2.768	44.005	
<i>Não sabe</i>	0	337	

Tabela 5.6: Distribuição do resultado do teste IgG segundo as variáveis sobre comorbidades.

Variáveis	Resposta IgG		P-valor
	Reagente	Não reagente	
<b>Diabetes</b>			
<i>Sim</i>	37.398	218.372	0,000
<i>Não</i>	218.629	2.300.680	
<i>Não sabe</i>	976	14.908	
<b>Hipertensão</b>			
<i>Sim</i>	77.756	600.288	0,002
<i>Não</i>	178.312	1.923.854	
<i>Não sabe</i>	935	9.819	
<b>Doença coração</b>			
<i>Sim</i>	16.447	130.432	0,408
<i>Não</i>	238.670	2.388.535	
<i>Não sabe</i>	1.887	14.993	
<b>Obesidade</b>			
<i>Sim</i>	18.015	83.491	0,000
<i>Não</i>	237.376	2.443.140	
<i>Não sabe</i>	1.612	7.330	
<b>Doença crônica</b>			
<i>Sim</i>	21.336	196.859	0,840
<i>Não</i>	234.338	2.326.793	
<i>Não sabe</i>	1.329	10.309	

O teste de hipóteses qui-quadrado tem sua estatística baseada no plano amostral, tem-se associação entre as variáveis quando rejeitamos a hipótese nula do teste de acordo com o nível de significância de 5%,  $p - valor < 0,05$ . A variável socioeconômica sexo ( $p - valor = 0,00$ ) e situação fora do mercado ( $p - valor = 0,02$ ) tiveram associação com a variável resposta do teste IgG. Os sintomas dos últimos 7 dias de febre ( $p - valor = 0,02$ ), tosse ( $p - valor = 0,00$ ), olfato/paladar ( $p - valor = 0,00$ ) e dificuldade de respirar ( $p - valor = 0,02$ ) tiveram associação com a variável resposta do teste IgG. A variáveis relacionadas as comorbidades diabetes ( $p - valor = 0,00$ ), hipertensão ( $p - valor = 0,01$ ) e obesidade ( $p - valor = 0,00$ ) tiveram associação com a variável resposta do teste IgG.

As variáveis monitoramento pelo portal corona ( $p - valor = 0,049$ ) estava próxima da significância estatística e as demais variáveis consideradas apresentaram ( $p - valor > 0,17$ ), e com isso não rejeitamos a hipótese nula do teste de independência entre as variáveis.

### 5.0.1 Regressão logística

Na análise multivariada foi considerada o método de regressão logística com cada variável explicativa categorizada. Inicialmente foi ajustado o modelo de regressão logística para cada uma das covariáveis apresentadas, conforme descrito na Seção 2. Foram mantidas nos passos subsequentes apenas aquelas que apresentaram um p-valor inferior a 0,30.

No processo de filtragem do modelo, o critério para remoção é a aceitação da hipótese nula do teste t, que o coeficiente é estatisticamente igual a zero,  $\beta_p = 0$ , sendo  $p$  o número de variáveis explicativas no ajuste. A ordem de remoção das variáveis segue uma ordem crescente de acordo com o  $p - valor$  para cada conjunto de categorias de cada variável independente.

Após cada tomada de decisão e remoção de uma variável, é feita o teste ANOVA para análise de desvio com o intuito de analisar a influência da variável no modelo. A hipótese nula do teste ANOVA é que a comparação entre o modelo com e sem a variável em discussão apresentam diferença estatística. Com isso, as variáveis sexo, obesidade, usou máscara, idade, sabe ler ou escrever, dificuldade de respirar, febre, hipertensão, e doença no coração foram removidas no ajuste de regressão.

Tabela 5.7: Estimativas dos parâmetros do modelo logístico e avaliação dos riscos

	Estimativa	Erro padrão	Valor t	Pr(> t )	
<b>Intercepto</b>	43,23	2,45	17,62	<2e-16	***
<b>Relação</b>					
<i>Cônjuge ou companheiro(a)</i>	0,81	0,33	2,43	0,02	*
<i>Filho(a), enteado(a)</i>	0,04	0,47	0,08	0,93	
<i>Pai, mãe, sogro(a)</i>	0,15	0,65	0,23	0,82	
<i>Neto(a), bisneto(a)</i>	-0,30	0,54	-0,55	0,58	
<i>Irmão(ã)</i>	0,04	0,57	0,07	0,94	
<i>Nora, genro</i>	15,75	0,31	50,30	<2e-16	***
<i>Outro parente</i>	0,38	0,49	0,78	0,44	
<i>Agregado</i>	-0,38	0,65	-0,59	0,56	
<i>Empregado doméstico</i>	14,35	1,06	13,57	<2e-16	***
<b>Grávida</b>					
<i>Não</i>	-0,46	1,02	-0,45	0,65	
<b>Vive com companheiro</b>					
<i>Não</i>	1,20	0,28	4,29	0,00	***
<i>Não sabe ou não quis responder</i>	15,93	1,16	13,74	<2e-16	***
<b>Situação fora do mercado</b>					
<i>Estudante</i>	-0,16	0,43	-0,38	0,70	
<i>Aposentado ou pensionista</i>	-0,06	0,32	-0,18	0,86	
<i>Não procura por trabalho</i>	2,23	1,15	1,94	0,05	.
<i>Com incapacidade temporária</i>	0,32	1,10	0,29	0,77	
<i>Com incapacidade permanente</i>	2,41	1,52	1,58	0,12	
<i>Outra</i>	0,58	0,55	1,04	0,30	
<i>Não sabe ou não quis responder</i>	-0,24	0,69	-0,35	0,73	
<b>Monitora</b>					
<i>Não</i>	-14,90	1,06	-14,06	<2e-16	***
<b>Corona Virus</b>					
<i>Não</i>	-13,21	1,23	-10,74	<2e-16	***
<b>Corona</b>					
<i>Não</i>	-16,16	0,85	-19,01	<2e-16	***
<b>Tosse</b>					
<i>Não</i>	0,80	0,40	2,00	0,05	*
<b>Olfato e paladar</b>					
<i>Não</i>	1,93	0,48	4,01	0,00	***
<b>Lavou ou álcool</b>					
<i>Às vezes</i>	-0,32	0,34	-0,93	0,35	
<i>Nunca</i>	1,01	1,10	0,91	0,36	
<i>Não sabe ou não respondeu</i>	15,22	1,13	13,44	<2e-16	***
<b>Diabetes</b>					
<i>Não</i>	0,61	0,33	1,86	0,06	
<i>Não sabe</i>	17,26	0,78	22,16	<2e-16	***

Na regressão final mantém-se as variáveis relação, grávida, vive com companheiro, monitora por aplicativo, aplicativo corona vírus, aplicativo corona, tosse, olfato e paladar, lavou as mãos ou fez uso de álcool, diabetes. As variáveis e categorias significantes para



o modelo foram relação com o responsável pelo domicílio e as categorias 'cônjuge' ( $p - valor = 0,02$ ), 'nora/genro' ( $p - valor = < 2e - 16$ ), e 'empregado doméstico' ( $p - valor = < 2e - 16$ ). A variável vive com o companheiro 'não' ( $p - valor = 0,00$ ) e 'não sabe' ( $p - valor = < 2e - 16$ ), monitora ( $p - valor = < 2e - 16$ ), corona vírus ( $p - valor = < 2e - 16$ ), corona ( $p - valor = < 2e - 16$ ), tosse ( $p - valor = 0,05$ ), olfato e paladar ( $p - valor = 0,00$ ), lavou ou álcool e a categoria 'não sabe' ( $p - valor = < 2e - 16$ ) e diabetes chegou próximo a significância com a categoria 'não' e 'não sabe' ( $p - valor = < 2e - 16$ ).

Tabela 5.8: Estimativas Razão de Chances e intervalo de confiança

Variáveis	$\exp(\beta)$	2,5%	97,5%
<b>Intercepto</b>			
<b>Relação</b>			
<i>Cônjuge ou companheiro(a)</i>	2,24	1,17	4,29
<i>Filho(a), enteado(a)</i>	1,04	0,42	2,59
<i>Pai, mãe, sogro(a)</i>	1,16	0,33	4,11
<i>Neto(a), bisneto(a)</i>	1,34	0,26	2,12
<i>Irmão(ã)</i>	1,04	0,34	3,17
<i>Outro parente</i>	1,05	0,56	3,84
<i>Agregado</i>	1,46	0,19	2,44
<b>Gravida</b>			
<i>Não</i>	1,58	0,09	4,65
<b>Vive com companheiro</b>			
<i>Não</i>	3,31	1,92	5,72
<b>Situação fora do mercado</b>			
<i>Estudante</i>	1,18	0,37	1,96
<i>Aposentado ou pensionista</i>	1,06	0,51	1,76
<i>Não procura por trabalho</i>	0,93	0,98	87,86
<i>Com incapacidade temporária ou em auxílio doença</i>	1,37	0,16	11,75
<i>Com incapacidade permanente</i>	1,11	0,56	220,30
<i>Outra</i>	1,78	0,60	5,28
<i>Não sabe ou não quis responder</i>	1,27	0,20	3,02
<b>Monitora</b>			
<i>Não</i>	0,00	0,00	0,00
<b>Corona Virus</b>			
<i>Não</i>	0,00	0,00	0,00
<b>Corona</b>			
<i>Não</i>	0,00	0,00	0,00
<b>Tosse</b>			
<i>Não</i>	2,21	1,02	4,83
<b>Olfato e paladar</b>			
<i>Não</i>	6,91	2,69	17,77
<b>Lavou ou álcool</b>			
<i>Às vezes</i>	1,37	0,37	1,43
<i>Nunca</i>	2,74	0,32	23,80
<b>Diabetes</b>			
<i>Não</i>	1,84	0,97	3,49

Os resultados da tabela 5.8 mostram que, dado o modelo escolhido, as seguintes evidências:

- A variável relação com o responsável pelo domicílio, categoria 'cônjuge', influencia positivamente na chance de receber diagnóstico positivo para COVID-19 com teste IgG. Os resultados mostram que aumenta em 2,24 vezes na chance de receber diag-

nóstico positivo para COVID-19 quando comparado a categoria base 'Responsável pelo domicílio'.

- A variável vive com o companheiro no domicílio, categoria 'não', influencia positivamente na chance de receber diagnóstico positivo para COVID-19 com teste IgG. Os resultados mostram que aumenta em 3,31 vezes na chance de receber diagnóstico positivo para COVID-19 quando comparado a categoria base 'sim'.
- A variável tosse, categoria 'não', influencia positivamente na chance de receber diagnóstico positivo para COVID-19 com teste IgG. Os resultados mostram que aumenta em 2,21 vezes na chance de receber diagnóstico positivo para COVID-19 quando comparado a categoria base 'sim'.
- A variável perda de olfato e paladar, categoria 'não', influencia positivamente na chance de receber diagnóstico positivo para COVID-19 com teste IgG. Os resultados mostram que aumenta em 6,71 vezes na chance de receber diagnóstico positivo para COVID-19 quando comparado a categoria sim 'sim'.
- As variáveis de monitoramento influenciam negativamente nas chances de receber diagnóstico positivo para COVID-19 com teste IgG. Porém, sobrestima a razão de chances.

O SARS-CoV-2 é altamente contagioso, traz complicações sintomáticas para pessoas com comorbidades e, sendo assim as estratégias para contenção do vírus são fundamentais para reduzir mortes, achatar a curva de crescimento e evitar o colapso do sistema de saúde. É o momento de atuação na divulgação de conhecimento em saúde para a população, para que a pandemia que chega ao país seja mitigada, preservando milhares de vidas com atitudes individuais.

O estudo realizado neste trabalho apresentou estimativas pontuais que apontam o reforço de todas as medidas preventivas e sobre a conscientização da população no enfrentamento a pandemia. Devemos lembrar que em situação de convívio domiciliar portadores assintomáticos, ou sintomáticos leves infectados pelo Novo Corona vírus podem contaminar outras pessoas em sua residência.

A semelhança dos sintomas do COVID19 a outras enfermidades comuns na região, também pode aumentar a negligência a doença. Indivíduos que apresentam algum sintoma e não buscam orientações médicas ou realizam o teste, se caracteriza como perda de informação para as estimativas analisadas diariamente. A Pesquisa Continuar Cuidando fez uma análise situacional e evolutiva do Estado da Paraíba, com relação a pandemia, reforça e recomendações quanto à necessária sustentação das medidas preventivas para impedir o crescimento do número de casos.

O presente trabalho demonstrou as variáveis que podem influenciar no resultado do teste de COVID19. Demonstrando que os sintomas recentes, sair de casa e a maioria das comorbidades consideradas na coleta dos dados, de fato, tem associação/relação com

resultado do teste IgG.

## REFERÊNCIAS BIBLIOGRÁFICAS

CAVALCANTE, J. R. et al. Covid-19 no Brasil: evolução da epidemia até a semana epidemiológica 20 de 2020. Epidemiologia e Serviços de Saúde, SciELO Brasil, v. 29, 2020.

COELHO H. F. C., et al., Relatório Final Continuar Cuidando. [S.l.]: UFPB-Secretaria de Saúde do Estado da Paraíba, 2021.

DUARTE, M. D. Q. et al. Covid-19 e os impactos na saúde mental: uma amostra do Rio Grande do Sul, Brasil. Ciência Saúde Coletiva SciELO Public Health, v. 25, p.3401–3411, 2020.

HÁJEK, P. On interpretability in set theories. Commentationes Mathematicae Universitatis Carolinae, Charles University in Prague, Faculty of Mathematics and Physics, v. 12, n. 1, p. 73–79, 1971.

HORVITZ, D. G.; THOMPSON, D. J. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association v. 47, n. 260, p. 663–685, 1952.

MAGNO, L. et al. Desafios e propostas para ampliação da testagem e diagnóstico para covid-19 no Brasil. Ciência Saúde Coletiva, SciELO Brasil, v. 25, p. 3355–3364, 2020.

MCCULLAGH, P.; NELDER, J. A. Generalized linear models. [S.l.]: Routledge, 2019.

NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. Journal of the Royal Statistical Society: Series A (General), Wiley Online Library, v. 135, n. 3, p. 370–384, 1972.

ROSEN, B. A users guide to Pareto ps sampling “.RD Report 2000: 6, Statistics Sweden, 2000.

SÄRNDAL, C. E.; SWENSSON, B.; WRETMAN, J. Model assisted survey sampling. [S.l.]: Springer Science Business Media, 2003.

SILVA P. L. N., et al., Amostragem: Teoria e Prática usando o R. [S.l.: s.n.], 2020.

TURKMAN, M. A. A.; SILVA, G. L. Modelos lineares generalizados—da teoria à prática. In: VIII Congresso Anual da Sociedade Portuguesa de Estatística, Lisboa. [S.l.: s.n.], 2000.

WASHINGTON, D. PanPan American health organization / world health organization. Epidemiological update: Coronavirus Disease (covid-19). <http://www.paho.org>, 2020

## 6.1 Anexo

### 6.1.1 Script

```

# Carrega pacotes requeridos
suppressMessages(library(descr))
suppressMessages(library(survey))
suppressMessages(library(srvyr))
suppressMessages(library(tidyverse))

# Lê arquivo com dados da pesquisa
pbccovid_calib1 <- readRDS(file=~"/TCC/pbccovid_calib1.rds")
# names(pbccovid_calib1$variables)

# Prepara objeto para permitir utilização de sintaxe tidyverse
pbccovid_calib1 <- as_survey_design(pbccovid_calib1)

# Cria objeto selecionando público relevante para análise
pbccovid_calib1_testadoIGG <-
  subset(pbccovid_calib1,
         resp_igg_tr != "Não testado" & resp_igg_tr !=
         "Inconclusivo (ou indeterminado)" & b04_idade > 17)

# Cria variável resposta recodificada para eliminar 'não testados'
pbccovid_calib1_testadoIGG <- pbccovid_calib1_testadoIGG %>%
  mutate(resp_igg_tr1 = as.numeric(resp_igg_tr),
         resp_igg_tr1 = ifelse(resp_igg_tr1 == 4, 2, resp_igg_tr1),
         resp_igg_tr3 = ifelse(resp_igg_tr1 == 3, 2, resp_igg_tr1),
         resp_igg_tr2 = factor(resp_igg_tr3,
                               labels=c("Positivo (reagente)",
                                         "Negativo (não reagente)")))

## resp_igg_tr
# Variáveis Socio-demográficas

# Tabela - sexo
svytable(~b02_sexo, pbccovid_calib1_testadoIGG)

```



```
# Tabela - relação
svytable(~b03_relacao, pbcovid_calib1_testadoIGG)

# Tabela - idade
svytable(~b04_idade, pbcovid_calib1_testadoIGG)

# Tabela - faixa de renda
svytable(~c01_faixa_de_renda, pbcovid_calib1_testadoIGG)

# Tabela - raça
svytable(~d04_raca_ou_cor, pbcovid_calib1_testadoIGG)

# Tabela - com gravida
svytable(~d01_gravida, pbcovid_calib1_testadoIGG)

# Tabela - teve filho
svytable(~d03_teve_filho, pbcovid_calib1_testadoIGG)

# Tabela - vive com companheiro
svytable(~d05_vive_com_companheiro, pbcovid_calib1_testadoIGG)

# Tabela - vive situação emprego
svytable(~d06_situacao_emprego, pbcovid_calib1_testadoIGG)

# Tabela - situação fora mercado
svytable(~d07_situacao_fora_mercado, pbcovid_calib1_testadoIGG)

# Tabela - sabe ler ou escrever
svytable(~d08_sabe_ler_escrever, pbcovid_calib1_testadoIGG)

# Tabela - frequentou escola
svytable(~d09_frequenta_frequentou_escola, pbcovid_calib1_testadoIGG)

# Tabela - acesso
svytable(~c02_acesso_internet, pbcovid_calib1_testadoIGG)

# Tabela - usa aplicativos
svytable(~d14_usa_aplicativos, pbcovid_calib1_testadoIGG)
```

```
# Tabela - Monitora
svytable(~d141_monitora, pbcovid_calib1_testadoIGG)

# Tabela - corana virus sus
svytable(~d142_coronavirus_sus, pbcovid_calib1_testadoIGG)

# Tabela - monitora d161
svytable(~d161_monitora, pbcovid_calib1_testadoIGG)

# Tabela - Corona
svytable(~d162_corona, pbcovid_calib1_testadoIGG)

# Tabela - febre
svytable(~e01_febre, pbcovid_calib1_testadoIGG)

# Tabela - tosse
svytable(~e02_tosse, pbcovid_calib1_testadoIGG)

# Tabela - olfato e paladar
svytable(~e03_olfato_paladar, pbcovid_calib1_testadoIGG)

# Tabela - dificuldade de respirar
svytable(~e04_dificuldade_respirar, pbcovid_calib1_testadoIGG)

# Tabela - estabelecimento medico
svytable(~e05_estabelecimento_medico, pbcovid_calib1_testadoIGG)

# Tabela - saiu de casa
svytable(~f01_saiu_de_casa, pbcovid_calib1_testadoIGG)

# Tabela - trabalhar
svytable(~f02a_trabalhar, pbcovid_calib1_testadoIGG)

# Tabela - comprar remedios
svytable(~f02c_comprar_remedios, pbcovid_calib1_testadoIGG)

# Tabela - encontrar amigos
svytable(~f02d_encontrar_amigos, pbcovid_calib1_testadoIGG)
```

```

# Tabela - atendimento a saude
svytable(~f02e_atendimento_saude, pbcovid_calib1_testadoIGG)

# Tabela - fazer exercicios
svytable(~f02f_fazer_exercicios, pbcovid_calib1_testadoIGG)

# Tabela - usou mascara
svytable(~f03_usou_mascara, pbcovid_calib1_testadoIGG)

# Tabela - lavou ou alcool
svytable(~f04_lavou_ou_alcool, pbcovid_calib1_testadoIGG)

# Tabela - diabetes
svytable(~h01_diabetes, pbcovid_calib1_testadoIGG)

# Tabela - hipertensao
svytable(~h02_hipertensao, pbcovid_calib1_testadoIGG)

# Tabela - Doença Coração
svytable(~h03_doenca_coracao, pbcovid_calib1_testadoIGG)

# Tabela - obesidade
svytable(~h04_obesidade, pbcovid_calib1_testadoIGG)

# Tabela - doença crônica
svytable(~h05_doenca_cronica, pbcovid_calib1_testadoIGG)

## resp_igg_tr
# Variáveis Socio-demográficas
# Testa independência da resposta com sexo
#svytable(~resp_igg_tr1+b02_sexo, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+b02_sexo, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+b02_sexo, pbcovid_calib1_testadoIGG)

mod.IGG <- svyglm(resp_igg_tr2 ~ b02_sexo,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

```



```

        family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta teve filho
svytable(~resp_igg_tr1+d03_teve_filho, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+d03_teve_filho, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+d03_teve_filho, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ d03_teve_filho,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta raça
svytable(~resp_igg_tr1+d04_raca_ou_cor, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+d04_raca_ou_cor, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+d04_raca_ou_cor, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ d04_raca_ou_cor,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta vive com companheiro
svytable(~resp_igg_tr1+d05_vive_com_companheiro, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+d05_vive_com_companheiro, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+d05_vive_com_companheiro, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ d05_vive_com_companheiro,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta vive situação emprego
svytable(~resp_igg_tr1+d06_situacao_emprego, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+d06_situacao_emprego, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+d06_situacao_emprego, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ d06_situacao_emprego,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

```

```

# Testa independência da resposta vive situação fora mercado
svytable(~resp_igg_tr1+d07_situacao_fora_mercado, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+d07_situacao_fora_mercado, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+d07_situacao_fora_mercado, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ d07_situacao_fora_mercado,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta com sabe ler ou escrever
svytable(~resp_igg_tr1+d08_sabe_ler_escrever, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+d08_sabe_ler_escrever, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+d08_sabe_ler_escrever, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ d08_sabe_ler_escrever,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta com frequentou escola
svytable(~resp_igg_tr1+d09_frequenta_frequentou_escola, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+d09_frequenta_frequentou_escola, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+d09_frequenta_frequentou_escola, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ d09_frequenta_frequentou_escola,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta rede
svytable(~resp_igg_tr1+d10_rede, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+d10_rede, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+d10_rede, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ d10_rede,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta com usa aplicativos
svytable(~resp_igg_tr1+d14_usa_aplicativos, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+d14_usa_aplicativos, pbcovid_calib1_testadoIGG)

```



```

        family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta com febre
svytable(~resp_igg_tr1+e01_febre, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+e01_febre, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+e01_febre, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ e01_febre,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta com tosse
svytable(~resp_igg_tr1+e02_tosse, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+e02_tosse, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+e02_tosse, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ e02_tosse,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta com olfato e paladar
svytable(~resp_igg_tr1+e03_olfato_paladar, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+e03_olfato_paladar, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+e03_olfato_paladar, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ e03_olfato_paladar,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta com dificuldade de respirar
svytable(~resp_igg_tr1+e04_dificuldade_respirar, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+e04_dificuldade_respirar, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+e04_dificuldade_respirar, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ e04_dificuldade_respirar,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

```



```

# Testa independência da resposta com estabelecimento medico
svytable(~resp_igg_tr1+e05_estabelecimento_medico, pbccovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+e05_estabelecimento_medico, pbccovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+e05_estabelecimento_medico, pbccovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ e05_estabelecimento_medico,
                  design = pbccovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

```

```

# Testa independência da resposta com saiu de casa
svytable(~resp_igg_tr1+f01_saiu_de_casa, pbccovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+f01_saiu_de_casa, pbccovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+f01_saiu_de_casa, pbccovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ f01_saiu_de_casa,
                  design = pbccovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

```

```

# Testa independência da resposta com trabalhar
svytable(~resp_igg_tr1+f02a_trabalhar, pbccovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+f02a_trabalhar, pbccovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+f02a_trabalhar, pbccovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ f02a_trabalhar,
                  design = pbccovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

```

```

# Testa independência da resposta com escolha
svytable(~resp_igg_tr1+f02b_escola, pbccovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+f02b_escola, pbccovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+f02b_escola, pbccovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ f02b_escola,
                  design = pbccovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

```

```

# Testa independência da resposta com comprar remedios
svytable(~resp_igg_tr1+f02c_comprar_remedios, pbccovid_calib1_testadoIGG)

```

```
svytable(~resp_igg_tr2+f02c_comprar_remedios, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+f02c_comprar_remedios, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ f02c_comprar_remedios,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)
```

```
# Testa independência da resposta com encontrar amigos
svytable(~resp_igg_tr1+f02d_encontrar_amigos, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+f02d_encontrar_amigos, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+f02d_encontrar_amigos, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ f02d_encontrar_amigos,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)
```

```
# Testa independência da resposta com atendimento a saude
svytable(~resp_igg_tr1+f02e_atendimento_saude, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+f02e_atendimento_saude, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+f02e_atendimento_saude, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ f02e_atendimento_saude,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)
```

```
# Testa independência da resposta com fazer exercicios
svytable(~resp_igg_tr1+f02f_fazer_exercicios, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+f02f_fazer_exercicios, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+f02f_fazer_exercicios, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ f02f_fazer_exercicios,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)
```

```
# Testa independência da resposta com usou mascara
svytable(~resp_igg_tr1+f03_usou_mascara, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+f03_usou_mascara, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+f03_usou_mascara, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ f03_usou_mascara,
```

```

        design = pbcovid_calib1_testadoIGG,
        family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta com lavou ou alcool
svytable(~resp_igg_tr1+f04_lavou_ou_alcool, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+f04_lavou_ou_alcool, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+f04_lavou_ou_alcool, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ f04_lavou_ou_alcool,
                 design = pbcovid_calib1_testadoIGG,
                 family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta com diabetes
svytable(~resp_igg_tr1+h01_diabetes, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+h01_diabetes, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+h01_diabetes, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ h01_diabetes,
                 design = pbcovid_calib1_testadoIGG,
                 family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta com hipertensao
svytable(~resp_igg_tr1+h02_hipertensao, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+h02_hipertensao, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+h02_hipertensao, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ h02_hipertensao,
                 design = pbcovid_calib1_testadoIGG,
                 family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta com Doença Coração
svytable(~resp_igg_tr1+h03_doenca_coracao, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+h03_doenca_coracao, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+h03_doenca_coracao, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ h03_doenca_coracao,
                 design = pbcovid_calib1_testadoIGG,
                 family = quasibinomial(link = "logit"))
summary(mod.IGG)

```

```

# Testa independência da resposta com obesidade
svytable(~resp_igg_tr1+h04_obesidade, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+h04_obesidade, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+h04_obesidade, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ h04_obesidade,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta com doença crônica
svytable(~resp_igg_tr1+h05_doença_cronica, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+h05_doença_cronica, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+h05_doença_cronica, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ h05_doença_cronica,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta com doença crônica
svytable(~resp_igg_tr1+usou_mascara, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+usou_mascara, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+usou_mascara, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ usou_mascara,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

# Testa independência da resposta com doença crônica
svytable(~resp_igg_tr1+usou_alcool, pbcovid_calib1_testadoIGG)
svytable(~resp_igg_tr2+usou_alcool, pbcovid_calib1_testadoIGG)
svychisq(~resp_igg_tr2+usou_alcool, pbcovid_calib1_testadoIGG)
mod.IGG <- svyglm(resp_igg_tr2 ~ usou_alcool,
                  design = pbcovid_calib1_testadoIGG,
                  family = quasibinomial(link = "logit"))
summary(mod.IGG)

## Regressão -----
mod.IGG <- svyglm(resp_igg_tr2 ~

```

```

b02_sexo +
b03_relacao +
b04_idade +
d01_gravida +
d05_vive_com_companheiro +
d07_situacao_fora_mercado +
d08_sabe_ler_escrever +
d141_monitora +
d142_coronavirus_sus +
d162_corona +
e01_febre +
e02_tosse +
e03_olfato_paladar +
e04_dificuldade_respirar +
f03_usou_mascara +
f04_lavou_ou_alcool +
h01_diabetes +
h02_hipertensao +
h03_doenca_coracao +
h04_obesidade,
design = pbccovid_calib1_testadoIGG,
family = quasibinomial(link = "logit"))
summary(mod.IGG)

```

```

mod.IGG1 <- svyglm(resp_igg_tr2 ~
  b03_relacao +
  b04_idade +
  d01_gravida +
  d05_vive_com_companheiro +
  d07_situacao_fora_mercado +
  d08_sabe_ler_escrever +
  d141_monitora +
  d142_coronavirus_sus +
  d162_corona +
  e01_febre +
  e02_tosse +
  e03_olfato_paladar +
  e04_dificuldade_respirar +
  f03_usou_mascara +

```

```

        f04_lavou_ou_alcool +
        h01_diabetes +
        h02_hipertensao +
        h03_doenca_coracao +
        h04_obesidade,
    design = pbcovid_calib1_testadoIGG,
    family = quasibinomial(link = "logit"))
summary(mod.IGG1)

anova(mod.IGG,mod.IGG1) # p = 0.96116 sex0
#_-----
mod.IGG2 <- svyglm(resp_igg_tr2 ~
                    b03_relacao +
                    b04_idade +
                    d01_gravida +
                    d05_vive_com_companheiro +
                    d07_situacao_fora_mercado +
                    d08_sabe_ler_escrever +
                    d141_monitora +
                    d142_coronavirus_sus +
                    d162_corona +
                    e01_febre +
                    e02_tosse +
                    e03_olfato_paladar +
                    e04_dificuldade_respirar +
                    f03_usou_mascara +
                    f04_lavou_ou_alcool +
                    h01_diabetes +
                    h02_hipertensao +
                    h03_doenca_coracao,
    design = pbcovid_calib1_testadoIGG,
    family = quasibinomial(link = "logit"))
summary(mod.IGG2)

anova(mod.IGG1,mod.IGG2) # p= 0.70455 obesidade
#_-----
mod.IGG3 <- svyglm(resp_igg_tr2 ~
                    b03_relacao +
                    b04_idade +

```

```

d01_gravida +
d05_vive_com_companheiro +
d07_situacao_fora_mercado +
d08_sabe_ler_escrever +
d141_monitora +
d142_coronavirus_sus +
d162_corona +
e01_febre +
e02_tosse +
e03_olfato_paladar +
e04_dificuldade_respirar +
f04_lavou_ou_alcool +
h01_diabetes +
h02_hipertensao +
h03_doenca_coracao,
design = pbcovid_calib1_testadoIGG,
family = quasibinomial(link = "logit"))
summary(mod.IGG3)

anova(mod.IGG2,mod.IGG3) # p= 0.25172  f03_usou_mascara
#-----
mod.IGG4 <- svyglm(resp_igg_tr2 ~
b03_relacao +
d01_gravida +
d05_vive_com_companheiro +
d07_situacao_fora_mercado +
d08_sabe_ler_escrever +
d141_monitora +
d142_coronavirus_sus +
d162_corona +
e01_febre +
e02_tosse +
e03_olfato_paladar +
e04_dificuldade_respirar +
f04_lavou_ou_alcool +
h01_diabetes +
h02_hipertensao +
h03_doenca_coracao,
design = pbcovid_calib1_testadoIGG,

```

```

        family = quasibinomial(link = "logit"))
summary(mod.IGG4)

anova(mod.IGG3,mod.IGG4) # p= 0.28011  idade
#-----
mod.IGG5 <- svyglm(resp_igg_tr2 ~
                    b03_relacao +
                    d01_gravida +
                    d05_vive_com_companheiro +
                    d07_situacao_fora_mercado +
                    d141_monitora +
                    d142_coronavirus_sus +
                    d162_corona +
                    e01_febre +
                    e02_tosse +
                    e03_olfato_paladar +
                    e04_dificuldade_respirar +
                    f04_lavou_ou_alcool +
                    h01_diabetes +
                    h02_hipertensao +
                    h03_doenca_coracao,
                    design = pbcovid_calib1_testadoIGG,
                    family = quasibinomial(link = "logit"))
summary(mod.IGG5)

anova(mod.IGG4,mod.IGG5) # p= 0.57005  d08_sabe_ler_escrever
#-----
mod.IGG6 <- svyglm(resp_igg_tr2 ~
                    b03_relacao +
                    d01_gravida +
                    d05_vive_com_companheiro +
                    d07_situacao_fora_mercado +
                    d141_monitora +
                    d142_coronavirus_sus +
                    d162_corona +
                    e01_febre +
                    e02_tosse +
                    e03_olfato_paladar +
                    f04_lavou_ou_alcool +

```



```

        h01_diabetes +
        h02_hipertensao +
        h03_doenca_coracao,
design = pbcovid_calib1_testadoIGG,
family = quasibinomial(link = "logit"))
summary(mod.IGG6)

anova(mod.IGG5,mod.IGG6) # p= 0.71728  e04_dificuldade_respirar
#_-----
mod.IGG7 <- svyglm(resp_igg_tr2 ~
        b03_relacao +
        d01_gravida +
        d05_vive_com_companheiro +
        d07_situacao_fora_mercado +
        d141_monitora +
        d142_coronavirus_sus +
        d162_corona +
        e02_tosse +
        e03_olfato_paladar +
        f04_lavou_ou_alcool +
        h01_diabetes +
        h02_hipertensao +
        h03_doenca_coracao,
design = pbcovid_calib1_testadoIGG,
family = quasibinomial(link = "logit"))
summary(mod.IGG7)

anova(mod.IGG6,mod.IGG7) # p= 0.37934  e01_febre
#_-----
mod.IGG8 <- svyglm(resp_igg_tr2 ~
        b03_relacao +
        d01_gravida +
        d05_vive_com_companheiro +
        d07_situacao_fora_mercado +
        d141_monitora +
        d142_coronavirus_sus +
        d162_corona +
        e02_tosse +
        e03_olfato_paladar +

```

```

        f04_lavou_ou_alcool +
        h01_diabetes +
        h03_doenca_coracao,
design = pbcovid_calib1_testadoIGG,
family = quasibinomial(link = "logit"))
summary(mod.IGG8)

```

```

anova(mod.IGG7,mod.IGG8) # p= 0.10991   h02_hipertensao
#_-----

```

```

mod.IGG9 <- svyglm(resp_igg_tr2 ~
        b03_relacao +
        d01_gravida +
        d05_vive_com_companheiro +
        d07_situacao_fora_mercado +
        d141_monitora +
        d142_coronavirus_sus +
        d162_corona +
        e02_tosse +
        e03_olfato_paladar +
        f04_lavou_ou_alcool +
        h01_diabetes ,
design = pbcovid_calib1_testadoIGG,
family = quasibinomial(link = "logit"))
summary(mod.IGG9)

```

```

anova(mod.IGG8,mod.IGG9) # p= 0.10056   h03_doenca_coracao
#_-----

```

```

mod.IGG <- svyglm(resp_igg_tr2 ~
        b03_relacao +
        d01_gravida +
        d05_vive_com_companheiro +
        d07_situacao_fora_mercado +
        d141_monitora +
        d142_coronavirus_sus +
        d162_corona +
        e02_tosse +
        e03_olfato_paladar +
        f04_lavou_ou_alcool +
        h01_diabetes ,

```

```
design = pbcovid_calib1_testadoIGG,  
family = quasibinomial(link = "logit")  
summary(mod.IGG)
```