



UNIVERSIDADE FEDERAL DO TOCANTINS
CÂMPUS DE PALMAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM
COMPUTACIONAL DE SISTEMAS

HUMBERTO DA COSTA FRIZZERA

**FATORES DE RISCO NA MORTALIDADE DE PACIENTES
BRASILEIROS HOSPITALIZADOS COM COVID-19:
CLÍNICOS, SOCIAIS, ECONÔMICOS E REGIONAIS.**

Palmas/TO
2023

HUMBERTO DA COSTA FRIZZERA

**FATORES DE RISCO NA MORTALIDADE DE PACIENTES
BRASILEIROS HOSPITALIZADOS COM COVID-19:
CLÍNICOS, SOCIAIS, ECONÔMICOS E REGIONAIS.**

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional de Sistemas da Universidade Federal do Tocantins. Foi avaliada para obtenção do título de Mestre em Modelagem Computacional de Sistemas e aprovada em sua forma final pelo orientador e pela Banca Examinadora.

Orientador: Dr. David Nadler Prata
Coorientador: Dr. Waldecy Rodrigues

Palmas/TO
2023

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal do Tocantins

F921f Frizzera, Humberto da Costa.

Fatores de risco na mortalidade de pacientes brasileiros hospitalizados com COVID-19: clínicos, sociais, econômicos e regionais. / Humberto da Costa Frizzera. – Palmas, TO, 2023.

100 f.

Dissertação (Mestrado Acadêmico) - Universidade Federal do Tocantins – Câmpus Universitário de Palmas - Curso de Pós-Graduação (Mestrado) em Modelagem Computacional de Sistemas, 2023.

Orientador: Dr. David Nadler Prata

Coorientador: Dr. Waldecy Rodrigues

1. Fatores clínicos, sociais, econômicos. 2. Fatores regionais. 3. Mortalidade em hospitalizados. 4. COVID-19. I. Título

CDD 004

TODOS OS DIREITOS RESERVADOS – A reprodução total ou parcial, de qualquer forma ou por qualquer meio deste documento é autorizado desde que citada a fonte. A violação dos direitos do autor (Lei nº 9.610/98) é crime estabelecido pelo artigo 184 do Código Penal.

Elaborado pelo sistema de geração automática de ficha catalográfica da UFT com os dados fornecidos pelo(a) autor(a).

FOLHA DE APROVAÇÃO

HUMBERTO DA COSTA FRIZZERA

Fatores de risco na mortalidade de pacientes brasileiros hospitalizados com COVID-19: clínicos, sociais, econômicos e regionais.

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional de Sistemas da Universidade Federal do Tocantins. Foi avaliada para obtenção do título de Mestre em Modelagem Computacional de Sistemas e aprovada em sua forma final pelo orientador e pela Banca Examinadora.

Data de aprovação: 16 / 02 / 2023

Banca Examinadora



Documento assinado digitalmente

DAVID NADLER PRATA

Data: 24/02/2023 11:41:31-0300

Verifique em <https://verificador.iti.br>

Prof. Dr. David Nadler Prata, UFT



Documento assinado digitalmente

WALDECY RODRIGUES

Data: 24/02/2023 19:29:34-0300

Verifique em <https://verificador.iti.br>

Prof. Dr. Waldecy Rodrigues, UFT



Documento assinado digitalmente

MARCELO LISBOA ROCHA

Data: 24/02/2023 19:41:43-0300

Verifique em <https://verificador.iti.br>

Prof. Dr. Marcelo Lisboa Rocha, UFT



Documento assinado digitalmente

MARCOS GONTIJO DA SILVA

Data: 24/02/2023 19:41:43-0300

Verifique em <https://verificador.iti.br>

Prof. Dr. Marcos Gontijo da Silva, UFT

Palmas – TO, 2023

*Dedico a minha esposa e meu pai que sempre
me apoiaram, me fortaleceram com palavras
de motivação e orações, para me manter de pé
na busca dos meus sonhos.*

*“O trabalho duro vence o talento, sempre que
o talento não trabalha duro.”
(Kevin Durant)*

AGRADECIMENTOS

Muitas pessoas me ajudaram ao longo desta caminhada, desde a decisão de se matricular no mestrado até a finalização deste trabalho. Sem dúvida, ter o apoio de muitos pais e amigos foi fundamental para minha determinação de terminar esta fase do meu desenvolvimento. Assim, quero agradecer a todos que se prontificaram a ajudar, seja com uma revisão de texto, uma frase de incentivo, ou mesmo apenas entendendo minha ausência em eventos festivos.

Primeiramente quero agradecer a Deus, por ter me dado forças pra sempre seguir e nunca desistir perante as dificuldades, mesmo nos momentos difíceis, esteve comigo me dando saúde mental e física para poder seguir no mestrado.

Gostaria de expressar minha gratidão à minha esposa, Mikaela, que me apoiou na decisão de me inscrever no mestrado do PPGMCS da UFT e que me demonstrou por meio de suas ações e palavras que compartilhava do meu objetivo de concluir o projeto. Quero agradecer a todos, que me apoiaram ao longo do mestrado pela paciência com minhas ausências em vários pontos, pelas inúmeras horas que passei ausente, enquanto eu me concentrava nas pesquisas, códigos, métodos e prazos, pela compreensão quando as datas de entrega de minhas tarefas coincidiram com os fins de semana mais curtos e para todo o resto.

Agradecer aos meus pais, Anselmo e Lena, que estiveram presente na minha educação e me incentivando nos meus sonhos, por entenderem e apoiarem minha ausência. Ao meu pai que foi insistente em me mostrar que apenas me esforçando no caminho da educação eu encontraria um futuro melhor; que muitas vezes abriram mão de seus próprios sonhos para permitir que eu e meu irmão pudéssemos construir os nossos.

Ao meu orientador Dr. David Nadler, quero agradecer por ter me apoiado desde que eu entrei no PPGMCS e me mostrou como funcionava todo o mestrado. Ao meu coorientador Waldecy Rodrigues, por estar participando e apoiando desde o início desse projeto aqui descrito.

Quero agradecer aos meus colegas de mestrados, em principal a Daniela, que fez parte desse projeto, e também em outros, apoiando e sempre incentivando nos estudos.

Por fim, agradeço a UFT por esse programa incrível e a todos os professores que instruiu com seus conhecimentos a nós alunos do programa.

Por fim peço perdão se esqueci de alguém, mas lembre-se que estará sempre nos meus agradecimentos e nas minhas orações. Obrigado família e amigos por tudo.

RESUMO

No ano de 2020, o Brasil ocupou o terceiro lugar mundial em números de casos e o segundo lugar em mortes por COVID-19. Entender os fatores de riscos, desigualdade social e étnicas em saúde é de suma importância, conforme a multiplicidade populacional e suas fragilidades políticas e econômicas. O contexto do Brasil tem apresentado gravidade no surto, pelos fatores sociais e econômicos e por sua composição racial complexa. O objetivo desta pesquisa foi definir os fatores de risco de mortalidade dos pacientes que necessitam de hospitalização durante a infecção do vírus COVID-19, a partir das suas características econômicas, sociais e epidemiológicas. Foram utilizados dados públicos de 1.956.350 registros epidemiológicos de pacientes brasileiros que procuraram assistência hospitalar na pandemia COVID-19 de fevereiro/2020 a maio/2021. A estatística descritiva permitiu quantificar o impacto da pandemia, demonstrando as diferenças entre regiões, etnias, idades e as comorbidades registradas. Para a análise de predição, foi utilizado o método de regressão logística. Foram avaliados os fatores relacionados de o paciente evoluir para o óbito, considerando suas comorbidades e seus fatores sociais, primordialmente sua renda relativa e se é analfabeto. As chances de morte são maiores entre pacientes com comorbidades, doenças neurológicas (1,99) e renal (1,97) e distúrbios da imunodeficiência (1,69). Enquanto a renda relativa (2,45) indica que os fatores sociais têm maior influência na mortalidade do que as comorbidades estudadas. Os pacientes residentes na região Norte têm maior chance de mortalidade em relação aos do Centro-Sul. Concluímos que as circunstâncias socioeconômicas, que também foram extremamente significativas, tiveram um impacto mais forte nos riscos de mortalidade COVID-19 no Brasil durante o período do estudo do que as comorbidades que ocorrem naturalmente (doenças neurológicas, renais e de imunodeficiência). Dado que mais pessoas são suscetíveis a circunstâncias empobrecidas, as variáveis regionais nas taxas de mortalidade são importantes.

Palavras-chaves: fatores sociais, fatores econômicos, fatores regionais, mortalidade em hospitalizados, COVID-19.

ABSTRACT

In the year 2020, Brazil occupied third place worldwide in case numbers and second place in deaths from Covid-19. Understanding the factors of risk, social inequality and ethnic health is of paramount importance, according to population multiplicity and its political and economic weaknesses. The context of Brazil has presented gravity in the outbreak, for social and economic factors and its complex racial composition. The objective of this research was to define the risk factors of mortality from patients who need hospitalization during the infection of the Covid-19 virus, based on their economic, social and epidemiological characteristics. Public data of 1,956,350 epidemiological records of Brazilian patients who sought hospital care at the COVID-19 pandemic from February/2020 to May/2021 were used. Descriptive statistics allowed to quantify the impact of the pandemic, demonstrating the differences between regions, ethnicities, ages and the registered comorbidities. For prediction analysis, the logistics regression method was used. The related factors of the patient evolve to death, considering his comorbidities and social factors, primarily his relative income and if illiterate were evaluated. Chances of death are higher among patients with comorbidities, neurological (1.99) and renal (1.97) diseases and immunodeficiency disorders (1.69). While relative income (2,45) indicates that social factors have greater influence on mortality than the studied comorbidities. Patients' resident in the northern region has a greater chance of mortality than those of the Center-South. We conclude that the socioeconomic circumstances, which were also extremely significant, had a stronger impact on the risks of covid-19 mortality in Brazil during the study period than naturally occurring (neurological, kidney and immunodeficiency diseases). Since more people are susceptible to impoverished circumstances, regional variables in mortality rates are important.

Key-words: social factors, economic factors, regional factors, mortality in hospitalized, COVID-19.

LISTA DE ILUSTRAÇÕES

Figura 1 – Termo epidemiologia derivado em grego.	23
Figura 2 – Principais paradigmas metodológicos na epidemiologia.	25
Figura 3 – Classificação de Problemas de Qualidade dos Dados.....	28
Figura 4 – Fases do processo do KDD.	30
Figura 5 - Principais etapas do pré-processamento	33
Figura 6 - Exemplo de Sistema Gerenciador de Banco de Dados.....	35
Figura 7 - Ilustração de conceito de Banco de Dados.	41
Figura 8 - Modelo de entidade-relacionamento do Banco de Dados.	42
Figura 9 - MER lógico do Banco de Dados.....	43
Figura 10 – Detalhamento das etapas desenvolvidas na criação e importação dos dados	45
Quadro 1 - Pseudo código do algoritmo Apriori.	32

LISTA DE TABELAS

Tabela 1 – Descrições das fases de pandemia da OMS e principais ações por fase.	18
Tabela 2 – Fatores de confiabilidade na qualidade de dados.	27

LISTA DE ABREVIATURAS E SIGLAS

PPGMCS	Programa de Pós-Graduação em Modelagem Computacional de Sistemas
UFT	Universidade Federal do Tocantins
SARS	Severe Acute Respiratory Syndrome
MERS	Middle East Respiratory Syndrome
OMS	Organização Mundial de Saúde
ONU	Organização das Nações Unidas
COVID-19	Corona Virus Disease 2019
SRAG	Síndrome Respiratória Aguda Grave Hospitalizados
SIDRA	Sistema IBGE de Recuperação Automática
AIDS	Acquired Immune Deficiency Syndrome
SIS	Sistemas de Informação de Saúde
KDD	Knowledge Discovery in Database
SGBD	Sistema Gerenciador de Banco de Dados
SQL	Structured Query Language
DDL	Data Definition Language
DML	Data Manipulation Language
BD	Banco de Dados
IBGE	Instituto Brasileiro de Geografia e Estatística
PNAD	Pesquisa Nacional por Amostragem de Domicílios
CSV	Comma-Separated Values
MS	Ministério da Saúde
SVS	Sistema de Vigilância em Saúde
SG	Síndrome Gripal
SIVEP-Gripe	Sistema de Informação da Vigilância Epidemiológica da Gripe
MER	Modelo de Entidade-Relacionamento
SAS	Statistical Analysis System
IRR	Índice de Renda Relativa
GLM	Generalized Linear Model

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Justificativa	15
1.2	Objetivos.....	16
1.2.1	Objetivo Geral	16
1.2.2	Objetivos Específicos	16
1.3	Estrutura Da Dissertação.....	16
2	FUNDAMENTAÇÃO TEÓRICA.....	17
2.1	Pandemia	17
2.1.1	Principais pandemias que assolaram o planeta.....	19
2.1.2	Pandemias ativas.....	20
2.2	Epidemiologia.....	22
2.3	Epidemiologia Computacional	23
2.4	Importância dos Dados	25
2.4.1	A importância dos dados na saúde.....	25
2.4.2	Qualidade dos dados	26
2.4.3	Problemas com a qualidade dos dados	27
2.5	Tratamento e armazenamento dos Dados	28
2.5.1	Descoberta de Conhecimento em Base de Dados.....	28
2.5.1.1	<i>Algoritmo APRIORI.....</i>	<i>30</i>
2.5.1.2	<i>Pré-processamento dos Dados</i>	<i>31</i>
2.5.2	Sistema Gerenciador de Banco de Dados.....	33
2.6	Trabalho importante da área	35
3	METODOLOGIA.....	36
3.1	Estruturação dos dados.....	36
3.1.1	Coleta de dados.....	36
3.1.1.1	<i>Dados do SRAG.....</i>	<i>36</i>
3.1.1.2	<i>Dados da PNAD</i>	<i>37</i>
3.1.2	Tratamento dos dados	38
3.1.2.1	<i>Pré-processamento</i>	<i>38</i>
3.1.3	Modelagem, Importação e exportação do banco de dados.....	39
3.1.3.1	<i>Projeto conceitual e lógico.....</i>	<i>40</i>
3.1.3.2	<i>Projeto físico e Importação e exportação dos dados</i>	<i>43</i>
3.2	Desenho de estudo e Análise estatística	45
3.2.1	Desenhos e visão geral de estudo	45
3.2.2	Análise de dados	46
3.2.3	Materiais Utilizados.....	48
4	ARTIGO: SOCIAL, ECONOMIC, AND REGIONAL DETERMINANTS OF MORTALITY IN HOSPITALIZED PATIENTS WITH COVID-19 IN BRAZIL	50
5	CONSIDERAÇÕES FINAIS.....	59
5.1	Contribuições da dissertação	61
5.2	Trabalhos futuros	61
	REFERÊNCIAS.....	62
	APÊNDICE A – SQL, PARA EXPORTAÇÃO DA PLANILHA.....	70
	APÊNDICE B – CÓDIGO FONTE REGRESSÃO LOGÍSTICA NO SAS... 71	
	APÊNDICE C – RESULTADOS DOS TESTES DE HIPÓTESES.....	72
	ANEXO A – DICIONÁRIO DE DADOS DAS INFORMAÇÕES EXPORTADAS.....	86

1 INTRODUÇÃO

Um estudo realizado por Jimenez *et al* (2003), evidenciou que havia mais de 30 doenças infecciosas com probabilidade de gerar impacto na saúde da população mundial, enfatizando que essas doenças atingiram a escala global, tornando-se pandemias. Estando presente vários tipos de doenças infecciosas, como as causadas por vírus, que desencadeiam quadros clínicos variáveis que variam desde um simples resfriado à quadros respiratórios graves, como exemplo o vírus SARS-CoV (*Severe Acute Respiratory Syndrome*) 1 e 2 e MERS-CoV (*Middle East Respiratory Syndrome*) (WU *et al*, 2020).

Recentemente, temos a pandemia causada pelo vírus do SARS-CoV-2, onde os primeiros casos suspeitos foram notificados em 31 de dezembro de 2019 na cidade chinesa de Wuhan, que para a OMS a doença que, até então, era definida como uma “pneumonia de origem desconhecida”, e que adiante rompeu a fronteira geográfica da China, tendo casos registrados inicialmente na Europa, nisso a OMS concluiu que se tratava de um surto patológico causado pelo SARS-CoV-2, responsável por causar a Síndrome Respiratória Aguda Grave (OLIVEIRA *et al*, 2020).

Meses depois, mais especificamente no dia 11 de março de 2020, devida a vasta expansão do vírus, pelo seu elevado potencial de transmissibilidade e letalidade, foi declarado, pela OMS (2020), como uma pandemia, já atingindo países de todo mundo (SOUZA-JÚNIOR *et al*, 2020). Segundo Waldman *et al* (1998), define-se pandemias quando há uma disseminação de uma doença para várias localidades geográficas. O então presidente da OMS, Tedros Adhanom Ghebreyesus, recomendou-se a todos os países que se realizasse variações de testagens para que seja confirmado a presença do vírus, como estratégia de enfrentamento (ONU News, 2020).

Anteriormente a pandemia causada pelo vírus SARS-CoV-2, o mundo já havia sofrido com outras pandemias, entre os anos de 2009 e 2010, causada por uma mutação do vírus Influenza surgiu a H1N1, conhecida também por Influenza A ou gripe suína, doença inicialmente identificada no México, mas que em dois meses já havia atingido cerca de 75 países. Com o fim da pandemia ficou registrado 650 mil casos e cerca de 18.400 mortes no mundo pela H1N1 (BUTANTAN, 2022). Já a COVID-19, nome dado pela OMS que se refere a (CO)rona (VI)rus (D)isease – 20(19), causada pelo vírus SARS-CoV-2, com mais de 588 milhões de casos e mais de 6,4 milhões de mortes (WHO, 2022a).

Os principais sintomas da COVID-19 são: Febre, Tosse, Cansaço e Perda de paladar ou Olfato. A diversas medidas que estão sendo tomadas para o enfrentamento, com intuito de evitar

novos casos e proteger a população, umas delas foi a utilização de medidas consideradas clássicas, como por exemplo, a quarentena (WILDER-SMITH *et al*, 2020).

Portanto, a COVID-19 é uma enfermidade que possui rápida transmissibilidade, o que é retratado nos dados epidemiológicos, além de ser de fácil transmissão, possui um alto grau de mortalidade. No Brasil o primeiro caso do COVID-19, foi registrado no dia 26 de fevereiro de 2020. Em maio de 2020, o Brasil se tornou o terceiro lugar mundial em números de casos e o segundo lugar em mortes por COVID-19 (HOPKINS, 2021).

Conforme publicado pelo Ministério da Saúde do Brasil (2021), são considerados grupo de risco para agravamento da COVID-19 a população com idade igual ou superior a 60 anos, que fazem uso do tabagismo, com obesidade, miocardiopatias de diferentes etiologias, hipertensão arterial, doença cerebrovascular, pneumopatias graves ou descompensadas, imunodepressão ou imunossupressão, doenças renais crônicas em estágio avançado, diabetes melito, doenças cromossômicas com estado de fragilidade imunológica, neoplasia maligna, cirrose hepática, algumas doenças hematológicas e em período de gestação.

Na pandemia, ficou evidente o abismo dos fatores sociais e econômicos deixando a disparidade social nacional e global mais evidente (LIMA *et al*, 2020; FARIAS *et al*, 2021; CARVALHO *et al*, 2020). Segundo Pires (2020), a população com renda mais baixa é a mais desfavorecida devido ao sistema de combate à pandemia, com menor acessibilidade à saúde e saneamento básico, e nas estratégias de prevenção com dificuldade de manter o isolamento social, por conta do emprego e renda.

Como enfrentamento e combate a COVID-19, foi criado no Brasil a LEI Nº 13.979, do dia 6 de fevereiro de 2020, que dispõe sobre as medidas para enfrentamento da emergência de saúde pública de importância internacional, definido no seu 2º artigo: Isolamento e Quarentena.

Na área da saúde, a realização da análise aprofundada dos dados com qualidade é de suma importância para o desenvolvimento, implementação e avaliação de intervenções clínicas e na instauração de políticas públicas. A maioria, das evidências na área da saúde, são resultados gerados pelo uso da análise de dados, como por exemplo, o estudo de coorte prospectivo realizado pelo (DOLL & HILL, 1956) com o primeiro resultado claro da ligação do câncer de pulmão e tabagismo. O Brasil é um país com suas peculiaridades relevantes para que seja realizado um estudo sobre o impacto da pandemia do COVID-19, por ser um país com gravidade no surto, pelos seus fatores sociais e econômicos e por sua complexa composição racial.

Pesquisas foram sugeridas pela The Lancet (2021), sobre a mortalidade causada pela COVID-19, considerando fatores sociais e biológicos, se atentando a disparidades étnicas e

socioeconômicas dos pacientes, além de enfatizar que a falta de associação entre etnia e mortalidade é preocupante, apontando uma lacuna de pesquisa.

Neste contexto, o objetivo deste trabalho foi verificar os fatores de risco de mortalidade dos pacientes que necessitam de hospitalização durante a infecção do vírus SARS-CoV-2 correlacionando com fatores econômicos, sociais e epidemiológicos, comparando os principais achados com outros trabalhos internacionais sobre o tema (BAQUI *et al*, 2020; RAVI, 2020).

1.1 Justificativa

Com à grande magnitude da pandemia causada pelo vírus SARS-CoV-2, mostrando que a COVID-19 é uma enfermidade que possui rápida transmissibilidade, o que é retratado nos dados epidemiológicos, além de ser de fácil transmissão, possuindo um alto grau de mortalidade. Em nível internacional, após a OMS ter declarado estado pandêmico pelo vírus, já foram notificados, de acordo com a Our World in Data (2022), aproximadamente 596 milhões de casos, sendo que deste total, já ocorreram em torno de 6,68 milhões de óbitos.

No Brasil, infelizmente não foi diferente, a COVID-19 tornou o país, o segundo em mortes em comparação com outros países com mais de 682 mil, ficando atrás somente do Estados Unidos da América com mais de 1,04 milhões de mortes (OWD, 2022). Portanto pesquisas são fundamentais para descobrir as principais causas das mortalidades causadas pela doença, levando em consideração dados sociais e biológicos, composição racial e socioeconômicas que podem interferir positivamente no agravamento dos casos e consequentemente aumentar a mortalidade.

Esses índices representam uma grande preocupação na saúde pública para o país, o que ocasionou a inclusão dessa doença, pelo Ministério da Saúde (2021), na base de dados das Síndrome Respiratória Aguda Grave Hospitalizados (SRAG) do Sistema de Informação de Vigilância Epidemiológica da Gripe, SIVEP-Gripe, onde os pacientes que foram notificados e hospitalizados com a doença foram registrados.

Sabendo da relevância dos dados dos pacientes hospitalizados com COVID-19 no Brasil, neste estudo foi realizada uma análise de dados do SIVEP-Gripe, com as notificações da doença, juntamente com dados do Sistema IBGE de Recuperação Automática (SIDRA, 2020), para as análises dos fatores sociais, fornecendo resultados dos indicadores determinantes sociais, econômicos e regionais na mortalidade de pacientes hospitalizados com COVID-19 no Brasil, sendo fundamentais para o desenvolvimento, implementação e avaliação de intervenções clínicas e seus resultados gerando apoio nas políticas públicas de saúde.

1.2 Objetivos

1.2.1 Objetivo Geral

Avaliação dos fatores de risco na mortalidade de pacientes brasileiros hospitalizados com COVID-19: sociais, econômicos e regionais a partir dos bancos de dados SRAG do SIVEP-Gripe e do SIDRA do IBGE no Brasil.

1.2.2 Objetivos Específicos

1. Coleta dos dados de pacientes hospitalizados (SRAG) e também os dados sociais e econômicos (SIDRA);
2. Tratar os dados, com pré-processamento dos mesmos, para que sejam importados para um banco de dados;
3. Modelagem do banco de dados;
4. Importação dos dados para banco de dados PostgreSQL e criação de Script SQL para selecionar e incorporar as bases de dados, com criações de variáveis para a análise para posterior exportação da seleção de dados;
5. Importação e criação de script para realização da Análise de dados no SAS;
6. Apresentação dos resultados para visualização das informações de forma gráfica e imagética;

1.3 Estrutura Da Dissertação

Os restantes capítulos desta dissertação estão estruturados da seguinte forma:

O capítulo 2 apresenta a fundamentação teórica, tomando uma posição sobre o significado dos dados e sua gestão eficaz, discutindo a questão da perda de dados, fazendo sugestões para o tratamento de dados biológicos, delineando conceitos para bancos de dados e dando uma visão geral do características da mineração de dados.

No capítulo 3, descreve a metodologia utilizada para este trabalho. Onde são descritos e justificados os procedimentos metodológicos utilizados na pesquisa, detalhando a análise exploratória dos dados e demais produtos desenvolvidos no decorrer do projeto.

Como resultados do estudo da análise de dados, realizada nos dados do SRAG, na qual avaliou-se os indicadores que levaram os pacientes hospitalizados por COVID-19 ao óbito, resultando na publicação de um artigo que está descrito no capítulo 4.

O capítulo 5, apresenta-se as considerações finais do projeto, assim como as perspectivas para trabalhos futuros, a importância dos resultados obtidos e a análise das dificuldades e contribuições alcançadas.

2 FUNDAMENTAÇÃO TEÓRICA

Os principais conceitos teóricos que sustentam este projeto serão apresentados neste capítulo de uma forma geral. A sessão inicia conceituando pandemia, epidemiologia e epidemiologia computacional. Em seguida discorrendo sobre a importância dos dados, a importância no setor da saúde, passando pela qualidade dos dados e os problemas em conter qualidade nos dados, após discorre sobre o tratamento e armazenamento dos dados e por fim sobre o trabalho importante da área.

2.1 Pandemia

Uma pandemia é a disseminação mundial de uma nova doença, termo que passa ser usado, quando uma epidemia de doença infecciosa afeta a população localizada numa região, se espalhando por diversos países ou continentes, usualmente afetando um grande número de pessoas (GREENBERG *et al*, 2005). A Organização Mundial da Saúde (OMS) é que define se uma doença se torna uma pandemia, ou seja, uma ameaça global (VAN-TAM; SELLWOOD, 2010). Em um plano desenvolvido pela OMS em casos de surgimento de surtos gripais a nível global, define-se as fases de uma pandemia e as recomendações para medidas nacionais antes e durante a mesma (BLYTHER, 2009).

Segundo WHO (2009), a três condições que fazem com que uma pandemia possa começar: Aparecimento de uma nova doença; O agente afeta humanos, causando doença séria; O agente transmite muito fácil entre humanos. As fases constam na Tabela 1.

Tabela 1 – Descrições das fases de pandemia da OMS e principais ações por fase.

	Probabilidade estimada de pandemia	Descrição	Principais ações nos países afetados	Principais ações em países ainda não afetados
Fase 1	Incerto	Nenhum vírus da gripe circulando entre os animais foi relatado como causador de infecção em humanos.	Produzir, implementar, exercitar e harmonizar os planos nacionais de preparação e resposta à pandemia de gripe com os planos nacionais de preparação e resposta a emergências.	
Fase 2		Sabe-se que um vírus da gripe que circula em animais domésticos ou selvagens causa infecção em humanos e, portanto, é considerado uma ameaça específica de pandemia potencial.		
Fase 3		Um vírus de influenza animal ou humano-animal causou casos esporádicos ou pequenos grupos de		

		doenças em pessoas, mas não resultou em transmissão de humano para humano suficiente para sustentar surtos em nível comunitário.		
Fase 4	Médio a alto	A transmissão de humano para humano de um vírus de influenza animal ou humano-animal capaz de sustentar surtos em nível comunitário foi verificada.	Contenção rápida.	Prontidão para resposta à pandemia.
Fase 5	Alto para certo	O mesmo vírus identificado causou surtos sustentados em nível comunitário em pelo menos dois países em uma região da OMS.	Resposta à pandemia: cada país deve implementar	Prontidão para resposta iminente.
Fase 6	Pandemia em andamento	Além dos critérios definidos na Fase 5, o mesmo vírus causou surtos sustentados em nível comunitário em pelo menos um outro país em outra região da OMS.	ações conforme solicitado em seus planos nacionais.	
Período pós-pico		Os níveis de gripe pandêmica na maioria dos países com vigilância adequada caíram abaixo dos níveis máximos.	Avaliação da resposta; recuperação; preparação para possível segunda onda	
Possível nova onda		O nível de atividade da gripe pandêmica na maioria dos países com vigilância adequada está aumentando novamente.	Resposta	
Período pós-pandemia		Os níveis de gripe voltaram aos níveis observados para a gripe sazonal na maioria dos países com vigilância adequada.	Avaliação da resposta; revisão de planos; recuperação.	

Fonte: Pandemic influenza preparedness and response: a WHO guidance document (OMS, 2009, pag. 28) ¹.

Na história, as pandemias que aconteceram e que foram registradas, causaram momentos de grande tensão e foram catalisadores de transformações em alguns casos. Tais

¹ https://apps.who.int/iris/bitstream/handle/10665/44123/9789241547680_eng.pdf.

acontecimentos colocaram sociedades inteiras sob ameaça, sendo objetos de estudos de historiadores. No tópico a seguir, conheça as principais pandemias que assolaram o planeta (RODRIGUES, 2020).

2.1.1 Principais pandemias que assolaram o planeta

Peste bubônica

Causada pela bactéria *Yersinia pestis*, a peste bubônica ou peste negra, e pode se disseminar por meio de contato com pulgas e roedores infectados (HAYS, 2020). Tendo seus principais sintomas incluem inchaço dos gânglios linfáticos, na virilha e/ou pescoço, dor de cabeça, fadiga, febre, calafrios e dores musculares (SCHUELER, 2021). A doença gerou pânico em toda Europa no século 14, matando entre 75 e 200 milhões de vítimas (RIBEIRO, 2022).

Varíola

É uma doença exantemática causada pelo vírus *Orthopoxvirus variolae*, que era transmitido de pessoa para pessoa, por meio do contato com o vírus, seja por meio do contato com pessoas infectadas, ou com objetos que possuam o vírus (REDEDOR, 2022; RODRIGUES, 2020), que foi responsável por várias epidemias e por milhares de mortes. Os sintomas são parecidos com o da gripe com febre, seguida de erupções na garganta, na boca e no rosto. Segundo a OMS, no ano de 1980 a doença foi erradicada, após uma campanha de vacinação maciça que envolveu o mundo inteiro (JÚNIOR, 2004).

Cólera

No rol das grandes pandemias que assolaram o mundo, temos as pandemias de cólera, que é causada pela bactéria *Vibrio cholerae*, onde sua transmissibilidade acontece a partir de consumo de água e/ou alimentos contaminados. Quando o paciente infectado pela cólera os principais sintomas são diarreia intensa, desidratação, cólicas e enjoos (COLEMAN, 1987; LEWINSOHN, 2003). No Brasil já teve vários surtos da doença, principalmente nas áreas mais pobres do Nordeste. No ano de 2019 no Iêmen, morreram mais 40 mil pessoas pela cólera (PICANÇO, 2020).

Gripe espanhola

Com os primeiros casos registrados nos Estados Unidos, na instalação militar localizada em Kansas no ano de 1918, o presidente da época censurou a imprensa para que as mortes não fossem noticiadas, estima-se que entre 40 a 50 milhões de mortes, causada pela mutação do

vírus influenza (RODRIGUES, 2020; ESTACIO, 2021). A gripe espanhola espalhou-se pelo mundo principalmente, por conta da movimentação de tropas no período da 1ª Guerra Mundial, como a Espanha não estava envolvida no embate, as notícias sobre a doença vinham de lá, por isso denominada como gripe espanhola. No Brasil em 1919, o atual presidente, Rodrigues Alves, morreu da doença. Os sintomas incluem febre, calafrios, dores musculares, tosse, congestão, coriza, dores de cabeça e fadiga (FILHO, 2022; SOUZA 2055).

Gripe suína

Doença que surgiu em porcos, causada pelo vírus Influenza A H1N1, onde em 2009 os sistemas de vigilância epidemiológica do México e Califórnia (EUA), receberam as primeiras notificações de pacientes infectados com o vírus H1N1, disseminando em ritmo acelerado para outros países, levando a óbito mais de 18 mil pessoas (LEITE *et al*, 2013; GRECO *et al*, 2009). No Brasil o primeiro caso foi em maio daquele ano, após um mês, 627 pessoas estavam infectadas no Brasil (RIBEIRO, 2022). A transmissão acontece por gotículas respiratórias no ar ou acesso a uma superfície contaminada com o vírus. Os sintomas se assemelham que de uma gripe comum: febre, dor de garganta, tosse, dor no corpo e calafrios (MORENS *et al*, 2009).

2.1.2 Pandemias ativas

No mundo, atualmente, existem duas pandemias em curso, uma é a AIDS, desde o ano de 1981 e a COVID-19, desde 2019.

HIV/AIDS

Nos Estados Unidos, a partir de 1981, foram registrados um aumento de mortes de pessoas por doenças consideradas raras, que com acompanhamento médico dos casos foram descobertos que o sistema imunológico havia entrado em colapso (MOSS & BACCHETTI, 1989). Em 1982, a doença foi enquadrada como Síndrome da Imunodeficiência Adquirida, ou AIDS (Acquired Immune Deficiency Syndrome), que posteriormente foi descoberta que é causada pelo vírus da imunodeficiência humana, HIV em inglês (Human Immunodeficiency Virus) (MERSON, 2008; VERGIS, 2000).

A AIDS, foi considerada pandemia devida a grande escala de propagação em todo o mundo (VERGIS, 2000). Desde o início da doença até o final do ano de 2018, cerca de 32 milhões de pessoas faleceram vítimas da doença e 75 milhões de pessoas contaminadas. No Brasil o primeiro caso foi registrado em São Paulo no ano de 1982 (SILVA, 2022). As principais

causas de transmissão são por relação sexual sem a devida proteção e/ou por transfusão de sanguínea (MELHUIH & LEWTHWAITE, 2018). A AIDS é uma doença que ainda não tem uma cura, mas com o tratamento é possível retardar o avanço da doença controlando a quantidade de vírus no corpo do infectado.

COVID-19

Doença que surgiu em dezembro de 2019, na China, que no início era definida como uma pneumonia de origem desconhecida, causada pelo vírus SARS-CoV-2, disseminando-se rapidamente pelo mundo (WHOa, 2019). Em março de 2020 a OMS classificou a COVID-19 como uma pandemia global (WHOb, 2020; BBC News, 2020). Seus principais sintomas são febres, tosse seca, fadiga, perda de paladar e olfato, falta de ar variando desde à constipação até pneumonia viral grave com insuficiência respiratória potencialmente fatal (WU & MCGOOGAN, 2020). No Brasil o primeiro caso notificado foi em fevereiro de 2020. Em maio do mesmo ano o Brasil assumiu o terceiro lugar em número de casos no mundo e o segundo em números de mortes (HOPKINS, 2021).

A transmissão da COVID-19, acontece quando as pessoas respiram o ar com gotículas que contêm o vírus. O risco de contaminação é maior quando pessoas estão próximas umas das outras, mas pode ocorrer em distâncias maiores, quando estão em ambiente fechados. Além da transmissão pela inalação do ar contaminado, a transmissão também pode ocorrer casos fluídos contaminados entrarem em contato com os olhos, nariz ou boca, raramente, através de superfícies contaminadas (HINDSON, 2020; CHANG, 2020). Quando a pessoas infectadas com o vírus, normalmente permanecem transmitindo a doença de 10 a 14 dias, esse contágio pode acontecer mesmo que a pessoa seja assintomática, ou seja, que não desenvolvam sintomas. Durante o passar do tempo ocorreram mutações do vírus onde produziram muitas variantes, com graus variados de infectividade e virulência (CDC, 2020; NHS, 2021).

Para combate a COVID-19 foram desenvolvidas várias vacinas e distribuídas ao redor do mundo desde dezembro de 2020. No entanto, 308 eram as candidatas as vacinas sendo que 73 em pesquisas clínicas, 24 em testes de fase I, 33 em testes fase II entraram na fase de avaliação. Destas, 16 foram para a fase 3 que é a fase final de desenvolvimento envolvendo testes em voluntários (WHO, 2020; VAC, 2021). De acordo com um estudo de junho de 2022 (WATSON, 2022), as vacinas reduziram o número de morte exponencialmente. Onde no primeiro ano de imunização, evitou 19,8 milhões de mortes de um potencial de 31,4 milhões de óbitos em decorrência da doença.

Diante deste cenário, este estudo utilizou os dados epidemiológicos disponibilizados pelo Ministério da Saúde do Brasil, para analisar e identificar os determinantes sociais, econômicos e regionais na mortalidade de pacientes hospitalizados com COVID-19 no Brasil.

2.2 Epidemiologia

Epidemiologia, dispõe-se em estudar quantitativamente a distribuição dos fenômenos de saúde/doença, e seus fatores condicionantes e determinantes, nas populações humanas (MACMAHON & PUGH, 2012; ALMEIDA FILHO & ROUQUAYROL, 2003). Segundo (PEREIRA, 2013), o termo epidemiologia deriva do grego, conforme Figura 1.

Figura 1 – Termo epidemiologia derivado em grego.



Fonte: autor baseado em Pereira (2013).

Segundo Rouquayrol & Goldbaum (2003) a epidemiologia pode ser definida como:

ciência que estuda o processo saúde-doença em coletividades humanas, analisando a distribuição e os fatores determinantes das enfermidades, danos à saúde e eventos associados à saúde coletiva, propondo medidas específicas de prevenção, controle ou erradicação de doenças, e fornecendo indicadores que sirvam de suporte ao planejamento, administração e avaliação das ações de saúde.

Estatística, ciências da saúde e ciências sociais compõem os três principais campos de conhecimento que compõem a epidemiologia. Sua área de atuação inclui educação e pesquisa em saúde, avaliação de práticas e serviços de saúde, vigilância de tendências epidemiológicas e diagnósticas e monitoramento das condições de saúde da população. O campo da epidemiologia é bastante amplo e abrange linhas gerais (PEREIRA, 2013):

- o ensino e pesquisa em saúde;
- a descrição das condições de saúde da população;
- a investigação dos fatores determinantes da situação de saúde;
- na avaliação do impacto das ações para alterar a situação de saúde.

O princípio fundamental da epidemiologia é que os eventos relacionados à saúde (como doenças, suas causas e o uso de serviços médicos) não se espalham por acaso entre as pessoas. Grupos populacionais podem ter mais casos de uma gravidade específica do que outros, e alguns podem sofrer mais mortes por uma doença específica (ROTHMAN *et al*, 2008; BEAGLEHOLE *et al*, 1993). Essas diferenças decorrem do fato de os fatores que afetam a saúde das pessoas não estarem distribuídos uniformemente pela população, favorecendo

determinados grupos em detrimento de outros (PEREIRA, 2013; ROTHMAN *et al*, 2008). Em geral, pode-se dizer que a distribuição de doenças de uma população é influenciada pelas características biológicas de seus membros, pelas condições socioculturais e econômicas de seu bairro e pelas características ambientais de seu entorno, fazendo com que o processo saúde-doença² se manifesta de forma diferente em diferentes populações.

Observando-se o desenvolvimento da epidemiologia ao longo dos anos revela que as ideias anteriormente descritas, já firmemente estabelecidas na comunidade científica, precisaram ser revisadas à medida que as descobertas científicas avançavam no campo da saúde, principalmente no que diz respeito ao processo de doença.

2.3 Epidemiologia Computacional

Discutimos os conceitos de epidemiologia no tópico anterior, bem como sua interdisciplinaridade e o fato de que seus estudos visam compreender melhor as doenças para controlá-las e erradicá-las. O campo da epidemiologia evoluiu de uma total falta de compreensão das doenças para um reconhecimento científico dos fatores subjacentes às doenças e, durante esse desenvolvimento, inúmeras metodologias foram desenvolvidas para abordar uma ampla gama de questões e estudos epidemiológicos.

Devido à relevância dessas questões e ao significado que a biologia tem e continuará a ter para a humanidade nas próximas décadas, é importante desenvolver modelos matemáticos e computacionais de fenômenos biológicos, fenômenos especialmente epidemiológicos. Com propósito epidemiológico, a modelagem tem dois objetivos claros. Quanto à simulação ou prevenção da propagação da doença. A segunda diz respeito à análise de métodos de prevenção, controle ou erradicação de doenças (RIZZI, 2008).

Segundo Jaworsky (2019) e Liu (2020), o campo da epidemiologia passou por uma série de mudanças de paradigma metodológico ao longo de seu desenvolvimento, que vão desde a pesquisa empírica que se baseia na observação de ocorrências até os estudos preditivos que usam modelagem matemática para prever a progressão das doenças (Figura 2).

² **Processo saúde-doença:** conceito utilizado para definir todas as variáveis associadas ao estado de "saúde" e "doença" de uma pessoa ou população, levando em consideração o fato de que ambos os estados estão interligados e resultam das mesmas causas. Segundo Laurell (1983), o termo "processo saúde-doença" refere-se à maneira como os processos biológicos de decomposição e reprodução, ocorrem em grupos sociais, destacando instâncias específicas em que uma função biológica diferente está presente e tem impacto no desenvolvimento regular das atividades diárias, ou seja, o surgimento da doença.

Figura 2 – Principais paradigmas metodológicos na epidemiologia.



Fonte: autor baseado em Liu (2020).

Esses três paradigmas servem de base para as metodologias epidemiológicas conhecidas como epidemiologia empírica, teórica e computacional:

- A metodologia empírica que é a fase inicial do estudo tenta abordar as questões de quando, onde e quem por meio da coleta de dados observacionais sobre a propagação da doença, o agente infeccioso, o ambiente e o indivíduo afetado;
- O modelo teórico emprega ferramentas matemáticas para caracterizar a progressão da transmissão da doença e suas relações com outros fatores, facilitando projeções de longo prazo e a tomada de decisões relacionadas à doença;
- O modelo computacional busca caracterizar e compreender a disseminação de doenças de forma tangível. Ele faz isso utilizando técnicas de inteligência artificial, análise de dados, visualização geoespacial, mineração de dados, aprendizado de máquina, entre outras coisas, para criar modelos e avaliar possíveis contramedidas. Isso dá a capacidade de considerar vários cenários e facilita a tomada de decisões.

Uma das tarefas mais difíceis e cruciais no campo da medicina é o diagnóstico, pois envolve vários fatores que, dependendo de seus valores, podem determinar um curso de tratamento (BONITA *et al*, 2010). Como resultado, é essencial usar os métodos mais formais para colaborar na realização desta atividade.

Por causa de seus recursos para automação e otimização de tarefas, os computacionais estão entre esses métodos que estão sendo bastante utilizados (GUSTAFSSON, 2011), com isso, é possível realizar diagnósticos mais rápidos e precisos.

As metodologias, desafios e oportunidades exclusivas de estudos epidemiológicos computadorizados foram geralmente identificados neste item. Reconhecendo que há espaço e uma necessidade contínua de novas técnicas e estudos voltados para a compreensão e controle

abrangentes tanto das doenças emergentes quanto daquelas que assolam a humanidade há décadas.

2.4 Importância dos Dados

O processamento de informações institucionais começou com um esforço para automatizar tarefas rotineiras e repetitivas para agilizar o trabalho em diversas áreas funcionais. A expansão e disseminação dos recursos computacionais levaram ao aumento constante do número de séries de dados.

Até o ano de 2020, foram gerados 40 trilhões de Gigabytes, ou 40 Zettabytes, de informações em todo o mundo, significando que 2,2 milhões de Terabytes de novos dados gerados todos os dias (EXAME, 2021).

Os dados são o resultado de observações documentadas e servem como uma representação de fatos, ideias e/ou instruções. O processamento dos dados e a disponibilização aos usuários são necessários porque a mera existência dos dados pode não ser suficiente para obter informações (DAVENPORT, 1998). Eles são os componentes fundamentais nos quais o computador executa as operações necessárias para concluir à tarefa em questão (COADIC & GOMES, 1996).

Segundo os autores (REZENDE & ABREU, 2013), um dado é transformado em informação que seus usuários possam entender quando for processado, tornando-o útil e agregando valor para auxiliar na tomada de decisão. As informações obtidas por meio do processamento de dados tornaram-se um dos recursos mais valiosos de uma instituição, influenciando significativamente os resultados de suas operações (FERNANDES, 2005).

Segundo Coadic e Gomes (1996), “Sem informação, a ciência não pode se desenvolver e viver. Sem informação a pesquisa seria inútil e o conhecimento não existiria”. Não só a informação precisa estar presente, mas também precisa ser de alta qualidade.

Assim, a coleta de dados primários é fundamental para a produção de informações de qualidade, o que exige que os instrumentos cadastrais (como formulários, planilhas, sistemas de informação e outros) sejam seguidos com atenção e os dados armazenados sejam guardados em local seguro.

2.4.1 A importância dos dados na saúde

Uma base de dados confiável é necessária para que as políticas públicas abordem os diversos problemas de saúde, sejam eles relacionados à morbidade ou à mortalidade. Essa base de informações deve apoiar e orientar o processo de tomada de decisão (ABOUZHR, 2007). A identificação dos fatores que influenciam o processo saúde-doença, as disparidades de saúde

e a eficácia de ações e programas para reduzir a carga de doença na população só são possíveis quando há informações precisas (MATHERS *et al*, 2005).

Os sistemas de informação em saúde (SIS) reúnem um conjunto de dados, informações e conhecimentos utilizados na área da medicina para subsidiar o planejamento, a melhoria e a tomada de decisão dos inúmeros profissionais de saúde envolvidos na assistência aos pacientes. Ter acesso a informações confiáveis é um requisito para realizar análises objetivas de situações sanitárias, tomar decisões baseadas em evidências e planejar ações relacionadas à saúde (RIPSA, 2002).

2.4.2 Qualidade dos dados

A qualidade dos dados é normalmente discutida em termos de sua precisão, que é sua adequação ao uso pretendido e a ausência de erros. Isso implica que é possível avaliar o dado a qualquer momento e, se necessário, corrigi-lo comparando-o com um valor válido e genuíno (CAPPIELLO *et al*, 2013).

Segundo Wang e Strong (1996), a qualidade dos dados depende de uma série de fatores que devem ser avaliados com base nos requisitos do usuário. Segundo Batini e Scannapieco (2016), a qualidade dos dados tem um conceito multidimensional devido à existência de muitos tipos de dados, métodos de representação e armazenamento, categorias de frequência de atualização e outras características únicas de cada sistema de informação.

Seis fatores de confiabilidade que impactam na qualidade dos dados foram identificados pelo autor do estudo (VEIGA, 2012), conforme Tabela 2.

Tabela 2 – Fatores de confiabilidade na qualidade de dados.

Fatores	Descrição
Compleitude	Um sinal de ter dados úteis suficientes para um determinado domínio;
Consistência	Sinal de ausência de inconsistências no banco de dados;
Credibilidade da fonte	Um sinal da reputação dos dados ou de sua fonte. Realiza um cálculo para determinar se os dados merecem crédito para serem considerados precisos;
Precisão	Uma medida da qualidade dos dados, definindo a medição ou a veracidade dos dados;
Preciso	Muitas vezes confundido com exatidão, mas ao contrário de exatidão, que tem a ver com erros, preciso tem a ver com o nível de detalhamento dos dados;
Confiabilidade	Uma medida da confiabilidade dos dados. É determinado pela análise dos resultados quanto à integridade, precisão, consistência e credibilidade da fonte.

Fonte: autor baseado em Veiga (2012).

Devido à importância da informação, é importante focar na sua qualidade. Quanto melhor o seu método de geração de dados e o material base, melhores serão os resultados que

você obterá dele. Adicionalmente, para a sua geração, aconselha-se a utilização de sistemas de informação para a sua gestão quanto mais estruturado for o processo (GUIMARÃES & ÉVORA, 2004).

2.4.3 Problemas com a qualidade dos dados

Os problemas com a qualidade dos dados geralmente são vistos como erros ou inconsistências nos dados armazenados. No entanto, levando em consideração o ambiente multidimensional, é mais apropriado definir um problema de qualidade como qualquer desafio descoberto em uma ou mais dimensões da qualidade. Por exemplo, valores ausentes que afetam a dimensão de integridade ou erros gramaticais que afetam a dimensão de consistência. Em outras palavras, os problemas de qualidade não devem estar ligados apenas à precisão das informações, mas também a outros fatores (WANG & STRONG, 1996; STRONG *et al*, 1997).

Existem várias propostas de definições hierárquicas de problemas de qualidade de dados na literatura. Entre eles está a classificação de problemas de Rahm e Do (2000) de acordo com o tipo de fonte de dados, sendo *Single-Source* ou *Multi-Source* e o nível do problema, sendo no esquema ou nos dados, que é mostrado na Figura 3.

Figura 3 – Classificação de Problemas de Qualidade dos Dados



Fonte: baseado conforme Rahm e Do (2000).

O trabalho de Kim *et al.* (2003), que utilizam o termo *dirty data* "dados sujos", apresenta uma classificação que inclui essencialmente três grandes categorias de problemas de dados: falta de dados (como um problema com valor insuficiente); dados presentes, mas incorretos (como um problema de correção); dados presentes e corretos, mas não utilizáveis (por exemplo, problema de dados duplicados, erros de sintaxe).

Müller e Freytag (2003), que utilizam a expressão *data anomalies* "anomalias de dados", categorizam os problemas de qualidade em três grupos: problemas estruturais, problemas semânticos e problemas de cobertura. Neste trabalho, os autores vão além dos outros e vinculam algumas das questões de qualidade às Dimensões de Qualidade de Consistência, Exatidão e Completude.

Apesar dos diversos termos e nomenclaturas utilizados na literatura (como erros, anomalias, sujeiras, pendências e problemas), Oliveira *et al.* (2005) argumentam que questões como dados ausentes, dados incorretos ou representações imprecisas comprometem a qualidade dos dados.

É crucial identificar as causas primárias desses problemas, além de classificar os problemas de qualidade de dados. Três fatores principais podem ser divididos na geração de dados de baixa qualidade (OLIVEIRA, 2007):

- 1) Erros humanos no momento da entrada dos dados, que podem incluir, principalmente, erros de digitalização, uso inadequado de abreviaturas e omissão de valores em atributos obrigatórios;
- 2) Erros de sistema que podem surgir durante a transmissão, processamento, transformação ou integração de dados;
- 3) Problemas organizacionais que podem surgir como resultado de processos falhos, trabalhadores sem treinamento e conscientização suficientes ou estrutura organizacional insuficiente.

2.5 Tratamento e armazenamento dos Dados

2.5.1 Descoberta de Conhecimento em Base de Dados

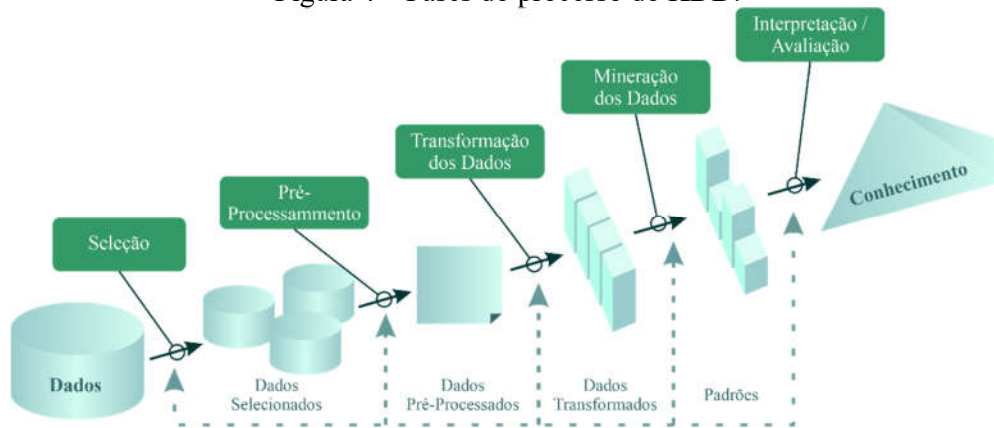
É uma tarefa difícil e frequentemente impossível para um ser humano encontrar conhecimento verdadeiramente útil contido nos enormes volumes de dados que estão disponíveis hoje; por isso, é imprescindível a utilização de métodos que deem suporte a essas análises, transformando os dados brutos em informações pertinentes e, em última instância, em conhecimento útil. Com isso surge a Descoberta de Conhecimento em Base de Dados (*Knowledge Discovery in Database - KDD*), que segundo FAYAD *et al.* (1996) é definido como “um processo de várias etapas, não trivial interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados”.

Mineração de Dados é definido por (FAYAD et al., 1996) como a etapa da descoberta do conhecimento que envolve a análise dos dados e a aplicação de algoritmos de descoberta que, apesar de algumas limitações computacionais, produzem um conjunto de regras sobre os dados.

É necessário realizar algumas etapas de forma breve e minuciosa para desvendar esses conhecimentos implícitos baseados em dados (KDD), garantindo que os resultados sejam precisos e simples de entender. Antes dessas etapas, FAYAD et al. (1996) observa a

necessidade de identificar o objetivo do processo de extração de informações por meio de entrevistas com especialistas, análise exploratória e escolha de hipóteses do modelo. Depois disso sendo possível iniciar as 5 fases do KDD, ilustradas na Figura 4.

Figura 4 – Fases do processo do KDD.



Fonte: autor adaptado Fayad et al. (1996).

Essas fases, abrangem uma série de tarefas interdependentes; o resultado de uma fase depende do resultado da próxima, e o analista participa de cada etapa, com a opção de repeti-la várias vezes, descritas abaixo:

- **Seleção** de dados, originados de uma ou mais fontes (planilhas, sistemas de gerenciamento, data warehouses) e potencialmente armazenados em vários formatos, levando em consideração as necessárias considerações de privacidade, ética e veracidade;
- **Pré-Processamento** ou tratamento dos dados, garantir a qualidade dos dados analisando dados ausentes, incorretos ou desnecessários e identificando anomalias (outliers);
- **Transformação dos dados**, entende como usar técnicas de normalização de dados, adição, criação de novos atributos, redução e sintetização;
- **Mineração dos dados**, concebe na construção de modelos técnicos na mineração de dados, onde essas técnicas visam confirmar uma hipótese ou descobrir novos padrões. Existem dois tipos de descoberta: predicativa e descritiva. Esses modelos são normalmente aplicados e revisados em várias ocasiões, dependendo do objetivo do projeto.
- Na **Interpretação / Avaliação**, por meio de métodos estatísticos ou avaliação de especialistas, é possível avaliar o desempenho do modelo.

A aplicação de algoritmos de mineração de dados para a descoberta de informações que não estão explicitamente declaradas na base só pode ser feita após a conclusão da etapa de pré-

processamento. Para fazer isso, pode-se usar redes neurais, algoritmos genéticos, modelos estatísticos e modelos probabilísticos. Existem inúmeras técnicas de mineração; estes devem ser escolhidos e aplicados na coleta de dados de acordo com o tipo de KDD que será executado (GOLDSCHMIDT et al, 2005; AMARAL, 2016).

Existem várias taxonomias utilizadas na literatura para categorizar as tarefas associadas à mineração de dados, mas uma das mais conhecidas é a de FAYAD et al. (1996), que apresenta uma divisão em dois níveis: as preditivas, que utilizam os valores dos dados combinados para prever valores futuros ou desconhecidos, e as descritivas, que buscam regras que descrevam os dados de forma compreensível para um ser humano. O autor incorpora classificação e regressão no conjunto preditivo, e entre as tarefas listadas estão agrupamento, resumo, dependência de modelagem e detecção de desvio.

Observou-se que a descoberta de conhecimento útil e prático é possibilitada pelo processo de KDD meticoloso e completo, que também inclui mineração de dados. Esse conhecimento é de extrema importância em muitas indústrias, inclusive na área da saúde, onde este trabalho está sendo realizado, pois é necessário coletar e analisar dados para produzir um grande volume de dados para realizar a vigilância nesta área. Como resultado, o uso da mineração de dados permite identificar, categorizar e prever padrões de doenças, auxiliando em análises clínicas e epidemiológicas, bem como na gestão de decisões e na avaliação de iniciativas de saúde pública.

2.5.1.1 Algoritmo APRIORI

O algoritmo APRIORI, um dos mais populares para mineração de dados, foi introduzido pela primeira vez por Agarwal, Imielinski e Swami (1993), como forma de criar regras para associação de dados. A execução do algoritmo funciona basicamente em duas etapas:

- Descobrir todos os grupos de itens frequentes com valores de suporte maiores ou iguais aos níveis mínimos de suporte especificados pelo usuário. Em termos de custo computacional, esta é a etapa de maior custo;
- Utilizar combinações comuns de itens constantes para transmitir regras associativas com valores de suporte e confiança que são maiores ou iguais àqueles que o usuário especificou.

Ao utilizar a pesquisa profunda, o algoritmo pode criar conjuntos de itens candidatos (conhecidos como padrão) com k elementos de conjuntos de candidatos com $k \neq 1$ elementos. A finalização só acontece após o último elemento no banco de dados processado e padrões

incomuns são excluídos. No Quadro 1, a seguir ilustra o pseudocódigo de como funciona o Apriori.

Quadro 1 - Pseudocódigo do algoritmo Apriori.

```

Input: ( $T, minSupport$ )
//  $T$  é o conjunto de dados e  $minSupport$  é o suporte mínimo
begin
   $L_1 = \{\text{itens frequentes}\};$ 
  for ( $k=2; L_{k-1} \neq \emptyset; k++$ ) do
     $C_k = \text{candidatos gerados a partir de } L_{k-1};$ 
    // é o produto cartesiano  $L_{k-1} \times L_{k-1}$  e elimina
    // qualquer conjunto de itens de tamanho  $k - 1$  que não seja frequente
    foreach transação  $t \in T$  do
      #incrementa o contador de todos candidatos em  $C_k$  que estão  $\subset$  em  $t$ ;
       $L_k = \text{candidatos em } C_k \text{ com } minSupport;$ 
    end
    // fim do for each
  end
  // fim do for
end
return  $\bigcup_k L_k;$ 

```

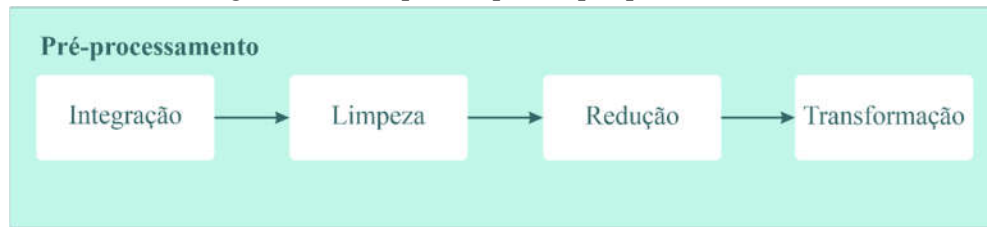
Fonte: baseado em Silva (2014).

Como é prática comum na mineração de regras, o algoritmo tenta identificar subconjuntos que são compartilhados por pelo menos um mínimo de C dos conjuntos de iterações dado um conjunto de conjuntos de iterações (por exemplo, cada lista de itens comprados em uma loja de moda). Uma abordagem bottom-up é usada pelo Apriori, na qual grupos de candidatos são testados contra os dados e subconjuntos frequentes são estendidos um item por vez (um processo conhecido como geração de candidatos).

2.5.1.2 Pré-processamento dos Dados

Devido ao seu tamanho, que geralmente é grande (geralmente vários terabytes ou mais), e sua provável origem de várias fontes heterogêneas, os bancos de dados do mundo real hoje são altamente suscetíveis a erros, ausências de dados e dados inconsistentes (SILVA, 2014). Como resultado, é possível que a qualidade desses dados seja comprometida e dados de baixa qualidade levem a saídas de mineração de dados de baixa qualidade. O pré-processamento, é uma fase anterior que envolve a preparação dos dados para que os algoritmos de mineração sejam aplicados (HAN *et al*, 2006). Na Figura 5, é possível visualizar as principais etapas do envolvidas no pré-processamento.

Figura 5 - Principais etapas do pré-processamento



Fonte: baseado em Silva (2014).

- **Integração dos dados:** normalmente, os dados vêm de uma variedade de fontes, incluindo bancos de dados, documentos de texto, planilhas, data warehouses, vídeos e imagens. Surge assim a necessidade de estes dados serem integrados num repositório único e consistente. Para tanto, é necessária uma análise minuciosa dos dados, levando em consideração redundâncias, dependências entre as variáveis e valores conflitantes, como categorias diferentes para os mesmos valores, regras divergentes para os mesmos dados, entre outros.
- **Limpeza dos dados:** essa fase de limpeza de dados visa eliminar problemas como registros incompletos, valores incorretos e dados inconsistentes para que não afetem o funcionamento dos algoritmos. As técnicas utilizadas nesta fase vão desde a remoção de registros problemáticos até a atribuição de valores padrão à aplicação de técnicas de agrupamento para auxiliar na descoberta dos melhores valores.
- **Redução dos dados:** a quantidade de dados usados na mineração é frequentemente grande; em certos casos, esse volume dificulta a análise dos dados e a realização da mineração propriamente dita. Como resultado, técnicas de redução de dados podem ser usadas para reduzir o volume de dados original, mantendo a representatividade dos dados originais. Isso permite que os algoritmos de mineração sejam operados de forma mais eficaz, mantendo a qualidade da saída. As estratégias utilizadas nesta etapa incluem a criação de estruturas otimizadas de dados (cubos de dados), a seleção de um subconjunto de atributos, a redução da dimensionalidade e a discretização.
- **Transformação dos dados:** nos casos em que alguns algoritmos trabalham apenas com valores numéricos e outros apenas com valores nominais, é necessário converter valores numéricos em valores nominais ou valores nominais em valores numéricos. Não existe um padrão único para transformação de dados, e uma variedade de técnicas podem ser empregadas dependendo dos resultados desejados. Algumas das técnicas utilizadas nesta fase incluem suavização,

agrupamento, generalização, normalização e criação de novos atributos baseados em atributos já existentes.

A implementação de uma ou mais técnicas de pré-processamento etapa por etapa pode ser a chave para resolver o problema. A diferença entre um modelo algorítmico bem-sucedido e um que falha pode ser o fator decisivo. Com base nisso, fica evidente que o pré-processamento incorreto pode resultar na inviabilidade dos dados que servirão de base para a mineração de dados.

2.5.2 Sistema Gerenciador de Banco de Dados

O modelo de dados físico é implementado em uma configuração que inclui usuários, software, hardware e operações do sistema. Essa arquitetura fornece a estrutura para suportar uma coleção de dados armazenados persistentemente que podem ser adicionados, excluídos, recuperados e usados em um sistema que protege a integridade dos dados e das operações (GRAVES & GOLDFARB, 2001).

Um banco de dados é uma coleção de dados armazenados que são vinculados para atender às necessidades de vários usuários em uma ou mais organizações ou, em outras palavras, uma coleção de bancos de dados interconectados com uma ampla variedade de tipos diferentes de tabelas (TEOREY *et al*, 2007).

Os dados que compõem uma organização são mantidos em um banco, que se encarrega de manter o controle sobre os dados de alto valor. Um banco de dados é um sistema cujo objetivo é armazenar, manter e disponibilizar informações pertinentes às operações de uma organização.

Em um esforço para eliminar a necessidade de gerenciamento de dados, os primeiros bancos de dados começaram como sistemas de gerenciamento de documentos. Esses sistemas tinham programas que podiam extrair e adicionar registros aos arquivos apropriados e armazenavam registros em vários arquivos (SILBERSCHATZ *et al*, 2010). No entanto, existem desvantagens em termos de armazenamento, redundância, uma variedade de formatos de arquivo, recuperação e segurança de dados.

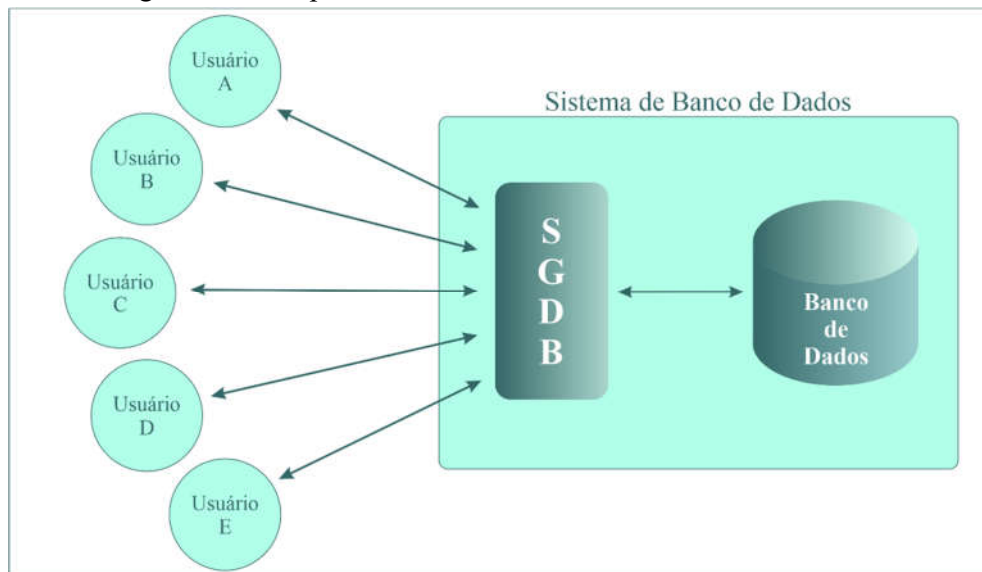
Estudos com o objetivo de desenvolver SGBDs, que, segundo ELMASRI e NAVATHE (2011), são compostos por um conjunto de dados vinculados a um conjunto de programas de acesso a esses dados, vêm sendo realizados para superar essas limitações. Os conceitos de Atomicidade, Consistência, Isolamento e Durabilidade – propriedades dos SGBDs – bem como operações fundamentais de banco de dados, como backup e restauração de dados, são aderidos pelos SGBDs.

Um sistema de software conhecido como Sistema Gerenciador de Banco de Dados (SGBD) é utilizado para manipular bancos de dados. Um SGBD aceita uma visão lógica (esquema e subesquema), uma visão física (métodos de acesso, agrupamento de dados), uma linguagem para definição de dados (SQL-DDL), uma linguagem para manipulação de dados (SQL-DML) e funções cruciais como transação gerenciamento e controle de concorrência, integridade de dados, recuperação de erros e segurança (TEOREY, 2007).

O software responsável pela criação e manutenção do BD é chamado de SGBD, e interage com o BD para possibilitar consultas do usuário e manipulação de informações, conforme mostra a Figura 6. Segundo Silberschatz *et al* (2010), um bom SGBD tem como principais objetivos:

- a garantia da privacidade dos dados através do isolamento;
- o compartilhamento de dados de forma segura;
- a garantia de integridade dos dados através de suas chaves;
- a abstração dos dados, de forma que as informações são disponibilizadas ao grupo de usuários que realmente tem necessidade de acessá-las.

Figura 6 - Exemplo de Sistema Gerenciador de Banco de Dados.



Fonte: autor baseado Teorey (2007).

Um modelo específico, que especifica como será a organização interna dos dados, é seguido por um SGBD. Os modelos de banco de dados desenvolvidos são: em redes, hierárquicos, relacionais, relacionados a objetos e orientados a objetos, nesta ordem (SILBERSCHATZ *et al*, 2010). Os SGBDs comerciais mais utilizados atualmente possuem estruturas organizacionais relacionais (MASSINO & ROLAND, 2015), onde a entidade é representada como um conjunto de relacionamentos. Os SGBDs que foram implementados

neste modelo de dados fornecem suporte para os tipos de dados básicos compostos de números, letras e caracteres.

O componente estrutural essencial do modelo relacional é a relação, que é composta por um ou mais atributos, cada um com um determinado tipo de dado. Cada linha na tabela é conhecida como "tupla" ou "registro", denota uma coleção de valores de dados relacionados e deve seguir as restrições de integridade referencial, regras e integridade das junções de relacionamento (SILBERSCHATZ *et al*, 2010).

A ordem física dos registros nas tabelas é irrelevante para o usuário, pois cada registro é identificado por um único valor, ou chave. Portanto, o usuário só precisa estar familiarizado com a estrutura organizacional do banco de dados e seus relacionamentos para recuperar os dados. A relação entre as duas tabelas é estabelecida por meio de um campo compartilhado que estabelece a exclusividade de cada registro. Este campo é conhecido como chave primária na tabela original e chave estrangeira na tabela relacionada na tabela original (ELMASRI & NAVATHE, 2011). Este modelo é particularmente adequado para sistemas de informação com dados não complexos, como os dados utilizados neste estudo.

2.6 Trabalho importante da área

Neste item após uma minuciosa revisão literária de referências bibliográficas, dentre as pesquisas analisadas e selecionadas para o desenvolvimento do trabalho, uma pesquisa de grande relevância, com qualidade nos resultados e abordando questões semelhantes à dessa pesquisa em questão.

O estudo realizado por Baqui *et al* (2020) sobre as variações étnicas e regionais na mortalidade hospitalar por COVID-19 no Brasil: um estudo observacional transversal. No estudo, foram utilizados 11321 dados de pacientes do SIVEP-Gripe, por estar no início da pandemia no Brasil. Onde foi realizado a análise de sobrevivência de regressão de Cox *mixed-effects* (efeitos mistos). Como resultado da pesquisa pode-se encontrar evidências de dois efeitos distintos, porém relacionados: aumento da mortalidade na zona norte (efeito regional) e entre as populações de pessoas de cor (pardo e preta) (efeito etnia). Diante desse cenário, torna-se imprescindível um estudo com dados mais atualizados da mortalidade hospitalar de pacientes com COVID-19 no Brasil, onde o resultado desta pesquisa apoie no fortalecimento de políticas públicas voltadas à proteção da saúde pública.

3 METODOLOGIA

Este capítulo apresenta todas atividades desenvolvidas para realização da coleta de dados, tratamento e importação para um banco de dados, onde posteriormente a realização da exportação dos dados preparados, para que seja realizada a análise dos dados.

3.1 Estruturação dos dados

Esta seção descreve os detalhes das atividades realizadas sobre os dados primários, como a coleta dos dados, compreensão da modelagem, estruturação e importação da base de dados de Síndrome Respiratória Aguda Grave do SIVEP-GRIPE e do Sistema IBGE de Recuperação Automática. Para isto, houve busca no entendimento da inserção das informações e identificação da chave primária para realizar relacionamento das bases de dados por aqui informadas.

A partir deste mapeamento, houve-se um pré-processamento dos dados, para qualificar os mesmos, e posteriormente importados no banco de dados, onde houve mineração dos dados e transformação, para posterior exportação da planilha pra análise de dados da mesma, dentre outros aspectos que serão discutidos nesta seção.

3.1.1 Coleta de dados

Esta seção discorre os detalhes das bases de dados coletadas para a realização das atividades do trabalho, está dividido em duas subseções onde a primeira descreve a base de dados do SRAG que consta as informações dos pacientes hospitalizados com COVID-19, e a segunda trata-se de dados da Pesquisa Nacional por Amostragem de Domicílios (PNAD) do Sistema IBGE de Recuperação Automática (SIDRA), com os coeficientes de nível de educação, gênero, idade e raça conjuntamente, considerando a divisão por estados brasileiros.

Todos os dados são disponíveis no formato *Comma-Separated Values* (CSV), que são arquivos de texto de formato regulamentado pelo RFC 4180, que faz uma ordenação de bytes ou um formato de terminador de linha, separando valores/colunas com vírgulas ou ponto e vírgula, esse último mais utilizado no Brasil (REPICI & BURTON, 2022).

3.1.1.1 Dados do SRAG

Para este estudo foram coletados os dados de pacientes hospitalizados com COVID-19 na base de Síndrome Respiratória Aguda Grave do SIVEP-Gripe, no período de 22 de fevereiro de 2020, quando foi registrado o primeiro caso de hospitalização de COVID-19 no Brasil, até o dia 10 de maio de 2021 o SIVEP-Gripe continha o total de 1.956.350 de registros

epidemiológicos de pacientes que procuraram assistência hospitalar em todos os estados brasileiros (MINISTÉRIO DA SAÚDE, 2020-2021b).

O Ministério da Saúde (MS), por meio da Secretaria de Vigilância em Saúde (SVS), desenvolve a vigilância do SRAG no Brasil, que vem se fortalecendo desde a pandemia de Influenza A(H1N1). Com isso, a vigilância SRAG foi implantada na rede de vigilância para influenza e outros vírus respiratórios, que antes operava apenas com a vigilância sentinela Síndrome Gripal (SG) (MINISTÉRIO DA SAÚDE, 2020-2021b).

Em 2020, a vigilância de COVID-19, infecção humana causadora de pandemia provocada pelo novo Coronavírus, foi incluído na rede de vigilância de influenza e outros vírus respiratórios. Atualmente, o sistema oficial para o registro dos casos e óbitos por SRAG é o Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe).

Os pacientes hospitalizados com SG que apresentavam dispneia/desconforto respiratório ou pressão persistente no tórax ou saturação de O₂ menor que 95% em ar ambiente ou coloração azulada dos lábios ou rosto, eram notificados no SIVEP-Gripe. Os dados SRAG são disponibilizados todas as semanas às quartas-feiras. Excepcionalmente, a atualização pode ocorrer em um dia diferente, dependendo do funcionamento do sistema de informação (MINISTÉRIO DA SAÚDE, 2020-2021b).

3.1.1.2 Dados da PNAD

Além dos dados do SRAG, foram coletados dados da Pesquisa Nacional por Amostra de Domicílios (PNAD), onde foram utilizados os coeficientes de nível de educação, gênero, idade e raça conjuntamente, considerando a divisão por estados brasileiros (SIDRA, 2020), possibilitando que sejam apontadas as evidências relacionadas ao fator social e à COVID-19. A importância para o assunto está na possibilidade de se revelar a existência de disparidades em relação à condição social do paciente e a sua probabilidade de evoluir para o óbito.

A Pesquisa Nacional por Amostra de Domicílio (PNAD) é um estudo realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE) sobre uma amostra de domicílios brasileiros que, por ter múltiplos objetivos, analisa uma variedade de características socioeconômicas da sociedade, como população, educação, emprego, renda, moradia, previdência social, migração, fecundidade, nupcialidade, saúde, nutrição, entre outros temas também.

O PNAD foi implantado no Brasil em caráter experimental no ano de 1967. A capacidade de monitorar regularmente as características socioeconômicas brasileiras sem incorrer em custos e com um intervalo de tempo relativamente curto entre as enumerações foi

a razão por trás da adoção da pesquisa por agrimensores domiciliares. Na época, o IBGE tentava realizar essas consultas trimestralmente (PNAD, 1969).

Os dados da PNAD estão disponíveis no Sistema IBGE de Recuperação Automática (SIDRA), que é um banco de tabelas estatísticas que tem como objetivo armazenar e disponibilizar os dados de pesquisas realizadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Os dados estão na forma de séries temporais, acompanhando seu comportamento ao longo do tempo, bem como ter os mesmos disponibilizados por níveis territoriais desagregados, como município, distrito e bairro, de modo a facilitar o conhecimento da realidade municipal.

3.1.2 Tratamento dos dados

Para melhorar a qualidade dos dados e realizar importação das informações para o banco de dados, foram realizadas algumas etapas do pré-processamento dos dados, citados na subseção a seguir, com principal objetivo de evitar erros na hora de realizar a análise dos dados.

3.1.2.1 *Pré-processamento*

O pré-processamento foi utilizado de acordo com as atividades de limpeza, transformação e integração descritas no Capítulo 0 para melhorar a qualidade dos dados quanto à verificação de duplicação, redundância e existência de relacionamento entre eles. Tendo em vista que esta tarefa estava relacionada à importação dos dados e todos os dados pertencentes ao conglomerado foram importados, a tarefa de pré-processamento de redução de dados não foi aplicada a esta coleta de dados. Os tópicos a seguir descrevem as especificidades das atividades realizadas nesta etapa.

O pré-processamento foi realizado somente na base de dados do SRAG, devido a mesma haver problema na estrutura dos dados. A base de dados do SIDRA, não houve problema, pode-se justificar devido que as informações serem geradas por sistema, sem entrada de dados diretamente pelo usuário.

Limpeza

Os campos que continham características especiais que o pesquisador responsável julgou fora da normal foram substituídos por nulo na atividade de limpeza de dados. Por exemplo, campos que seu tipo era numérico continham caractere “-“, sendo inválido. Neste caso, pelo o valor ser inválido o valor da coluna é substituído por nulo ou vazio.

Nesta mesma atividade, de valores inválidos em campos numéricos, houve algumas informações com “XX” que é uma string em campo de valor numérico. Neste quesito o valor por ser diferente de valor numérico, foi substituído o valor da coluna por nulo ou vazio.

Seguindo esta lógica de trabalho nos dados, algumas informações com campo de observação por exemplo, que é um campo do tipo string, havia valores com caractere “;” que no caso da planilha do SRAG, é o caractere utilizado como separador de colunas/campos, acarretando em erro na estrutura dos dados. Então neste cenário, o caractere “;” foram substituídos por caractere “ ” limpo por se tratar de uma string. Algumas informações de campos do tipo string continham caractere “\”, que seguindo o trabalho realizado com o caractere “;”, citado acima, neste caso também foram substituídos por caractere “ ” limpo.

Transformação

Os dados foram transformados para que pudessem ser mapeados para um conjunto de tabelas levando em consideração seus tipos de acordo com o uso pretendido. Na planilha eletrônica alguns campos tinham o mesmo tipo de dados, que era texto. No mapeamento, foi necessário considerar um tipo numérico para campos com valores numerais, por exemplo as chaves, e o tipo string para campos com textos.

Integração

Dada a disponibilidade dos dados primários, foi necessário que esses dados passassem por um procedimento de integração e relacionamento para serem armazenados em um único repositório. Para isso, foi necessária uma análise minuciosa dos dados, durante a qual foram anotadas suas características para posteriormente estruturá-los e relacioná-los.

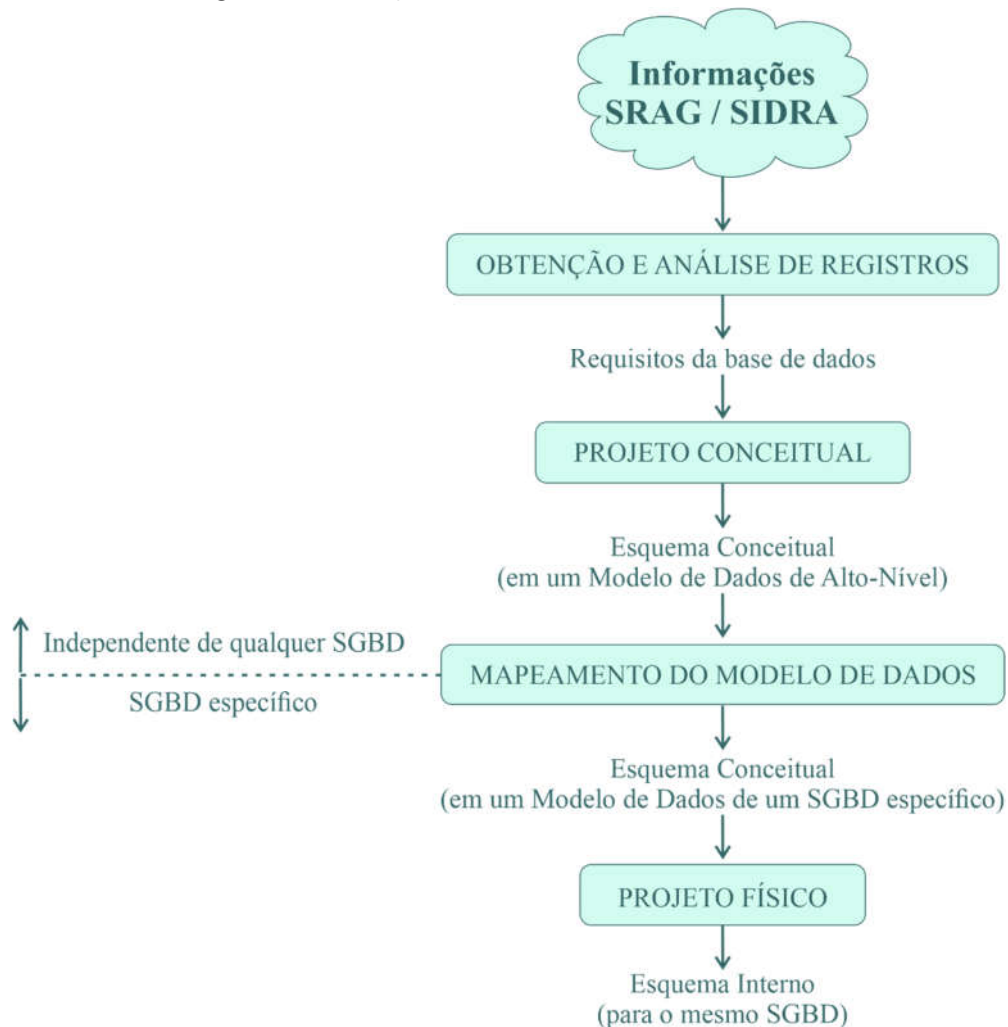
O resultado do pré-processamento ajudou de alguma forma no desenvolvimento do projeto do banco de dados, bem como na atividade de importação de dados para a base incorporada.

3.1.3 Modelagem, Importação e exportação do banco de dados

Para relacionar as informações dos dados do SRAG e SIDRA, foi estruturado um banco de dados para importação das planilhas CSV com os dados. As atividades envolvidas na estruturação do banco de dados seguiram as etapas recomendadas por Elmasri e Navathe (2011) e apresentadas na Figura 7. Essas fases envolvem levantamento e análise de requisitos, projeto conceitual, projeto lógico e projeto físico.

As atividades descritas na Figura 7 foram amplamente realizadas como parte dessas tarefas, além do modelo proposto pelos autores e, principalmente, porque existiam dados primários. Essas atividades envolviam ajustar os dados primários e importá-los para o modelo.

Figura 7 - Ilustração de conceito de Banco de Dados.



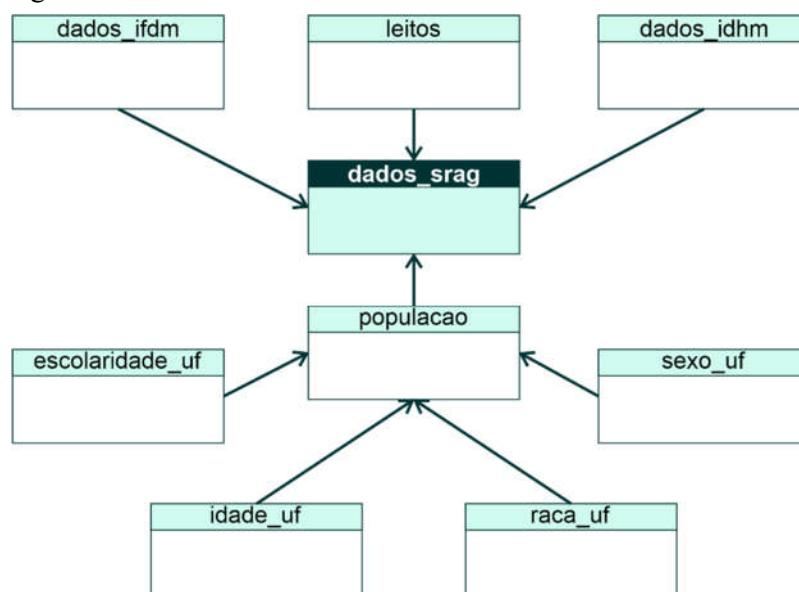
Fonte: autor baseado em Elmasri e Navathe (2011)

Os detalhes de cada um dos processos que compõem o projeto do banco de dados serão apresentados nas seções a seguir.

3.1.3.1 Projeto conceitual e lógico

Na fase da projeção conceitual, sem levar em consideração os detalhes de implementação e armazenamento, foi desenvolvido o um esquema Modelo de Entidade-Relacionamento (MER), ilustrado na Figura 8. Esquema este que foi criado usando os requisitos de dados que foram elicitados.

Figura 8 - Modelo de entidade-relacionamento do Banco de Dados.



Fonte: próprio autor.

É possível observar no MER apresenta a entidade *dados_srag* onde é a entidade que receberá as informações do SRAG coletadas do SIVEP-GRIPE disponibilizado pelo MS. As entidades que se relacionam a ela são informações que foram coletadas e que estão disponíveis no SIDRA do IBGE.

O processo de construção da estrutura do banco de dados começou após a conclusão da modelagem conceitual e envolveu o mapeamento do modelo conceitual para o modelo lógico.

Nesta fase, foi determinado se o modelo relacionado demonstrou adequação à estrutura organizacional do projeto. Isso era especialmente verdadeiro porque o modelo lidava com dados diretos e incluía garantias de integridade de dados, gerenciamento de concorrência, recuperação de erros, segurança e otimização de consultas.

Na ilustração Figura 9 a seguir, é possível visualizar a estrutura lógica de todo o banco de dados, com as entidades/tabelas e seus relacionamentos. As entidades que se relacionam com a entidade com os dados do SRAG, *dados_srag*, são feitas pela chave primária de código de município, *co_mun_res*. Já as entidades que se relacionam com a entidade população, *populacao_municipio*, são relacionadas pela chave de código de estado *uf*.

3.1.3.2 Projeto físico e Importação e exportação dos dados

O projeto físico foi elaborado conforme modelo conceitual e lógico anteriormente citados para criação da estrutura do banco de dados. As fases de criação do projeto físico, foram realizadas utilizando a linguagem de padrão universal para manipulação de dados *Structured Query Language* (SQL), utilizando o SGBD Postgres³, nas etapas desenvolvidas a seguir:

- Criação do banco de dados;
- Criação das tabelas para importação dos dados do SRAG e dados disponibilizados do SIDRA.
- Execução de comando SQL com filtros, relacionamentos e criação de colunas para preparação das informações para exportação dos dados;
- Exportação dos dados gerado via comando SQL, para a estatística descritiva e para Análise de dados utilizando o SAS® OnDemand for Academics⁴.

Após o processo de criação do banco de dados, foi iniciado a importação dos dados já citados anteriormente aqui nesse capítulo. A Figura 10 a seguir, detalha as atividades que foram realizadas no desenvolvimento de todo processo na importação do banco de dados e exportação da planilha para a análise de dados.

³ O PostgreSQL é um sistema de gerenciamento de banco de dados objeto-relacional (SGBDOR) baseado no POSTGRES Versão 4.2 desenvolvido pelo Departamento de Ciência da Computação da Universidade da Califórnia em Berkeley.

⁴ Trata-se de um sistema integrado de aplicações para o processamento e análise estatística de dados, consistindo em módulos de Acesso e Recuperação de Dados, Gerenciamento de Arquivos, rotinas de Geração de Gráficos e Geração de Relatórios.

Figura 10 – Detalhamento das etapas desenvolvidas na criação e importação dos dados



Fonte: próprio autor

Criação das tabelas

Na etapa da criação das tabelas, a estrutura de criação foi realizada seguindo a estrutura realizada na Figura 9, das tabelas e campos, conforme informações coletadas.

Importação dos dados

A fase de importação dos dados coletados do SRAG e SIDRA, foram realizadas após criação do projeto físico e realização de todo tratamento dos dados. Onde os campos e chaves primárias seguiram o projeto lógico.

Criação comando SQL

A realização do script comando SQL dos dados já importados, foi de suma importância para o tratamento dos mesmos, onde dados que continham valores nulos, eram substituídos para 0 em caso de numérico e vazio em caso de string. Os filtros dos dados realizados nessa consulta com relacionamento com os dados sociais, ocorreu de todos os pacientes com informações válidas, e que eram diagnosticadas e confirmadas com exame positivo para portadora do vírus SARS-COV-2, causadora do COVID-19.

Nesta etapa, também ocorreu criação de campos de rendas relativas calculadas a partir dos dados das tabelas aqui importadas, realizadas via chave código município e chave de estado. Informações que foram de suma importância na atividade de estatística descritiva e análise de dados da planilha exportada.

Exportação da planilha

Nessa fase após realização dos tratamentos e transformação dos dados em uma única planilha com todas as informações importadas, coletadas dos dados de pacientes hospitalizados com COVID-19 e também das informações dos coeficientes de nível de educação, gênero, idade e raça conjuntamente, considerando a divisão por estados brasileiros.

Em seguida, houve a realização do processamento para a exportação da planilha filtrada com os dados de pacientes com COVID-19 e seus índices sociais do seu estado de residência, em formato CSV seguindo o padrão americano, ou seja, separado por “;”, para que posteriormente serem importadas para o SAS e proceder na realização na análise de dados.

3.2 Desenho de estudo e Análise estatística

Nesta seção será apresentado o desenho do estudo pressuposto e análises estatísticas utilizadas e todos os materiais que foram usados para produção dos resultados obtidos neste trabalho.

3.2.1 Desenhos e visão geral de estudo

A análise foi realizada no conjunto dos dados públicos de Casos de Síndrome Respiratória Aguda Grave Hospitalizados (SRAG) do SIVEP-Gripe⁵, a partir de 22 de fevereiro de 2020, quando foi registrado o primeiro caso de hospitalização de COVID-19 no Brasil, até o dia 10 de maio de 2021 o SIVEP-Gripe continha o total de 1.956.350 de registros epidemiológicos de pacientes que procuraram assistência hospitalar em todos os estados brasileiros. Para o estudo, foram considerados os dados dos pacientes que tiveram a classificação final da doença registrada com o COVID-19, com indicador de evolução do caso definida como óbito ou sobrevivência, e as informações de idade, etnia e sexo registradas. Nos dados do SIVEP-Gripe além das informações já citadas também constam as comorbidades e sintomas dos pacientes.

Para compor o conjunto de dados, foram coletados indicadores referentes a fatores sociais, caracterizados em idade, raça, sexo e escolaridade por estado. Este indicador foi atribuído para cada paciente conforme suas características demográficas e local de residência

⁵ Ano de 2020: <https://opendatasus.saude.gov.br/dataset/srag-2020>.
A partir de 2021: <https://opendatasus.saude.gov.br/dataset/srag-2021-a-2023>.

no país de acordo ao SIVEP-Gripe e os dados da Pesquisa Nacional por Amostra e Domicílios-PNAD (2020)⁶.

Nos dados coletados referentes a renda média da população por estado, foi realizado o cálculo de índice de renda relativa seguindo a Fórmula 1 a seguir:

$$IRR = \frac{\text{Valor Renda Estado}}{\text{Média Renda Brasil}} \quad (1)$$

O resultado do índice calculado é o que foi utilizado para indicação de renda relativa do paciente de acordo com o seu estado (Unidade Federativa) de residência.

Embora o Brasil seja dividido em cinco regiões, para este trabalho foi criada uma divisão do país em duas macrorregiões, considerando as características semelhantes e a abordagem de Baqui et al. (2020)⁷. A macrorregião Norte, abrange as regiões Norte e Nordeste; e a macrorregião Centro-Sul, abrange as regiões Centro-Oeste, Sudeste e Sul. Esta divisão regional foi realizada para fins de modelagem e análise de dados nos modelos estatísticos escolhidos. Para a análise das chances de mortalidade entre as raças, para fins de análise no modelo estatístico, os pacientes foram agrupados em duas classes, brancos e não brancos.

3.2.2 Análise de dados

Na análise foi usada estatística descritiva para quantificar o impacto da pandemia de COVID-19 no Brasil, de forma a demonstrar as diferenças entre regiões, etnias, idades e as comorbidades registradas no banco de dados utilizado. Para a análise de predição, foi utilizado o método de regressão logística, baseando-se no trabalho de Rodrigues e Parreira (2013).

O modelo estatístico de regressão logística permite estimar a probabilidade de ocorrência de uma determinada categoria de falha (Y) em relação a um ou mais preditores (X), que podem ser contínuos ou categóricos. Uma regressão é dita binária quando a variável dependente exibe apenas dois níveis ou classes. Multinomial é usado quando há vários níveis ou classes. Dessa forma, é possível pensar na regressão logística como um complemento da regressão linear aplicada a variáveis categóricas a partir de uma função de ligação, uma generalização do teste do Qui-quadrado ou, de forma mais geral, um exemplo específico de um

⁶ Dados PNAD: <https://sidra.ibge.gov.br/home/pnadm/brasil>.

⁷ BAQUI, Pedro et al. Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. *The Lancet Global Health*, v. 8, n. 8, p. e1018-e1026, 2020. [https://doi.org/10.1016/S2214-109X\(20\)30285-0](https://doi.org/10.1016/S2214-109X(20)30285-0).

modelo linear generalizado (GLM) que incorpora uma ligação *logit* (BARKER & SHAW, 2015).

A função utilizada é definida por $\text{logit}(p)$ e está relacionada à probabilidade de o paciente evoluir para o óbito, considerando suas comorbidades e seus fatores sociais, neste caso, utilizando a variável renda relativa.

O modelo *logit* é normalmente utilizado quando a variável dependente é binária, fundamentalmente para medir a probabilidade de um determinado fenômeno. Como resposta de interesse para a variável dependente, o modelo logístico binário pode ser representado conforme mostrado na Fórmula 2:

$$P_i = E(Y = 1 | X_{1i}, X_{ni}) = \frac{e^{b_0 + b_1 X_{1i} + \dots + b_n X_{ni}}}{1 + e^{b_0 + b_1 X_{1i} + \dots + b_n X_{ni}}} \quad (2)$$

A probabilidade máxima foi utilizada para estimar as probabilidades de óbitos de indivíduos hospitalizados com COVID-19. O objetivo foi maximizar a função de probabilidade (ou seu logaritmo), ou seja, obter (através de um processo iterativo) os valores dos parâmetros do modelo de que forma a probabilidade de observar os valores de Y_i é a maior (máxima) possível.

Na regressão logística, o tamanho da amostra é fundamental (FERNANDES *et al*, 2021). Para aplicar o modelo, também é relevante considerar alguns aspectos relacionados ao desenho estatístico. O método é sensível à multicolinearidade (altos níveis de correlação entre variáveis independentes), e a solução mais comum para superar essa limitação é expandir a amostra (KENNEDY, 2008).

A função é definida pelo $\text{logit}(p)$ e está relacionada à probabilidade de casos de pacientes evoluírem para óbito, considerando suas comorbidades e fatores sociais, ou seja, utilizando o índice de renda relativa. O risco p é calculado a seguir:

$$\text{logit}(p) = \ln \left[\frac{p}{1-p} \right] \quad (3)$$

Os valores são aplicados na Fórmula 3 e a probabilidade de morte é encontrada como $1-p$. Usando esse método, calculamos coeficientes que indicam riscos para caso de COVID-19 com intervalos de confiança de Wald em 99% para as razões de chances de sobrevivência.

Este modelo estatístico permitiu observar quais variáveis são fatores de influência na chance de mortalidade do indivíduo e em que medida estes efeitos se comportam a partir do coeficiente obtido. Onde o valor de p é $<.0001$ para todas as variáveis, afirmando que elas podem ser consideradas explicativas para o cálculo da probabilidade de um indivíduo não sobreviver.

3.2.3 Materiais Utilizados

Esta seção apresenta as informações, dos materiais utilizados para o desenvolvimento das atividades desse trabalho, desde a preparação dos dados, até a etapa de análise estatística dos dados.

Sublime Text

O editor de código é uma ferramenta que todo desenvolvedor precisa em sua caixa de ferramentas. Somos apresentados a uma variedade de opções de programas quando começamos a trabalhar neste campo, e isso pode nos confundir na hora de tomar uma decisão (HILLMAN, 2022). Neste artigo, usamos o *Sublime Text*, para a parte inicial do pré-processamento dos dados da planilha CSV.

PostgreSQL

O PostgreSQL é um poderoso sistema de banco de dados objeto-relacional de código aberto com mais de 35 anos de desenvolvimento ativo que lhe rendeu uma forte reputação de confiabilidade, robustez de recursos e desempenho. Esse SGBD foi utilizado na etapa de criação de todo projeto de banco de dados, desde a importação dos dados SRAG e SIDRA, tratamento e transformação da planilha para exportação para o software SAS.

Microsoft Office Excel

Lançado em 1985, o Microsoft Office Excel, conhecido também por Office Excel ou somente Excel, é um software da Microsoft. É um programa que gera planilhas, o que antigamente só era feito por papel. Nas atividades de criação de tabelas dos resultados e gráficos foi utilizado o Excel, com as informações que foram exportadas das planilhas preparadas no Postgres.

SAS® OnDemand for Academics⁸

O software SAS foi criado na Universidade Estadual da Carolina do Norte, nos EUA, como um projeto de análise para pesquisa agrícola e inicialmente a sigla significava “Statistical Analysis System”. O SAS é um sistema integrado de aplicativos para análise de dados, que inclui recuperação de dados, gerenciamento de documentos, análise de dados estatísticos Acesso ao Banco de Dados, Geração de Gráficos e Geração de Relatórios.

O software SAS, foi utilizado nas atividades de análise de dados para a análise de predição, utilizando o método de regressão logística e de criação de tabelas e gráficos dos resultados obtidos da análise de dados.

⁸ SAS® OnDemand for Academics: Trata-se de um sistema integrado de aplicações para o processamento e análise estatística de dados, consistindo em módulos de Acesso e Recuperação de Dados, Gerenciamento de Arquivos, rotinas de Geração de Gráficos e Geração de Relatórios.

4 ARTIGO: SOCIAL, ECONOMIC, AND REGIONAL DETERMINANTS OF MORTALITY IN HOSPITALIZED PATIENTS WITH COVID-19 IN BRAZIL

Devido a devastação da pandemia, levando mais de 600 mil pessoas a óbito no Brasil, pela a doença denominada COVID-19, causada pelo vírus SARS-COV-2, este trabalho analisou as variáveis de comorbidades, idade, sexo, etnia e região geográfica de residência, isoladamente. Os dados para as análises foram coletados do sistema SIVEP-GRIPE do Ministério da Saúde do Brasil, também se utilizou dados de fatores sociais mais detalhados, disponibilizado pelo Sistema IBGE de Recuperação Automática (SIDRA), que resultou em um artigo publicado dia 31 de março de 2022, no periódico científico, Frontier Public Health, ISSN: 2296-2565. A revista eletrônica possui um fator de impacto de 6.461 (2021), pontuação na CAPES de B1 para a área interdisciplinar e o artigo encontra-se disponível no endereço: <https://doi.org/10.3389/fpubh.2022.856137>.

RESUMO

Em 10 de maio de 2021, o Brasil ocupava o segundo lugar no mundo em mortes por COVID-19. Compreender os fatores de risco, ou desigualdade social e étnica na atenção à saúde de acordo com a população de uma determinada cidade e fragilidade política ou econômica é de suma importância. O Brasil teve um surto de COVID-19 de gravidade à luz de fatores sociais e econômicos e sua complexa demografia racial. O objetivo deste estudo foi verificar as chances de mortalidade de pacientes hospitalizados durante a infecção por COVID-19 com base em suas características econômicas, sociais e epidemiológicas. Constatamos que as chances de óbito são maiores entre os pacientes com comorbidades, doenças neurológicas (1,99) e renais (1,97) e imunodeficiências (1,69). Enquanto a renda relativa (2,45) indica que os fatores sociais têm maior influência na mortalidade do que as comorbidades estudadas. Pacientes residentes na macrorregião Norte do Brasil apresentam maior chance de mortalidade em comparação aos da região Centro-Sul do Brasil. Concluímos que, no período estudado, as chances de mortalidade por COVID-19 no Brasil foram mais fortemente influenciadas por condições de pobreza socioeconômica do que por comorbidades naturais (distúrbios neurológicos, renais e de imunodeficiência), que também foram muito relevantes. Fatores regionais são relevantes nas taxas de mortalidade, visto que mais indivíduos estão vulneráveis às condições de pobreza.

Autores: Waldecy Rodrigues, Humberto da Costa Frizzera, Daniela Mascarenhas de Queiroz Trevisan, David Prata, Geovane Rossone Reis, Raulison Alves Resende.



Social, Economic, and Regional Determinants of Mortality in Hospitalized Patients With COVID-19 in Brazil

OPEN ACCESS

Edited by:

Fatima Saleh,
Beirut Arab University, Lebanon

Reviewed by:

Carlos Sanchez-Piedra,
Instituto de Salud Carlos III
(ISCIII), Spain
Elié Sokhn,
Beirut Arab University, Lebanon

*Correspondence:

Waldecy Rodrigues
waldecy@uft.edu.br

†ORCID:

Waldecy Rodrigues
orcid.org/0000-0002-5584-6586
Humberto da Costa Frizzera
orcid.org/0000-0002-7944-7571
Daniela Mascarenhas de Queiroz
Trevisan
orcid.org/0000-0003-3677-0851
David Prata
orcid.org/0000-0002-1414-4000
Geovane Rossonne Reis
orcid.org/0000-0002-6549-2371
Raulison Alves Resende
orcid.org/0000-0002-9063-7237

Specialty section:

This article was submitted to
Public Health Policy,
a section of the journal
Frontiers in Public Health

Received: 16 January 2022

Accepted: 09 March 2022

Published: 31 March 2022

Citation:

Rodrigues W, da Costa Frizzera H,
Trevisan DMdQ, Prata D, Reis GR and
Resende RA (2022) Social, Economic,
and Regional Determinants of
Mortality in Hospitalized Patients With
COVID-19 in Brazil.
Front. Public Health 10:856137.
doi: 10.3389/fpubh.2022.856137

Waldecy Rodrigues^{1,2*}, Humberto da Costa Frizzera^{1†},
Daniela Mascarenhas de Queiroz Trevisan^{1†}, David Prata^{1†}, Geovane Rossonne Reis^{2†} and
Raulison Alves Resende^{3†}

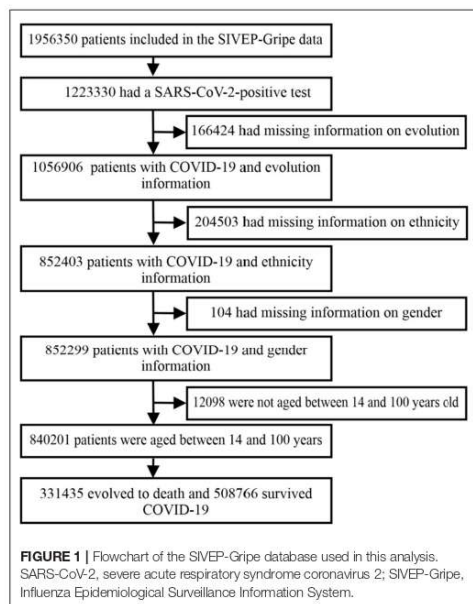
¹ Program of Computational Modelling, Federal University of Tocantins, Palmas, Brazil, ² Program of Institute of Regional Development, Federal University of Tocantins, Palmas, Brazil, ³ Insper—Institute of Education and Research, São Paulo, Brazil

On May 10, 2021, Brazil ranked second in the world in COVID-19 deaths. Understanding risk factors, or social and ethnic inequality in health care according to a given city population and political or economic weakness is of paramount importance. Brazil had a seriousness COVID-19 outbreak in light of social and economic factors and its complex racial demographics. The objective of this study was to verify the odds of mortality of hospitalized patients during COVID-19 infection based on their economic, social, and epidemiological characteristics. We found that odds of death are greater among patients with comorbidities, neurological (1.99) and renal diseases (1.97), and immunodeficiency disorders (1.69). While the relative income (2.45) indicates that social factors have greater influence on mortality than the comorbidities studied. Patients living in the Northern macro-region of Brazil face greater chance of mortality compared to those in Central-South Brazil. We conclude that, during the studied period, the chances of mortality for COVID-19 in Brazil were more strongly influenced by socioeconomic poverty conditions than by natural comorbidities (neurological, renal, and immunodeficiency disorders), which were also very relevant. Regional factors are relevant in mortality rates given more individuals being vulnerable to poverty conditions.

Keywords: social factors, economic factors, regional factors, mortality in hospitalized, COVID-19

INTRODUCTION

On December 31, 2019, the first cases of the international rapidly expanding SARS-CoV-2 virus were reported to the World Health Organization (WHO), which until then was defined as a “pneumonia of unknown origin.” After migrating beyond China’s borders, the WHO established that it was a pathological outbreak caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) virus, which caused Severe Acute Respiratory Syndrome (1, 2). Months later, on 11 March, 2020, the WHO declared a pandemic, since the virus had spread exponentially to countries all over the world (3, 4). Brazil became the third most affected country in the world, in terms of cases, and second in deaths, from COVID-19 (5).



Statistical Analysis

We used statistical data to quantify the impact of the COVID-19 pandemic on the number of deaths in Brazil. Our main question is to demonstrate the differences between regions, ethnicities, ages and, comorbidities of patients recorded in the database. The model shown the mortality rates based on variables with a fixed effect for the patients, which were comorbidities and social factors.

For this analysis, we used the logistic regression since the central question of this research is related to a dichotomous variable, whether the hospitalized individual evolved to death or not. The prediction analysis was based on the study by Rodrigues and Parreira (15).

Logistic regression is a useful tool for answering the odds of an event occurring. It can predict the binomial outcome of a dependent variable (target) using one or a set of independent variables (predictive) (16). Binary categorical variables are often used in empirical research in health sciences, such as History of Abortion: yes or no (17), Food Security Status: Insecure Food or Secure Food and Food Stability: < 3 times or 3 times (18), Post-Traumatic Stress Disorder: yes or no (19).

The logit model is usually used when the dependent variable is binary, fundamentally to measure the probability of a given phenomenon. As the response of interest to the dependent variable, the binary logistic model can be represented as shown

in Formula 1:

$$P_i = E(Y = 1 | X_{i1}, X_{ni}) = \frac{e^{b_0 + b_1 X_{i1} + \dots + b_n X_{ni}}}{1 + e^{b_0 + b_1 X_{i1} + \dots + b_n X_{ni}}} \quad (1)$$

The maximum likelihood was used to estimate the probabilities of deaths of individuals hospitalized with COVID-19. The objective was to maximize the likelihood function (or its logarithm), that is, to obtain (through an iterative process) the values of the model parameters in what way the probability of observing the values of Y_i is the highest (maximum) possible.

In logistic regression, the sample size is fundamental (20). To apply the model, is also relevant to consider some aspects related to the statistical design. The method is sensitive to multicollinearity (high levels of correlation between independent variables), and the most common solution to overcome this limitation is to expand the sample (21).

The function is defined by logit (p) and is related to the probability of patient cases evolving to death, considering their comorbidities and social factors, i.e., using the relative income index. The p risk is calculated below:

The values are applied in the formula $\text{logit}(p) = \ln \left[\frac{p}{1-p} \right]$ and the odds of death is found as $1-p$. Using this method, we calculated coefficients that indicate risks for COVID-19 case with Wald confidence intervals at 99% for the odds ratios.

RESULTS

The patients were registered in the system according to their city of residence. The absolute numbers are shown in Figure 2A, and the numbers of cases per 100,000 people are shown in Figure 2B.

Based on research from Baqui et al. (11), Tables 1, 2 show the demographic and comorbidity data between survivors and non-survivors, and their ethnic composition at each stage of the progression of COVID-19.

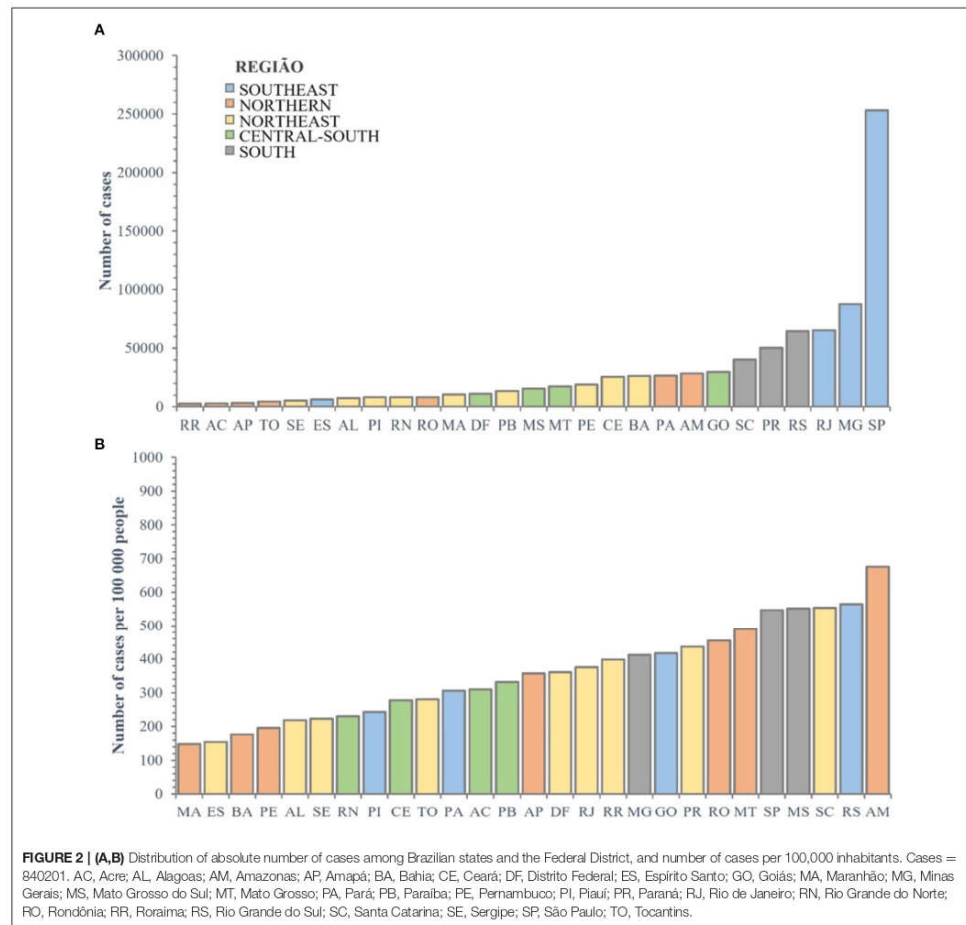
Regarding patients that have comorbidities, there were more registered deaths in the North macro-region in all comorbidities compared to the Central-South. Anova test ($p < 0.0001$ for all comorbidities; F value for each comorbidity in the Supplemental Material), suggesting structural health disparities. Even with a different number of hospitalized patients and death in the regions, the previous quantities are relative to their population, which makes the comparison statistically plausible.

The older patients had a greater mortality percentage, especially in the North ($p < 0.0001$; $F = 17489.1$). In that region the non-survivors people over 51 years old are 55.8% and in Central-South they are 44.9%.

Table 2 shows a higher proportion of hospitalized deaths in the North (46.7%) than Central-South (37.2%), suggesting a regional effect. Anova test ($p < 0.0001$; $F = 5755.17$), shown in the Supplemental Material

In addition, there were higher proportions of black ethnicities in deaths in the Central-South region (43.1%), suggesting an ethnic effect. Anova test ($p < 0.0001$; $F = 153.00$) shown in the Supplemental Material with the Tukey test result.

Figures 3A,B show comorbidity prevalence calculated considering patient, ethnicity, and region, in this case, the



Northern or Central-South, for both COVID survivors and deaths. One can see that there is substantial asymmetry between the results, with more deaths in the North. Furthermore, black Brazilians without comorbidities had the highest death rates in both the North (0.34) and the Central-South (0.42).

For groups with comorbidities in the North, most deaths were among indigenous (0.67) ethnicities. In the Central-South, when patients had one or more comorbidity, most deaths were among the black people (0.54).

The distribution of non-survivors, by age and ethnicity, that was higher for patients aged above 51 in the North (0.56) and Central-South (0.46), as shown **Figures 3C,D**. For patients aged

14–50, there was greater probability for survival in both regions. Analyzing only the descriptive data, one cannot determine large disparities in survival rates among racial groups according to number of comorbidities and age group.

The statistical model allowed us to observe which variables, and to what degree, they influence the individual's odds of mortality based on the coefficient obtained. The validity of the model is shown in **Table 3**, with $p < 0.0001$ for all variables, showing that they are explanatory for calculating the probability of an individual not surviving the disease.

One can infer that the probability of a patient not surviving are greater for patients with some comorbidities, arranged in

TABLE 1 | Demographic characteristics and comorbidities among COVID-19 survivors and non-survivors.

	Survivors (<i>n</i> = 508,766)	Non-survivors (<i>n</i> = 331,435)
North (<i>n</i> = 199,285)	106,266 (53.3%)	93,019 (46.7%)
Age (years)		
14–50 (<i>n</i> = 58,747)	44,201 (75.2%)	14,546 (24.8%)
51–100 (<i>n</i> = 140,538)	62,064 (44.2%)	78,474 (55.8%)
Gender		
Women (<i>n</i> = 88,225)	48,182 (54.6%)	40,043 (45.4%)
Men (<i>n</i> = 111,060)	58,084 (52.3%)	52,976 (47.7%)
Ethnic group		
White (<i>n</i> = 28,998)	14,757 (50.9%)	14,241 (49.1%)
Mixed-race (<i>n</i> = 157,676)	848,85 (53.8%)	72,791 (46.2%)
Black (<i>n</i> = 8,481)	4,281 (50.5%)	4,200 (49.5%)
Asian (<i>n</i> = 2,821)	1,642 (58.2%)	1,179 (41.8%)
Indigenous (<i>n</i> = 1,309)	701 (53.6%)	608 (46.4%)
Comorbidities		
Renal disease (<i>n</i> = 7,276)	2,405 (33.1%)	4,871 (66.9%)
Liver disease (<i>n</i> = 1,623)	586 (36.1%)	1,037 (63.9%)
Immunosuppression (<i>n</i> = 4,228)	1,822 (43.1%)	2,406 (56.9%)
Obesity (<i>n</i> = 11,017)	5,366 (48.7%)	5,651 (51.3%)
Neurological disease (<i>n</i> = 4890)	1,761 (36.0%)	3,129 (64.0%)
Pulmonary disease (<i>n</i> = 4,648)	1,663 (35.8%)	2,985 (64.2%)
Hematologic disease (<i>n</i> = 1,198)	530 (44.2%)	668 (55.8%)
Diabetes (<i>n</i> = 50,146)	22,397 (44.7%)	27,749 (55.3%)
Cardiovascular disease (<i>n</i> = 60,834)	27,351 (45.0%)	33,483 (55.0%)
Central-South (<i>n</i> = 640,916)	402,500 (62.8%)	238,416 (37.2%)
Age (years)		
14–50 (<i>n</i> = 183,849)	152,460 (82.9%)	31,389 (17.1%)
51–100 (<i>n</i> = 457,067)	250,040 (55.1%)	207,027 (44.9%)
Gender		
Women (<i>n</i> = 288,383)	183,063 (63.5%)	105,320 (36.5%)
Men (<i>n</i> = 352,533)	219,437 (62.2%)	133,096 (37.8%)
Ethnic group		
White (<i>n</i> = 406,146)	256,977 (63.3%)	149,169 (36.7%)
Mixed-race (<i>n</i> = 188,455)	118,534 (62.9%)	69,921 (37.1%)
Black (<i>n</i> = 37,682)	21,438 (56.9%)	16,244 (43.1%)
Asian (<i>n</i> = 7,591)	4,870 (64.2%)	2,721 (35.8%)
Indigenous (<i>n</i> = 1,042)	681 (65.4%)	361 (34.6%)
Comorbidities		
Renal disease (<i>n</i> = 23,699)	9,464 (39.9%)	14,235 (60.1%)
Liver disease (<i>n</i> = 5,426)	2,379 (43.8%)	3,047 (56.2%)
Immunosuppression (<i>n</i> = 15,469)	7,352 (47.5%)	8,117 (52.5%)
Obesity (<i>n</i> = 53,845)	30,364 (56.4%)	23,481 (43.6%)
Neurological disease (<i>n</i> = 25,113)	9,984 (39.8%)	15,129 (60.2%)
Pulmonary disease (<i>n</i> = 25,134)	10,856 (43.2%)	14,278 (56.8%)
Hematologic disease (<i>n</i> = 4,540)	2,290 (50.4%)	2,250 (49.6%)
Diabetes (<i>n</i> = 158,467)	84,854 (53.5%)	73,613 (46.5%)
Cardiovascular disease (<i>n</i> = 225,237)	121,774 (54.1%)	103,463 (45.9%)

order of least to greatest, in Table 3. The three worst factors are neurological diseases, renal diseases, and immunodeficiency disorders. Among the social variables analyzed separately, illiterate patients are more likely to die from the disease, along

with patients aged over 50. Patients residing in the Northern macro-region have coefficients that indicate a probability of mortality relative to patients in the Central-South regions, along with women relative to men. The relative income variable, relevant and central to this work, indicates that the higher the index, the lower the chance of patient death.

Figure 4 shows the odds ratio for each variable, statistically significant at a 99% confidence interval. According to these indexes, the three comorbidities that most lead the patients to death are neurological disease, renal disease, and immunodeficiency disorders, increased by 1.99, 1.97, and 1.69 times, respectively.

One can see that social factors are very significant in increasing risk of death among hospitalized patients. For individuals with lower relative income, for example, odds of death increased 2.45 times. Also, regional factors are relevant, since patients residing in the North and Northeast regions of Brazil are 1.37 times more likely to die, possibly associated with the urgent care and emergency hospital conditions available to patients.

DISCUSSION

This paper presented a study on patient mortality for hospitalizations with COVID-19 in Brazil. To the best of our knowledge, this study gives the highest number of deaths from COVID-19 from both health and social factors using Brazilian data. We were able to verify that social factors were very decisive in determining COVID-19 mortality in Brazil, even higher than the comorbidity factors.

This analysis allowed us to observe that patient survival is higher for younger females with fewer comorbidities, in line with results worldwide (22–24). Also, other important and new trends for social factors were found in Brazil. Similar to Baqui et al. (11), significant regional variations were found both in terms of case characteristics and results. São Paulo, Minas Gerais, and Rio de Janeiro states had the most absolute cases. Amazonas, Rio Grande do Sul, and Santa Catarina states had the most cases per 100,000 inhabitants. These states are important gateways to Brazil.

The distribution of hospitalized patients with COVID-19, i.e., the North, at 24% of all patients, and the Central-South, at 76%, is disproportional with the population sizes of these regions. The North holds 36% of the Brazilian population, and the Central-South region 64% (14). This difference highlights national heterogeneity, and we can therefore hypothesize that hospitalization rates in the North are lower than the Central-South, although this considers disproportionate increases of patients with COVID-19 in Amazonas in January 2021, in a new wave of cases following a new variant of the virus (25), leading us to infer that the hospital structure in the North was inadequate in attending patients.

Some substantial index variations were observed in each region. States in the Northern macro-region tended to have higher risk factors than states in the Central-South region, in alignment with the highest percentage of deaths in the North, according to Table 1. Many black Brazilians

TABLE 2 | Ethnic characteristics of patients at each stage of hospitalization from COVID-19.

	Brazilian population*	Hospital admission	ICU admission	Death	Death/hospitalization	Death (Not ICU)	Death (ICU)
Northern (n = 199,285)							
White	27.8%	28,998 (14.6%)	10,173 (16.2%)	14,241 (15.3%)	49.1%	6,867 (14.3%)	7,374 (16.4%)
Mixed-race	61.5%	157,676 (79.1%)	48,667 (77.6%)	72,791 (78.3%)	46.2%	37,960 (79.0%)	34,831 (77.4%)
Black	8.8%	8,481 (4.3%)	2,880 (4.6%)	4,200 (4.5%)	49.5%	2,139 (4.5%)	2,061 (4.6%)
Asian	1.2%	2,821 (1.4%)	811 (1.3%)	1,179 (1.3%)	41.8%	613 (1.3%)	566 (1.3%)
Indigenous	0.7%	1,309 (0.7%)	194 (0.3%)	608 (0.7%)	46.4%	460 (1.0%)	148 (0.3%)
Central-South (n = 640,916)							
White	58.7%	406,146 (63.4%)	136,622 (64.1%)	149,169 (62.6%)	36.7%	66,164 (61.2%)	83,005 (63.7%)
Mixed-race	33.2%	188,455 (29.4%)	61,157 (28.7%)	69,921 (29.3%)	37.1%	32,456 (30.0%)	37,465 (28.7%)
Black	6.8%	37,682 (5.9%)	12,655 (5.9%)	16,244 (6.8%)	43.1%	8,076 (7.5%)	8,168 (6.3%)
Asian	1.1%	7,591 (1.2%)	2,564 (1.2%)	2,721 (1.1%)	35.8%	1,175 (1.1%)	1,546 (1.2%)
Indigenous	0.3%	1,042 (0.2%)	295 (0.1%)	361 (0.2%)	34.6%	171 (0.2%)	190 (0.1%)

*Census of the Brazilian population.

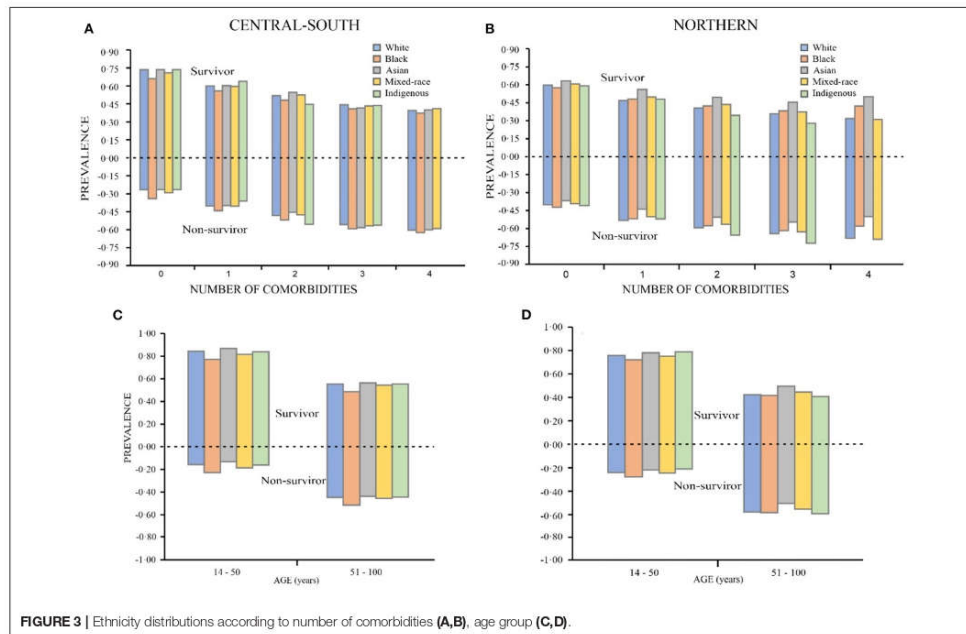


FIGURE 3 | Ethnicity distributions according to number of comorbidities (A,B), age group (C,D).

identify themselves as mixed-race (26). According to the analysis, both ethnic groups share high-risk factors and high death rates.

Of all patients diagnosed with COVID-19 (hospitalized or not) who died in Brazil, the Central-South region has the

highest mortality rate per 100,000 inhabitants (27). However, when the mortality rate is calculated for hospitalized patients only, the numbers are different. The logistic regression results suggested that risk of death is 1.37 times greater in the Northern macro-region than in the Central-South macro-region when a patient is admitted to hospital with COVID-19. It strongly indicates the high complexity health system in the Central-South macro-region. That macro-region is the richer part of the country, despite also have great incoming and poverty inequality.

Additionally, the study corroborates information stating that the Brazilian risk group comprises the elderly with comorbidities (28). Furthermore, it has been demonstrated that hospitalized Brazilian black or mixed-race patients, or those who live in the North, are at greater risk of death from COVID-19. In short, relative income associated with race, illiteracy, and regional issues, are relevant indicators for social aspects related to the COVID-19 pandemic.

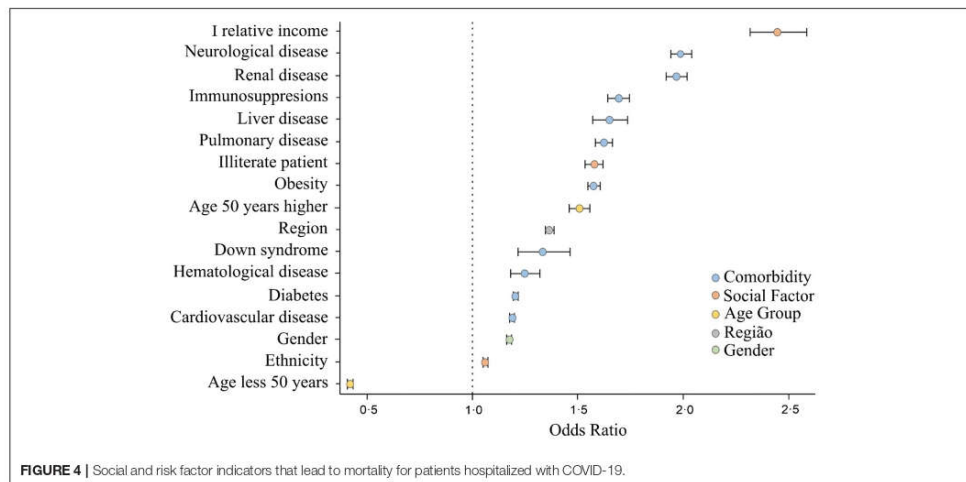
In Brazil, measures to contain the spread of the virus were managed by local cities. However, concerns about mortality rates are still growing, especially after confirmation of the first case of a new variant from India in Maranhão state on May 22, 2021, which is 50 times more communicable (29). Across the country, urgent political attention is required, directed toward the importance of vaccination in preventing COVID-19 mortality, especially when the official number of deaths has surpassed the 600,000 mark in Brazil.

The estimated relative income index was the most intense result in the statistical model performed, demonstrating that the worse the socioeconomic conditions of the individuals, the greater their chance of death when hospitalized with COVID-19. This index (relative income)

TABLE 3 | Model for mortality rates.

Parameter	Estimate	Wald qui-square	Pr > ChiSq	
Intercept	-1.5534	4143.2854	<0.0001	
Age_less_50_years	-0.8672	2678.2748	<0.0001	
Ethnicity	Non-white	0.0603	128.3516	<0.0001
Gender	Male	0.1614	1127.7744	<0.0001
Cardiovascular_disease		0.1729	1114.5035	<0.0001
Diabetes		0.1875	1137.8771	<0.0001
Hematological_disease		0.2217	59.5684	<0.0001
Down_syndrome		0.2888	37.9069	<0.0001
Region	Northern	0.3127	2061.7896	<0.0001
Age_50_years_higher		0.4106	652.1627	<0.0001
Obesity		0.4556	2596.4568	<0.0001
Illiterate_patient		0.4566	1096.2094	<0.0001
Pulmonary_disease		0.4858	1513.6012	<0.0001
Liver_disease		0.5021	376.4814	<0.0001
Immunosuppressions		0.5278	1157.7189	<0.0001
Renal_disease		0.6774	2906.2803	<0.0001
Neurological_disease		0.6880	3007.4068	<0.0001
I_relative_income*		0.8959	1029.9471	<0.0001
Number of observations used		840,201		

*I relative income means the inverse of the relative income. The higher that indicator, the worse the individual's economic condition.



was even greater than the main comorbidity found (neurological disease).

Additionally, individuals living in less populated regions located further North of the country were also more likely to die when they required hospitalization. This information led us to conclude that it is essential to strengthen the conditions of high-complexity care in the public health system, especially for the poorest individuals. Besides, it is important to consider in the public health system planning, more equitable distribution of equipment and medical support.

This study had some limitations especially regarding to the existing database. To avoid missing information, we clipped 43.5% from the initial database. These records were about evolution, ethnicity and, gender. Besides, the social database does not consider people under 14 years old, which exclude the possibility to analyze this group. We suggest that in future studies this could be better explored and considered.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: <https://opendatasus.saude.gov.br/dataset/bd-srag-2021>.

REFERENCES

- Oliveira AC, Lucas TC, Iquiapaza RA. O que a pandemia da Covid-19 tem nos ensinado sobre adoção de medidas de precaução? *Texto Contexto Enferm.* (2020) 29:1–15. doi: 10.1590/1980-265x-tce-2020-0106
- Souza-Júnior JR, Cruz RCR, Brito VC, Santos ELS, Dutra ME, Freitas IM, et al. COVID-19 e a promoção da saúde em tempos de pandemia. *Revista Eletrônica Acervo Saúde.* (2020) 46:1–7. doi: 10.25248/reas.e3837.2020
- OMS. Organização Mundial da Saúde. *Folha Informativa – COVID-19, doença causada pelo novo Coronavírus.* Organização Pan-Americana da Saúde (OPAS). (2020). Available online at: <https://www.paho.org/pt/covid19> (accessed May 18, 2021).
- WHO. Coronavirus disease 2019. *(COVID-19) Situation Report – 51.* Geneva (2020).
- Hopkins J. *University and Medicine. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University and Medicine. Coronavirus Resource Center.* (2021). Available online at: <https://coronavirus.jhu.edu/map.html> (accessed May 20, 2021).
- Lima JC, Rangel F, Pires AS. *Combate à COVID-19 e a falácia da política economicista. Boletim Coletividade – Sociologia na Pandemia – PPGS/UFSCar.* (2020). Available online at: <http://www.ppgs.ufscar.br/sociologia-na-pandemia-1> (accessed May 23, 2021).
- Farias MN, Leite JD. *Social Vulnerability and Covid-19: Considerations Based on Social Occupational Therapy.* Scielo (2021). doi: 10.1590/2526-8910.ctoen2099
- Pires LL, Carvalho L, Xavier LL. *COVID-19 e desigualdade no Brasil.* (2020). Available online at: https://www.researchgate.net/publication/340452851_COVID19_e_Desigualdade_no_Brasil (accessed May 23, 2021).
- Pires RRC. *Os efeitos sobre grupos sociais e territórios vulnerabilizados das medidas de enfrentamento à crise sanitária da COVID-19: propostas para o aperfeiçoamento da ação pública.* Instituto de Pesquisa Econômica Aplicada – IPEA. (2020). <http://www.ipea.gov.br/portal/publicacoes> (accessed June 23, 2021).
- Mathur R, Rentsch CT, Morton CE, Hulme WJ, Schultze A, MacKenna B, et al. Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: an observational cohort study using the OpenSAFELY platform. *Lancet.* (2021) 397:1711–24. doi: 10.1016/S0140-6736(21)00634-6

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

WR, DP, DT, and RR developed the research question and reviewed the statistical results. DT and HC collected and pre-processed the epidemiological and socioeconomic data. WR, DP, DT, and HC conducted analysis with descriptive statistics and regression logistic. GR participated in the theoretical foundations. All authors designed the study, analysis plan, reviewed early, and final versions of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2022.856137/full#supplementary-material>

- Baqi P, Bica I, Marra V, Ercole A, van der Schaar M. Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. *Lancet Global Health.* (2020) 8:e1018–26. doi: 10.1016/S2214-109X(20)30285-0
- Ravi K. Ethnic disparities in COVID-19 mortality: are comorbidities to blame?. *Lancet.* (2020) 396:22. doi: 10.1016/S0140-6736(20)31423-9
- Ministry of Health. *SIVEP-Gripe public dataset, 2020-2021.* Available online at: <https://opendatasus.saude.gov.br/dataset/> (accessed May 10, 2021).
- Brazilian Institute of Geography and Statistics. *SIDRA – Banco de tabelas de estatísticas. Consulta de dados em Tabela.* (2020). Available online at: <https://sidra.ibge.gov.br/home/pimptbr/brasil> (accessed June 23, 2021).
- Rodrigues W, Parreira LA. Análise do risco ao desemprego entre grupos demográficos no município de Palmas-TO no ano de 2008: uma aplicação do modelo de regressão logística binomial. *Informe GEPEC.* (2014) 17:23–33. doi: 10.48075/igpec.v17i1.5905
- Adeyemo A, Wimmer H, Powell LM. Effects of normalization techniques on logistic regression in data science. *J Inform Syst Appl Res.* (2019) 12:37.
- Poisson and Logistics Regression Analysis on Electromagnetic Field Radiation: A Case of Environmental Pollution.
- Modelling Status Food Security Households Disease Sufferers Pulmonary Tuberculosis Uses the Method Regression Logistics Binary.
- Zhang R, Hou T, Kong X, Wang G, Wang H, Xu S, et al. PTSD among healthcare workers during COVID-19 outbreak: A study raises concern for non-medical staff in low-risk areas. *Front Psychiatry.* (2021) 12:e696200. doi: 10.3389/fpsy.2021.696200
- Fernandes AAT, Figueiredo Filho DB, Rocha ECD, Nascimento WDS. Leia este artigo se você quiser aprender regressão logística. *Revista de Sociologia e Política.* (2021) 2021:28. doi: 10.1590/1678-987320287406en
- Kennedy P. *A Guide to Econometrics.* John Wiley and Sons (2008).
- The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. *China CDC Weekly.* (2020) 2:113–22. doi: 10.46234/ccdcw2020.032
- Wenham C, Smith J, Morgan R. COVID-19: the gendered impacts of the outbreak. *Lancet.* (2020) 395:846–48. doi: 10.1016/S0140-6736(20)30526-2
- Halpin DMG, Faner R, Sibila O, Badia JR, Agusti A. Do chronic respiratory diseases or their treatment affect the risk of SARS-CoV-2 infection?. *Lancet Respir Med.* (2020) 8:436–38. doi: 10.1016/S2213-2600(20)0167-3

25. USP, Jornal da. *Cientistas apontam hipóteses para novo surto de covid-19 em Manaus*. (2021). Available online at: <https://jornal.usp.br/ciencias/cientistas-apontam-hipoteses-para-novo-surto-de-covid-19-em-manaus/> (accessed June 11, 2021).
26. Bailey SR, Telles EE. Multiracial versus collective black categories: examining census classification debates in Brazil. *Ethnicities*. (2006) 6:74–101. doi: 10.1177/1468796806061080
27. Brazil, Ministry of Health. COVID-19 - Painel Coronavirus. Coronavirus Brasil. Brasília-DF (2020). Available online at: <https://covid.saude.gov.br/> (accessed May 24, 2021).
28. UFRGS. *COVID-19 e os Grupos de Risco*. (2021). Available online at: <http://www.ufrgs.br/das/campanhas/grupos-de-risco-covid-19>. (accessed June 11, 2021).
29. Ibrahim Y. *Covid-19: o que se sabe sobre a cepa indiana já identificada no Brasil* (2021). Available online at: <https://www.correiobraziliense.com.br/ciencia-e-saude/2021/05/4926039-covid-19-o-que-se-sabe-sobre-a-cepa-indiana-ja-identificada-no-brasil.html> (accessed June 11, 2021).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Rodrigues, da Costa Frizzera, Trevisan, Prata, Reis and Resende. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

5 CONSIDERAÇÕES FINAIS

Uma vez que os resultados da pesquisa são apoiados por uma rede de processos que se coordenam para produzir respostas baseadas em informações, a informação há muito tem status de importância e agora é uma das principais atividades de uma instituição.

Esse fato faz com que os pesquisadores percebam que a simples coleta de dados pode não ser suficiente para obter resultados de alta qualidade para seus trabalhos. Ao lidar com bases de dados, inclusive de conteúdo biológico, quantidades significativas de dados são coletadas, e o valor dessas coletas de dados está frequentemente vinculado a estudos de ambientes que sofreram mudanças de longo prazo.

Além disso, é de conhecimento que o calibre dos dados que são utilizados para compilar as informações é de extrema importância. Se a base da informação fosse instável, ela não contribuiria favoravelmente para as descobertas do estudo. Com base nessas características, esta pesquisa descobriu que a concepção e o tratamento dos dados durante o processo de coleta e durante a manutenção dos dados constituem os principais pilares do desenvolvimento da informação. Isso significa que é importante ficar atento ao planejamento e aos recursos utilizados para a gestão dos bancos de dados no que diz respeito à organização e preservação desses conglomerados, bem como ao apuramento preciso dos campos modelo-específicos.

Com isto, este trabalho apresentou um estudo sobre a mortalidade de pacientes hospitalizados com COVID-19 no Brasil de fevereiro/2020 à maio/2021. Até o momento, até onde conhecemos, é o estudo com o maior número de mortes por COVID-19 que envolve tanto fatores de saúde quanto sociais com dados do Brasil. A análise permitiu observar que a probabilidade de sobrevivência dos pacientes são os mais jovens, do sexo feminino e ter menos comorbidades, em concordância com os resultados mundiais (TEAM, 2020; WENHAM *et al*, 2020; HALPIN *et al*, 2020). Entretanto, no Brasil foram encontradas outras tendências importantes de fatores sociais.

Como no estudo de Baqui *et al.* (2020), foram encontradas variações regionais significativas tanto nas características dos casos quanto nos resultados. Os estados de São Paulo, Minas Gerais e Rio de Janeiro foram os que tiveram o maior número de casos absolutos. Amazonas, Rio Grande do Sul e Santa Catarina tiveram o maior número de casos por 100 mil habitantes. Esses estados são importantes portas de entradas para o Brasil

É importante citar que, a distribuição de pacientes com COVID-19 hospitalizados entre as regiões Norte, com 24% dos pacientes, e Centro-Sul, com 76%, é discordante com o tamanho populacional dessas regiões, onde Norte tem 36% habitantes, e a região Centro-Sul possui 64% (SIDRA, 2020). Essa diferença evidencia a heterogeneidade nacional e, com isso, é possível

levantar a hipótese de que a taxa de hospitalização na região Norte é menor do que na região Centro-Sul, embora contabilize o aumento desproporcional de pacientes com COVID-19 no Amazonas em janeiro de 2021, na denominada nova onda de casos transmitidas pela nova variante do vírus (USP, 2021), levando a inferir que a estrutura hospitalar do Norte do país não tem atendido na mesma proporção da necessidade de internação.

Foi observado variação substancial no fator de risco por região. Os estados da região Norte tendem a apresentar fatores de riscos mais elevados do que os da região Centro-Sul, concordando com o maior percentual de não sobreviventes na região Norte. Muitos brasileiros de raça preta acabam se identificando como de raça parda (BAILEY & TELLES, 2006). Conforme apresentado na análise, ambos os grupos étnicos compartilham fatores de risco elevados e porcentagens elevadas de não sobreviventes.

Dentre todos os pacientes diagnosticados com COVID-19 que evoluíram para óbito no Brasil, a região Centro-Sul é a que apresenta maior índice de mortalidade por 100 mil habitantes (MINISTÉRIO DA SAÚDE, 2022c), no entanto, quando a taxa de mortalidade é calculada sobre pacientes hospitalizados os números são diferentes. Os resultados sugeriram que na macrorregião Norte, quando o paciente é admitido em hospital com COVID-19, a chance de não sobrevivência é 1,37 vezes maior do que na macrorregião Centro-Sul.

Para realização do estudo foram encontradas algumas limitações que podem ser discutidas. Limitações e possíveis vieses, comuns em trabalhos em conjunto de dados e pesquisas observacionais. No conjunto de dados do SIVEP-Gripe de pacientes hospitalizados, foram encontradas informações faltantes, como de evolução do caso (166424), etnia (204503), sexo (104) e idade (10098) de um total de 1223330 pacientes com COVID-19. É importante frisar também que a análise foi realizada sobre os dados de pacientes que foram hospitalizados, que constam no conjunto de dados, disponibilizado pelo SIVEP-Gripe.

De forma adicional, o estudo corroborou com a informação de que os brasileiros do grupo de risco são os de maior idade e com comorbidade para não sobrevivência (UFRGS, 2021). Além disso, permitiu demonstrar que os pacientes brasileiros de raças pretas ou pardas hospitalizados, ou aqueles que vivem na macrorregião Norte, correm maior risco de não sobrevivência com COVID-19. Em suma, a renda relativa associadas à raça, ao analfabetismo e à questão regional são indicadores relevantes no quesito social relacionado à pandemia COVID-19.

No Brasil, as medidas de contenção da disseminação do vírus foram geridas pelas cidades individualmente, no entanto, a preocupação com a mortalidade ainda é crescente, especialmente com a confirmação do primeiro caso da nova variante da Índia encontrada no

estado do Maranhão, que é 50 vezes mais transmissível, no dia 22 de maio de 2021 (IBRAHIM, 2021). Em todo país, é requerida atenção política urgente direcionada para a importância da vacinação na prevenção a mortalidade ao COVID-19, especialmente atualmente, quando o número oficial de mortes ultrapassou a marca de 600 mil mortes no país.

5.1 Contribuições da dissertação

RODRIGUES, W.; FRIZZERA, H. C.; TREVISAN, D. M. Q.; PRATA, D.; REIS, G. R.; RESENDE, R. A. Social, Economic, and Regional Determinants of Mortality in Hospitalized Patients With COVID-19 in Brazil. **Front Public Health**. 2022 Mar 31;10:856137. doi: 10.3389/fpubh.2022.856137. PMID: 35433600; PMCID: PMC9008306.

5.2 Trabalhos futuros

Ao longo do desenvolvimento deste trabalho, puderam ser identificadas algumas possibilidades de melhoria e de continuação a partir de futuras pesquisas, sugerimos que novos trabalhos sejam desenvolvidos considerando a região e o seu clima. Trabalhos neste sentido podem permitir observar se o fator regional é apenas influenciado pelas políticas públicas ou também pelo fator climático.

REFERÊNCIAS

- ABOUZAHAR, C; ADJEI, S; KANCHANACHITRA, C. From data to policy: good practices and cautionary tales. **The Lancet**, v. 369, n. 9566, p. 1039-1046, 2007.
- AGARWAL, R.; IMIELINSKI, T.; SWAMI, A. Database mining: A performance perspective. *Knowledge and Data Engineering, IEEE Transactions on*, **IEEE**, v. 5, n. 6, p. 914-925, 1993.
- ALMEIDA FILHO, N.; ROUQUAYROL, M. Z. Elementos da metodologia epidemiológica. **In: Rouquayrol, M.Z. Epidemiologia & Saúde**. 6. ed. Rio de Janeiro. MEDSI, 2003.
- AMARAL, F. Aprenda Mineração de dados: Teoria e Prática – Rio de Janeiro, **RJ Atlas Books**, 2016.
- BAILEY S. R; TELLES E. E. Multiracial versus collective black categories: examining census classification debates in Brazil. **Ethnicities**, 2006; 6: 74–101.
- BAQUI, Pedro; BICA, Loana; MARRA, Valerio; ERCOLE, Ari; SCHAAR Mihaela V. D. Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. **The Lancet Global Health**, 2020b, v. 8, n. 8, p. e1018-e1026.
- BARKER, L. E; SHAW, K. M. (2015). Best (but oft-forgotten) practices: Checking assumptions concerning regression residuals. **The American Journal of Clinical Nutrition**, 102(3), 533–539. <https://doi.org/10.3945/ajcn.115.113498>
- BATINI, C.; SCANNAPIECO, M. Data and Information Quality: Dimensions, Principles and Techniques. **Springer**, 2016.
- BBC News. Coronavirus confirmed as pandemic. BBC News. BBC. 11 de março de 2020. Consultado em 11 de março de 2020
- BEAGLEHOLE, R.; BONITA, R.; KJELLSTRÖM, T. (1993). Basic epidemiology (pp. 133-142). **Geneva: World Health Organization**.
- BLYTHER, Tiaji S. The 2009 Influenza Pandemic: U.S. Responses to Global Human Cases. (2009) [S.l.]: **Congressional Research Service**.
- BONITA, R.; BEAGLEHOLE, R.; KJELLSTROM, T. **Epidemiologia Básica**. 2. ed. São Paulo: Grupo Editorial Nacional, 2010.
- BUTANTAN. Instituto Butantan. **Antes da Covid-19: conheça 3 doenças que também fizeram o mundo tremer neste século**. [Internet]. <https://butantan.gov.br/covid/butantan-tira-duvida/tira-duvida-noticias/antes-da-covid-19-conheca-3-doencas-que-tambem-fizeram-o-mundo-tremer-neste-seculo>. Acesso em: 11 de agosto de 2022.
- CAPPIELLO, C.; CARO, A.; RODRIGUEZ, A.; CABALLERO, I. An approach to design business processes addressing data quality issues. **In: ECIS**. [S.l.: s.n.], 2013. p. 216.
- CARVALHO, Laura; PIRES, Luiza N.; XAVIER, Laura de L. COVID-19 e Desigualdade no Brasil. **Researchgate**, 2020. 10.13140/RG.2.2.27014.73282.

CDC. **Coronavirus Disease 2019 (COVID-19)**. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/index.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Ffaq.html. Acesso em: 14 de setembro de 2022.

CHANG, Sheryl L. et al. Modelling transmission and control of the COVID-19 pandemic in Australia. **Nature communications**, v. 11, n. 1, p. 1-13, 2020.

COADIC, Y.-F. L.; GOMES, M. Y. F. A ciência da informação. [S.l.]: **Briquet de lemos Livros**, 1996.

COADIC, Y.-F. L.; GOMES, M. Y. F. A ciência da informação. [S.l.]: **Briquet de lemos Livros**, 1996.

COLEMAN, W. Koch's comma bacillus: the first year. **Bulletin of the History of Medicine** 61, no. 3 (1987): 315-342.

DOU, Diário Oficial da União. **Lei nº 13.979, de 6 de fevereiro de 2020**. Imprensa Nacional Governo Federal - Brasil. 2020. Disponível em: <https://www.in.gov.br/en/web/dou/-/lei-n-13.979-de-6-de-fevereiro-de-2020-242078735>. Acesso em: 23 de maio de 2022.

ELMASRI, R.; NAVATHE, S. Sistemas de Banco de Dados—Fundamentos e aplicações, tradução da 6a. ed.[por] Daniel Vieira. [S.l.]: **São Paulo, Pearson Addison Wesley**, 2011.

ESTACIO. **Saiba quais foram as maiores pandemias na história mundial!**. (2021) Blog Estácio. <https://blog.estacio.br/medicina/pandemias-na-historia/>. Acesso em: 20 de outubro de 2022.

EXAME. **Temos mais dados do que nunca. Como usá-los a nosso favor?**. (2021). Disponível em: <https://exame.com/carreira/dados-uso-favor/>. Acesso em: 20 de outubro de 2022.

FARIAS, Magno N.; LEITE JUNIOR, Jaime D. Social vulnerability and Covid-19: considerations based on social occupational therapy. **Cadernos Brasileiros de Terapia Ocupacional**, v. 29, 2021.

FAYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery: An Overview. Knowledge Discovery and Data Mining, **Menlo Park, AAAI Press**, 1996.

FERNANDES A. A. T.; FIGUEIREDO F. D. B.; ROCHA E. C. D; NASCIMENTO W. D. S. Leia este artigo se você quiser aprender regressão logística. **Revista de Sociologia e Política**. (2021) 2021:28. doi: 10.1590/1678-987320287406en.

FERNANDES, A. C. Gestão de Sistemas de Informação. Tese (Doutorado) — **Universidade Técnica de Lisboa**, 2005.

FILHO, Lauro A. C.; **Gripe espanhola: a mãe das pandemias**. (2022). Hospital do Coração. <https://hospitaldcoracao.com.br/novo/midias-e-artigos/artigos-nomes-da-medicina/gripe-espanhola-a-mae-de-todas-pandemias/>. Acesso em: 20 de outubro de 2022.

GOLDSCHMIDT, R. R.; PASSOS, E. **Data Mining: Um Guia Prático**. Rio de Janeiro, Campus, 2005.

GRAVES, M.; GOLDFARB, C. F. Designing XML databases. [S.l.]: **Prentice Hall PTR**, 2001.

GRECO, Dirceu B.; TUPINAMBÁS Unai; FONSECA Marise. Influenza A (H1N1): histórico, estado atual no Brasil e no mundo, perspectivas. (2009). **Revista Médica de Minas Gerais**. Vol. 19.2. ISSN 2238-3182. <http://rmmg.org/artigo/detalhes/467>. Acesso em: 20 de outubro de 2022.

GREENBERG, Raymond S.; DANIELS, Stephen R.; FLANDERS, Dana W.; ELEY, John W. Boring, III, John R (2005). **Epidemiologia Clínica**. 3ª ed. Porto Alegre: Artmed. p. 18. ISBN 85-363-0159-7.

GUIMARÃES, E. M. P.; ÉVORA, Y. D. M. Sistema de informação: instrumento para tomada de decisão no exercício da gerência. Ciência da Informação, Brasília, **SciELO Brasil**, v. 33, n. 1, p. 72–80, 2004.

GUSTAFSSON, B. Fundamentals of scientific computing. **Springer Science & Business Media**, 2011.

HALPIN, DMG; FANER, R; SIBILA, O; BADIA, J.R; AGUSTI A. Do chronic respiratory diseases or their treatment affect the risk of SARS-CoV-2 infection?. **Lancet Respir Med** 2020; 8: 436–38.

HAN, J.; KAMBER, M.; PEI, J. Data mining: concepts and techniques. [S.l.]: **Morgan kaufmann**, 2006.

HAYS, J. Epidemics and pandemics. **Their impacts on Human History**. (2005). Austin, Texas: Fundação Kahle.

HILLMAN Mônica M. **Sublime Text - conheça um dos melhores editores de código**. (2022). Disponível em: <https://www.alura.com.br/artigos/sublime-text-um-dos-melhores-editores-de-codigo>. Acesso em: 28 de outubro de 2022.

HINDSON, Jordan. COVID-19: faecal–oral transmission?. **Nature Reviews Gastroenterology & Hepatology**, v. 17, n. 5, p. 259-259, 2020.

HOPKINS, Johns. University & Medicine. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University & Medicine. [Internet]. **Coronavirus Resource Center**. 2021. Disponível em: <https://coronavirus.jhu.edu/map.html>. Acesso em: 20 de maio de 2021.

IBRAHIM, Y. **Covid-19: o que se sabe sobre a cepa indiana já identificada no Brasil**. 2021. <https://www.correiobraziliense.com.br/ciencia-e-saude/2021/05/4926039-covid-19-o-que-se-sabe-sobre-a-cep-indiana-ja-identificada-no-brasil.html>. Acesso em: 24 de novembro de 2022.

JAWORSKY, Denise. An allied research paradigm for epidemiology research with Indigenous peoples. **Archives of public health**, v. 77, n. 1, p. 1-12, 2019.

JIMÉNEZ, Villamil; CARLOS, Luis. (2013). Epidemias y pandemias: Una realidad para el siglo XXI. Um mundo y una salud. **Revista Lasallista de Investigación**, v. 10, n. 1, p. 7-8.

JÚNIOR, Antonio C. C. T.; História da varíola. (2004). **Revista Médica de Minas Gerais**. Vol. 15.1. ISSN 2238-3182. <http://rmmg.org/artigo/detalhes/1461>. Acesso em: 20 de outubro de 2022.

KENNEDY P. A Guide to Econometrics. **John Wiley and Sons** (2008).

LAURELL, A. C. A saúde-doença como processo social. **In: NUNES, E. D. (Org.). Medicina social: aspectos históricos e teóricos**. São Paulo: Global, 1983. p. 133-158. (Coleção Textos, 3).

LEITE, Cleber A.; LEITÃO, Aurileide A.; LEITE, Aurilucia A.; Influenza A (H1N1): histórico, cenário atual e perspectivas. **Ensaio e Ciência C Biológicas Agrárias e da Saúde**, v. 17, n. 1, 2013.

LEWINSOHN, Rachel. Três epidemias: lições do passado. **In: Três epidemias: lições do passado**. 2003. p. 318-318.

LIMA, Jacob C.; RANGEL, Felipe; PIRES, Aline. S. Combate à COVID-19 e a falácia da política economicista. **Contemporânea, seção especial: Sociologia na Pandemia**. v. 10 n. 1, p. 411-416. Janeiro - Abril de 2020.

MACMAHON, B.; PUGH, T. F. (2012). Principios y métodos de epidemiología. **In Principios y métodos de epidemiología** (pp. 339-339).

MASSINO, E. G.; ROLAND, C. E. de F. Banco de dados objeto-relacional para aplicações web. **Revista Eletrônica de Sistemas de Informação e de Gestão Tecnológica**, v. 5, n. 1, 2015.

MATHERS, D.C; FAT D. M; INOUE M; RAO C; LOPEZ A. D. Counting the dead and what they died from: an assessment of the global status of cause of death data. **Bulletin of the World Health Organization** 2005; 83:171-177.

MATHUR, Rohini et al. Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: an observational cohort study using the OpenSAFELY platform. **The Lancet**, 2021 May 8. 397(10286):1711-1724.

MELHUIISH, Anne; LEWTHWAITE, Penny. Natural history of HIV and AIDS. **Medicine**, v. 46, n. 6, p. 356-361, 2018.

MERSON, Michael H.; O'MALLEY, J.; SERWADDA, D.; APISUK, C. The history and challenge of HIV prevention. **The lancet**, v. 372, n. 9637, p. 475-488, 2008.

MINISTÉRIO DA SAÚDE. **Atendimento e fatores de risco**. Ministério da Saúde, Governo Federal - Brasil. 2021. Disponível em: <https://www.gov.br/saude/pt-br/coronavirus/atendimento-tratamento-e-fatores-de-risco>. Acesso em: 23 de Julho de 2022a.

MINISTÉRIO DA SAÚDE. **SIVEP-Gripe public dataset**. Ministério da Saúde, Governo Federal - Brasil. 2020-2021. <https://opendatasus.saude.gov.br/dataset/>. Acesso em: 10 de agosto de 2021b.

MINISTÉRIO DA SAÚDE. **COVID-19 - Painel Coronavírus. Coronavírus Brasil**. [Internet]. BRASIL. Brasília-DF, 2020. Disponível em: <https://covid.saude.gov.br/>. Atualizado em 24 de maio de 2021. Acesso em: 24 de novembro de 2022c.

MORENS, D.; TAUBENBERGER, J.; FAUCI, A. (2009). The Persistent Legacy of the 1918 Influenza Virus. **New England Journal of Medicine**, 361 (3), 225-229 DOI: 10.1056/NEJMp0904819.

MOSS, Andrew R.; BACCHETTI, Peter. Natural history of HIV infection. **Aids**, v. 3, n. 2, p. 55-62, 1989.

NHS, Inform. **Coronavirus (COVID-19): General advice**. <https://www.nhsinform.scot/illnesses-and-conditions/infections-and-poisoning/coronavirus-covid-19/coronavirus-covid-19/>. Acesso em: 14 de setembro de 2022.

OLIVEIRA, Adriana C.; LUCAS, Thabata C.; IQUIAPAZA, Robert A. O que a pandemia da Covid-19 tem nos ensinado sobre adoção de medidas de precaução?. **Texto Contexto Enferm**. [Internet], v. 29, p. 1-15, 2020.

OLIVEIRA, P.; RODRIGUES, F.; HENRIQUES, P. R. A formal definition of data quality problems. In: **Proceedings of the International Conference on Information Quality**, MIT. p. 13–26.2005

OLIVEIRA, Paulo J. M. **Deteção e correção de problemas de qualidade dos dados: modelo, sintaxe e semântica**. 2007. Tese de Doutorado. Universidade do Minho (Portugal).

OMS. Organização Mundial da Saúde. **Folha Informativa – COVID-19, doença causada pelo novo Coronavírus**. [Internet]. Organização Pan-Americana da Saúde (OPAS), 2020. Disponível em: <https://www.paho.org/pt/covid19>. Acesso em: 18 de maio de 2021.

ONU News. **OMS diz que combate ao novo coronavírus tem de incluir testagem de casos**. ONU Brasil. 2020. <https://news.un.org/pt/story/2020/03/1707472>. Acesso em: 11 de agosto de 2022.

OWD, Our World in Data. **COVID-19 Data Explorer**. Our World in Data. 2022. Disponível em: <https://ourworldindata.org/explorers/coronavirus-data-explorer>. Acesso em: 21 de agosto de 2022.

PEREIRA, M. G. Epidemiologia: teoria e prática. **Rio de Janeiro: Guanabara Koogan**, 2013.

PICANÇO, J.; A Cólera, uma Pandemia Imperialista. (2020). PaleoMundo. Blog Unicamp. <https://www.blogs.unicamp.br/paleoblog/2020/04/08/a-colera-uma-pandemia-imperialista/>. Acesso em: 20 de outubro de 2022.

PIRES, Roberto R. C. **Os efeitos sobre grupos sociais e territórios vulnerabilizados das medidas de enfrentamento à crise sanitária da COVID-19: propostas para o**

aperfeiçoamento da ação pública. Instituto de Pesquisa Econômica Aplicada – IPEA. 2020. Disponível em: <http://www.ipea.gov.br/portal/publicacoes>. Acesso em: 23 de maio de 2022.

PNAD. **Pesquisa Nacional por Amostra de Domicílios-Dados Gerais.** 4º Trimestre de 1967. Região I -Estados da Guanabara e Rio de Janeiro. 2ª Tiragem, fevereiro de 1969. p 2. https://biblioteca.ibge.gov.br/visualizacao/periodicos/59/pnad_1967_n2_sp.pdf.

RAHM, E.; DO, H. H. Data cleaning: Problems and current approaches. **IEEE Data Eng. Bull.**, v. 23, n. 4, p. 3–13, 2000.

RAVI, Krithi. Ethnic disparities in COVID-19 mortality: are comorbidities to blame?. June 19, 2020. **The Lancet**, 2020, volume 396, issue 10243, p22, july 04.

REDE DOR; **Variola.** (2022). Rede dor. São Luiz. <https://www.rededorsaoluiz.com.br/doencas/variola>. Acesso em: 20 de outubro de 2022.

REPICI, Dominic J.; BURTON D. **The Comma Separated Value (CSV) File Format.** (2022). Disponível em: <https://creativyst.com/Doc/Articles/CSV/CSV01.shtml>. Acesso em: 28 de outubro de 2022.

REZENDE, D.; ABREU, A. D. Tecnologia da Informação: Aplicada a Sistemas de Informação Empresariais. [S.l.]: **ATLAS**, 2013. ISBN 9788522475483.

RIBEIRO, W.; **5 principais pandemias da história.** (2022). Instituto de Ciência, Tecnologia e Qualidade – ICTQ. <https://ictq.com.br/farmacia-clinica/2485-5-principais-pandemias-da-historia>. Acesso em: 20 de outubro de 2022.

RIPSA, Rede Interagencial de Informações para a Saúde. **Indicadores básicos de saúde no Brasil: conceitos e aplicações.** Brasília: Organização Pan-Americana da Saúde; 2002.

RIZZI, Claudia B. et al. Epidemio-sistema computacional web integrado sobre eventos epidemiológicos. **ECA-Encontro em Computação Aplicada**, UNIOESTE, Cascavel, 2008.

RODRIGUES, L.; **Conheça as 5 maiores pandemias da história.** (2020). Revista Galileu. <https://revistagalileu.globo.com/Ciencia/Saude/noticia/2020/03/conheca-5-maiores-pandemias-da-historia.html>. Acesso em: 20 de outubro de 2022.

RODRIGUES, Waldecy; PARREIRA, Livian A. Análise do risco ao desemprego entre grupos demográficos no município de Palmas-TO: uma aplicação do modelo de regressão logístico binomial. **Informe Gepec** 17.1 (2013).

ROTHMAN, Kenneth J.; SANDER Greenland; TIMOTHY L. L.; Modern epidemiology. Vol. 3. **Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins**, 2008.

ROUQUAYROL, M. Z.; GOLDBAUM, M. Epidemiologia, História Natural e Prevenção de doenças. In: **ROUQUAYROL, M.Z. Epidemiologia & Saúde.** 6. ed. Rio de Janeiro. MEDSI, 2003.

SCHUELER, P.; **O que é uma pandemia.** (2021). Notícias e Artigos. Fiocruz. <https://www.bio.fiocruz.br/index.php/br/noticias/1763-o-que-e-uma-pandemia>. Acesso em: 20 de outubro de 2022.

SIDRA, IBGE. SIDRA - Banco de tabelas de estatísticas. **Consulta de dados em Tabela**. 2020. <https://sidra.ibge.gov.br/home/pimpfbr/brasil>. Acesso em: 26 de abril de 2021.

SILBERSCHATZ, A.; HENRY F. K.; SHASHANK S. Database system concepts. Vol. 5. **New York: McGraw-Hill**, 2002.

SOUZA, Christiane M. C. de. A gripe espanhola em Salvador, 1918: cidade de becos e cortiços. (2005). **História, ciências, saúde-Manguinhos**, v. 12, p. 71-99, 2005.

SOUZA-JÚNIOR, José Roberto de; CRUZ, Roberta Cristina Ribeiro; BRITO, Vinícius Cardoso; SANTOS, Estela Larissa Silva dos; DUTRA, Mellaine Fontes; FREITAS, Izabella Muraro de et al. COVID-19 e a promoção da saúde em tempos de pandemia. **Revista Eletrônica Acervo Saúde**, v. 46, n. 3837, p. 1-7, 2020.

STRONG, D. M.; LEE, Y. W.; WANG, R. Y. Data quality in context. **Communications of the ACM**, ACM, v. 40, n. 5, p. 103–110, 1997.

TEAM, Epidemiology. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. **China CDC weekly**, v. 2, n. 8, p. 113, 2020.

TEOREY, T.; LIGHTSTONE, S.; NADEAU, T. Projeto e Modelagem de Banco de Dados. 4. ed. Tradução de Daniel Vieira. Rio de Janeiro: **Editora Elsevier**, 2007.

UFRGS. **COVID-19 e os Grupos de Risco**. Universidade Federal do Rio Grande do Sul. 2021. <http://www.ufrgs.br/das/campanhas/grupos-de-risco-covid-19>. Acesso em: 24 de novembro de 2022.

USP, Jornal da. **Cientistas apontam hipóteses para novo surto de covid-19 em Manaus**. 2021. <https://jornal.usp.br/ciencias/cientistas-apontam-hipoteses-para-novo-surto-de-covid-19-em-manaus/>. Acesso em 24 de novembro de 2022.

VAC. COVID-19 vaccine development pipeline. **Vaccine Centre**, London School of Hygiene and Tropical Medicine. 1 de março de 2021. Acesso em: 14 de setembro de 2022.

VAN-TAM, Jonathan; SELLWOOD, Chloe. Brief History and Epidemiological Features of Pandemic Influenza (2010). **Introduction to Pandemic Influenza**. Wallingford, Oxfordshire: CABI. p. 41-56. ISBN 978-1-84593-578-8.

VEIGA, Allan K. **Um estudo sobre qualidade de dados em biodiversidade: aplicação a um sistema de digitalização de ocorrências de espécies**. 2012. Dissertação (Mestrado em Sistemas Digitais) - Escola Politécnica, Universidade de São Paulo, São Paulo, 2012. doi:10.11606/D.3.2012.tde-03052012-120914. Acesso em: 2022-11-20.

VERGIS, Emanuel N.; MELLORS, John W. Natural history of HIV-1 infection. **Infectious Disease Clinics**, v. 14, n. 4, p. 809-825, 2000.

WANG, R. Y.; STRONG, D. M. Beyond accuracy: What data quality means to data consumers. **Journal of management information systems**, Taylor & Francis, v. 12, n. 4, p. 5–33, 1996.

WANG, R. Y.; STRONG, D. M. Beyond accuracy: What data quality means to data consumers. **Journal of management information systems**, Taylor & Francis, v. 12, n. 4, p. 5–33, 1996.

WATSON, Oliver J. et al. Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. **The Lancet Infectious Diseases**, v. 22, n. 9, p. 1293-1302, 2022.

WENHAM, C; SMITH, J; MORGAN R. COVID-19: the gendered impacts of the outbreak. **Lancet**, 2020; 395: 846–48.

WHO. **Draft landscape of COVID-19 candidate vaccines**. 12 de novembro de 2020. <https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines>. Acesso em: 14 de setembro de 2022.

WHO. World Health Organization. **Coronavirus disease 2019 (COVID-19)**. [Internet]. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10. Acesso em: 11 de agosto de 2022.

WHO; Global Influenza Programme & World Health Organization. (2009). **Pandemic influenza preparedness and response: a WHO guidance document**. World Health Organization. <https://apps.who.int/iris/handle/10665/44123>. Acesso em: 20 de outubro de 2022.

WHOa. **Statement Regarding Cluster of Pneumonia Cases in Wuhan, China**. WHO. 31 de dezembro de 2019. <https://www.who.int/china/news/detail/09-01-2020-who-statement-regarding-cluster-of-pneumonia-cases-in-wuhan-china>. Acesso em: 12 de março de 2022

WHOb. **Director-General's opening remarks at the media briefing on COVID-19**. WHO. 11 de março de 2022. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2022>. Acesso em: 12 de março de 2022

WU, Zunyou; MCGOOGAN, Jennifer M. (2020). Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72,314 Cases From the Chinese Center for Disease Control and Prevention. **JAMA**. 323(13):1239-1242. doi: 10.1001/jama.2020.2648.

APÊNDICE A – SQL, PARA EXPORTAÇÃO DA PLANILHA.

```

select
  dados.*,
  round(dados.renda / (select max(renda) from vw_dados_paciente),3) as renda_relativa,
  round(dados.renda_p2 / (select max(renda_p2) from vw_dados_paciente),3) as renda_relativa_p2,
  round(dados.renda_p3 / (select max(renda_p3) from vw_dados_paciente),3) as renda_relativa_p3,
  --populacao por municipio/mortes/taxa de mortes/sobrevivente/ taxa de sobreviventes - calc por 1 mil habitantes
  coalesce(vms.populacao,0) as populacao,
  coalesce(vms.mortes,0) as mortes,
  round(coalesce(vms.taxa_morte,0),2) as taxa_morte,
  coalesce(vms.sobrevivente ,0) as sobrevivente,
  round(coalesce(vms.taxa_sobrevivente,0),2) as taxa_sobrevivente,
  coalesce(vms.mortes,0) + coalesce(vms.sobrevivente ,0) as total_hospitalizados,
  round((coalesce(vms.taxa_morte,0) + coalesce(vms.taxa_sobrevivente,0)),2) as taxa_total_hospitalizados,
  -- dados sobre raça por municipio (mortos e sobreviventes) - calc por 1 mil habitantes
  coalesce(vqr.raca_amarela, 0) + coalesce(vqr.raca_amarela_mortos, 0) as nr_raca_amarela_total,
  round(coalesce(((coalesce(vqr.raca_amarela, 0) + coalesce(vqr.raca_amarela_mortos, 0)) / vms.populacao::numeric)*1000,0),2) as taxa_raca_amarela_total,
  coalesce(vqr.raca_branca, 0) + coalesce(vqr.raca_branca_mortos, 0) as nr_raca_branca_total,
  round(coalesce(((coalesce(vqr.raca_branca, 0) + coalesce(vqr.raca_branca_mortos, 0)) / vms.populacao::numeric)*1000,0),2) as taxa_raca_branca_total,
  coalesce(vqr.raca_indigena, 0) + coalesce(vqr.raca_indigena_mortos, 0) as nr_raca_indigena_total,
  round(coalesce(((coalesce(vqr.raca_indigena, 0) + coalesce(vqr.raca_indigena_mortos, 0)) / vms.populacao::numeric)*1000,0),2) as taxa_raca_indigena_total,
  coalesce(vqr.raca_parda, 0) + coalesce(vqr.raca_parda_mortos, 0) as nr_raca_parda_total,
  round(coalesce(((coalesce(vqr.raca_parda, 0) + coalesce(vqr.raca_parda_mortos, 0)) / vms.populacao::numeric)*1000,0),2) as taxa_raca_parda_total,
  coalesce(vqr.raca_preta, 0) + coalesce(vqr.raca_preta_mortos, 0) as nr_raca_preta_total,
  round(coalesce(((coalesce(vqr.raca_preta, 0) + coalesce(vqr.raca_preta_mortos, 0)) / vms.populacao::numeric)*1000,0),2) as taxa_raca_preta_total,
  -- dados sobre sexo por municipio (mortos e sobreviventes) - calc por 1 mil habitantes
  coalesce(vqs.sexo_feminino, 0) + coalesce(vqs.sexo_feminino_mortos, 0) as nr_sexo_feminino_total,
  round(coalesce(((coalesce(vqs.sexo_feminino, 0) + coalesce(vqs.sexo_feminino_mortos, 0)) / vms.populacao::numeric)*1000,0),2) as taxa_sexo_feminino_total,
  coalesce(vqs.sexo_masculino, 0) + coalesce(vqs.sexo_masculino_mortos, 0) as nr_sexo_masculino_total,
  round(coalesce(((coalesce(vqs.sexo_masculino, 0) + coalesce(vqs.sexo_masculino_mortos, 0)) / vms.populacao::numeric)*1000,0),2) as taxa_sexo_masculino_total
  -- dados idhm por municipio
  round(coalesce(di.idhm, 0),2) as idhm,
  round(coalesce(di.area_territorial, 0),2) as area_territorial,
  round(coalesce(di.densidade_demografica, 0),2) as densidade_demografica,
  round(coalesce(di.escolarizacao, 0),2) as escolarizacao,
  round(coalesce(di.pib_per_capita, 0),2) as pib_per_capita,
  -- dados ifdm por municipio
  round(coalesce(di2.ifdm, 0),2) as ifdm,
  round(coalesce(di2.educacao, 0),2) as educacao,
  round(coalesce(di2.emplo_renda, 0),2) as emprego_renda,
  round(coalesce(di2.saude, 0),2) as saude,
  -- dados de profissionais da saúde
  coalesce(vqlme.medicos, 0) as nr_medicos,
  round(coalesce(vqlme.medicos / vms.populacao::numeric)*1000,0),2) as taxa_medicos,
  coalesce(vqlme.enfermeiros, 0) as nr_enfermeiros,
  round(coalesce(vqlme.enfermeiros / vms.populacao::numeric)*1000,0),2) as taxa_enfermeiros,
  -- dados koppen
  coalesce(k.koppen, '') as koppen
from
  vw_dados_paciente_2_atual dados
  left outer join vw_mortes_sobrev_atual vms on vms.ibge::text = substring(dados.cod_ibge,1,6)
  left outer join vw_qtd_comorbidade_atual vqc on vqc.ibge::text = substring(dados.cod_ibge,1,6)
  left outer join vw_qtd_leitos_med_enf vqlme on vqlme.ibge::text = substring(dados.cod_ibge,1,6)
  left outer join vw_qtd_raca_atual vqr on vqr.ibge::text = substring(dados.cod_ibge,1,6)
  left outer join vw_qtd_sexo_atual vqs on vqs.ibge::text = substring(dados.cod_ibge,1,6)
  left outer join dados_idhm di on di.cod_ibge = dados.cod_ibge::integer
  left outer join dados_ifdm di2 on di2.cod_ibge = dados.cod_ibge::integer
  left outer join koppen k on k.ibge = dados.cod_ibge::integer

```

APÊNDICE B – CÓDIGO FONTE REGRESSÃO LOGÍSTICA NO SAS.

```

LIBNAME mydata "/home/u49148566/ " access=readonly;
DATA new; set home.covid_paciente; /*database released on May/15/2021*/

if cs_raca in (2,3,4,5) then cs_raca = 0; /*Raça não branca*/

if macro_regiao = 'S' THEN macro_regiao = 1;
if macro_regiao = 'N' THEN macro_regiao = 0;
macro_regiao1 = input(macro_regiao, 4.); /*casting the variable*/

/*We selected the patients that were between 14 and 100 years old*/
if idade_paciente >= 14 or idade_paciente >100;

/*Divided the patients between Age_less_50_years and Age_50_years_higher*/
if idade_paciente <=50 then idade_menor_40 = 1;
if idade_paciente > 50 then idade_maior_50 = 1;

/*The relative income means the inverse of the relative income. The higher that indicator, the worse the individual's economic condition.*/
renda_relational = 1-renda_relativa_p3;

/*Renaming the variables*/
RENAME evolucao = Death renal = Renal_disease macro_regiao1 = Region neurologic = Neurological_disease imunodepre = Immunosuppressions
hepatica = Liver_disease pneumopati = Pulmonary_disease cs_escol_n_analfabeto = Illiterate_patient idade_maior_60 = Age_50_years_higher
Obesidade = Obesity hematologi = Hematological_disease diabetes = Diabetes cs_sexo = Gender cardiopati = Cardiovascular_disease
cs_raca = Ethnicity idade_menor_40 = Age_less_50_years renda_relational = I_relative_income sind_down = Down_syndrome;

/*Applying the logistic regression*/
PROC LOGISTIC descending plots(only)=oddsratio(order = descending)simple ; /*range = (0,2)*/

class Ethnicity(ref='1') Gender(ref='F') Region(ref='1') / param=ref;

model Death = Age_50_years_higher Age_less_50_years Down_syndrome Region Neurological_disease Renal_disease Immunosuppressions
Liver_disease Pulmonary_disease Illiterate_patient Obesity Hematological_disease Diabetes Gender Cardiovascular_disease
Ethnicity I_relative_income / ctable;

run;

```


APÊNDICE C – RESULTADOS DOS TESTES DE HIPÓTESES.

Macrorregião

O teste mostra que houve mais óbitos na macrorregião Norte.

* Considera N para a macrorregião Norte e S para a macrorregião Centro-Sul.



Figura 1 Diferença entre as Macrorregiões Centro-Sul e Norte.

O teste mostra que foram maiores as proporções de etnia negra nas mortes na região Centro-Sul.

* Considere os seguintes valores para cada etnia:

- 1: Branca
- 2: Preta
- 3: Amarela
- 4: Parda
- 5: Indígena



Figura 2 óbitos por etnia

Tabela 1 Resultados do teste Tukey:

Comparisons significant at the 0.05 level are indicated by ***.				
Ethnicity_South Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Black – Mixed-race	0.060059	0.052622	0.067495	***
Black – White	0.063802	0.056705	0.070899	***
Black – Asian	0.072630	0.056051	0.089209	***
Black – Indigenous	0.084632	0.043247	0.126017	***
Mixed-race – White	0.003743	0.000070	0.007416	***
Mixed-race – Asian	0.012571	-0.002855	0.027998	
Mixed-race – Indigenous	0.024573	-0.016364	0.065510	
White – Asian	0.008828	-0.006438	0.024094	
White – Indigenous	0.020830	-0.020047	0.061707	
Asian – Mixed-race	-0.012571	-0.027998	0.002855	
Asian – White	-0.008828	-0.024094	0.006438	
Asian – Indigenous	0.012002	-0.031535	0.055538	

O teste mostra que os pacientes mais velhos tiveram maior mortalidade, principalmente na região Norte.
 * Considere 0 representando pacientes com menos de 51 anos e 1 representando pacientes com idade igual ou superior a 51 anos.

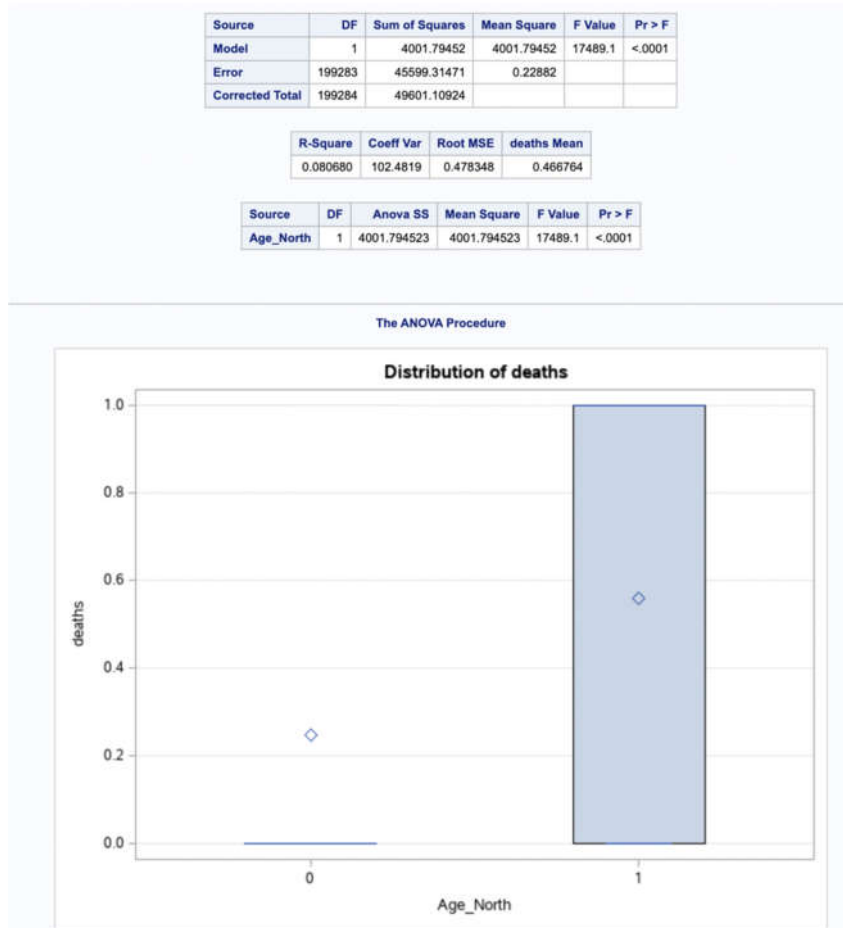


Figura 3 Mortalidade na macrorregião Norte por idade

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	10442.3395	10442.3395	48050.1	<.0001
Error	640914	139284.6762	0.2173		
Corrected Total	640915	149727.0157			

R-Square	Coef Var	Root MSE	deaths Mean
0.069743	125.3192	0.466178	0.371993

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Age_South	1	10442.33955	10442.33955	48050.1	<.0001

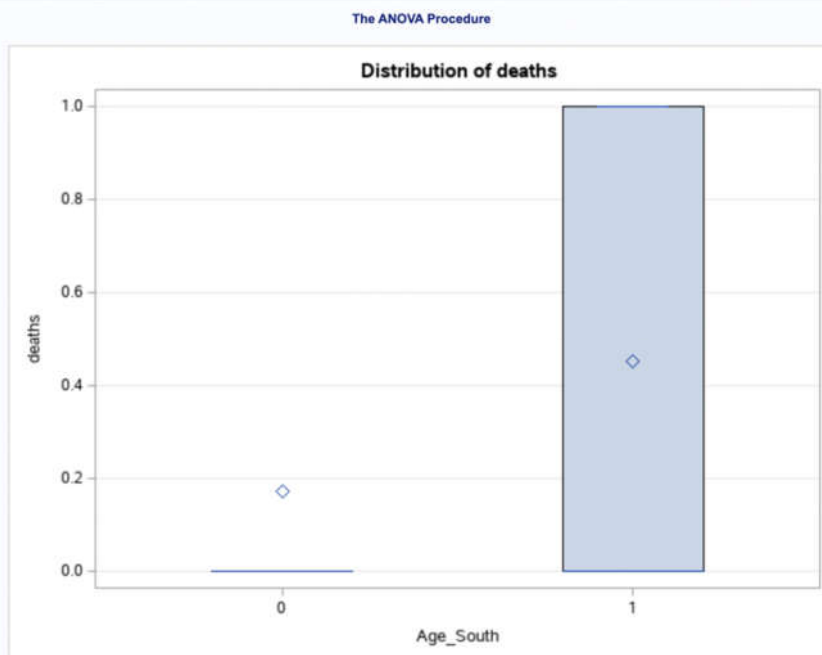


Figura 4 Mortalidade na macrorregião Centro-Sul por faixa etária.

Comorbidades

O teste mostra que houve mais óbitos na região Norte em todas as comorbidades.

*Considere 0 representando o Centro-Sul e 1 representando a macrorregião Norte.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	555.0144	555.0144	2330.00	<.0001
Error	840199	200138.4615	0.2382		
Corrected Total	840200	200693.4760			

R-Square	Coeff Var	Root MSE	death Mean
0.002765	123.7254	0.488061	0.394471

Source	DF	Anova SS	Mean Square	F Value	Pr > F
renal	1	555.0144454	555.0144454	2330.00	<.0001

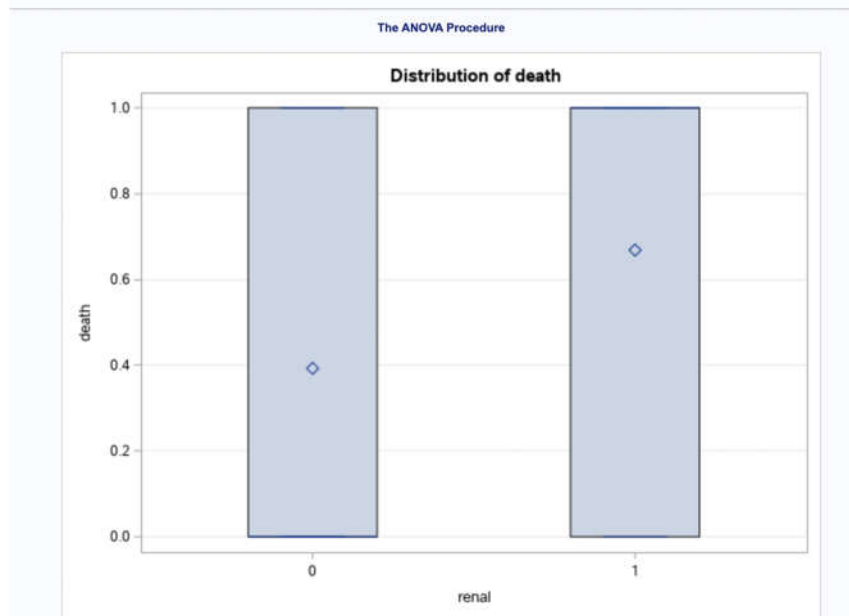


Figura 5 Teste anova de doença renal

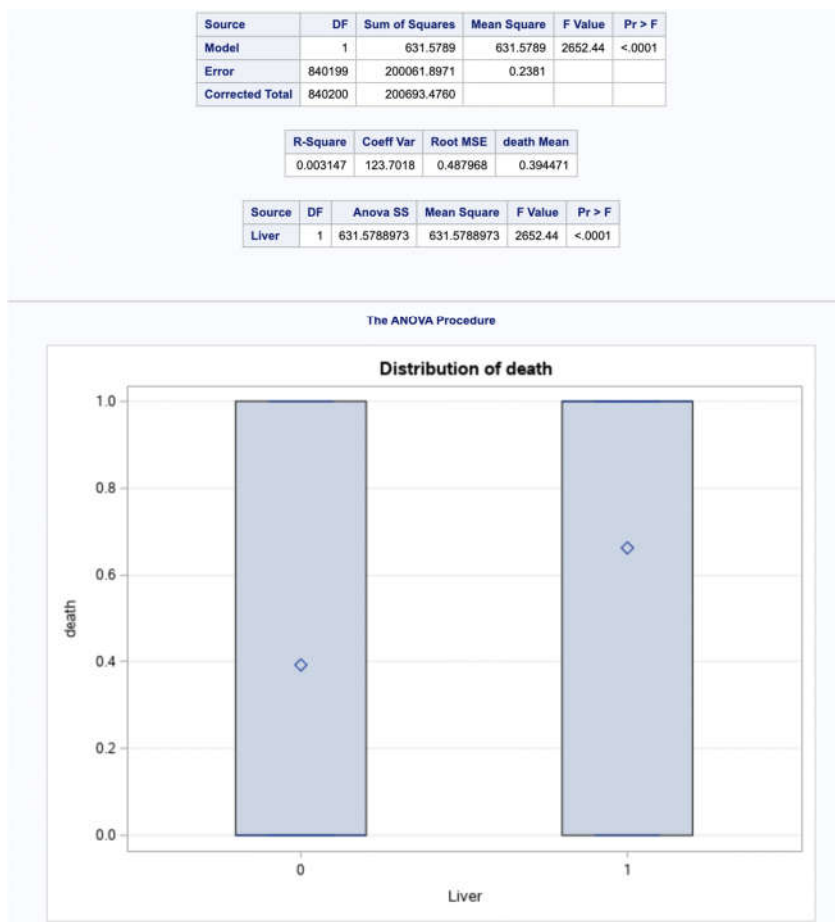


Figura 6 Teste anova de doença hepática

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	129.5317	129.5317	542.63	<.0001
Error	840199	200563.9443	0.2387		
Corrected Total	840200	200693.4760			

R-Square	Coeff Var	Root MSE	death Mean
0.000645	123.8569	0.488580	0.394471

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Immunosuppression	1	129.5317117	129.5317117	542.63	<.0001

The ANOVA Procedure

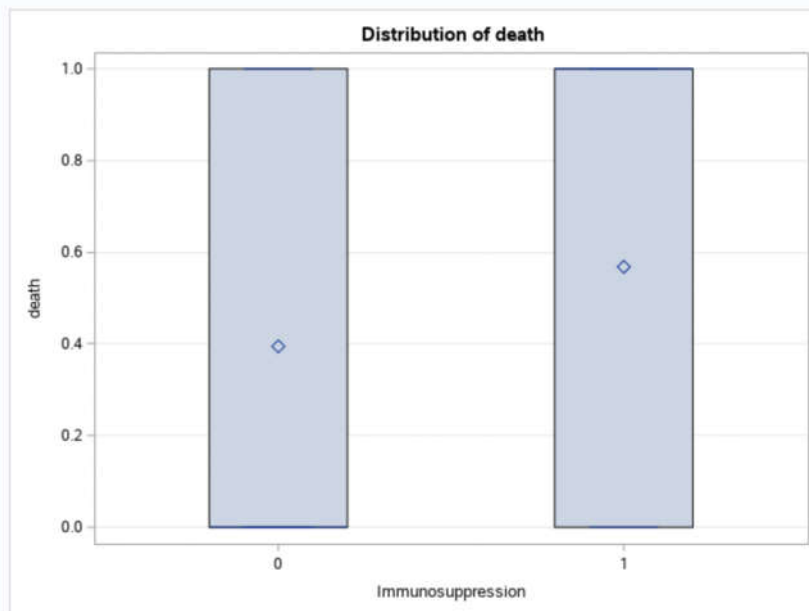


Figura 7 Teste anova da doença de imunossupressão

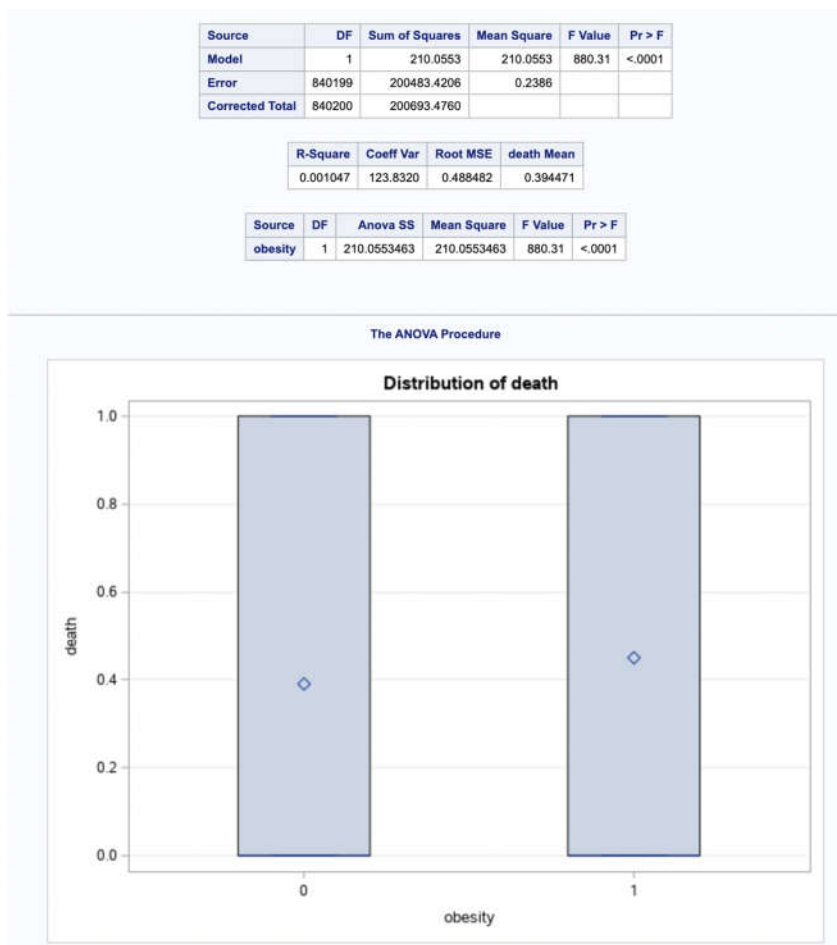


Figura 8 Teste anova de obesidade



Figura 9 Teste anova de doença neurológica

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	286.8598	286.8598	1202.65	<.0001
Error	840199	200406.6161	0.2385		
Corrected Total	840200	200693.4760			

R-Square	Coeff Var	Root MSE	death Mean
0.001429	123.8083	0.488388	0.394471

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Pulmonary	1	286.8598372	286.8598372	1202.65	<.0001

The ANOVA Procedure

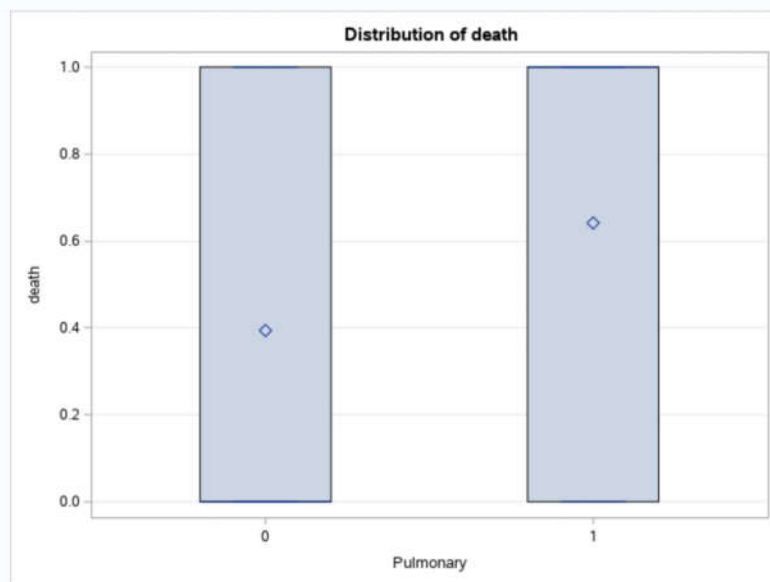


Figura 10 Teste anova de doença pulmonar

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	31.9240	31.9240	133.67	<.0001
Error	840199	200661.5520	0.2388		
Corrected Total	840200	200693.4760			

R-Square	Coeff Var	Root MSE	death Mean
0.000159	123.8870	0.488698	0.394471

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Hematologic	1	31.92398222	31.92398222	133.67	<.0001

The ANOVA Procedure

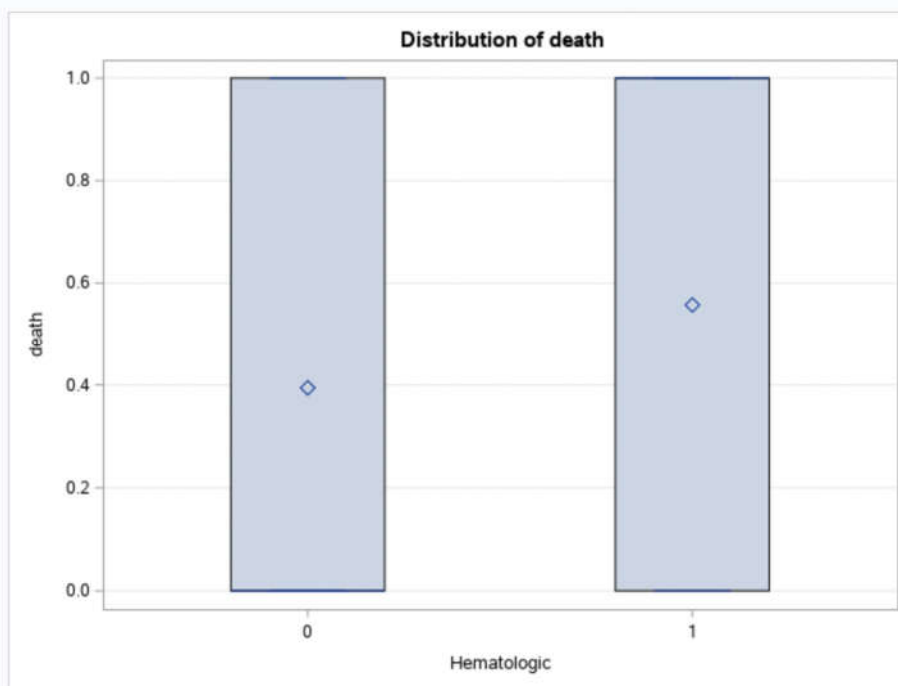


Figura 11 Teste anova de doença hematológica



Figura 12 Teste de diabetes anova

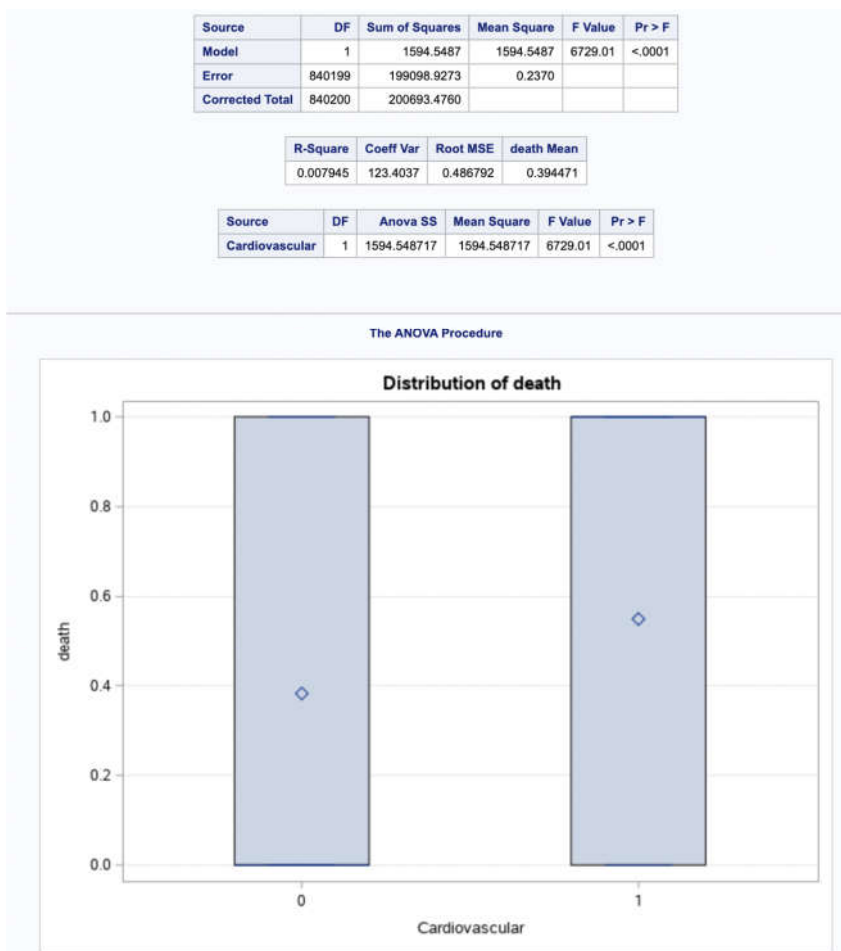


Figura 13 Teste anova de doença cardiovascular

ANEXO A – DICIONÁRIO DE DADOS DAS INFORMAÇÕES EXPORTADAS.

Nome do campo	Tipo	Categoria	Descrição	Características	DBF
Data do preenchimento da ficha de notificação	Date DD/MM/AAA A		Data de preenchimento da ficha de notificação	Campo Obrigatório: Data deve ser <= a data da digitação.	DT_NOTIFIC
Semana Epidemiológica do preenchimento da ficha de notificação	Varchar2(6)		Semana Epidemiológica do preenchimento da ficha de notificação	Campo Interno: Calculado a partir da data dos Primeiros Sintomas. (SS)	SEM_NOT
Data de 1ºs sintomas	Date DD/MM/AAA A		Data de 1º sintomas do caso	Campo Obrigatório: Data deve ser <= a data da digitação e data do preenchimento da ficha de notificação.	DT_SIN_PRI
Semana Epidemiológica dos Primeiros Sintomas	Varchar2(6)		Semana Epidemiológica dos Primeiros Sintomas	Campo Interno: Calculado a partir da data dos Primeiros Sintomas. (SS)	SEM_PRI
Unidade de Saúde Código (CNES)	Varchar2(7)	Tabela com códigos CNES e nomes das Unidades Sentinelas cadastradas no sistema SUS.	Unidade Sentinela que realizou o atendimento, coleta de amostra e registro do caso.	Campo Obrigatório: Preenchendo o nome da unidade, o código é preenchido automaticamente, e vice-versa;	ID_UNIDADE OU CO_UNI_NOT
Sexo	Varchar2 (1)	1-Masculino 2-Feminino 9-Ignorado	Sexo do paciente.	Campo Obrigatório	CS_SEXO
Data de nascimento	Date DD/MM/AAA A		Data de nascimento do paciente.	Campo Essencial: Data deve ser <= a data dos primeiros sintomas.	DT_NASC
(ou) Idade	Varchar2(3)		Idade informada pelo paciente quando não se sabe a data de nascimento.	Campo Obrigatório: Idade deve ser <= 150.	NU_IDADE_N
(ou) Tipo/Idade	Varchar2(1)	1-Dia 2-Mês 3-Ano		Campo Obrigatório: Se digitado a data de nascimento, o campo Idade/Tipo é calculado e preenchido automaticamente pelo sistema	TP_IDADE
Raça/Cor	Varchar2(2)	1-Branca 2-Preta 3-Amarela 4-Parda 5-Indígena 9-Ignorado	Cor ou raça declarada pelo paciente	Campo Obrigatório	CS_RACA
Escolaridade	Varchar2(1)	0- Sem escolaridade/ Analfabeto 1- Fundamental 1º ciclo (1ª a 5ª série) 2- Fundamental 2º ciclo (6ª a 9ª série) 3- Médio (1º ao 3º ano) 4-Superior 5- Não se aplica 9-Ignorado	Nível de escolaridade do paciente.	Campo Essencial	CS_ESCOL_N
Regional de Saúde de Residência Código (IBGE)	Varchar2 (6)	Tabela com código e nomes das Regionais de Saúde dos municípios de residência padronizados pelo IBGE.	Regional de Saúde onde está localizado o Município de residência do paciente.	Campo Interno	ID_RG_RESI OU CO_RG_RESI

Município Código (IBGE)	Varchar2(6)	Tabela com código e nome dos Municípios padronizados pelo IBGE.	Município de residência do paciente.	Campo Obrigatório	ID_MN_RESI OU CO_MUN_RES
Zona	Varchar2(1)	1-Urbana 2-Rural 3-Periurbana 9-Ignorado	Zona geográfica do endereço de residência do paciente.	Campo Essencial	CS_ZONA
Sinais e Sintomas/Febre	Varchar2 (1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou febre?	Campo Essencial	FEBRE
Sinais e Sintomas/Tosse	Varchar2 (1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou tosse?	Campo Essencial	TOSSE
Sinais e Sintomas/Dor de Garganta	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou dor de garganta?	Campo Essencial	GARGANTA
Sinais e Sintomas/Dispneia	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou dispneia?	Campo Essencial	DISPNEIA
Sinais e Sintomas/Desconforto Respiratório	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou desconforto respiratório?	Campo Essencial	DESC_RESP
Sinais e Sintomas/Saturação O2 < 95%	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou saturação O2 < 95%?	Campo Essencial	SATURACAO
Sinais e Sintomas/Diarreia	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou diarreia?	Campo Essencial	DIARREIA
Sinais e Sintomas/Vômito	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou vômito?	Campo Essencial	VOMITO
Sinais e Sintomas/Dor abdominal	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou dor abdominal?	Campo Essencial	DOR_ABD
Sinais e Sintomas/Fadiga	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou fadiga?	Campo Essencial	FADIGA
Sinais e Sintomas/Perda do Olfato	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou perda do olfato?	Campo Essencial	PERD_OLFT
Sinais e Sintomas/Perda do Paladar	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou perda do paladar?	Campo Essencial	PERD_PALA
Sinais e Sintomas/Outros	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou outro(s) sintoma(s)?	Campo Essencial	OUTRO_SIN
Fatores de risco	Varchar2 (1)	1-Sim 2-Não 9-Ignorado	Paciente apresenta algum fator de risco	Campo Essencial	FATOR_RISC
Fatores de risco/ Puérpera	Varchar2 (1)	1-Sim 2-Não 9-Ignorado	Paciente é puérpera ou parturiente (mulher que pariu recentemente – até 45 dias do parto)?	Campo Essencial Habilitado se selecionado no campo 8- Sexo Feminino.	PUERPERA
Fatores de risco/ Doença Cardiovascular Crônica	Varchar2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui Doença Cardiovascular Crônica?	Campo Essencial	CARDIOPATI
	Varchar2 (1)	1-Sim 2-Não	Paciente possui Doença	Campo Essencial	HEMATOLOG I

Fatores de risco/ Doença Hematológica Crônica		9-Ignorado	Hematológica Crônica?		
Fatores de risco/ Síndrome de Down	Varchar2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui Síndrome de Down?	Campo Essencial	SIND_DOWN
Fatores de risco/ Doença Hepática Crônica	Varchar2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui Doença Hepática Crônica?	Campo Essencial	HEPATICA
Fatores de risco/ Asma	Varchar2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui Asma?	Campo Essencial	ASMA
Fatores de risco/ Diabetes mellitus	Varchar2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui Diabetes <i>mellitus</i> ?	Campo Essencial	DIABETES
Fatores de risco/ Doença Neurológica Crônica	Varchar2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui Doença Neurológica?	Campo Essencial	NEUROLOGIC
Fatores de risco/ Outra Pneumatopatia Crônica	Varchar2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui outra pneumopatia crônica?	Campo Essencial	PNEUMOPATI
Fatores de risco/ Imunodeficiência ou Imunodepressão	Varchar2 (1)	1-Sim 2-Não 9- Ignorado	Paciente possui Imunodeficiência ou Imunodepressão (diminuição da função do sistema imunológico)?	Campo Essencial	IMUNODEPRE
Fatores de risco/ Doença Renal Crônica	Varchar2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui Doença Renal Crônica?	Campo Essencial	RENAL
Fatores de risco/ Obesidade	Varchar2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui obesidade?	Campo Essencial	OBESIDADE
Fatores de risco/ Outros	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Paciente possui outro(s) fator(es) de risco?	Campo Essencial	OUT_MORBI
Houve internação?	Varchar2(1)	1-Sim 2-Não 9-Ignorado	O paciente foi internado?	Campo Essencial Caso o campo não seja igual a 1 – Sim o sistema emitirá um aviso indicando que não atende a definição de caso.	HOSPITAL
Data da internação por SRAG	Date DD/MM/AAA A		Data em que o paciente foi hospitalizado.	Campo Obrigatório Data deve ser maior ou igual a 2- Data de 1ºs sintomas e menor ou igual a data da digitação (atual).	DT_INTERNA
Internado em UTI?	Varchar2(1)	1-Sim 2-Não 9-Ignorado	O paciente foi internado em UTI?	Campo Essencial	UTI
Data da entrada na UTI	Date DD/MM/AAA A		Data de entrada do paciente na unidade de Terapia intensiva (UTI).	Campo Essencial Habilitado se campo 47- Internado em UTI? for igual a 1. Data deve ser maior ou igual a 2-Data de 1ºs sintomas da SRAG e menor ou igual a data da digitação (atual).	DT_ENTUTI
Data da saída da UTI	Date DD/MM/AAA A		Data em que o paciente saiu da intensiva (UTI).	Campo Essencial Habilitado se campo 47- Internado em UTI? for igual a 1.	DT_SAIDUTI

				Data deve ser maior ou igual a 48-Data da entrada na UTI e menor ou igual a data da digitação (atual).	
Uso de suporte ventilatório?	Varchar2(1)	1- Sim, invasivo 2- Sim, não invasivo 3-Não 9-Ignorado	O paciente fez uso de suporte ventilatório?	Campo Essencial	SUPPORT_VEN
Raio X de Tórax	Varchar2(1)	1- Normal 2- Infiltrado intersticial 3- Consolidação 4-Misto 5-Outro 6-Não realizado 9- Ignorado	Informar resultado de Raio X de Tórax.	Campo Essencial	RAIOX_RES
Aspecto Tomografia	Number(3)	1- Típico COVID-19 2- Indeterminado COVID-19 3- Atípico COVID-19 4- Negativo para Pneumonia 5- Outro 6- Não realizado 9-Ignorado	Informar o resultado da	Campo Essencial	TOMO_RES
Coletou amostra?	Varchar2 (1)	1-Sim 2-Não 9-Ignorado	Foi realizado coleta de amostra para realização de teste diagnóstico?	Campo Essencial	AMOSTRA
Tipo de amostra	Varchar2(30)	1- Secreção de Naso- orofaringe 2- Lavado Broco-alveolar 3-Tecido <i>post-mortem</i> 4-Outra, qual? 5- LCR 9-Ignorado	Tipo da amostra clínica coletada para o teste diagnóstico.	Campo Essencial Habilitado de campo 55- Coletou amostra? = 1.	TP_AMOSTRA
Resultado da RT-PCR/outro método por Biologia Molecular	Varchar2 (1)	1- Detectável 2- Não Detectável 3-Inconclusivo 4- Não Realizado 5- Aguardando Resultado 9- Ignorado	Resultado do teste de RT-PCR/outro método por Biologia Molecular.	Campo Essencial Este campo virá marcado com 5-Aguardando Resultado e estará habilitado se o campo 55-Coletou amostra? = 1.	PCR_RESUL
Cassificação final do caso	Varchar2(1)	1- SRAG por influenza 2- SRAG por outro vírus respiratório 3- SRAG por outro agente etiológico, qual:	Diagnóstico final do caso.	Campo Obrigatório	CLASSI_FIN

		4- SRAG não especificado 5- SRAG por COVID-19			
Evolução do caso	Varchar2(1)	1-Cura 2-Óbito 3- Óbito por outras causas 9-Ignorado	Evolução do caso	Campo Essencial	EVOLUCAO
Data da alta ou óbito	Date DD/MM/AAA A		Data da alta ou óbito	Campo Essencial Data da alta ou do óbito deve ser > ou = a data dos primeiros sintomas e <= a data da digitação (atual). Habilitado se campo 74- Evolução do caso = 1 ou 2.	DT_EVOLUCA
Tipo de verificação do município		1 – Município Residência do Paciente 2 – Município, onde foi notificado.		A utilização desse campo, se dar por motivo que a dados de internação de paciente onde não consta informação do município de residência, então os calculos por município passam a ser pelo município de notificação, caso tenha a informação de município residência os cálculos é feito pelo município residência, essa verificação é só para mostrar de qual município estar sendo realizado os cálculos.	TIPO_MUNICIPIO
Número de População do Município residência do paciente	Inteiro		Número da população do município residência		POPULAÇÃO
Número de Mortes por COVID no município	Inteiro		Número de mortes por COVID no município de residência do paciente		MORTES
Taxa de mortes no município por COVID	Double		Taxa de mortes no município de residência do paciente por COVID	Os dados gerados nesse campo é composto da seguinte forma: MORTES / POPULAÇÃO * 1000 Onde a taxa é calcula a cada 1 mil habitantes.	TAXA_MORTE
Número de sobreviventes do município por COVID	Double		Número de sobrevivente por COVID no município de residência do paciente		SOBREVIVENTE
Taxa de sobrevivente no município por COVID			Taxa de sobrevivente no município de residência do paciente por COVID	Os dados gerados nesse campo é composto da seguinte forma: SOBREVIVENTE / POPULAÇÃO * 1000 Onde a taxa é calcula a cada 1 mil habitantes.	TAXA_SOBREVIVENTE
Numero de hospitalizados por COVID no município			Número de hospitalizados por COVID no município de residência do paciente		TOTAL_HOSPITALIZADOS

Taxa de hospitalizados no município por COVID			Taxa de hospitalizados no município de residência do paciente por COVID	Os dados gerados nesse campo é composto da seguinte forma: TOTAL_HOSPITALIZADOS / POPULACAO * 1000	TAXA_TOTAL_HOSPITALIZADOS
Número de pacientes com COVID que tenham asma por município			Número de pacientes com COVID que tenham asma por município de residência		NR_ASMA
Taxa de pacientes com asma e com COVID por município.			Taxa de pacientes com asma e com COVID por município.	Os dados gerados nesse campo é composto da seguinte forma: NR_ASMA / POPULACAO * 1000	TAXA_NR_ASMA
Número de pacientes mortos por COVID que tenham asma por município			Número de pacientes mortos por COVID que tenham asma por município		NR_ASMA_MORTOS
Taxa de pacientes mortos por COVID que tenham asma por município			Taxa de pacientes mortos por COVID que tenham asma por município	Os dados gerados nesse campo é composto da seguinte forma: NR_ASMA_MORTOS / POPULACAO * 1000	TAXA_NR_ASMA_MORTOS
Número de pacientes com COVID que tenham cardiopatia por município			Número de pacientes com COVID que tenham cardiopatia por município		NR_CARDIOPATI
Taxa de pacientes com COVID que tenham cardiopatia por município			Taxa de pacientes com COVID que tenham cardiopatia por município	Os dados gerados nesse campo é composto da seguinte forma: NR_CARDIOPATI / POPULACAO * 1000	TAXA_NR_CARDIOPATI
Número de pacientes mortos por COVID que tenham cardiopatia por município			Número de pacientes mortos por COVID que tenham cardiopatia por município		NR_CARDIOPATI_MORTOS
Taxa de pacientes mortos por COVID que tenham cardiopatia por município			Taxa de pacientes mortos por COVID que tenham cardiopatia por município	Os dados gerados nesse campo é composto da seguinte forma: NR_CARDIOPATI_MORTOS / POPULACAO * 1000	TAXA_NR_CARDIOPATI_MORTOS
Número de pacientes com COVID que tenham diabetes por município			Número de pacientes com COVID que tenham diabetes por município		NR_DIABETES
Taxa de pacientes com COVID que tenham diabetes por município			Taxa de pacientes com COVID que tenham diabetes por município	Os dados gerados nesse campo é composto da seguinte forma: NR_DIABETES / POPULACAO * 1000	TAXA_NR_DIABETES
Número de pacientes mortos por COVID que tenham diabetes por município			Número de pacientes mortos por COVID que tenham diabetes por município		NR_DIABETES_MORTOS
Taxa de pacientes mortos por COVID que tenham diabetes por município			Taxa de pacientes mortos por COVID que tenham diabetes por município	Os dados gerados nesse campo é composto da seguinte forma: NR_DIABETES_MORTOS / POPULACAO * 1000	TAXA_NR_DIABETES_MORTOS

Número de pacientes com COVID que tenham hematologia por município			Número de pacientes com COVID que tenham hematologia por município		NR_HEMATOLOGI
Taxa de pacientes com COVID que tenham hematologia por município			Taxa de pacientes com COVID que tenham hematologia por município	Os dados gerados nesse campo é composto da seguinte forma: NR_HEMATOLOGI / POPULACAO * 1000	TAXA_NR_HEMATOLOGI
Número de pacientes mortos por COVID que tenham hematologia por município			Número de pacientes mortos por COVID que tenham hematologia por município		NR_HEMATOLOGIS
Taxa de pacientes mortos por COVID que tenham hematologia por município			Taxa de pacientes mortos por COVID que tenham hematologia por município	Os dados gerados nesse campo é composto da seguinte forma: NR_HEMATOLOGI_MORTOS / POPULACAO * 1000	TAXA_NR_HEMATOLOGI_MORTOS
Número de pacientes com COVID que são hepática por município			Número de pacientes com COVID que são hepática por município		NR_HEPATICA
Taxa de pacientes com COVID que são hepática por município			Taxa de pacientes com COVID que são hepática por município	Os dados gerados nesse campo é composto da seguinte forma: NR_HEPATICA / POPULACAO * 1000	TAXA_NR_HEPATICA
Número de pacientes mortos por COVID que são hepática por município			Número de pacientes mortos por COVID que são hepática por município		NR_HEPATICA_MORTOS
Taxa de pacientes mortos por COVID que são hepática por município			Taxa de pacientes mortos por COVID que são hepática por município	Os dados gerados nesse campo é composto da seguinte forma: NR_HEPATICA_MORTOS / POPULACAO * 1000	TAXA_NR_HEPATICA_MORTOS
Número de pacientes com COVID que são imunodepre por município			Número de pacientes com COVID que são imunodepre por município		NR_IMUNODEPRE
Taxa de pacientes com COVID são imunodepre por município			Taxa de pacientes com COVID são imunodepre por município	Os dados gerados nesse campo é composto da seguinte forma: NR_IMUNODEPRE / POPULACAO * 1000	TAXA_NR_IMUNODEPRE
Número de pacientes mortos por COVID que são imunodepre por município			Número de pacientes mortos por COVID que são imunodepre por município		NR_IMUNODEPRE_MORTOS
Taxa de pacientes mortos por COVID que são imunodepre por município			Taxa de pacientes mortos por COVID que são imunodepre por município	Os dados gerados nesse campo é composto da seguinte forma: NR_IMUNODEPRE_MORTOS / POPULACAO * 1000	TAXA_NR_IMUNODEPRE_MORTOS
Número de pacientes com COVID que são neurológicas por município			Número de pacientes com COVID que são neurológicas por município		NR_NEUROLOGIC
Taxa de pacientes com COVID são imunodepre por município			Taxa de pacientes com COVID são imunodepre por município	Os dados gerados nesse campo é composto da seguinte forma: NR_NEUROLOGIC / POPULACAO * 1000	TAXA_NR_NEUROLOGIC

Número de pacientes mortos por COVID que são neurológicas por município			Número de pacientes mortos por COVID que são neurológicas por município		NR_NEUROLOGIC_MORTOS
Taxa de pacientes mortos por COVID que são neurológicas por município			Taxa de pacientes mortos por COVID que são neurológicas por município	Os dados gerados nesse campo é composto da seguinte forma: NR_NEUROLOGIC_MORTOS / POPULACAO * 1000	TAXA_NR_NEUROLOGIC_MORTOS
Número de pacientes com COVID que obesas por município			Número de pacientes com COVID que obesas por município		NR_OBESIDADE
Taxa de pacientes com COVID obesas por município			Taxa de pacientes com COVID obesas por município	Os dados gerados nesse campo é composto da seguinte forma: NR_OBESIDADE / POPULACAO * 1000	TAXA_NR_OBESIDADE
Número de pacientes mortos por COVID obesos por município			Número de pacientes mortos por COVID obesos por município		NR_OBESIDADE_MORTOS
Taxa de pacientes mortos por COVID obesos por município			Taxa de pacientes mortos por COVID obesos por município	Os dados gerados nesse campo é composto da seguinte forma: NR_OBESIDADE_MORTOS / POPULACAO * 1000	TAXA_NR_OBESIDADE_MORTOS
Número de pacientes com COVID com outras comorbidades por município			Número de pacientes com COVID com outras comorbidades por município		NR_OUT_MORBI
Taxa de pacientes com COVID com outras comorbidades por município			Taxa de pacientes com COVID com outras comorbidades por município	Os dados gerados nesse campo é composto da seguinte forma: NR_OUT_MORBI / POPULACAO * 1000	TAXA_NR_OUT_MORBI
Número de pacientes mortos por COVID com outras comorbidades por município			Número de pacientes mortos por COVID com outras comorbidades por município		NR_OUT_MORBI_MORTOS
Taxa de pacientes mortos por COVID com outras comorbidades por município			Taxa de pacientes mortos por COVID com outras comorbidades por município	Os dados gerados nesse campo é composto da seguinte forma: NR_OUT_MORBI_MORTOS / POPULACAO * 1000	TAXA_NR_OUT_MORBI_MORTOS
Número de pacientes com COVID que tenham pneumopatia por município			Número de pacientes com COVID que tenham pneumopatia por município		NR_PNEUMOPATI
Taxa de pacientes com COVID que tenham pneumopatia por município			Taxa de pacientes com COVID que tenham pneumopatia por município	Os dados gerados nesse campo é composto da seguinte forma: NR_PNEUMOPATI / POPULACAO * 1000	TAXA_NR_PNEUMOPATI
Taxa de pacientes mortos por COVID que tenham pneumopatia por município			Taxa de pacientes mortos por COVID que tenham pneumopatia por município	Os dados gerados nesse campo é composto da seguinte forma: NR_PNEUMOPATI_MORTOS / POPULACAO * 1000	TAXA_NR_PNEUMOPATI_MORTOS
Número de pacientes com COVID que tenham puerpera por município			Número de pacientes com COVID que tenham puerpera por município		NR_PUERPERA

Taxa de pacientes com COVID que tenham puerpera por município			Taxa de pacientes com COVID que tenham puerpera por município	Os dados gerados nesse campo é composto da seguinte forma: NR_PUERPERA / POPULACAO * 1000	TAXA_NR_PUERPERA
Número de pacientes mortos por COVID que tenham puerpera por município			Número de pacientes mortos por COVID que tenham puerpera por município		NR_PUERPERA_MORTOS
Taxa de pacientes mortos por COVID que tenham puerpera por município			Taxa de pacientes mortos por COVID que tenham puerpera por município	Os dados gerados nesse campo é composto da seguinte forma: NR_PUERPERA_MORTOS / POPULACAO * 1000	TAXA_NR_PUERPERA_MORTOS
Número de pacientes com COVID que tenham problema renal por município			Número de pacientes com COVID que tenham problema renal por município		NR_RENAL
Taxa de pacientes com COVID que tenham problema renal por município			Taxa de pacientes com COVID que tenham problema renal por município	Os dados gerados nesse campo é composto da seguinte forma: NR_RENAL / POPULACAO * 1000	TAXA_NR_RENAL
Número de pacientes mortos por COVID que tenham problema renal por município			Número de pacientes mortos por COVID que tenham problema renal por município		NR_RENAL_MORTOS
Taxa de pacientes mortos por COVID que tenham problema renal por município			Taxa de pacientes mortos por COVID que tenham problema renal por município	Os dados gerados nesse campo é composto da seguinte forma: NR_RENAL_MORTOS / POPULACAO * 1000	TAXA_NR_RENAL_MORTOS
Número de pacientes com COVID que tenham síndrome de down por município			Número de pacientes com COVID que tenham síndrome de down por município		NR_SIND_DOWN
Taxa de pacientes com COVID que tenham síndrome de down por município			Taxa de pacientes com COVID que tenham síndrome de down por município	Os dados gerados nesse campo é composto da seguinte forma: NR_SIND_DOWN / POPULACAO * 1000	TAXA_NR_SIND_DOWN
Número de pacientes mortos por COVID que tenham síndrome de down por município			Número de pacientes mortos por COVID que tenham síndrome de down por município		NR_SIND_DOWN_MORTOS
Taxa de pacientes mortos por COVID que tenham síndrome de down por município			Taxa de pacientes mortos por COVID que tenham síndrome de down por município	Os dados gerados nesse campo é composto da seguinte forma: NR_SIND_DOWN_MORTOS / POPULACAO * 1000	TAXA_NR_SIND_DOWN_MORTOS
Número de pacientes com COVID são da raça amarela por município			Número de pacientes com COVID são da raça amarela por município		NR_RACA_AMARELA
Taxa de pacientes mortos por COVID são da raça amarela por município			Taxa de pacientes mortos por COVID são da raça amarela por município	Os dados gerados nesse campo é composto da seguinte forma: NR_RACA_AMARELA / POPULACAO * 1000	TAXA_RACA_AMARELA
Número de pacientes mortos por COVID são da raça amarela por município			Número de pacientes mortos por COVID são da raça amarela por município		NR_RACA_AMARELA_MORTOS

Taxa de pacientes mortos por COVID são da raça amarela por município			Taxa de pacientes mortos por COVID são da raça amarela por município	Os dados gerados nesse campo é composto da seguinte forma: NR_RACA_AMARELA_MORTOS / POPULACAO * 1000	TAXA_RACA_AMARELA_MORTOS
Número de pacientes com COVID são da raça branca por município			Número de pacientes com COVID são da raça branca por município		NR_RACA_BRANCA
Taxa de pacientes mortos por COVID são da raça branca por município			Taxa de pacientes mortos por COVID são da raça branca por município	Os dados gerados nesse campo é composto da seguinte forma: NR_RACA_BRANCA / POPULACAO * 1000	TAXA_RACA_BRANCA
Número de pacientes mortos por COVID são da raça branca por município			Número de pacientes mortos por COVID são da raça branca por município		NR_RACA_BRANCA_MORTOS
Taxa de pacientes mortos por COVID são da raça branca por município			Taxa de pacientes mortos por COVID são da raça branca por município	Os dados gerados nesse campo é composto da seguinte forma: NR_RACA_BRANCA_MORTOS / POPULACAO * 1000	TAXA_RACA_BRANCA_MORTOS
Número de pacientes com COVID são da raça indefinida por município			Número de pacientes com COVID são da raça indefinida por município		NR_RACA_INDEFINIDA
Taxa de pacientes mortos por COVID são da raça indefinida por município			Taxa de pacientes mortos por COVID são da raça indefinida por município	Os dados gerados nesse campo é composto da seguinte forma: NR_RACA_INDEFINIDA / POPULACAO * 1000	TAXA_RACA_INDEFINIDA
Número de pacientes mortos por COVID são da raça indefinida por município			Número de pacientes mortos por COVID são da raça indefinida por município		NR_RACA_INDEFINIDA_MORTOS
Taxa de pacientes mortos por COVID são da raça indefinida por município			Taxa de pacientes mortos por COVID são da raça indefinida por município	Os dados gerados nesse campo é composto da seguinte forma: NR_RACA_INDEFINIDA_MORTOS / POPULACAO * 1000	TAXA_RACA_INDEFINIDA_MORTOS
Número de pacientes com COVID são da raça indígena por município			Número de pacientes com COVID são da raça indígena por município		NR_RACA_INDIGENA
Taxa de pacientes com COVID são da raça indígena por município			Taxa de pacientes com COVID são da raça indígena por município	Os dados gerados nesse campo é composto da seguinte forma: NR_RACA_INDIGENA / POPULACAO * 1000	TAXA_RACA_INDIGENA
Número de pacientes mortos por COVID são da raça indígena por município			Número de pacientes mortos por COVID são da raça indígena por município		NR_RACA_INDIGENA_MORTOS
Taxa de pacientes mortos por COVID são da raça indígena por município			Taxa de pacientes mortos por COVID são da raça indígena por município	Os dados gerados nesse campo é composto da seguinte forma: NR_RACA_INDIGENA_MORTOS / POPULACAO * 1000	TAXA_RACA_INDIGENA_MORTOS
Número de pacientes com COVID são da raça parda por município			Número de pacientes com COVID são da raça parda por município		NR_RACA_PARDA

Taxa de pacientes por COVID são da raça parda por município			Taxa de pacientes mortos por COVID são da raça parda por município	Os dados gerados nesse campo é composto da seguinte forma: NR_RACA_PARDA / POPULACAO * 1000	TAXA_RACA_PARDA
Número de pacientes mortos por COVID são da raça parda por município			Número de pacientes mortos por COVID são da raça parda por município		NR_RACA_PARDA_MORTOS
Taxa de pacientes mortos por COVID são da raça parda por município			Taxa de pacientes mortos por COVID são da raça parda por município	Os dados gerados nesse campo é composto da seguinte forma: NR_RACA_PARDA_MORTOS / POPULACAO * 1000	TAXA_RACA_PARDA_MORTOS
Número de pacientes com COVID são da raça preta por município			Número de pacientes com COVID são da raça preta por município		NR_RACA_PRETA
Taxa de pacientes com COVID são da raça preta por município			Taxa de pacientes com COVID são da raça preta por município	Os dados gerados nesse campo é composto da seguinte forma: NR_RACA_PRETA / POPULACAO * 1000	TAXA_RACA_PRETA
Número de pacientes mortos por COVID são da raça preta por município			Número de pacientes mortos por COVID são da raça preta por município		NR_RACA_PRETA_MORTOS
Taxa de pacientes mortos por COVID são da raça preta por município			Taxa de pacientes mortos por COVID são da raça preta por município	Os dados gerados nesse campo é composto da seguinte forma: NR_RACA_PRETA_MORTOS / POPULACAO * 1000	TAXA_RACA_PRETA_MORTOS
Número de pacientes com COVID do sexo feminino por município			Número de pacientes com COVID do sexo feminino por município		NR_SEXO_FEMININO
Taxa de pacientes com COVID do sexo feminino por município			Taxa de pacientes com COVID do sexo feminino por município	Os dados gerados nesse campo é composto da seguinte forma: R_SEXO_FEMININO / POPULACAO * 1000	TAXA_SEXO_FEMININO
Número de pacientes mortos com COVID do sexo feminino por município			Número de pacientes mortos com COVID do sexo feminino por município		NR_SEXO_FEMININO_MORTOS
Taxa de pacientes mortos com COVID do sexo feminino por município			Taxa de pacientes mortos com COVID do sexo feminino por município	Os dados gerados nesse campo é composto da seguinte forma: NR_SEXO_FEMININO_MORTOS / POPULACAO * 1000	TAXA_SEXO_FEMININO_MORTOS
Número de pacientes com COVID do sexo indefinido por município			Número de pacientes com COVID do sexo indefinido por município		NR_SEXO_INDEFINIDO
Taxa de pacientes com COVID do sexo indefinido por município			Taxa de pacientes com COVID do sexo indefinido por município	Os dados gerados nesse campo é composto da seguinte forma: NR_SEXO_INDEFINIDO / POPULACAO * 1000	TAXA_SEXO_INDEFINIDO
Número de pacientes mortos com COVID do sexo indefinido por município			Número de pacientes mortos com COVID do sexo indefinido por município		NR_SEXO_INDEFINIDO_MORTOS

Taxa de pacientes mortos com COVID do sexo indefinido por município			axa de pacientes mortos com COVID do sexo indefinido por município	Os dados gerados nesse campo é composto da seguinte forma: NR_SEXO_INDEFINIDO_MORTOS / POPULACAO * 1000	TAXA_SEXO_INDEFINIDO_MORTOS
Número de pacientes com COVID do sexo masculino por município			Número de pacientes com COVID do sexo masculino por município		NR_SEXO_MASCULINO
Taxa de pacientes com COVID do sexo masculino por município			Taxa de pacientes com COVID do sexo masculino por município	Os dados gerados nesse campo é composto da seguinte forma: NR_SEXO_MASCULINO / POPULACAO * 1000	TAXA_SEXO_MASCULINO
Número de pacientes mortos com COVID do sexo masculino por município			Número de pacientes mortos com COVID do sexo masculino por município		NR_SEXO_MASCULINO_MORTOS
Taxa de pacientes mortos com COVID do sexo masculino por município			Taxa de pacientes mortos com COVID do sexo masculino por município	Os dados gerados nesse campo é composto da seguinte forma: NR_SEXO_MASCULINO_MORTOS / POPULACAO * 1000	TAXA_SEXO_MASCULINO_MORTOS
Número de Leitos existentes privados de suporte ventilatório pulmonar – COVID			Número de Leitos existentes privados de suporte ventilatório pulmonar – COVID		NR_LEITOS_EXIST_PRIV_SPC
Taxa de Leitos existentes privados de suporte ventilatório pulmonar – COVID			Taxa de Leitos existentes privados de suporte ventilatório pulmonar – COVID	Os dados gerados nesse campo é composto da seguinte forma: NR_LEITOS_EXIST_PRIV_SPC / POPULACAO * 1000	TAXA_LEITOS_EXIST_PRIV_SPC
Número de Leitos existentes privados de Unidade de Cuidados Intermediários			Número de Leitos existentes privados de Unidade de Cuidados Intermediários		NR_LEITOS_EXIST_PRIV_UCI
Taxa de Leitos existentes privados de Unidade de Cuidados Intermediários			Taxa de Leitos existentes privados de Unidade de Cuidados Intermediários	Os dados gerados nesse campo é composto da seguinte forma: NR_LEITOS_EXIST_PRIV_UCI / POPULACAO * 1000	TAXA_LEITOS_EXIST_PRIV_UCI
Número de Leitos existentes privados de Unidade de Terapia Intensiva - UTI			Número de Leitos existentes privados de Unidade de Terapia Intensiva - UTI		NR_LEITOS_EXIST_PRIV_UTI
Taxa de Leitos existentes privados de Unidade de Terapia Intensiva - UTI			Taxa de Leitos existentes privados de Unidade de Terapia Intensiva - UTI	Os dados gerados nesse campo é composto da seguinte forma: NR_LEITOS_EXIST_PRIV_UTI / POPULACAO * 1000	TAXA_LEITOS_EXIST_PRIV_UTI
Número de Leitos existentes públicos de suporte ventilatório pulmonar – COVID			Número de Leitos existentes públicos de suporte ventilatório pulmonar – COVID		NR_LEITOS_EXIST_PUB_SPC

Taxa de Leitos existentes públicos de suporte ventilatório pulmonar – COVID			Taxa de Leitos existentes públicos de suporte ventilatório pulmonar – COVID	Os dados gerados nesse campo é composto da seguinte forma: NR_LEITOS_EXIST_PUB_SPC / POPULACAO * 1000	TAXA_LEITOS_EXIST_PUB_SPC
Número de Leitos existentes públicos de Unidade de Cuidados Intermediários			Número de Leitos existentes públicos de Unidade de Cuidados Intermediários		NR_LEITOS_EXIST_PUB_UCI
Taxa de Leitos existentes públicos de Unidade de Cuidados Intermediários			Taxa de Leitos existentes públicos de Unidade de Cuidados Intermediários	Os dados gerados nesse campo é composto da seguinte forma: NR_LEITOS_EXIST_PUB_UCI / POPULACAO * 1000	TAXA_LEITOS_EXIST_PUB_UCI
Número de Leitos existentes públicos de Unidade de Terapia Intensiva - UTI			Número de Leitos existentes públicos de Unidade de Terapia Intensiva - UTI		NR_LEITOS_EXIST_PUB_UTI
Taxa de Leitos existentes públicos de Unidade de Terapia Intensiva - UTI			Taxa de Leitos existentes públicos de Unidade de Terapia Intensiva - UTI	Os dados gerados nesse campo é composto da seguinte forma: NR_LEITOS_EXIST_PUB_UTI / POPULACAO * 1000	TAXA_LEITOS_EXIST_PUB_UTI
Número de leitos normais privados (não UTI, UCI e SPC)			Número de leitos normais privados (não UTI, UCI e SPC)		NR_LEITOS_NORMAL_EXIST_PRIV
Taxa de leitos normais privados (não UTI, UCI e SPC)			Taxa de leitos normais privados (não UTI, UCI e SPC)	Os dados gerados nesse campo é composto da seguinte forma: NR_LEITOS_NORMAL_EXIST_PRIV / POPULACAO * 1000	TAXA_LEITOS_NORMAL_EXIST_PRIV
Número de leitos normais públicos (não UTI, UCI e SPC)			Número de leitos normais públicos (não UTI, UCI e SPC)		NR_LEITOS_NORMAL_EXIST_PUB
Taxa de leitos normais públicos (não UTI, UCI e SPC)			Taxa de leitos normais públicos (não UTI, UCI e SPC)	Os dados gerados nesse campo é composto da seguinte forma: NR_LEITOS_NORMAL_EXIST_PUB / POPULACAO * 1000	TAXA_LEITOS_NORMAL_EXIST_PUB
Número de leitos normais privados, pagos pelo MS (não UTI, UCI e SPC)			Número de leitos privados, pagos pelo MS (não UTI, UCI e SPC)		NR_LEITOS_NORMAL_SUS_PRIV
Taxa de leitos normais privados, pagos pelo MS (não UTI, UCI e SPC)			Taxa de leitos normais privados, pagos pelo MS (não UTI, UCI e SPC)	Os dados gerados nesse campo é composto da seguinte forma: NR_LEITOS_NORMAL_SUS_PRIV / POPULACAO * 1000	TAXA_LEITOS_NORMAL_SUS_PRIV
Número de leitos privados pagos pelo MS de suporte ventilatório pulmonar – COVID			Número de leitos pagos pelo MS de suporte ventilatório pulmonar – COVID		NR_LEITOS_SUS_PRIV_SPC
Taxa de leitos privados pagos pelo MS de suporte ventilatório pulmonar – COVID			Taxa de leitos pagos pelo MS de suporte ventilatório pulmonar – COVID	Os dados gerados nesse campo é composto da seguinte forma: NR_LEITOS_SUS_PRIV_SPC / POPULACAO * 1000	TAXA_LEITOS_SUS_PRIV_SPC

Número de leitos privados pagos pelo MS Unidade de Cuidados Intermediários			Número de leitos privados pagos pelo MS Unidade de Cuidados Intermediários		NR_LEITOS_SUS_PRIV_UCI
Taxa de leitos privados pagos pelo MS Unidade de Cuidados Intermediários			Taxa de leitos privados pagos pelo MS Unidade de Cuidados Intermediários	Os dados gerados nesse campo é composto da seguinte forma: NR_LEITOS_SUS_PRIV_UCI / POPULACAO * 1000	TAXA_LEITOS_SUS_PRIV_UCI
Número de leitos privados pagos pelo MS, Unidade de Terapia Intensiva - UTI			Número de leitos privados pagos pelo MS, Unidade de Terapia Intensiva - UTI		NR_LEITOS_SUS_PRIV_UTI
Taxa de leitos privados pagos pelo MS, Unidade de Terapia Intensiva - UTI			Taxa de leitos privados pagos pelo MS, Unidade de Terapia Intensiva - UTI	Os dados gerados nesse campo é composto da seguinte forma: NR_LEITOS_SUS_PRIV_UTI / POPULACAO * 1000	TAXA_LEITOS_SUS_PRIV_UTI
Números do IDHM do município			Números do IDHM do município		IDHM
Área territorial do município			Área territorial do município		AREA_TERRITORIAL
Densidade Demográfica do município			Densidade Demográfica do município		DENSIDADE_DEMOGRAFICA
Índice de escolarização do município			Índice de escolarização do município		ESCOLARIZACAO
Índice de PIB e PER CAPITA do município			Índice de PIB e PER CAPITA do município		PIB_PER_CAPITA
IFDM do município			IFDM do município		IFDM
Índice de Educação do município			Índice de Educação do município		EDUCACAO
Índice de emprego e renda por município			Índice de emprego e renda por município		EMPREGO_RENDA
Índice de saúde do município			Índice de saúde do município		SAUDE
Número de médicos por município			Número de médicos por município		NR_MEDICOS
Taxa de número de médicos por município			Taxa de número de médicos por município	Os dados gerados nesse campo é composto da seguinte forma: NR_MEDICOS / POPULACAO * 1000	TAXA_MEDICOS
Número de enfermeiros por município			Número de enfermeiros por município		NR_ENFERMEIROS
Taxa de número de enfermeiros por município			Taxa de número de enfermeiros por município	Os dados gerados nesse campo é composto da seguinte forma: NR_ENFERMEIROS / POPULACAO * 1000	TAXA_ENFERMEIROS
Informação climática Köppen do município			Informação climática Köppen do município		KOPPEN