Universidade Estadual de Maringá
Departamento de Estatística
Programa de Pós-graduação em Bioestatística

THAÍS BARBOSA DE OLIVEIRA

MINERAÇÃO DE REGRAS DE ASSOCIAÇÃO EM DADOS DE COVID-19 NO BRASIL

Maringá – Paraná
2022

THAÍS BARBOSA DE OLIVEIRA

MINERAÇÃO DE REGRAS DE ASSOCIAÇÃO EM DADOS DE COVID-19 NO BRASIL

Dissertação apresentada ao Programa de Pós-graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá como requisito parcial para obtenção do título de mestre em Bioestatística.

Orientador: Prof. Dr. Josmar Mazucheli

Universidade Estadual de Maringá - UEM

Departamento de Estatística - DES

Programa de Pós-Graduação em Bioestatística

Maringá – Paraná

2022

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central - UEM, Maringá - PR, Brasil)

O48m

Oliveira, Thaís Barbosa de
Mineração de regras de associação em dados de COVID-19 no Brasil / Thaís Barbosa de Oliveira. -- Maringá, PR, 2023.
16 f.

Orientador: Prof. Dr. Josmar Mazucheli.
Dissertação (Mestrado) - Universidade Estadual de Maringá, Centro de Ciências Exatas, Departamento de Estatística, Programa de Pós-Graduação em Bioestatística, 2023.

1. Mineração de dados (Computação). 2. COVID-19. 3. Pandemia. I. Mazucheli, Josmar, orient. II. Universidade Estadual de Maringá. Centro de Ciências Exatas. Departamento de Estatística. Programa de Pós-Graduação em Bioestatística. III. Título.

CDD 23.ed. 519.5

Márcia Regina Paiva - CRB-9/1267

THAÍS BARBOSA DE OLIVEIRA

Mineração de regras de associação em dados de Covid-19 no Brasil

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

BANCA EXAMINADORA

Prof. Dr. Josmar Mazucheli
Universidade Estadual de Maringá – PBE/UEM

Willian Luís de Oliveira

Prof. Dr. Willian Luis de Oliveira
Universidade Estadual de Maringá – PBE/UEM

Profª. Dra. Elza Kimura Grimshaw
Universidade Estadual de Maringá – DFA/UEM

Maringá, 15 de dezembro de 2022.

Para minha mãe, Sandra, que foi membro de população ribeirinha no interior do Pará, não teve mãe, tampouco acesso à educação e à saúde durante sua infância e juventude. Apesar disso, é a pessoa mais doce e bondosa deste mundo, e ensinou-me a valorizar o conhecimento como um bem precioso. Eu quero sempre honrar o seu nome e a sua história.

Agradecimentos

A Deus que guiou os meus passos até a Universidade Estadual de Maringá (UEM). À minha família, sem a qual eu não teria chegado até aqui. À minha amiga-irmã, Caroliny Victória, que nunca deixou de me amparar e acolher durante este tempo.

Ao meu orientador, Dr.Josmar Mazucheli, — cujo entusiasmo pela ciência, integridade e inteligência eu admiro —, pela paciência e parceria na construção destes estudos.

Aos meus amigos, Karla Sacani, Maria Oliveira, Larissa Intrebartoli e Mateus Faccin. A nossa amizade foi essencial durante o mestrado e acredito que, nem nos meus maiores sonhos, eu imaginaria amigos melhores do que vocês.

Aos laços que formei no Programa de Residência da Escola Superior de Ciências da Saúde (ESCS), os quais compartilham do meu amor pela ciência e participaram deste percurso como meus companheiros de jornada profissional: Lorrany Rodrigues, Sara Meneses, Maurício Hirata e Wendy Fernandes.

Ao Departamento de Saúde da Família (DESF) do Ministério da Saúde, por me proporcionar um crescimento profissional extraordinário neste ano e apoiar a conclusão dos meus estudos.

Aos professores da banca, Dra.Elza Kimura e Dr.William Luis, por aceitarem o convite para avaliar esta dissertação e pelas contribuições que, certamente, a enriqueceram.

Por fim, agradeço ao povo brasileiro; à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); à Fundação de Apoio à Fiocruz (Fiotec); e ao Ministério da Saúde, por subsidiarem meus estudos até aqui. Espero que minha devolutiva, presente e futura, sempre faça jus a todo o investimento feito em minha formação.

Resumo

A Mineração de Regras de Associação (MRA) consiste em buscar padrões em conjuntos de dados por meio de algoritmos especializados. A literatura científica aponta que este método pode ser utilizado na identificação de padrões de comportamento de doenças. Considerando-se que a COVID-19 é uma emergência global em saúde pública, esta dissertação apresenta uma revisão integrativa sobre o método MRA aplicado a dados de casos da doença, além de aplicá-lo em dados reais advindos de plataforma pública e publicizável do Brasil, composto por grupos de indivíduos em idade produtiva e idosos com esquema vacinal completo e hospitalizados. O estudo de revisão apontou que a MRA é útil quanto à identificação de padrões de ocorrência da doença e seus determinantes, assim como apresenta potencial de auxílio no planejamento em saúde pública. Corroborando com essa premissa, quando aplicado aos dados do Brasil, o método demonstrou-se efetivo no reconhecimento de perfis epidemiológicos associados à COVID-19.

Palavras-chave: Mineração de dados; COVID-19; Pandemia.

Abstract

Association Rule Mining (MRA) consists of looking for patterns in datasets through specialized algorithms. The scientific literature indicates that this method can be used to identify disease behavior patterns. Considering that COVID-19 is a global public health emergency, this dissertation presents an integrative review of this method applied to data from cases of the disease, in addition to applying it to real data from a public and publishable platform in Brazil, composed of groups of individuals of working age and elderly with a complete vaccination schedule and hospitalized. The review study pointed out that ARM is useful when identifying patterns of disease occurrence and its determinants, as well as having the potential to aid public health planning. Corroborating this premise, when applied to data from Brazil, the method proved to be effective in recognizing epidemiological profiles associated with COVID-19.

Keywords: Data mining; COVID-19; Pandemic.

Lista de Abreviaturas e Siglas

Coronavirus disease 2019 (COVID-19)

Equipamento de Proteção Individual (EPI)

Fundação Oswaldo Cruz (Fiocruz)

Mineração de Regras de Associação (MRA)

Organização Mundial da Saúde (OMS)

Programa de Pós-graduação em Bioestatística (PBE)

Research, Society and Development (RSD)

Revista Eletrônica Acervo Saúde (REAS)

Universidade Estadual de Maringá (UEM)

Universidade de Brasília (UnB)

Unidade de Terapia Intensiva (UTI)

Sumário

1 Resumo Geral 10
1.1 Introdução 10
1.2 Justificativa 12
1.3 Objetivos 12
1.3.1 Objetivo Geral 12
1.3.2 Objetivos Específicos 12
1.4 Metodologia 13
1.5 Resultados 13
1.6 Conclusão 13
1.7 Produtos desta dissertação 13
1.8 Referências 15

Capítulo 1

Resumo Geral

1.1 Introdução

A COVID-19, causada pelo novo coronavírus (SARS-CoV-2), pode ser classificada como uma doença infectocontagiosa de emergência em saúde pública global, sendo considerada uma pandemia (CRUZ et al., 2021). A doença pode ser transmitida por meio de contato com secreções respiratórias, gotículas ou aerossóis expelidos pelo nariz ou boca de pessoas infectadas (REINHARDT, 2022). Os sintomas mais comuns são: febre, tosse, fadiga, dor de garganta e cefaleia, podendo evoluir para quadros mais graves, tais como desconforto respiratório, choque séptico e até mesmo ao óbito (PASCARELLA et al., 2020; SILVA et al., 2021). Os grupos populacionais mais vulneráveis à doença, ou seja, — mais suscetíveis ao desenvolvimento de um quadro clínico grave —, são idosos, portadores de doenças crônicas não transmissíveis e imunossuprimidos (LIU et al., 2020; MOHAMMED et al., 2020; NEUMANN-PODCZASKA et al., 2020).

A pandemia pode também dar origem a mutações genéticas dos agentes etiológicos das doenças ao longo do tempo, devido a sua interação com hospedeiros humanos. Segundo Mendes et al. (2022), essas variantes podem ser classificadas como de interesse ou de preocupação. Ainda de acordo com os mesmos autores, as variantes de interesse carecem de monitoramento contínuo e podem ser classificadas como causadoras de transmissão comunitária ou como responsáveis por múltiplos casos. As de preocupação, além de causarem esses efeitos, também são responsáveis por alterações no perfil epidemiológico da doença, aumento de transmissibilidade, mudança de quadro clínico e redução dos efeitos de medidas sociais de contenção da doença.

Até dezembro de 2021, a Organização Mundial de Saúde (OMS) descreveu oito variantes de interesse da COVID-19, como a Épsilon, que surgiu pela primeira vez por volta de junho

de 2020 nos EUA; a Zeta, detectada no Brasil em abril de 2020; a Eta e a Iota, detectadas pela primeira vez nos EUA em novembro de 2020; a Teta, identificada nas Filipinas e no Japão em fevereiro de 2021; a Kappa, detectada pela primeira vez na Índia em dezembro de 2021; a Lambda, detectada pela primeira vez no Peru em junho de 2021; e a Mu, identificada na Colômbia em agosto de 2021 (CASCELLA et al., 2022). Por sua vez, as variantes de preocupação da COVID-19, classificadas pela OMS foram: Alfa, relatada no Reino Unido em dezembro de 2020; Beta, detectada na África do Sul em outubro de 2020; Gama, identificada em dezembro de 2020 no Brasil; a Delta, inicialmente identificada em dezembro de 2020 na Índia; e a Ômicron, identificada pela primeira vez na África do Sul em novembro de 2021 (CASCELLA et al., 2022).

Em outubro de 2022, esta lista sofreu atualização e não houve listagem de variante de interesse em circulação, mas de variante de preocupação: a Ômicron e suas linhagens descendentes (BA.1, BA.2, BA.3, BA.4, BA.5), assim como suas formas recombinantes circulantes (BA.1, BA.2, XE) (OMS, 2022a).

Dentre as medidas mais comuns adotadas para a contenção da COVID-19 e suas variantes em escala global desde o início da pandemia em 2020, tem-se o isolamento social, a utilização de Equipamento de Proteção Individual (EPI), tais como máscara de proteção respiratória, e o reforço de lavagem de mãos e uso de álcool em gel. Como método de imunização passiva, conta-se com as vacinas, implementadas a partir do ano de 2021. As vacinas registradas e aprovadas para aplicação em seres humanos por laboratórios são: Pfizer/BioNTech, AstraZeneca/Oxford, Janssen, Moderna, Sinopharm, Sinovac, Bharat, Novavax, Casino, e Valneva (OMS, 2022b). No entanto, destaca-se que nem todas as vacinas listadas foram adotadas pelos países, uma vez que alguns países escolheram duas ou mais em detrimento de outras, ou tiveram acesso limitado por questões socioeconômicas (LOEMBÉ; NKENGASONG, 2021).

Todo esse processo de disseminação, contenção e desenvolvimento de medidas de imunização passiva de análise da distribuição da COVID-19 em populações do mundo, estabeleceu desafios para o trabalho com grandes bancos de dados epidemiológicos, em especial, a respeito da melhor escolha para tratá-los (ALAMO et al., 2020). Uma das técnicas inovadoras de análise de dados observacionais é a Mineração de Regras de Associação (MRA). A MRA permite a identificação de padrões de ocorrência de atributos de variáveis que implicam na presença de outros atributos em um conjunto de dados, além de ter a vantagem de baixa complexidade interpretativa e a não exigência de critérios de distribuição em função do tempo (KATRAGADDA et al., 2021). Dessa forma, o método de MRA pode ser usado para fornecer padrões populacionais de distribuição de doenças que poderiam passar despercebidos com a utilização de outras técnicas de tratamento de dados (DOMADIYA; RAO, 2021).

A partir disso, o presente trabalho aplicou este método em dados de pacientes hospitaliza-

dos por COVID-19 no Brasil, assim como realizou uma revisão integrativa sobre a temática.

1.2 Justificativa

A realização de estudos com MRA aplicado à COVID-19 é urgente e apresenta uma dimensão de inovação e pioneirismo no Brasil. Soma-se a isso, o compromisso do auxílio científico de programas educacionais para com o bem-estar da população nos espaços geográficos nos quais estão inseridos. O Programa de Pós-graduação em Bioestatística (PBE) da Universidade Estadual de Maringá (UEM) destaca-se no cenário de pesquisa a respeito de conjunturas epidemiológicas por ser, até o presente momento, o único no país e na América Latina com o título "Bioestatística". O programa também insere alunos formandos e formados em cenários de assistência e gestão de serviços de saúde. Como a eticidade de uma pesquisa implica, não apenas, mas também, assegurar benefícios às comunidades envolvidas em sua pesquisa, a realização deste estudo configura-se como uma postura ética a ser assumida pelo PBE/UEM mediante o histórico epidemiológico da COVID-19 no Brasil.

1.3 Objetivos

1.3.1 Objetivo Geral

Apresentar uma revisão integrativa a respeito da Mineração de Regras de Associação em dados de casos de COVID-19, assim como aplicá-la em dados reais de pacientes hospitalizados devido à doença no Brasil.

1.3.2 Objetivos Específicos

- Investigar a produção científica a respeito da utilização de MRA em dados de casos de COVID-19 por meio de revisão integrativa;
- Identificar padrões de condições clínicas a partir de dados de pacientes em idade produtiva, com esquema vacinal completo contra a doença, hospitalizados devido à sua forma grave e com apresentação de status de admissão em Unidade de Terapia Intensiva (UTI) no Brasil;
- Identificar padrões e relações entre sintomas, comorbidades, gênero, internação em Unidade de Terapia Intensiva (UTI) e estado de sobrevida de idosos, totalmente vacinados e hospitalizados devido à COVID-19 no Brasil.

1.4 Metodologia

Para a investigação da literatura, utilizou-se da revisão integrativa. Para o estudo de padrões clínicos de sintomatologias e demais fatores associados à hospitalização por COVID-19 de pacientes em idade produtiva e idosos, utilizou-se de dados secundários de base de dados pública e publicizável com aplicação do método da MRA.

1.5 Resultados

Verificou-se que a MRA é útil para o estudo de padrões de distribuição de viroses emergentes e seus determinantes, tal qual a COVID-19. Quando aplicado a dados reais, o método mostra-se superior à aplicação de medidas estatísticas descritivas simples, pois aponta associações entre variáveis e apresenta potencial de utilização quando em situações de crise global mediante o surgimento de patologias com comportamento de distribuição espacial e temporal desconhecido.

1.6 Conclusão

A partir dos estudos realizados, observou-se que a MRA pode fornecer subsídios para os processos de gerenciamento epidemiológico e planejamento em saúde, assim como para o fortalecimento de políticas públicas de saúde já existentes.

1.7 Produtos desta dissertação

Os artigos com aplicação de MRA em dados reais de COVID-19 foram publicados em revistas científicas. Para isso, pesquisou-se por periódicos que apresentavam publicações recentes sobre COVID-19, dentre os quais, encontrou-se a Revista Eletrônica Acervo Saúde (REAS) e a Research Society and Development (RSD). O artigo relativo a dados de pacientes em idade produtiva foi publicado na REAS e, o de pacientes idosos, na RSD. No que se refere ao artigo que trata da revisão integrativa sobre o método MRA, este será submetido em revista científica.

Os artigos foram idealizados pelo professor Josmar Mazucheli e a mestranda Thaís Barbosa de Oliveira. A elaboração dos trabalhos contou com o apoio de pesquisadores de três instituições externas à UEM: Lorrany Santos Rodrigues e Caroliny Victoria dos Santos Silva (Universidade de Brasília - UnB), Maurício Yukio Hirata (Fundação Oswaldo Cruz - Fiocruz) e Wendy Rayanne Fernandes dos Santos (Ministério da Saúde, da Secretaria de Atenção Primária à Saúde).

Segundo a Lei nº 9.610 de 19 de fevereiro de 1998, de direitos autorais, a reprodução total dos artigos submetidos e publicados só deve ocorrer por meio das revistas que os detêm. Diante disso, os artigos publicados podem ser acessados e identificados por meio das seguintes referências:

- OLIVEIRA, Thaís Barbosa de et al. Clinical condition patterns of fully vaccinated brazilian patients of working age hospitalized due to severe Covid-19 according to ICU admission status. Revista Eletrônica Acervo Saúde, v. 15, n. 10, p. e10997-e10997, 2022. DOI: https://doi.org/10.25248/reas.e10997.2022

- OLIVEIRA, Thaís Barbosa et al. Description of health patterns of fully vaccinated older adults hospitalized due to COVID-19 in Brazil through association rules. Research, Society and Development, v. 11, n. 16, p. e36111637666-e36111637666, 2022.DOI: https://doi.org/10.33448/rsd-v11i16.37666

1.8 Referências

AGUIAR, Ícaro Cardoso et al. Aspectos físicos, mentais e sociais dos profissionais da área da saúde no enfrentamento à pandemia da COVID-19: uma revisão bibliográfica da literatura. Brazilian Journal of Health Review, v. 5, n. 1, p. 693-701, 2022.

ALAMO, Teodoro et al. Covid-19: Open-data resources for monitoring, modeling, and forecasting the epidemic. Electronics, v. 9, n. 5, p. 827, 2020.

BAI, Francesca et al. Female gender is associated with long COVID syndrome: a prospective cohort study. Clinical Microbiology and Infection, v. 28, n. 4, p. 611. e9-611. e16, 2022.

BEGA, Maria Tarcisa Silva; SOUZA, M. N. Pandemia e efeito-território: a desigualdade social como catalisadora da Covid-19. Revista Brasileira de Sociologia, v. 9, n. 21, p. 25-54, 2021.

BONIFÁCIO, Lívia P. et al. Long-Term Symptoms among COVID-19 Survivors in Prospective Cohort Study, Brazil. Emerging infectious diseases, v. 28, n. 3, p. 730, 2022.

BHUIYAN, Mejbah Uddin et al. Epidemiology of COVID-19 infection in young children under five years: A systematic review and meta-analysis. Vaccine, v. 39, n. 4, p. 667-677, 2021.

CASCELLA, Marco et al. Features, evaluation, and treatment of coronavirus (COVID-19). StatPearls Publishing, 2022.

CRUZ, M. Palacios et al. COVID-19, a worldwide public health emergency. Revista Clínica Española (English Edition), v. 221, n. 1, p. 55-61, 2021.

DOMADIYA, Nikunj; RAO, Udai Pratap. Privacy preserving association rule mining on distributed healthcare data: Covid-19 and breast cancer case study. SN Computer Science, v. 2, n. 6, p. 1-9, 2021.

FARUQUE, Fazlay S. Geospatial Technology for Human Well-Being and Health: An Overview. Geospatial Technology for Human Well-Being and Health, 2022.

FUCHS, Anna et al. The Impact of COVID-19 Pandemic on Female Sexual Function. International Journal of Environmental Research and Public Health, v. 19, n. 6, p. 3349, 2022.

GARCIA, Leila Posenato; DUARTE, Elisete. Intervenções não farmacológicas para o enfrentamento à epidemia da COVID-19 no Brasil. Epidemiologia e Serviços de Saúde, v. 29, 2020.

KATRAGADDA, Satya et al. Association mining based approach to analyze COVID-19 response and case growth in the United States. Scientific Reports, v. 11, n. 1, p. 1-12, 2021.

LIU, Hong et al. Comorbid chronic diseases are strongly correlated with disease severity among COVID-19 patients: a systematic review and meta-analysis. Aging and disease, v. 11, n. 3, p. 668, 2020.

LOEMBÉ, Marguerite Massinga; NKENGASONG, John N. COVID-19 vaccine access in Africa: Global distribution, vaccine platforms, and challenges ahead. Immunity, v. 54, n. 7, p. 1353-1362, 2021.

MENDES, João et al. Atualização das recomendações da Sociedade Portuguesa de Cuidados Intensivos e do Grupo de Infeção e Sépsis para a abordagem da COVID-19 em Medicina Intensiva. Revista Brasileira de Terapia Intensiva, v. 33, p. 487-536, 2022.

MOHAMMED, Ali Haider et al. The risk and impact of COVID-19 pandemic on immunosuppressed patients: Cancer, HIV, and solid organ transplant recipients. AIDS Rev, v. 22, n. 3, p. 151-7, 2020.

NEUMANN-PODCZASKA, Agnieszka et al. COVID 19-clinical picture in the elderly population: a qualitative systematic review. Aging and disease, v. 11, n. 4, p. 988, 2020.

ORGANIZAÇÃO MUNDIAL DA SAÚDE – OMS. Tracking SARS-CoV-2 variants. World Health Organization, 2022a. Disponível em: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/. Acesso em: 16, outubro, 2022.

ORGANIZAÇÃO MUNDIAL DA SAÚDE – OMS. Perguntas frequentes: vacinas contra a COVID-19. World Health Organization, 2022b. Disponível em: https://www.paho.org/pt/vacinas-contra-covid-19/perguntas-frequentes-vacinas-contra-covid-19. Acesso em: 16, outubro, 2022.

PASCARELLA, Giuseppe et al. COVID-19 diagnosis and management: a comprehensive review. Journal of internal medicine, v. 288, n. 2, p. 192-206, 2020.

PÉREZ-RIERAA, Andrés Ricardo et al. COVID-19 e influenza: implicações no contexto da saúde pública. Journal of Human Growth and Development, v. 32, n. 1, p. 6, 2022.

REINHARDT, Érica Lui. Transmissão da COVID-19: um breve reexame das vias de transmissão por gotículas e aerossóis. Revista Brasileira de Saúde Ocupacional, v. 47, 2022.

SILVA, Maria Isabel Caetano et al. Diagnósticos de enfermagem em casos de COVID-19 com evolução clínica para sepse. Research, Society and Development, v. 10, n. 1, p. e17410111232-e17410111232, 2021.

SILVA, Severino Jefferson Ribeiro; PENA, Lindomar. Collapse of the public health system and the emergence of new variants during the second wave of the COVID-19 pandemic in Brazil. One Health, v. 13, p. 100287, 2021.

SOBRAL, Marcos Felipe Falcão et al. A retrospective cohort study of 238,000 COVID-19 hospitalizations and deaths in Brazil. Scientific Reports, v. 12, n. 1, p. 1-11, 2022.

ZEISER, Felipe André et al. First and second COVID-19 waves in Brazil: A cross-sectional study of patients' characteristics related to hospitalization and in-hospital mortality. The Lancet Regional Health-Americas, v. 6, p. 100107, 2022.