



Pib-CnPq

AVALIAÇÃO DOS FATORES DE RISCO E SINTOMAS DE PACIENTES INTERNADOS COM SÍNDROME RESPIRATÓRIA AGUDA GRAVE NAS DIFERENTES UNIDADES FEDERATIVAS DO BRASIL



BIO-IA

Autores: Cristian Pigato, Marcos Vinicius Rosseto, Scheila de Avila e Silva (Orientador(a))

INTRODUÇÃO / OBJETIVO

A mais recente pandemia de vírus CoV iniciou-se em março de 2019, em Wuhan, na China. No início de 2020, o Centro de Controle de Doenças e Prevenção Chines detectou em pessoas hospitalizadas o novo CoV o que levou à nomenclatura do vírus como CoV-19. A rápida disseminação e contágio do vírus levou à OMS a decretar, em 11 de março de 2020 a pandemia COVID-19. Em 21 de junho de 2021, já registravam-se 178.118.587 casos confirmados com 3.864.180 mortes. A utilização de mineração de dados mostra-se relevante para realizar-se uma visão exploratória dos dados disponíveis em repositórios públicos, a fim de trazer à tona padrões e informações que possam ser relevantes à cerca da COVID-19 desta forma possibilitando que futuros pesquisadores possam utilizar essas informações em suas pesquisas. Assim, através de algoritmos que retornam padrões e tendências nos dados analisados é possível realizar-se inferências que são analisadas e validadas por especialistas de domínio. Nesta pesquisa, procurou-se através de mineração de dados no DATA SUS via o software R, realizou-se revisão de artigos para elucidação de cronologia, alterações moleculares do vírus CoV, fisiopatologia e sobre inteligência artificial. Na mineração de dados, apontou-se para correlação de doenças prevalentes (diabetes, cardiopatias, pneumopatias...) com a gravidade de casos com a SARS-CoV-2 em pacientes internados em hospitais de unidades federativas do Brasil com o objetivo de encontrar-se correções entre as doenças prevalentes, taxa de mortalidade em relação a essas doenças e padrões encontrados em unidades federativas do Brasil bem como o comportamento por Regiões Brasileiras.

MATERIAL E MÉTODOS

Os dados foram analisados por meio de estatística descritiva e algoritmos de mineração de dados. Os softwares a serem utilizados para esse fim foram softwares de programação estatística como R, Python. A análise de regressão foi utilizada para prever resultados para as variáveis dependentes, através de variáveis preditores, explicativas ou independentes. Para que com isso, seja possível avaliar o impacto de cada variável independente sobre o desfecho. Para realizar a regressão, foi necessário construir um modelo preditivo para os dados e, posteriormente, utilizar esse modelo para prever os valores das variáveis dependentes a partir das variáveis independentes. Quando existe apenas uma variável independente, definimos a técnica como regressão simples, em contrapartida, na regressão múltipla, mais de uma variável independente é utilizada para prever o desfecho. Os algoritmos de classificação e regressão, foram aplicados utilizando a técnica de validação cruzada (*k-fold-cross-validation*, *k-FCV*). Posteriormente, foram realizadas as etapas de treinamento e validação, as quais foram repetidas *k* vezes, sendo utilizado para o treinamento *k-1* arquivos. Para cada interação, o arquivo de validação possui um *k* diferente.

Os dados atualizados sobre COVID-19 no Brasil foram obtidos do repositório de dados público OpenDataSUS (<https://opendatasus.saude.gov.br/dataset/bd-srag-2020>). Adquiridos em formato CSV e para facilitar-se a manipulação foi criado um banco de dados relacional. Foi escolhido o gerenciador de bancos de dados relacionais PostgreSQL, pois possui robustez nos dados e possibilita utilização através de várias linguagens de programação, como, Python, R, C# etc.

Para a construção do banco de dados foi necessário realizar-se o mapeamento das variáveis, uma vez que elas estão disponíveis em uma planilha csv e a descrição dos campos encontra-se em um dicionário dados, o qual pode ser obtido junto com os dados. De modo a automatizar a importação dos dados, tendo em vista que esses são atualizados periodicamente, foi criado um importador de dados a fim de automatizar o processo de obtenção e atualização dos dados. As doenças que foram associadas com o desfecho mortalidade foram: asma, cardiopatias, diabetes, hematológicas, hepáticas, neurológicas, obesidade, pneumopatias, renais.

Além disso, realizou-se revisão bibliográfica com enfoque na cronologia, diferença das moléculas e avanço do vírus CoV pelo mundo.

RESULTADOS

(Nota: Utilizou-se a maioria das tabelas com o padrão da região sudeste para melhor comparação entre os quadros de doenças.)

As doenças mais prevalentes trouxeram resultados distintos e com diferença de comportamentos. Encontrou-se como resultados:

Asma, Doenças Hematológicas e Hepáticas: As unidades federativas tiveram dados semelhante ao demonstrado na tabela abaixo (Tabela 1), com um padrão de baixa mortalidade em asmáticos, doenças hematológicas e hepáticas e o mantimento da porcentagem de óbitos baixo e estável durante os anos de 2020, 2021 e 2022, com variação de 1-2% dentre esse percentual durante os 3 anos, mantendo-se um padrão semelhante dentre os 3 perfis de doenças.

Região-Estado	2020		2021		2022	
	Porcentagem	Total de Casos	Porcentagem	Total de Casos	Porcentagem	Total de Casos
Sul		n 13930		n 22845		n 3334
Paraná	05,13%	5183	05,05%	10816	04,54%	1254
Rio Grande do Sul	06,58%	6047	07,16%	9212	06,64%	1459
Santa Catarina	04,48%	2700	04,96%	2817	04,66%	621

Tabela 1: Dados de mortalidade em portadores de asma na região sul

Cardiopatias: Como esperava-se doenças crônicas importantes como cardiopatias na pandemia seriam a maior causa de mortalidade as pessoas que seriam infectadas pelo Covid, como visto na tabela abaixo que mostra a maior porcentagem de mortalidade dentre todas as doenças em Espírito Santo (tabela 2). Tivera uma discrepância dentre as unidades federativas em relação a porcentagem de mortalidade de 25,58% no Acre em 2022 a 97,46% em 2021 no Estado do Espírito do Santo. Contudo, uma variação de 50-70% de mortalidade na maioria dos estados em casos de SARS-CoV com cardiopatia associada. Com o avanço da vacinação os casos de internação com cardiopatia diminuíram drasticamente em 2022.

RESULTADOS

	2020		2021		2022	
	Porcentagem	Total de Casos	Porcentagem	Total de Casos	Porcentagem	Total de Casos
Sudeste		n 55839		n 72831		n 9721
Espírito Santo	89,23%	1849	97,46%	1504	85,96	114
Minas Gerais	59,67%	9954	59,98%	16940	59,31%	1927
Rio de Janeiro	84,11%	11203	86,29%	10556	85,83%	1241
São Paulo	71,44%	32833	71,19%	43831	70,95%	6439

Tabela 2: Tabela de cardiopatias na região sudeste

Diabetes: Na diabetes tivera um padrão importante e que mostrou a maior influência da vacinação dentre as doenças. Na maioria das regiões e estados teve-se um decréscimo na mortalidade mesmo com a diminuição de casos internados com o COVID-19. A tabela 3 fora mantido os estados do sudeste para notar-se a porcentagem alta de mortalidade semelhante às cardiopatias, contudo um padrão positivo e com diminuição da mortalidade com o avanço da vacinação dentre as regiões e unidades federativas e também a diminuição da internação no ano de 2022 com as políticas de vacinação mais solidificadas.

	2020		2021		2022	
	Porcentagem	Total de Casos	Porcentagem	Total de Casos	Porcentagem	Total de Casos
Sudeste		n 50852		n 59382		n 7885
Espírito Santo	70,28%	1336	88,16%	769	54,65%	86
Minas Gerais	42,94%	9652	44,27%	16484	38,68%	1830
Rio de Janeiro	72,20%	9246	68,26%	1336	58,64%	162
São Paulo	55,82%	30618	54,73%	40793	48,07%	5807

Tabela 3: Tabela diabetes na região sudeste

Obesidade: Teve-se um padrão diferente de mortalidade com a obesidade. Mesmo com a diminuição dos casos após 2021, criou-se um padrão de mortalidade semelhante a uma Curva de Gauss com concavidade para baixo (ou seja, teve-se um padrão semelhante em 2020 e 2022, mas em 2021 teve-se um aumento da mortalidade) como visto na Tabela 4.

	2020		2021		2022	
	Porcentagem	Total de Casos	porcentagem	Total de Casos	Porcentagem	Total de Casos
Sudeste		n 40061		n 55565		n 6942
Espírito Santo	23,36%	779	69,78%	364	10,52%	57
Minas Gerais	12,83%	9176	19,57%	15299	09,43%	1632
Rio de Janeiro	23,76%	5470	35,39%	5339	15,38%	520
São Paulo	14,77%	24636	27,30%	34563	10,39%	4733

Tabela 4: Tabela Obesidade na região sudeste

Pneumopatias e Doenças Renais: As pneumopatias em grande parte dos estados e regiões tivera uma curva contrária a de obesidade e semelhante às doenças renais com uma curva de Gauss com concavidade para cima em número de mortalidades como visto na tabela 5 (ou seja, com um padrão em 2020, que decaiu em 2021 e voltou a crescer em 2022).

	2020		2021		2022	
	Porcentagem	Total de Casos	Porcentagem	Total de Casos	Porcentagem	Total de Casos
Sudeste		N: 40768		N: 52784		N: 7321
Espírito Santo	28,90%	858	44,48%	254	33,82%	68
Minas Gerais	12,48%	9308	09,86%	15102	19,14%	1682
Rio de Janeiro	20,08%	5363	14,83%	4470	24,47%	568
São Paulo	12,45%	25239	08,86%	32958	17,62%	5003

Tabela 5: Tabela Pneumopatias na região sudeste

Doenças Neurológicas: Apresentou-se um padrão de mortalidade diferente dos demais, mudando em cada unidade federativa e com padrões decrescentes, crescentes e curvas de Gauss com concavidades diferentes, não demonstrando qualidade dos dados apresentados.

CONSIDERAÇÕES FINAIS

Demonstrou-se no estudo alguns fatores importantes. O primeiro fora a importância da vacinação e seu avanço a partir de 2021 para a diminuição de internações por SARS-COV. Mostrou-se pela taxa de mortalidade, que a maioria das doenças comportaram-se, certa estabilidade na porcentagem de mortalidade em doenças hematológicas e hepáticas com porcentagem abaixo de 10%, Cardiopatias e Diabetes com porcentagens altas de mortalidade e diminuição dos casos totais de internados, obesidade com diminuição de internados com PCR positivo para CoV, pneumopatias e doenças renais criando-se uma hipótese de seletividade dos casos mais graves após a vacinação e que junto a internação provou-se taxa de mortalidade maior que no ano de 2020 e doenças neurológicas que teve um padrão anômalo e pouco específico.

Um ponto relevante a comentar-se é o sistema de saúde sobrecarregado e a subnotificação de doenças através da informatização e CID-11 na hora da informatização. Dessa forma, acredita-se que esses dados do DATASUS são apenas uma amostra da real totalidade de casos com comorbidades associadas e servem para demonstrar a gravidade que cada doença teve associada ao CoV e sua respectiva mortalidade. Outro ponto que discutiu-se fora o uso do CID-11 sobre SARS-COV 2 para angariar recursos as saúde pública e centros de alta complexidade devido o alto déficit orçamentário que os hospitais sofrem com a desatualização da tabela de pagamento a procedimentos do SUS.

REFERÊNCIAS BIBLIOGRÁFICAS

- Kirtipal, N., Bharadwaj, S., & Kang, S. G. (2020). From SARS to SARS-CoV-2, insights on structure, pathogenicity and immunity aspects of pandemic human coronaviruses. *Infection, Genetics and Evolution*, 104502. doi:10.1016/j.meegid.2020.104502
- Muralidar, S., Ambi, S. V., Sekaran, S., & Krishnan, U. M. (2020). The emergence of COVID-19 as a global pandemic: Understanding the epidemiology, immune response and potential therapeutic targets of SARS-CoV-2. *Biochimie*.
- SILVA, Leandro Augusto; PERES, Sarajane Marques; BOSCARIOLI, Clodis. *Introdução a Mineração de dados com aplicação em R*. 1ª Edição. Rio de Janeiro: Elsevier, 2016
- PubMed (nih.gov)
- (<https://opendatasus.saude.gov.br/dataset/bd-srag-2020>).

APOIO: CnPq