UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE ESTATÍSTICA

Alexander Bandeira Lira

UMA ANÁLISE ESTATÍSTICA DA MORTALIDADE POR COVID-19 NO MUNDO

Recife
2024

Alexander Bandeira Lira

Uma análise estatística da mortalidade por COVID-19 no mundo

Trabalho de conclusão de curso apresentado ao Curso de Bacharelado em Estatística da Universidade Federal de Pernambuco como requisito parcial para obtenção do título de Bacharel em Estatística.

Orientador: Prof. Dr. Francisco Cribari Neto

Recife-PE
2024

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Lira, Alexander Bandeira.
Uma análise estatística da mortalidade por COVID-19 no mundo / Alexander Bandeira Lira. - Recife, 2024.
20 : il., tab.

Orientador(a): Francisco Cribari Neto
Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Ciências Exatas e da Natureza, Estatística - Bacharelado, 2024.
Inclui referências.

1. Análise exploratória de dados. 2. COVID-19. 3. Distribuição beta. 4. Regressão beta. I. Neto, Francisco Cribari. (Orientação). II. Título.

310 CDD (22.ed.)

Alexander Bandeira Lira

UMA ANÁLISE ESTATÍSTICA DA MORTALIDADE POR COVID-19 NO MUNDO

Trabalho de conclusão de curso apresentado ao Curso de Bacharelado em Estatística da Universidade Federal de Pernambuco como requisito parcial para obtenção do título de Bacharel em Estatística.

Aprovado em: 08/03/2024

BANCA EXAMINADORA

_____

Prof. Dr. Francisco Cribari Neto (Orientador)

Universidade Federal de Pernambuco

_____

Profa. Dra. Fernanda De Bastiani (Examinador Interno)

Universidade Federal de Pernambuco

_____

Profa. Dra. Maria Cristina Falcão Raposo (Examinador Interno)

Universidade Federal de Pernambuco

Agradecimentos

Gostaria de expressar minha profunda gratidão a todas as pessoas e entidades que contribuíram significativamente para a realização deste trabalho e para minha jornada acadêmica. Sem dúvida, essa conquista não teria sido possível sem o apoio generoso e as bênçãos de Deus e o suporte de muitas pessoas especiais.

Em primeiro lugar, expresso minha mais sincera gratidão a Deus, cuja graça e orientação estiveram presentes em todos os momentos desta jornada. A Ele dedico este trabalho e agradeço por ser minha luz e meu guia em cada passo dado.

Como membro consagrado da comunidade católica Shalom, expresso minha profunda gratidão por ser um lugar de encontro com Deus e comigo mesmo. O apoio espiritual e a comunhão compartilhada foram fontes de força e inspiração ao longo deste caminho acadêmico e pessoal.

Aos professores que foram fontes de conhecimento, orientação e inspiração, expresso minha sincera gratidão. Em especial, desejo expressar minha gratidão ao meu orientador, Dr. Francisco Cribari Neto, cuja orientação sábia e incentivo foram fundamentais para o desenvolvimento deste trabalho. À professora Dra. Fernanda De Bastiani, mesmo que ela desconheça profundamente a importância do seu apoio em minha jornada, sou profundamente grato. À professora Dra. Caliteia Santana de Sousa, que, em meio a dificuldades, ajudou-me a permanecer de pé e seguir no meu desenvolvimento acadêmico e pessoal.

À minha mãe, Irene Bandeira, verdadeira guerreira na minha criação e de meus irmãos, expresso minha eterna gratidão. Seu amor incondicional e sacrifício foram o alicerce que me sustentou em todos os momentos, e sou profundamente grato por sua presença constante em minha vida. Também expresso minha gratidão a minha avó Ines Do Nascimento, que sempre foi uma grande mãe para mim, e a minha noiva, Carollina Arêde, cujo apoio e amor foram essenciais ao longo deste tempo.

Aos amigos José Helio, Michell Ferreira, Barbara Souza, Rondinelly, João Vitor e Renata Kelly, minha profunda gratidão. Suas palavras de encorajamento, apoio e amizade foram um grande conforto nos momentos de desafio e um lembrete constante do valor da amizade e do apoio mútuo.

Por último, agradeço a mim mesmo por minha determinação, perseverança e por nunca desistir, mesmo diante dos desafios. Esta conquista é uma celebração de todos os esforços investidos e uma lembrança do poder da fé, do amor e do apoio daqueles que estiveram ao meu lado.

A todos que contribuíram de alguma forma, minha mais profunda gratidão. Que possamos continuar compartilhando e crescendo juntos em nossas futuras jornadas.

Resumo

Este trabalho visa analisar as taxas de mortalidade por COVID-19 em 144 países durante os anos de 2021 e 2022. A análise combina análise descritiva, correlacional e modelagem estatística para identificar os preditores capazes de explicar as diferenças nas taxas de mortalidade entre os países. Os resultados revelam heterogeneidade apreciável nas taxas de mortalidade, com algumas nações enfrentando taxas mais elevadas do que outras. Identificamos correlações entre a taxa de mortalidade e variáveis demográficas, socioeconômicas e de saúde, destacando a complexidade dos fatores que influenciam a dinâmica da pandemia. A modelagem de regressão beta utilizada mostra que a estrutura etária da população, o nível socioeconômico e a composição religiosa podem desempenhar papéis importantes na variação da mortalidade entre os países. Este estudo conduz a conclusões capazes de subsidiar políticas de saúde pública e destaca a importância de abordagens integradas na compreensão da pandemia de COVID-19. Adicionalmente, as conclusões do presente estudo fornecem informações valiosas para o delineamento de políticas de saúde pública e estratégias de intervenção, contribuindo para a compreensão do desenrolar de pandemias e ressaltando a importância de abordagens integradas e multidisciplinares na compreensão e enfrentamento de desafios de saúde pública de escala global.

Palavras-chave: Análise exploratória de dados, COVID-19, Distribuição beta, Regressão beta

Abstract

This study aims to analyze COVID-19 mortality rates in 144 countries during the years 2021 and 2022. The analysis combines descriptive, correlational, and statistical modeling to identify predictors capable of explaining differences in mortality rates among countries. The results reveal considerable heterogeneity in mortality rates, with some nations facing higher rates than others. We identify correlations between mortality rate and demographic, socioeconomic, and health variables, highlighting the complexity of factors influencing the pandemic dynamics. The beta regression modeling used shows that population age structure, socioeconomic level, and religious composition may play important roles in the variation of mortality among countries. This study leads to conclusions capable of informing public health policies and emphasizes the importance of integrated approaches in understanding the COVID-19 pandemic. In addition, the conclusions of this study provide valuable information for the design of public health policies and intervention strategies, contributing to the understanding of the course of pandemics and highlighting the importance of integrated and multidisciplinary approaches in understanding and tackling public health challenges on a global scale.

Keywords: Beta distribution, Beta regression, COVID-19, Exploratory data analysis

Lista de ilustrações

Figura 1 – Histograma e boxplot da taxa de mortalidade por COVID-19 dos 144 países em 2021/22 11
Figura 2 – Matriz de correlação da taxa de mortalidade por COVID-19 e as demais variáveis consideradas em nosso estudo 15
Figura 3 – Gráfico residual semi-normal com envelopes simulados. 17

Lista de tabelas

Tabela 1 – Variáveis, descrição e período 6
Tabela 2 – Variáveis e URLs (fonte) 7
Tabela 3 – Estatísticas descritivas da taxa de mortalidade por COVID-19 dos 144 países em 2021/22 10
Tabela 4 – Países com as maiores e menores taxas de mortalidade por COVID-19 dos 144 países em 2021/22 12

Sumário

1 INTRODUÇÃO 1

2 OBJETIVOS 2
2.1 Objetivo geral 2
2.2 Objetivos específicos 2

3 DADOS E MÉTODO 3
3.1 Dados e área de estudo 3
3.2 Modelagem de regressão 5

4 PROGRAMA COMPUTACIONAL 9

5 RESULTADOS E DISCUSSÃO 10
5.1 Análise descritiva dos dados 10
5.2 Modelagem de regressão 14

6 CONCLUSÃO 19

REFERÊNCIAS 20

1 Introdução

COVID-19 é uma doença infecciosa causada pelo vírus SARS-CoV-2, que foi identificada pela primeira vez na China em dezembro de 2019. O vírus se espalha facilmente pelo ar e pode causar uma variedade de sintomas, incluindo febre, tosse, falta de ar e fadiga. Em casos graves, a COVID-19 pode levar à morte Wu et al. (2020). A doença se espalhou para todos os continentes e já infectou mais de 700 milhões de pessoas em todo o mundo. O número de mortos pela COVID-19 é estimado em cerca de 7 milhões Terra Notícias (2023), sendo que mais de 700 mil dessas mortes ocorreram no Brasil Ministério da Saúde (2023). A pandemia também causou uma série de impactos econômicos e sociais, incluindo a paralisação de atividades econômicas, a perda de empregos e o aumento da desigualdade.

Em maio de 2023, o Comitê de Emergência da OMS se reuniu e recomendou que a entidade declare o fim da crise do coronavírus como uma emergência de saúde pública de interesse internacional, seu nível mais elevado de alerta e que se encontrava em vigor desde 30 de janeiro de 2020. Em consequência, a OMS encerrou o status de emergência global para COVID-19, mais de três anos após sua declaração original. Agência Brasil (2023)

O esforço de vacinação contra a COVID-19 no mundo foi um dos maiores e mais bem-sucedidos da história. Em menos de dois anos, mais de 11 bilhões de doses de vacinas contra a COVID-19 foram administradas em todo o mundo, cobrindo mais de 60% da população mundial. O esforço de vacinação contra a COVID-19 foi liderado pela OMS, que trabalhou em estreita colaboração com governos, fabricantes de vacinas e organizações não governamentais Mathieu et al. (2021).

O COVAX, iniciativa global liderada pela OMS, Gavi e CEPI, foi fundamental para garantir o acesso equitativo às vacinas contra a COVID-19. Através dos mecanismos COVAX AMC e Autofinanciamento, o programa já entregou mais de 1 bilhão de doses para 140 países, enfrentando desafios como a insuficiência de recursos, desigualdade na distribuição e hesitação vacinal.

O alcance e a intensidade da pandemia de COVID-19 não foram uniformes no panorama internacional. Os países têm diferentes níveis de desenvolvimento econômico, de densidade demográfica, de globalização, de aceitação de medidas de mitigação, de alcance de campanhas de vacinação, havendo ainda diferenças em suas estruturas etárias. Em consequência, há variabilidade apreciável nas taxas de mortalidade por COVID-19 em diferentes países que serão descritos no desenvolvimento deste trabalho.

2 Objetivos

2.1 Objetivo geral

O objetivo amplo do presente estudo é identificar possíveis fatores explicativos de taxas de mortalidade por COVID-19 em um amplo número de países nos anos de 2021 e 2022. A taxa de mortalidade é definida como o número de mortes por COVID-19 dividido pela população vezes 100. Trata-se, assim, de taxa de mortalidade por 100 habitantes.

2.2 Objetivos específicos

Os objetivos específicos do presente trabalho são os seguintes:

- Coleta e sistematização de dados sobre mortalidade por COVID-19 e outras variáveis para um amplo corte transversal de países;
- Realização de análise descritiva dos dados coletados;
- Determinação de níveis de correlação (positivos e negativos) entre a taxa de mortalidade por COVID-19 e outras variáveis;
- Ajuste de um modelo de regressão para a taxa de mortalidade por COVID-19 e avaliação de sua capacidade preditiva.

Dessa forma, objetivamos identificar variáveis correlacionadas com as taxas de mortalidade por COVID-19 em um amplo corte transversal de países. Tais variáveis podem ser utilizadas em análises preditivas, ou seja, para prever níveis de mortalidade a partir das características e realidades prevalentes em diferentes países.

3 Dados e método

3.1 Dados e área de estudo

Nosso objetivo reside na análise das taxas de mortalidade por COVID-19 a partir de informações sobre os volumes totais de mortes em uma corte transversal de 144 países de 1 de janeiro de 2021 a 31 dezembro de 2022. A taxa de mortalidade é definida em relação à população e não com base no número de testes positivos de COVID-19, uma vez que essa variável é medida com acentuada imprecisão. Em particular, a disponibilidade de testes e os volumes de sub-notificações variam entre países. Trabalharemos com a seguinte definição de taxa de mortalidade: número total de mortes por COVID-19 no período dividido pela população do país vezes 100. Ou seja, trabalharemos com o número de mortes decorrentes de COVID-19 por 100 habitantes. A Tabela 1 apresenta informações sobre as variáveis consideradas em nossa análise. Essa tabela contém os nomes das variáveis, uma breve descrição e o período dos dados. As URLs das páginas de onde foram obtidos os dados estão listadas na Tabela 2.

Os dados utilizados no presente Trabalho de Conclusão de Curso (TCC) abrangem uma ampla variedade de variáveis relacionadas às situações socioeconômicas, demográficas e de saúde da população em diferentes países, com foco nos anos de 2021 e 2022. Esses dados englobam informações importantes, como `paises`, que se refere aos país, isto é, o país de origem dos dados, `continente`, que é o continente no qual o país está localizado, `liberdade_imprensa`, índice de liberdade de imprensa que faz parte de um Ranking Mundial de Liberdade de Imprensa (<https://rsf.org/pt-br/ranking>), que avalia as condições do jornalismo em 180 países e territórios, e é publicada por ocasião do Dia Mundial da Liberdade de Imprensa, 3 de maio.

Tem-se também `total_novas_mortes`, que são as novas mortes atribuídas à COVID-19. As contagens podem incluir mortes prováveis, quando relatadas. Nos raros casos em que a fonte reporta uma alteração diária negativa devido a uma correção de dados, esta métrica é definida como NA. `população_total` é a população total de 2021 e 2022, que é o mesmo período dos dados, `taxa_mortalidade` é a taxa de mortalidade calculada a partir da divisão de duas variáveis já obtidas (`total_novas_mortes` / `população_total`) vezes 100, `idade_mediana_da_populacao` refere-se à idade mediana da população do país, projeção da ONU em 2020, `porcentagem_populacao_diabetes` é a porcetagem da população com diabetes (% da população entre 20 e 79 anos) em 2017, `expectativa_vida` é a expectativa de vida ao nascer em 2019, `indice_gini` é o índice/coeficiente de Gini; é um número entre 0 e 1, onde 0 corresponde à igualdade perfeita (onde todos têm o mesmo rendimento) e 1 corresponde à desigualdade perfeita (onde uma

pessoa tem todo o rendimento, e todas as outras não têm rendimento). `PIB_percapita` é o Produto Interno Bruto (PIB) per capita, que é uma medida que descreve o valor total de bens e serviços de uma localização geográfica e como ele se relaciona com a população da região. O PIB per capita é uma evolução desta métrica e é obtido dividindo o PIB de um país pela sua população. O valor que resulta deste cálculo é o PIB per capita do país.

Também foi obtido o `IDH_2021`, Índice de Desenvolvimento Humano (IDH), é uma métrica compilada pelo Programa das Nações Unidas para o Desenvolvimento e usada para quantificar a "realização média de um país em três dimensões básicas do desenvolvimento humano: uma vida longa e saudável, conhecimento e um padrão de vida específica". `proporcao_nao_religioso` é a proporção da população que se recusa a praticar uma religião. `QI_pais`, Quociente de Inteligência (QI) médio, é uma medida da inteligência humana. As pessoas que desejam ter seu QI medido fazem testes padronizados e recebem uma pontuação que classifica seu nível de inteligência. Quanto maior a pontuação de QI, mais inteligente essa pessoa é considerada. `porcentagem_populacao_urbana` que é a porcentagem da população urbana por país (% da população total), e, por fim, `porcentagem_vacinacao` é a taxa de vacinação, isto é, a porcentagem de vacinas efetivamente aplicadas em cada país para qualquer doença, não apenas COVID-19.

Por fim, também obtivemos dados sobre as porcetagens das populações locais com até 14 anos representado pela variável (`porcentagem_populacao_0_a_14`), entre 15 e 64 anos representado pela variável (`porcentagem_populacao_15_a_64`) e com 65 anos ou mais (`porcentagem_populacao_65+`).

As fontes de dados foram cuidadosamente selecionadas, priorizando reputação e credibilidade. As informações sobre número de mortos por COVID-19, Índice de Liberdade de Imprensa, idade mediana da população no período em decorrência de COVID-19 em cada país, a porcentagem da população com diabetes, expectativa de vida e outras variáveis demográficas foram obtidas do repositório de dados Our World in Data (<https://ourworldindata.org>). Os dados referentes ao índice de Gini de desigualdade de renda abrangem o período de 1992 a 2022. Nota-se que, devido à indisponibilidade de informações mais recentes, alguns países possuem dados datados de 1992. Entretanto, a maioria das informações é mais recente, conforme evidenciado na fonte dos dados obtidos em <https://en.wikipedia.org/wiki/List_of_countries_by_income_equality>, considerando o registro mais recente disponível por país.

Os indicadores econômicos, incluindo o PIB per capita de cada país, foram extraídos do World Population Review. Informações sobre IDH de 2021 por país foram obtidas da mesma fonte. Dados sobre a proporção de pessoas não religiosas ou ateias, QI médio, porcentagem da população urbana por país e porcentagem de vacinação também foram obtidos no mesmo repositório (<https://worldpopulationreview.com/>). A fonte da liberdade de imprensa é da organização não governamental Reporters Without Borders (<https://rsf.org/pt-br/ranking>) e as informações referentes às porcentagem

da população em cada faixas etárias ou estrutura etária foram obtidas em(<https://en.wikipedia.org/wiki/List_of_countries_by_age_structure>). A fonte original dos dados é The World Factbook. As técnicas estatísticas utilizadas são medidas descritivas, figuras, tabelas, correlação linear de Pearson e regressão beta.

3.2 Modelagem de regressão

A análise empírica utilizou modelagem de regressão beta. A classe de modelos de regressão beta foi introduzida por Ferrari e Cribari-Neto (2004) para uso com variáveis aleatórias que assumem valores no intervalo unitário padrão, $(0, 1)$, e cujas médias são impactadas por outras variáveis, utilizadas como variáveis explicativas ou independentes no modelo. O modelo foi posteriormente generalizado por Simas, Barreto-Souza e Rocha (2010) através da incorporação de um submodelo para o parâmetro de precisão.

Seja Y variável aleatória com distribuição beta, denotado $Y \sim \mathcal{B}(\mu, \phi)$. A função densidade de probabilidade de Y é

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1,$$

$0 < \mu < 1$ e $\phi > 0$, $\Gamma(\cdot)$ denotando a função gama, i.e., $\Gamma(w) = \int_0^\infty z^{w-1}e^{-z}dz$. A média e a variância de Y são, respectivamente, $\mathbb{E}(Y) = \mu$ e $\text{Var}(Y) = \mu(1-\mu)/(1+\phi)$. O parâmetro ϕ é chamado de parâmetro de precisão, uma vez que a variância de Y diminui quando o valor de ϕ aumenta. A densidade beta assume formas acentuadamente distintas dependendo dos valores dos parâmetros, podendo ser simétrica, assimétrica à esquerda ou assimétrica à direita, podendo também assumir forma de 'J', de 'J invertido' ou de 'U'. Adicionalmente, a distribuição uniforme é um caso particular da lei beta, ocorrendo quando $\mu = 1/2$ e $\phi = 2$.

Sejam $y_1, \ldots, y_n$ variáveis aleatórias independentes tais que $y_t \sim \mathcal{B}(\mu_t, \phi_t)$, $t = 1, \ldots, n$. O modelo de regressão beta é dado por

$$g_1(\mu_t) = \sum_{j=1}^{p} x_{tj}\beta_j = \eta_{1t} \quad \text{e} \quad g_2(\phi_t) = \sum_{j=1}^{q} z_{tj}\gamma_j = \eta_{2t},$$

em que $\beta = (\beta_1, \ldots, \beta_p)' \in^p$ e $\gamma = (\gamma_1, \ldots, \gamma_q)' \in^q$, $(p + q < n)$ são vetores de parâmetros, η_{1t} e η_{2t} são preditores lineares de $x_{t1}, \ldots, x_{tp}$ e são $z_{t1}, \ldots, z_{tq}$ observações de p e q variáveis independentes, respectivamente. As covariáveis em cada submodelo devem ser linearmente independentes. Em geral, $x_{t1} = z_{t1} = 1\,\forall t$, ou seja, os preditores lineares dos dois submodelos incluem interceptos. Adicionalmente, as funções de ligação $g_1 : (0, 1) \to$ e $g_2 :_+ \to$ devem ser estritamente monótonas e duas vezes continuamente diferenciáveis. As funções de ligação mais utilizadas no submodelo da média são logit, probit, loglog, cloglog e cauchit; no submodelo da precisão, são usadas as ligações log e raiz quadrada.

Tabela 1 – Variáveis, descrição e período

Variável	**Descrição**	**Período**
`paises`	País	2021-2022
`continente`	Continente da localização geográfica	2021-2022
`porcentagem_liberdade_imprensa`	Porcentagem de Liberdade de Imprensa	2020
`total_novas_mortes`	Novas mortes por COVID-19	2021 - 2022
`população_total`	População (dados mais recentes)	2021 - 2022
`taxa_mortalidade`	Cálculo: `total_novas_mortes` / `população_total` ×100	2021 - 2022
`idade_mediana_da_populacao`	Idade mediana da população (projeção da ONU)	2020
`porcentagem_populacao_diabetes`	Prevalência de diabetes (% da população entre 20 e 79 anos)	2017
`expectativa_vida`	Expectativa de vida ao nascer	2019
`indice_gini`	Índice de Gini de desigualdade de renda	1992-2022
`PIB_percapta`	PIB per capita dado em dólares	2023
`IDH_2021`	Índice de Desenvolvimento Humano (IDH)	2023
`proporcao_nao_religioso`	Proporção de pessoas não religiosas ou ateus	2023
`porcentagem_cristaos`	Percentual de cristãos	2023
`QI_pais`	Quociente de Inteligência médio	2023
`porcentagem_populacao_urbana`	Porcentagem de população urbana por país (% da população total)	2022
`porcentagem_vacinacao`	Taxas de vacinação geral	2023
`porcentagem_populacao_0_a_14`	Porcentagem da população entre os 0 e os 14 anos	2020
`porcentagem_populacao_15_a_64`	Porcentagem da população entre os 0 e os 14 anos	2020
`porcentagem_populacao_65`	Porcentagem da população com mais de 65 anos	2020

Tabela 2 – Variáveis e URLs (fonte)

Variável	**URL**
paises	<https://ourworldindata.org/covid-deaths>
continente	<https://ourworldindata.org/covid-deaths>
porcentagem_liberdade_imprensa	<https://rsf.org/pt-br/ranking>
total_novas_mortes	<https://ourworldindata.org/covid-deaths>
população_total	<https://ourworldindata.org/covid-deaths>
taxa_mortalidade	Campo Calculado
idade_mediana_da_populacao	<https://ourworldindata.org/covid-deaths>
porcentagem_populacao_diabetes	<https://ourworldindata.org/covid-deaths>
expectativa_vida	<https://ourworldindata.org/covid-deaths>
indice_gini	<https://en.wikipedia.org/wiki/List_of_countries_by_income_equality>
PIB_percapta	<https://worldpopulationreview.com/country-rankings/gdp-per-capita-by-country>
IDH_2021	<https://worldpopulationreview.com/country-rankings/hdi-by-country>
proporcao_nao_religioso	<https://worldpopulationreview.com/country-rankings/religion-by-country>
porcentagem_cristaos	<https://en.wikipedia.org/wiki/Christianity_by_country#References>
QI_pais	<https://worldpopulationreview.com/country-rankings/average-iq-by-country>
porcentagem_populacao_urbana	<https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS>
porcentagem_vacinacao	<https://worldpopulationreview.com/country-rankings/vaccination-rates-by-country>
porcentagem_populacao_0_a_14	<https://en.wikipedia.org/wiki/List_of_countries_by_age_structure>
porcentagem_populacao_15_a_64	<https://en.wikipedia.org/wiki/List_of_countries_by_age_structure>
porcentagem_populacao_65	<https://en.wikipedia.org/wiki/List_of_countries_by_age_structure>

A estimação dos parâmetros do modelo de regressão beta é feita por máxima verossimilhança. Para tanto, maximiza-se numericamente o logarítmo da função de verossimilhança utilizando um algoritmo de otimização não linear Newton ou quasi-Newton. A significância dos regressores é em geral avaliada via testes z e da razão de verossimilhanças. Para mais detalhes sobre regressão beta, ver Cribari-Neto e Zeileis (2010) e Douma e Weedon (2019).

4 Programa computacional

No desenvolvimento do trabalho, foi utilizado o ambiente de computação estatística R; ver <https://www.R-project.org>. Tal plataforma foi utilizada para filtragem dos dados, consolidação dos dados, análise exploratória de dados, análise de regressão e produção de gráficos. Trata-se de software livre. Utilizamos também a interface gráfica RStudio; ver <https://posit.co>. R é um software livre e de código aberto desenvolvido para computação estatística e análise de dados. É uma linguagem de programação poderosa e flexível que pode ser usada para uma ampla gama de tarefas, desde análises de dados simples até modelagens complexas. R é amplamente utilizado por cientistas, engenheiros, pesquisadores e outros profissionais que trabalham com dados. RStudio é uma interface gráfica para R que fornece uma variedade de recursos para ajudar os usuários a escrever código, visualizar dados e compartilhar seus resultados.

Os seguintes pacotes para R foram utilizados no decorrer de nossa análise empírica: *betareg*, *corrplot*, *dplyr*, *ggcorrplot*, *ggplot2*, *moments*, *readr*.

5 Resultados e discussão

5.1 Análise descritiva dos dados

Inicialmente, realizamos uma análise descritiva da variável de nosso interesse, a saber: taxa de mortalidade por COVID-19. Apresentamos na Tabela 3 estatísticas descritivas sobre tal variável. Os dados, como notado anteriormente, referem-se aos anos de 2021 e 2022 de maneira cumulativa. As taxas de mortalidades mínimas e máximas (média e mediana) são 0.0001 e 0.4558 (0.0993 e 0.0727), respectivamente. O desvio-padrão é 0.1052 e o primeiro e terceiro quartis são, respectivamente, 0.0100 e 0.1467. Note que há assimetria à direita (coeficiente de assimetria: 1.1981) e excesso de curtose (coeficiente de curtose: 3.7902). Sabemos, assim, que metade (um quarto) [três quartos] dos países apresenta mortalidade acima de 0.0727 (0.0100) [0.1467]. A assimetria positiva indica maior concentração de dados na parte inferior da escala.

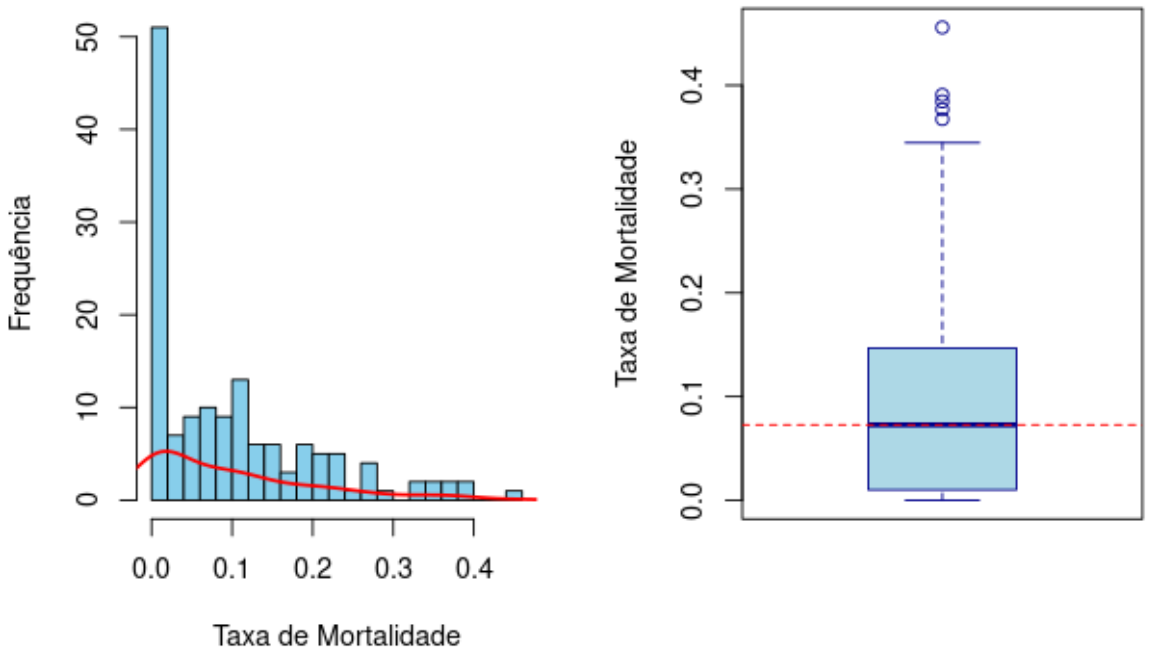
Tabela 3 – Estatísticas descritivas da taxa de mortalidade por COVID-19 dos 144 países em 2021/22

Estatística	**Valor**
Mínimo	0.0001
Primeiro quartil	0.0101
Mediana	0.0727
Média	0.1000
Terceiro quartil	0.1488
Máximo	0.4558
Desvio-padrão	0.1057
Assimetria	1.1877
Curtose	3.7486

A Figura 1 contém o histograma (painel esquerdo) e o boxplot das taxas de mortalidade (painel direito). A linha vermelha sobreposta ao histograma representa uma estimativa não paramétrica da densidade da variável de interesse. Nota-se do boxplot que há países com taxas de mortalidade por COVID-19 consideravelmente elevadas, ou seja, com um volume substancial de óbitos em relação à sua população. Entre os outliers destacados no boxplot estão os seguintes países: Bulgária, Hungria e Geórgia.

As estatísticas descritivas e os gráficos apresentados fornecem informações importantes para a compreensão da heterogeneidade prevalente nas experiências dos diferentes países diante da pandemia de COVID-19. A existência de países com taxas de mortalidade muito baixas e de outros com taxas de mortalidade mais elevadas indica a importância de investigar as diferentes práticas de saúde pública adotadas, as infraestruturas de saúde

Figura 1 – Histograma e boxplot da taxa de mortalidade por COVID-19 dos 144 países em 2021/22



e outros fatores que possam contribuir para diferenças de patamares de mortalidade, subsidiando estratégias mais eficazes de combate a pandemias.

A Tabela 4 apresenta os países com as dez maiores e as dez menores taxas de mortalidade por COVID-19. Há padrões distintos entre diferentes regiões do mundo. Entre os países com as maiores taxas de mortalidade, notamos predominância de nações europeias. Há, nessa lista, nove países europeus e um país sul-americano (Peru). Os países com as taxas de mortalidade mais elevadas são Bulgária, Hungria e Geórgia. A concentração dos países mais impactados pela pandemia na Europa sugere a possível influência de fatores regionais específicos na gestão da pandemia, tais como sistemas de saúde, políticas de resposta, e dinâmicas populacionais.

Por outro lado, os dez países com as menores taxas de mortalidade são predominantemente africanos. Nessa lista, há oito países africanos, um norte-americano (Nicarágua) e um asiático (Tajiquistão). Os países com as menores taxas de mortalidade são Burúndi, Tajiquistão e Chade.

Países africano apresentaram tipicamente taxas de mortalidade por COVID-19 mais baixas relativamente a outras regiões do mundo, como a Europa, devido a um conjunto de fatores interligados. Em primeiro lugar, a população africana é notavelmente mais jovem que a de muitas outras partes do mundo, com idade média significativamente menor. Isso pode desempenhar um papel crucial nas taxas de mortalidade por COVID-19 dos países africanos, uma vez que a doença atinge com mais severidade pessoas mais velhas,

Tabela 4 – Países com as maiores e menores taxas de mortalidade por COVID-19 dos 144 países em 2021/22

Maiores taxas de mortalidade		
País	**Continente**	**Taxa de mortalidade**
Bulgária	Europa	0.4558
Hungria	Europa	0.3909
Geórgia	Europa	0.3845
Bósnia e Herzegovina	Europa	0.3768
Peru	América do Sul	0.3674
Letônia	Europa	0.3447
Croácia	Europa	0.3405
Eslováquia	Europa	0.3311
Montenegro	Europa	0.3306
Grécia	Europa	0.2917
Menores taxas de mortalidade		
País	**Continente**	**Taxa de mortalidade**
Nicarágua	América do Norte	0.0012
República Centro-Africana	África	0.0009
Benin	África	0.0009
Nigéria	África	0.0009
Níger	África	0.0008
Sudão do Sul	África	0.0007
Serra Leoa	África	0.0006
Chade	África	0.0005
Tajiquistão	Ásia	0.0004
Burúndi	África	0.0001

especialmente aquelas com 65 anos ou mais. Além disso, as respostas rápidas e eficazes dos governos africanos no início da pandemia, incluindo medidas de saúde pública, como distanciamento social, uso de máscaras e 'lockdowns' rigorosos, contribuíram para conter a propagação do vírus. A colaboração e adesão da população a essas medidas também foram fundamentais para mitigar os impactos da pandemia. Em contraste, a Europa enfrentou desafios apreciáveis na implementação de medidas de controle da propagação da doença, resultando em taxas de mortalidade mais altas, especialmente durante as primeiras ondas da pandemia.

Essas diferenças destacam a importância das respostas governamentais e da demografia da população na dinâmica da pandemia. Enquanto a África se beneficiou de uma população mais jovem e de ações rápidas para conter a propagação do vírus, a Europa enfrentou dificuldades devido à sua população mais envelhecida e atrasos na implementação de medidas eficazes de saúde pública.

Países mais desenvolvidos tendem a ter pirâmides etárias mais concentradas em

populações mais idosas, tornando-os mais vulneráveis à mortalidade por COVID-19. Isso se deve a diversos fatores, como maior acesso à saúde, que aumenta a expectativa de vida, e estilo de vida, que pode incluir maior prevalência de doenças crônicas. Ademais, tais países geralmente possuem áreas urbanas com elevadas densidades populacionais, o que facilita a disseminação do vírus SARS-CoV-2. Em nossa análise, consideraremos uma ampla gama de variáveis auxiliares a fim de capturar diferentes facetas da relação entre o patamar de desenvolvimento de um país e sua taxa esperada de mortalidade por COVID-19.

A seguir, analisaremos os patamares de correlação linear entre a taxa de mortalidade por COVID-19 e as demais variáveis consideradas em nosso estudo. A Figura 2 contém a matriz de correlação das variáveis. Tal matriz fornece uma visão abrangente das inter-relações entre tais variáveis. Observamos que `taxa_mortalidade` correlaciona negativamente com `proporcao_populacao_0_a_14` (-0.63), `indice_gini` (-0.15) e `populacao_total` (-0.11). É possível notar também correlações positivas fortes com `idade_mediana_da_populacao` (0.66) e `porcentagem_populacao_65` (0.65).

Notamos também haver correlação positiva moderada entre a `taxa_mortalidade` com `QI_pais` e `expectativa_vida` sendo (0.47) e (0.49) respectivamente. Isso indica que países com populações com o QI mais alto em média tendem a ser mais impactadas pelo vírus SARS-CoV-2 assim como a expectativa de vida mais alta que também é um fator de impacto negativo para a taxa de mortalidade no país.

Em relação à liberdade de imprensa (variável `porcentagem_liberdade_imprensa`), identificamos correlação positiva moderada (0.37) com a `taxa_mortalidade`. Isso indica que a taxa de mortalidade tende a ser mais elevada em países com maior liberdade de imprensa. Note-se que países mais desenvolvidos, onde a população tende a ser mais idosa, contam, tipicamente, com patamares mais elevados de liberdade de imprensa.

Ao abordar a desigualdade econômica, observamos uma correlação positiva fraca (0.19) da `taxa_mortalidade` com a variável `PIB_percapita` indica que países com maior renda podem enfrentar desafios adicionais na gestão da pandemia, destacando a complexidade dessas relações.

As variáveis relacionadas ao ambiente urbano e à vacinação também mostram associações interessantes. A `taxa_mortalidade` identificamos uma correlação positiva moderada de (0.35) com a `porcentagem_populacao_urbana`, já com `proporcao_vacinacao` observamos uma correlação positiva fraca de (0.13). Países com maior urbanização e taxas de vacinação mais elevadas podem enfrentar desafios distintos na gestão da propagação do vírus, ressaltando a importância de abordagens adaptadas a contextos urbanos e estratégias eficazes de vacinação.

Além disso, fatores religiosos e educacionais desempenham papéis significativos na dinâmica da pandemia. A porcentagem de cristãos, representada pela variável `porcentagem_cristaos`, mostra uma correlação positiva moderada de (0.37). Da mesma forma, o QI médio do país, representado por `QI_pais`, apresenta uma correlação positiva

moderada de (0.48) com a taxa de mortalidade. Essas associações destacam a necessidade de considerar múltiplos contextos culturais e educacionais na gestão da pandemia.

As possíveis variáveis para o modelo são aquelas com maiores correlações absolutas, a saber: as porcentagens da população nas faixas etárias de 0 a 14 anos e acima de 65 anos, revelando variações na distribuição etária; o QI médio do país, que pode indicar o nível educacional e a inteligência média da população; a expectativa de vida, refletindo as condições de saúde e qualidade de vida; o patamar de liberdade de imprensa, indicando a liberdade de expressão e transparência do governo; e a porcentagem de cristãos na população, refletindo aspectos culturais e religiosos.

A expectativa de vida varia consideravelmente, indo de 53.28 anos a 84.63 anos, indicando discrepâncias nos sistemas de saúde e nas condições socioeconômicas entre os países estudados. Além disso, a porcentagem de cristãos varia de 0.01% a 98%, evidenciando a complexidade cultural e religiosa. A distribuição da população em diferentes faixas etárias também exibe variações marcantes, com a porcentagem da população acima de 65 anos variando de 1.19% a 29.18%. Tais números destacam a interconexão dos fatores que influenciam não apenas a mortalidade por COVID-19, mas também outros indicadores de saúde pública e bem-estar.

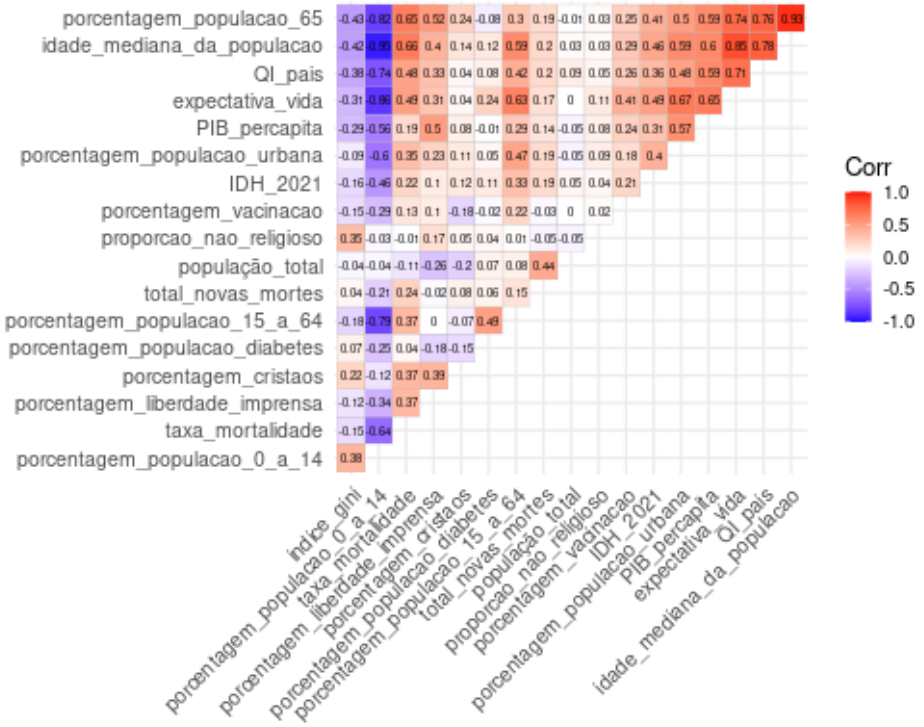
As médias e medianas dessas variáveis fornecem uma visão geral interessante. Por exemplo, a média da expectativa de vida é 72.76 anos, enquanto a mediana é 74.48 anos, sugerindo uma distribuição levemente assimétrica à esquerda. Da mesma forma, a média da porcentagem de cristãos é aproximadamente 53.08%, com uma mediana de 59.17%, indicando uma distribuição também assimétrica à esquerda. Essas medidas ressaltam a importância de considerar não apenas os valores centrais, mas também a dispersão e a forma das distribuições ao interpretar os dados.

5.2 Modelagem de regressão

A seguir apresentaremos uma modelagem de regressão beta em que a variável resposta é da taxa de mortalidade por COVID-19 e as demais variáveis são utilizadas como condicionantes, ou seja, como potenciais variáveis explicativas. O objetivo reside na obtenção de um modelo de regressão com estrutura relativamente simples, mas com boa capacidade de prever os patamares de mortalidade dos diferentes países a partir dos valores das variáveis condicionantes utilizadas na modelagem.

A variável resposta (y_t) é a taxa de mortalidade e as demais variáveis podem ser consideradas como variáveis condicionantes. O processo de seleção de modelos foi baseado em: (i) critérios de informação de Akaike e bayesiano (AIC e BIC, respectivamente), (ii) pseudo-R^2 de Nagelkerke (NAGELKERKE et al., 1991), (iii) análise gráfica de resíduos, (iv) gráfico meio-normal de resíduos com envelopes simulados, (v) teste RESET de especificação correta do modelo (PEREIRA; CRIBARI-NETO, 2014). Utilizamos o

Figura 2 – Matriz de correlação da taxa de mortalidade por COVID-19 e as demais variáveis consideradas em nosso estudo



pseudo-R^2 de Nagelkerke porque considera o ajuste obtido a partir dos dois submodelos, diferentemente do pseudo-R^2 proposto por Ferrari e Cribari-Neto (2004), que se baseia apenas no submodelo da média. Os envelopes simulados nos gráficos residuais meio-normais correspondem aos quantis de 0.025 e 0.975 dos resíduos obtidos a partir de 100 conjuntos de dados simulados. Foi utilizado o resíduo ponderado padronizado 2 proposto por Espinheira, Ferrari e Cribari-Neto (2007). Não foi utilizado procedimento automatizado de busca de modelos. Muitos modelos foram estimados e avaliados individualmente. Iniciamos com um modelo de precisão fixa com todas as possíveis variáveis explicativas. Em seguida, as variáveis regressoras não significativas ao nível de significância de 5% foram, individual e sequencialmente, removidas do modelo seguindo ordem decrescente de p-valores dos testes z individuais. Numa segunda etapa, modelou-se a precisão de forma conjunta à média, avaliando-se novamente significâncias estatísticas dos regressores do submodelo da média.

Com base no procedimento de busca de modelos detalhado acima, chegou-se ao seguinte modelo de regressão beta:

$$\begin{aligned} \text{logit}(\mu_t) &= \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \beta_5 x_{t5} + \beta_6 (x_{t4} \times x_{t6}) \\ \log(\phi_t) &= \gamma_1 + \gamma_2 x_{t6} + \gamma_3 x_{t3} + \gamma_4 x_{t4}, \end{aligned}$$

em que, para o t-ésimo país, x_{t2} é a porcentagem da população entre até entre 15 e 64 anos, x_{t3} é a porcentagem da população com 65 anos ou mais, x_{t4} é o PIB per capita, x_{t5} é a porcentagem de cristãos, x_{t6} é a porcentagem da população até 14 anos. As funções de ligação da média e da precisão são, respectivamente, logit e log. Os valores dos critérios

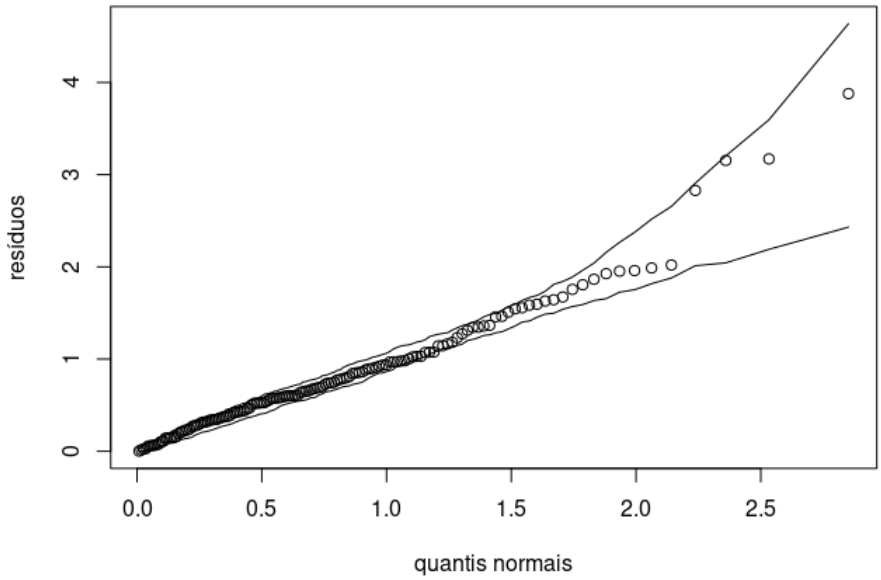
de informação para o modelo estimado são AIC $= -550.6110$ e BIC $= -523.9454$. O pseudo-R^2 é 0.7227, o que indica que o modelo estimado possui elevado poder explicativo. A hipótese nula de que o modelo estimado está corretamente especificado não é rejeitada pelo teste RESET aos níveis usuais de significância (p-valor $= 0.7018$). A Figura 3 contém o gráfico meio normal dos resíduos com envelopes simulados. Todos os resíduos se encontram entre os envelopes inferior e superior. Isso constitui evidência de que a especificação do modelo está correta, indicando também ausência de observações atípicas. As estimativas de máxima verossimilhança dos parâmetros são (erros-padrão entre parênteses) $\hat{\beta}_1 = -14.1296$ (0.7673), $\hat{\beta}_2 = 0.1526$ (0.0002), $\hat{\beta}_3 = 0.1398$ (0.0001), $\hat{\beta}_4 = -0.0001$ (< 0.0001), $\hat{\beta}_5 = 0.0087$ (< 0.0001), $\hat{\beta}_6 = 2.8413 \times 10^{-6}$ (6.0382×10^{-7}), $\hat{\gamma}_1 = -2.3481$ (0.7061), $\hat{\gamma}_2 = 0.1560$ (0.0005), $\hat{\gamma}_3 = 0.1340$ (0.0011) and $\hat{\gamma}_4 = 1.4413 \times 10^{-5}$ (6.5908×10^{-6}). As hipóteses nulas de todos os testes z individuais são rejeitadas ao nível de significância de 1%.

Notamos que as estimativas de β_2 e β_3 são positivas, implicando que a taxa média de mortalidade por COVID-19 tende a aumentar quando as proporções da população entre 15 e 64 anos e acima de 64 anos são aumentadas. A estimativa de β_4 é negativa. Não podemos, todavia, concluir pela existência de relação inversa entre a resposta média e o patamar de riqueza por habitante dado haver no modelo variável de interação envolvendo esse regressor. Encontramos também relação direta entre a porcentagem de cristãos na população e a taxa média de mortalidade, uma vez que $\hat{\beta}_5 > 0$. Por fim, a estimativa de β_6 é positiva. Concluímos, assim, que a taxa média de mortalidade cresce com o produto entre a porcentagem da população até 14 anos e o PIB per capita. No que tange ao submodelo da precisão, as estimativas de γ_2, γ_3 e γ_4 são postivas. Assim, a precisão aumenta com as variáveis regressoras correspondentes. Há, assim, menor variabilidade da resposta em sociedades com maiores porcentagens de jovens (até 14 anos) e idosos (65 anos ou mais) e também em sociedades com rendas per capita mais elevadas.

A partir do modelo de regressão beta proposto para a taxa de mortalidade por COVID-19, podemos avançar na interpretação dos resultados obtidos. A primeira observação relevante é o elevado poder explicativo do modelo, conforme indicado pelo pseudo-R^2 de Nagelkerke, que atinge o valor de 0.7227. Isso sugere que as variáveis incluídas na modelagem são capazes de explicar parcela apreciável da variação na taxa de mortalidade entre os países considerados. Além disso, os critérios de informação AIC e BIC também apontam para a adequação do modelo, visto que apresentam valores relativamente baixos comparativamente a outros modelos considerados, sinalizando bom compromisso entre ajuste e complexidade.

Ao analisar as estimativas dos parâmetros do modelo, observamos que as variáveis relacionadas à estrutura etária da população têm impacto significativo sobre a taxa de mortalidade por COVID-19. As estimativas positivas de β_2 e β_3 indicam que as proporções de indivíduos entre 15 e 64 e de idosos na população estão associadas a uma maior taxa média de mortalidade. Esse resultado está em linha com as evidências epidemiológicas que

Figura 3 – Gráfico residual semi-normal com envelopes simulados.



apontam para uma maior vulnerabilidade desses grupos à doença.

Por outro lado, a estimativa negativa de β_4 sugere uma relação inversa entre o PIB per capita e a taxa de mortalidade, embora devamos interpretá-la com cautela devido à presença de uma interação com outra variável. A estimativa do coeficiente associado à interação entre a porcentagem da população até 14 anos e o PIB per capita, representada por x_{t6}, indica que o efeito do PIB sobre a mortalidade varia conforme a estrutura etária da população. Isso sugere que, em países com proporção expressiva de jovens, um aumento no PIB per capita está associado a um aumento adicional na taxa média de mortalidade.

A presença no modelo da variável representando a porcentagem de cristãos na população (x_{t5}) também é interessante, com estimativa positiva para o coeficiente de regressão correspondente (β_5). Isso sugere que países com uma maior proporção de cristãos tendem a apresentar taxas médias de mortalidade por COVID-19 mais elevadas. Essa relação pode ser explicada por diversos fatores, incluindo aspectos culturais, comportamentais e sociais, que podem influenciar na disseminação e na gravidade da doença.

No que diz respeito ao submodelo da precisão, as estimativas positivas dos coeficientes γ_2, γ_3 e γ_4 indicam que a precisão do modelo aumenta com a proporção de jovens e idosos na população, assim como com o aumento do PIB per capita. Isso sugere que a variabilidade da taxa de mortalidade entre os países é menor em sociedades com uma estrutura etária mais homogênea e com maior poder econômico.

O modelo de regressão estimado apresenta bom poder preditivo. As previsões obtidas para muitos países são bastante acuradas. Por exemplo, taxas de mortalidade por COVID-19 observadas e preditas: (i) Finlândia: 0.1468 e 0.1325, (ii) França: 0.1447 e 0.1447, (iii) Israel: 0.0924 e 0.0914, (iv) Suécia: 0.1194 e 0.1233, (v) Suíça: 0.0727 e

0.0735. Ademais, a comparação entre taxas observadas e preditas de mortalidade permite identificar países que tiveram gerenciamento deficiente da pandemia de COVID-19, que são aqueles com mortalidade efetiva notadamente superior à esperada com base nos patamares de seus preditores. Esse é o caso do Brasil, que tem as seguintes taxas observada e esperada de mortalidade: 0.2327 e 0.1918. Ou seja, no Brasil registrou-se mortalidade mais de 20% superior ao que seria esperado a partir dos preditores do país. O gerenciamento da crise sanitária também foi inefetivo no Reino Unido (observado: 0.1819, predito: 0.1281). O Peru apresentou resultado muito desfavorável segundo o modelo estimado, com taxas de mortalidade efetivas e esperadas iguais a 0.3674 e 0.1234, respectivamente. Ou seja, o patamar de mortalidade observado foi quase 200% superior ao esperado. Em contraste, há países em que houve gerenciamento eficaz da pandemia e onde as taxas observadas de mortalidade ficaram notadamente abaixo das preditas. Por exemplo (observado e previsto): Austrália (0.0645 e 0.1080), Canadá (0.0880 e 0.1265), Espanha (0.1366 e 0.1896), Islândia (0.0424 e 0.1410), Japão (0.0436 e 0.1019), Noruega (0.0810 e 0.1082), Nova Zelândia (0.0445 e 0.1164) e Portugal (0.1857 e 0.2612).

Em suma, os resultados obtidos por meio da análise de regressão beta fornecem informações importantes sobre os preditores da taxa de mortalidade por COVID-19, destacando a relevância das características demográficas e socioeconômicas na sua variação entre os países. No entanto, é importante ressaltar que a interpretação dos resultados deve considerar as limitações do modelo e a complexidade do fenômeno estudado, sendo necessárias análises complementares para uma compreensão mais abrangente dos fatores envolvidos.

6 Conclusão

Realizamos uma análise descritiva e de regressão abrangente das taxas de mortalidade por COVID-19 em uma ampla variedade de países durante os anos de 2021 e 2022 que conduziu a conclusões importantes. Primeiramente, identificamos heterogeneidade apreciável nas taxas de mortalidade entre os países, com algumas nações apresentando taxas muito baixas e outras enfrentando taxas mais elevadas. Essas disparidades ressaltam a importância de investigar os fatores capazes de prever tais diferenças, a fim de informar estratégias eficazes de combate a pandemias futuras.

A análise descritiva dos dados revelou padrões interessantes nas taxas de mortalidade por COVID-19. Observamos uma predominância de países europeus entre os que apresentam as maiores taxas de mortalidade, enquanto nações africanas compõem a maioria dos países com as menores taxas. Além disso, identificamos correlações expressivas entre a taxa de mortalidade e várias variáveis demográficas, socioeconômicas e de saúde, destacando a complexidade dos fatores que influenciam a dinâmica da pandemia ao nível global.

A modelagem de regressão beta permitiu uma análise mais aprofundada dos preditores da taxa de mortalidade por COVID-19. O modelo estimado apresentou alto poder explicativo e indicou que a estrutura etária da população, o nível socioeconômico e a composição religiosa podem desempenhar papéis importantes na variação da mortalidade entre os países. Além disso, a precisão do modelo mostrou-se influenciada pela estrutura demográfica e pelo nível de desenvolvimento econômico dos países, evidenciando a complexidade das relações envolvidas.

Em suma, os resultados da análise realizada no presente Trabalho de Conclusão de Curso destacam a importância de considerar uma ampla gama de fatores na compreensão da dinâmica da pandemia de COVID-19 ao nível global. A abordagem adotada neste estudo, que combinando análise descritiva, correlacional e modelagem estatística, proporcionou informações valiosas sobre os determinantes da mortalidade por COVID-19. Pesquisas futuras podem beneficiar-se da continuidade dessa abordagem integrada para entender melhor os fatores que influenciam a mortalidade por COVID-19 e informar políticas de saúde pública mais eficazes.

Referências

Agência Brasil. *OMS declara fim da emergência em saúde por COVID-19*. 2023. <https://agenciabrasil.ebc.com.br/saude/noticia/2023-05/oms-declara-fim-da-emergencia-em-saude-por-covid-19>. Citado na página 1.

CRIBARI-NETO, F.; ZEILEIS, A. Beta regression in R. *Journal of Statistical Software*, v. 34, n. 2, p. 1–24, 2010. Citado na página 8.

DOUMA, J. C.; WEEDON, J. T. Analysing continuous proportions in ecology and evolution: A practical introduction to beta and dirichlet regression. *Methods in Ecology and Evolution*, v. 10, n. 9, p. 1412–1430, 2019. Citado na página 8.

ESPINHEIRA, P. L.; FERRARI, S. L. P.; CRIBARI-NETO, F. On beta regression residuals. *Journal of Applied Statistics*, v. 35, n. 4, p. 407–419, 2007. Citado na página 15.

FERRARI, S. L. P.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, v. 31, n. 7, p. 799–815, 2004. Citado 2 vezes nas páginas 5 e 15.

MATHIEU, E. et al. A global database of COVID-19 vaccinations. *Nat Hum Behav*, 2021. Disponível em: <https://ourworldindata.org/covid-vaccinations>. Citado na página 1.

Ministério da Saúde. *Brasil chega a marca de 700 mil mortes por COVID-19*. 2023. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/noticias/2023/marco/brasil-chega-a-marca-de-700-mil-mortes-por-covid-19>. Citado na página 1.

NAGELKERKE, N. J. et al. A note on a general definition of the coefficient of determination. *biometrika*, Oxford University Press, v. 78, n. 3, p. 691–692, 1991. Citado na página 14.

PEREIRA, T. L.; CRIBARI-NETO, F. Detecting model misspecification in inflated beta regressions. *Communications in Statistics - Simulation and Computation*, v. 43, n. 3, p. 631–656, 2014. Citado na página 14.

SIMAS, A. B.; BARRETO-SOUZA, W.; ROCHA, A. V. Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis*, v. 54, n. 2, p. 348–366, 2010. Citado na página 5.

Terra Notícias. *Número de mortes por COVID no mundo supera 7 milhões, diz OMS*. 2023. Disponível em: <https://www.terra.com.br/noticias/mundo/numero-de-mortes-por-covid-no-mundo-supera-7-milhoes-diz-oms,0f8543ea0ead8a1235e3677ab6add089313ze8pn.html#:~:text=Entidade%20acredita%20que%20n%C3%BAmero%20de%20v%C3%ADtimas%20seja%20at%C3%A9%203x%20maior&text=A%20Organiza%C3%A7%C3%A3o%20Mundial%20da%20Sa%C3%BAde,em%20dezembro%20do%20ano%20passado.> Citado na página 1.

WU, C. et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in wuhan, china. *JAMA Internal Medicine*, v. 180, n. 7, p. 934–943, 2020. Citado na página 1.