UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE FÍSICA - ESCOLA DE ENGENHARIA
CURSO DE ENGENHARIA FÍSICA

JESUS DANIEL YEPEZ ROJAS

IMPLEMENTAÇÃO DE PCA PARA AVALIAÇÃO DE DESEMPENHO DOS PAÍSES FRENTE AO SARS-CoV-2

Porto Alegre
2021

JESUS DANIEL YEPEZ ROJAS

IMPLEMENTAÇÃO DE PCA PARA AVALIAÇÃO DE DESEMPENHO DOS PAÍSES FRENTE AO SARS-CoV-2

Trabalho de conclusão de curso de graduação apresentado à Universidade Federal do Rio Grande do Sul como parte dos requisitos para a obtenção do título de Bacharel em Engenharia Física.

Orientador: Prof. Dr. Sebastián Gonçalves

Porto Alegre

2021

AGRADECIMENTOS

Primeiramente quero dedicar meus agradecimentos a meus pais, Rosangel Rojas e José Yepez, pelo apoio e pelo suporte ao longo de todos os anos de graduação e por me ensinar que é possível sonhar grande, mesmo começando a vida de zero num novo país. Ao meu irmão José Mauricio Yepez, pela companhia e o apoio sempre que precisei.

Agradeço ao professor Sebastián Gonçalves pela orientação, sempre disposto a ajudar com conselhos e sugestões que auxiliaram na construção do presente trabalho.

Aos professores integrantes da banca orientadora Cristiano Krug e Leonardo Brunnet, pelas correções e sugestões no desenvolvimento deste trabalho.

Aos meus amigos da época do cursinho Eduardo, Gabriel, Matheus e Alfredo, por todos os anos de amizade.

A meus colegas de graduação pelo suporte e companheirismo.

Ao Rubens e o Deomar, pela ajuda incondicional que ajudaram a superar os momentos mais difíceis do curso.

E finalmente agradecer ao Brasil, por me dar acesso à educação pública de qualidade, mesmo como estrangeiro, me permitindo hoje ter oportunidades além das que poderia sonhar no meu país de origem.

RESUMO

A COVID-19 é uma doença produzida pelo vírus SARS-CoV-2. Esse vírus se espalhou rapidamente pelo mundo, o que levou a Organização Mundial da Saúde a declará-lo uma pandemia no dia 11 de março de 2020. Cada país tem adotado diferentes estratégias de contenção e de testagem do vírus, isso e a diferença de condições socioeconômicas dificultam a comparação direta dos números de óbitos. Foi desenvolvida uma metodologia que permitiu avaliar 158 países em relação ao número de óbitos per capita na primeira onda do COVID -19. Primeiramente foi proposta uma metodologia para definir, por meio do número de reprodução, uma janela de tempo, específica para cada país, correspondente à primeira onda. A partir de um conjunto de variáveis demográficas e econômicas publicamente disponíveis, foi analisada a capacidade dessas variáveis de separar os países em grupos com mais ou menos óbitos per capita. Para isso usamos agrupamento por faixas, metodologia inspirada em técnicas da área de engenharia de qualidade. A análise individual levou a conclusões contraditórias no caso de variáveis como PIB e IDH, que foram atribuídas à alta correlação com variáveis demográficas. A técnica de análise de componente principal foi então aplicada para gerar componentes independentes. O resultado final da análise sugere que os números elevados de mortes per capita de alguns países podem ser pela distribuição etária dos mesmos.

Palavras-chave: comparação, algoritmo, PCA, COVID-19.

ABSTRACT

COVID-19 is a disease produced by the SARS-CoV-2 virus. This virus spread rapidly worldwide, which led the World Health Organization to declare it a pandemic on March 11, 2020. Each country has adopted different containment and testing strategies for the virus, joined with the different socioeconomic conditions making difficult to compare the numbers of deaths directly. Therefore, a methodology was developed to assess 158 countries in relation to the number of deaths *per capita* in the first wave of COVID-19. First, a methodology was proposed to define, using the reproduction number, a specific time window for each country corresponding to the first wave. Then, was tested the capability of a set of demographic and economic variables for separate countries into groups with different values of deaths *per capita*. For this, we use grouping by bands, a methodology inspired by techniques from the field of quality engineering. The individual analysis led to contradictory conclusions in the case of variables such as GDP and HDI, which were attributed to a high correlation with demographic variables. The PCA technique was then applied to generate independent components. The final result of the analysis suggests that the high numbers of *per capita* deaths in some countries are justified by their age distribution.

Keywords: comparison, algorithm, PCA, COVID-19.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1. INTRODUÇÃO

O SARS-CoV-2 faz parte de uma família de vírus que podem causar doenças em humanos e animais. Desde a detecção do primeiro caso na China, no final de 2019, o vírus espalhou-se rapidamente pelo mundo, com níveis alarmantes de propagação e gravidade. No dia 30 de janeiro de 2020, a Organização Mundial de Saúde (OMS) classificou a doença produzida pelo vírus como uma emergência de saúde internacional e, pouco tempo depois, no dia 11 de março, foi declarada pandemia, OMS [1].

O Brasil, que registrou seu primeiro caso no dia 26 de fevereiro de 2020, em meados do mês de julho se posiciona como o segundo país com maior número de casos no mundo, Cândido et al. [2]. Um ano depois o país chegou ao seu pior momento na pandemia, sendo a média móvel de mortes na primeira semana de março de 2021 a maior desde a chegada do vírus. Ao longo desse período de pouco mais de um ano, o país passou por uma série de picos, períodos com número elevado de leitos de UTI ocupados seguidos de períodos de descida do número de contagens. Isto é, o desempenho na administração/mitigação da epidemia foi variável, possivelmente associado também ao comportamento da população. Períodos de relativa tranquilidade deram lugar a períodos de estresse dos sistemas de saúde, o que também foi observado em outros países.

A antecipação desses períodos de estresse, mediante parâmetros de aferição objetiva, teria permitido a elaboração de medidas capazes de evitar a saturação da capacidade hospitalar minimizando o número de óbitos. Diferentes países (em casos de países continentais, diferentes regiões também) em diferentes períodos têm apresentado desempenhos muito diversos em relação à administração da epidemia de COVID-19. Isso aponta para o objetivo do trabalho: definir um parâmetro ou um conjunto de parâmetros que permitam a comparação objetiva e direta dos países a partir das suas evoluções individuais em relação à pandemia. Passada a epidemia de COVID-19, essa avaliação pode permitir retrospectivamente analisar falhas e acertos nas estratégias para conter o avanço dessa e de futuras epidemias.

Observando a Figura 1, vemos a evolução da quantidade de óbitos por milhão de habitantes, nos primeiros dias de pandemia, em cinco países: Bélgica, Itália, Nova Zelândia, Estados Unidos e Brasil. Os dados foram alinhados usando como referência o dia de registro do 100º caso. A priori é possível afirmar que, entre os países apresentados, a Bélgica teve

um desempenho pior no período selecionado, todavia, existem alguns vetores escondidos: Na Bélgica o critério de contagem de casos e óbtios é menos rigoroso em comparação com os outros países, BBC News [3]. O tamanho do Brasil em relação a alguns dos outros países pode gerar viés mesmo em comparação com números normalizados por população. Um dos motivos para isso é a diferença do tempo de "espalhamento" da doença no território.

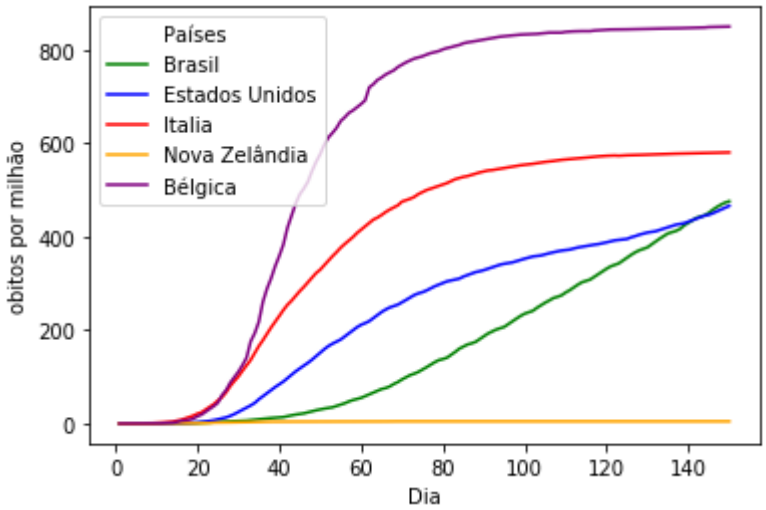


Figura 1: Óbitos per capita vs dias após o 100º caso, na Bélgica, Itália, Estados Unidos, Nova Zelândia e Brasil. Fonte: Our World in Data.

O exemplo comparativo anterior foi feito ilustrar a quantidade de elementos que podem ser considerados ao comparar a evolução do vírus em um grupo de países, evidenciando a necessidade de realizar uma análise aprofundada para atingir avaliações mais adequadas. Existem alguns fatores que dificultam a comparação direta entre países, Rutger et al. [4] os descreve como se indica a continuação:

(i) O vírus não chegou simultaneamente em todos os países;

(ii) Os números absolutos são incomparáveis por causa dos diferentes tamanhos de população;

(iii) Os números normalizados pelo tamanho da população (casos ou mortes por milhão de habitantes) também podem ser difíceis de comparar, pois a infecção não se espalha homogeneamente em todos os países. Na China, por exemplo, a província de Hubei foi severamente afetada, diferente do resto do país; assim, uma correção pelo tamanho total da população chinesa não seria representativa;

(iv) Diferentes distribuições etárias podem causar diferente susceptibilidade a óbitos por Covid-19. Japão, por exemplo, possui uma das populações mais idosas no mundo;

(v) O número de óbitos acumulado e a taxa de letalidade são influenciados pelas diferenças nas políticas de realização de testes;

2. OBJETIVO

Construção de uma ferramenta, parâmetros ou conjunto de parâmetros, adequados para classificar e comparar o desempenho dos países frente à pandemia do COVID-19, identificando quais países possuem valores de óbitos compreensíveis em função das suas condições demográficas. Consequentemente, identificar que países evitaram chegar nesses valores e quais os ultrapassaram.

3. ALCANCE

A metodologia computacional permitirá a comparação mútua entre países das suas capacidades potenciais, ou seja, identificar os países com condições de proporcionar uma resposta diferente à sua resposta real. O método será desenvolvido a partir das informações da primeira onda.

4. FUNDAMENTAÇÃO TEÓRICA

4.1. COVARIÂNCIA

A covariância é uma medida do grau de interdependência numérica entre duas variáveis aleatórias. Assim, variáveis independentes têm covariância zero. Dado dois conjuntos de variáveis X e Y, a covariância entre X e Y é dada pela Equação 1:

$$\sum_{i=1}^{n} \frac{(x_i - \mu_x)(y_i - \mu_y)}{n} \quad (1)$$

Onde μ_x é o valor médio do conjunto X e μ_y o valor médio do conjunto Y.

4.2. CORRELAÇÃO

O coeficiente de correlação indica o quanto as variações de duas variáveis *X* e *Y* estão conexas. O coeficiente de correlação pode assumir valores numéricos entre -1 e 1. Quando a correlação assume valores positivos, indica que as variáveis *X* e *Y* estão correlacionadas positivamente, ou seja, quando *X* aumenta, *Y* também aumenta. Quando o coeficiente for negativo as variáveis estão relacionadas inversamente, assim, quando *X* aumenta, *Y* diminui e vice-versa.

O coeficiente de correlação de Pearson é um dos coeficientes mais comumente usados, o mesmo indica a existência de uma relação linear entre duas variáveis, Cohen et al. [5]. O coeficiente de correlação de Pearson das variáveis X e Y pode ser calculado com a Equação 2, pela razão entre a covariância de X e Y e o produto dos desvios padrão individuais de ambas variáveis.

$$Corr(X,Y) = \frac{Cov\ (X,Y)}{\sigma_Y \sigma_X} \tag{2}$$

4.3. PADRONIZAÇÃO

A técnica de padronização consiste na transformação de um conjunto de variáveis X (onde já é sabido que a variável possui uma distribuição normal), em uma distribuição normal de média e variância igual a 1. A transformação é feita pela Equação 3:

$$z = \frac{x - \mu}{\sigma} \tag{3}$$

Onde x pertence ao conjunto de valores de X, μ é a média de X, e σ o desvio padrão do conjunto X.

4.4. ANÁLISE DE COMPONENTE PRINCIPAL (PCA)

Grandes conjuntos de dados são cada vez mais comuns e muitas vezes difíceis de interpretar. Jolliffe e Cadima [6] definem a Análise de Componente Principal ou PCA como uma técnica para reduzir a dimensionalidade de cada conjunto de dados, aumentando a interpretabilidade, mas, ao mesmo tempo, minimizando a perda de informações. Neste trabalho definimos o PCA como um tipo de transformação linear onde conjuntos de dados são transformados para um novo conjunto de coordenadas. Essas novas coordenadas são chamadas de componentes principais e representam um novo conjunto de variáveis artificiais (variáveis sintéticas) que são funções lineares daquelas originais. A característica principal das novas componentes resultantes da PCA é a maximização da variância, onde a primeira coordenada possui a maior variância por qualquer projeção dos dados, Maćkiewicz and Ratajczak [7].

4.5. MODELO SIR

O modelo SIR (e variantes) é o modelo mais conhecido e utilizado para o estudo de epidemias, Weber et al. [8]. Segundo o modelo SIR, um indivíduo suscetível (S) em contato com um indivíduo infectado (I) possui uma probabilidade por unidade de tempo, beta, de

virar infetado. A evolução do grupo de suscetíveis em função do tempo é dada pelas Equações 4 até 6:

$$\frac{dS}{dt} = -\beta SI \quad (4)$$

$$\frac{dI}{dt} = +\beta SI - \gamma I \quad (5)$$

$$\frac{dR}{dt} = \gamma I \quad (6)$$

Onde β é a taxa de transmissão e γ é a taxa de recuperação.

5. METODOLOGIA

5.1. SELEÇÃO DE PAÍSES

Dos 195 países no mundo, foram escolhidos 158 com população acima do milhão de habitantes e pelo menos 1 óbito registrado, ver Anexo 1. Os 158 países representam 99,8% da população do mundo. O maior país analisado é a China, com 1.434 milhões; o menor país analisado é Djibuti, com estimativa de 1 milhão em 2020. Em média, os 158 países possuem uma população de 49,5 milhões de habitantes.

Dados dos óbitos por milhão de habitantes foram usados para o estudo dos países. Todos os dados utilizados no projeto foram obtidos de Our World in Data [9].

5.2. ESTRUTURA DE BASE DE DADOS

A base de dados era disponibilizada em formato .csv, onde cada linha corresponde a um dia específico para um país específico. A seguir, a lista de informações utilizadas com respectivas fontes de origem agrupadas segundo Our World in Data.

- Óbitos confirmados por dia por país: repositório de dados do centro de sistemas em ciências e engenharia (CSSE) Johns Hopkins University (JHU) [10].
- Número de reprodução: Francisco Arroyo Marioli, Banco Central de Chile [11].
- População: Nações Unidas, Departamento de Economia e Segurança Social 2019 [12].
- Densidade populacional: World Bank Group [13].
- Mediana de Idade: Nações Unidas, Divisão de População, 2017, [12].
- Parcela da população acima de 65 anos: World Bank Group [13].
- Parcela da população acima de 70 anos: Nações Unidas, Divisão de População, 2017 [12].

- PIB per capita: World Bank Group [13].
- IDH, Índice de Desenvolvimento Humano: World Bank Group [13].
- Índice de pobreza extrema: World Bank Group [13].
- Expectativa de vida: Nações Unidas, Divisão de População [12].
- Pacientes em ICU por milhão: Fontes Individuais para cada país.
- Camas de hospital por milhão: OECD, Eurostat, World Bank [13].
- Mortes por doenças cardiovasculares; Global Burden of Disease Collaborative Network.
- Fumantes femininos: World Bank World Development Indicators [13].
- Fumantes Masculinos: World Bank World Development Indicators [13].
- Facilidade de acesso lavagem de mãos: United Nations Statistics Division [12].
- Excesso de mortalidade: Human Mortality Database (2021) [14].
- Prevalência de Diabetes: World Bank World Development Indicators [12].

5.3. PRÉ-PROCESSAMENTO

O pré-processamento da base de dados contemplou, num primeiro momento, o tratamento de algumas informações faltantes. Não todos os 158 países possuíam as informações disponíveis; por exemplo, dos 158 países, não consta informação do IDH para 4 países. Nesses casos foi inserido o valor da média de IDH dos outros 154 países. Variáveis com muitos países sem informação foram descartadas, como será discutido na etapa de escolha de variáveis.

As variáveis utilizadas possuem ordens de grandeza diferentes entre elas, isto pode causar vieses indesejados na implementação da técnica de PCA, assim, cada variável foi normalizada, ou seja, cada valor individual foi dividido pela média, fazendo o novo valor da média ser 1. Seguidamente o valor da variável foi centralizada em 0. A literatura indica que para a correta implementação do PCA deve ser realizada uma Padronização das variáveis, Jolliffe and Cadima [6]. Contudo, a técnica de padronização deve ser aplicada a distribuições normais, condição que não é satisfeita no conjunto de dados. Na Figura 2 vemos como as variáveis de PIB per capita e fumantes femininos possuem distribuições assimétricas.

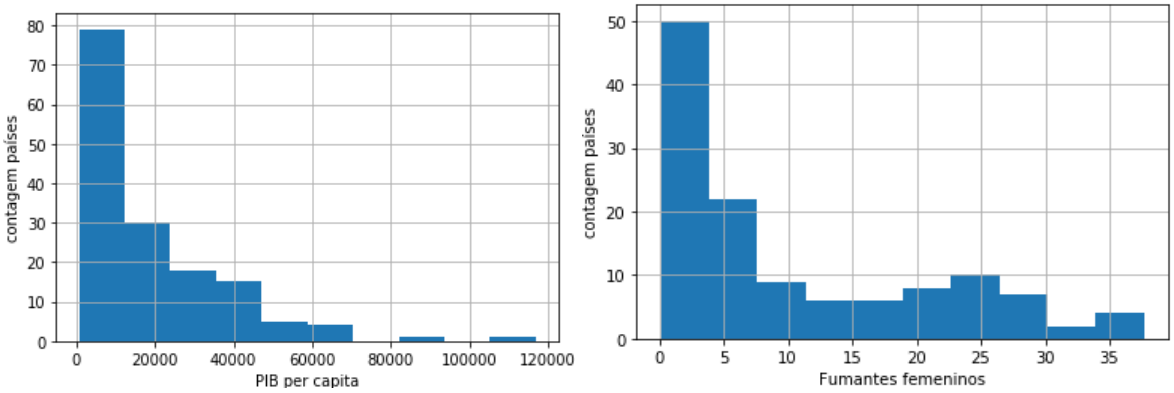


Figura 2: na esquerda, histograma de países por faixa de PIB (US dollars). Na direita, histograma de países por faixa de fumantes femininos (%). Em ambos os casos se observa distribuição assimétrica. Fonte: Própria.

5.4. DEFINIÇÃO DA PRIMEIRA ONDA

Com o objetivo de avaliar a evolução da pandemia em uma janela de tempo equivalente para todos os países utilizamos o número efetivo de reprodução R, definido como o número médio de casos secundários produzidos a partir de um caso, para identificar esse ponto de partida comum.

Os valores do número R usados nesta pesquisa foram obtidos a partir do trabalho de Arroyo et al. [15], os quais desenvolvem uma nova maneira de estimar o número de reprodução eficaz de uma doença infecciosa (R). Esse método de estimativa explora um mapeamento estrutural entre R e a taxa de crescimento do número de indivíduos infectados derivado do modelo SIR básico. Esse último é um modelo matemático de doenças infecciosas, definido pelos acrônimos SIR (Suscetível, Infeccioso, Recuperado) cuja ordem mostra os padrões de fluxo entre os compartimentos.

Por definição R >1 significa que cada indivíduo infectado pode contagiar pela sua vez mais de uma pessoa, assim, enquanto R>1 o número de indivíduos infectados continuará crescendo.

É possível estimar o número de R por meio do modelo SIR a partir da taxa de transmissão e da taxa de recuperação [15]. Na Figura 3 se mostra o número efetivo de reprodução para China, Itália e Estados Unidos. Note-se que para os três países a R começa acima de 1 o que indica um crescimento acelerado do número de infectados, mas eventualmente estabiliza em um valor próximo de 1.

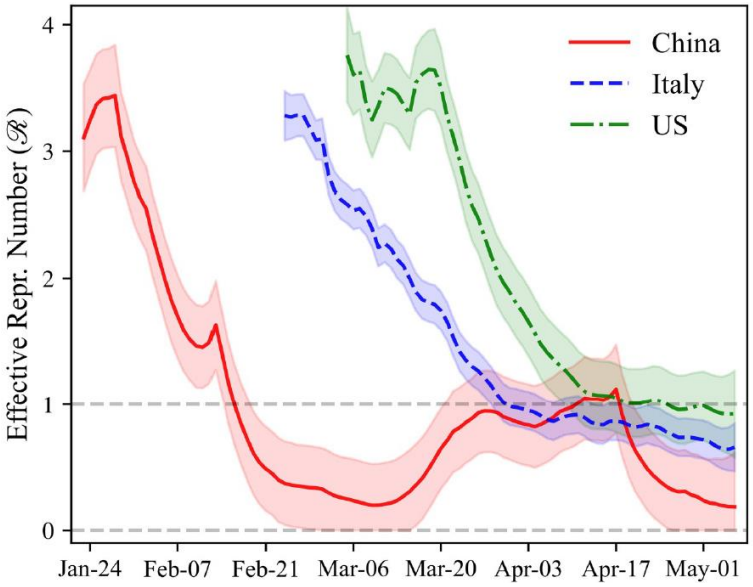


Figura 3: Número de reprodução efetivo R para China, Itália e Estados Unidos. Fonte: Arroyo, et al. [15]

No presente trabalho foi proposto definir como primeira onda o período de tempo do início das contagens de casos até o dia onde o número efetivo de reprodução R for igual ou menor a 1, pela primeira vez. Na Figura 4 descreve-se o algoritmo criado nesta pesquisa para definir o período da primeira onda de cada país.

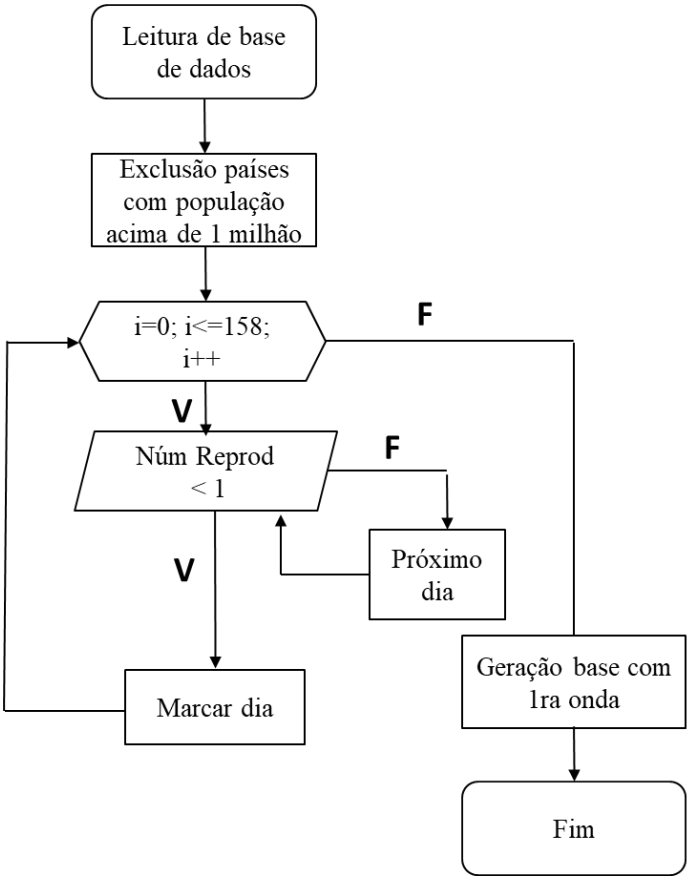


Figura 4: Algoritmo gerado para determinar o período da primeira onda. Fonte: Própria.

Também foram geradas as curvas dos novos casos por dia (média móvel semanal) para Itália, Estados Unidos e Brasil, até o dia onde o respectivo R de cada país atingiu o critério de R <= 1 (ver os gráficos na Figura 5).

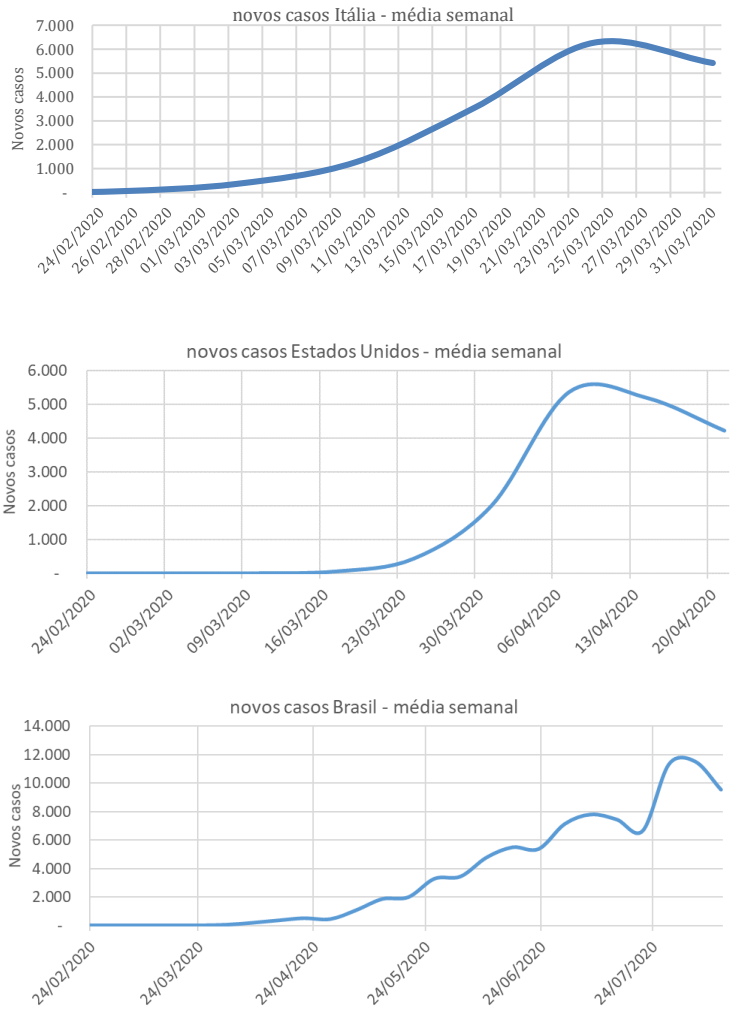


Figura 5: Curva de novos casos por dia (média móvel semanal) para Itália, Estados Unidos e Brasil. Fonte: Própria.

Em média, os 158 países tiveram um tempo de 96 dias desde o primeiro caso até o número de reprodução ficar abaixo de 1. Na Figura 6 se observa a distribuição de países por janelas de tempos da primeira onda.

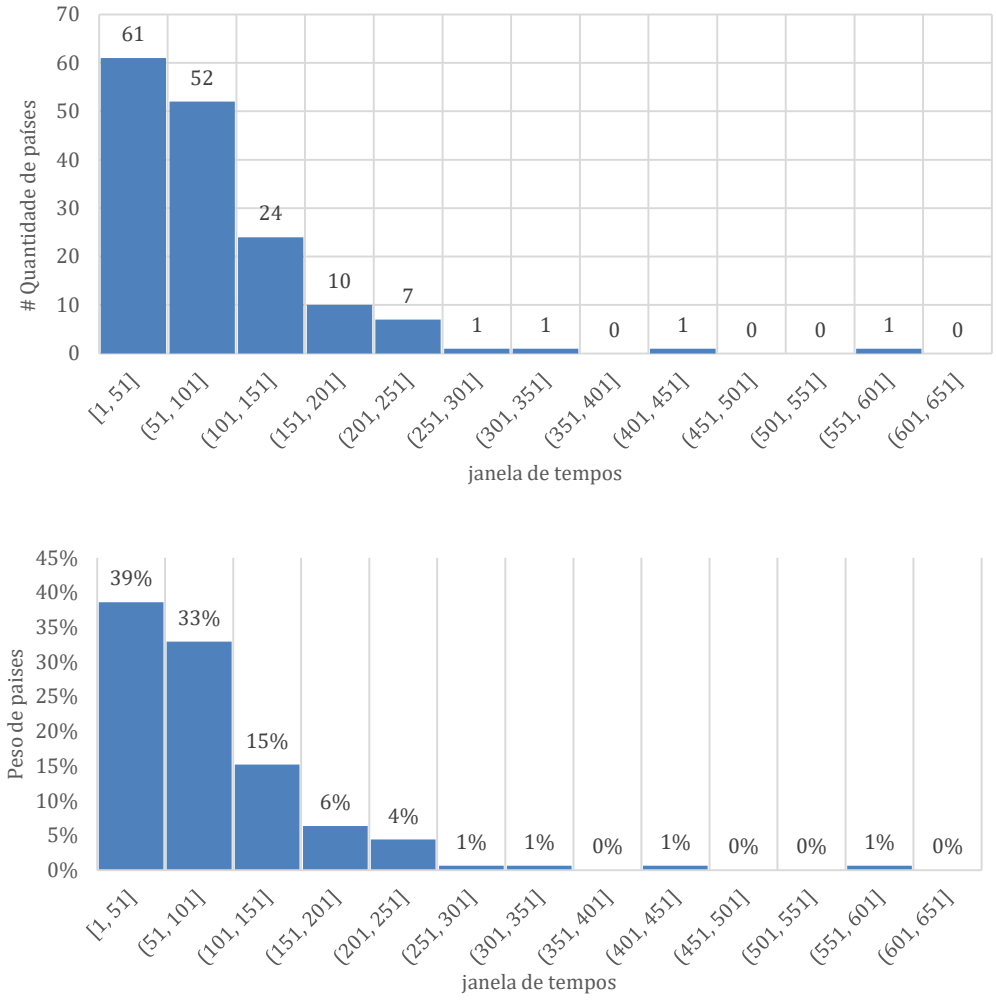


Figura 6: Distribuição de países por janelas de tempos da primeira onda, na figura superior em quantidades, na figura inferior em porcentagem sobre total de países. Fonte: Própria.

5.5. SELEÇÃO DE VARIÁVEIS

Os dados utilizados são disponibilizados por Our World in Data [9], onde são coletadas informações da evolução diária do COVID-19 de todos os países, além de disponibilizadas informações demográficas sobre cada país.

A Tabela 1 contém as informações adicionais disponibilizadas na coluna variáveis, a correlação de tais informações com os óbitos per capita (usando como critério de tempo a primeira onda) e a quantidade de países que possuem tais informações. Por um lado, procurou-se incluir o maior número de variáveis de potencial significância para com a pandemia. Por outra parte, queria-se considerar o maior número de países dentro do critério de inclusão. Como resultado desse compromisso, porcentagem de pacientes em ICU, excesso de mortalidade e disponibilidade de camas em hospitais foram descartadas. Fumantes masculinos, prevalência de diabetes, densidade populacional e camas de hospital

disponíveis foram descartados por possuírem uma correlação muito baixa com o número de mortes per capita por país.

Tabela 1: Correlação de variáveis com óbitos per capita. Fonte: Própria.

Variáveis	Correlação com óbitos	Países
Expectativa de vida	0,51	157
Pacientes ICU por milhão	0,82	16
Camas de Hospital por milhão	0,86	19
Densidade da população	0,05	155
Mediana de Idade	0,49	157
Parcela acima de 65 anos	0,53	155
Parcela acima de 70 anos	0,52	156
PIB por habitante	0,46	154
Índice de pobreza extrema	-0,47	112
Parcela de mortes por doenças cardiovasculares	-0,28	156
Prevalência de diabetes	0,16	155
Fumantes femininos	0,48	125
Fumantes Masculinos	-0,03	123
Facilidade de acesso a lavagem de mãos	0,34	84
Camas de hospital por mil habitantes	0,35	139
Expectativa de vida	0,51	157
Índice de desenvolvimento Humano	0,53	155
Excesso de mortalidade	0,40	76

O fato de não existir correlação com a densidade da população é inesperado, pois existia uma expectativa de que centros urbanos com maior densidade populacional, que têm uma consequente taxa de contágios maior, possuíssem valores maiores de óbitos [16]. Análise semelhante desenvolvida por pesquisadores da Bulgária levou à mesma conclusão sobre a densidade populacional e o número de óbitos per capita [17].

5.6. ANÁLISE DE VARIÁVEIS

A escolha dos parâmetros foi feita com o intuito de separar os países em grupos com valores de óbitos por habitantes diferentes. Na Figura 7 temos todos os países com número de óbitos por milhão no eixo vertical e a mediana da idade da população na horizontal. Os dados usados correspondem aos dados da primeira onda.

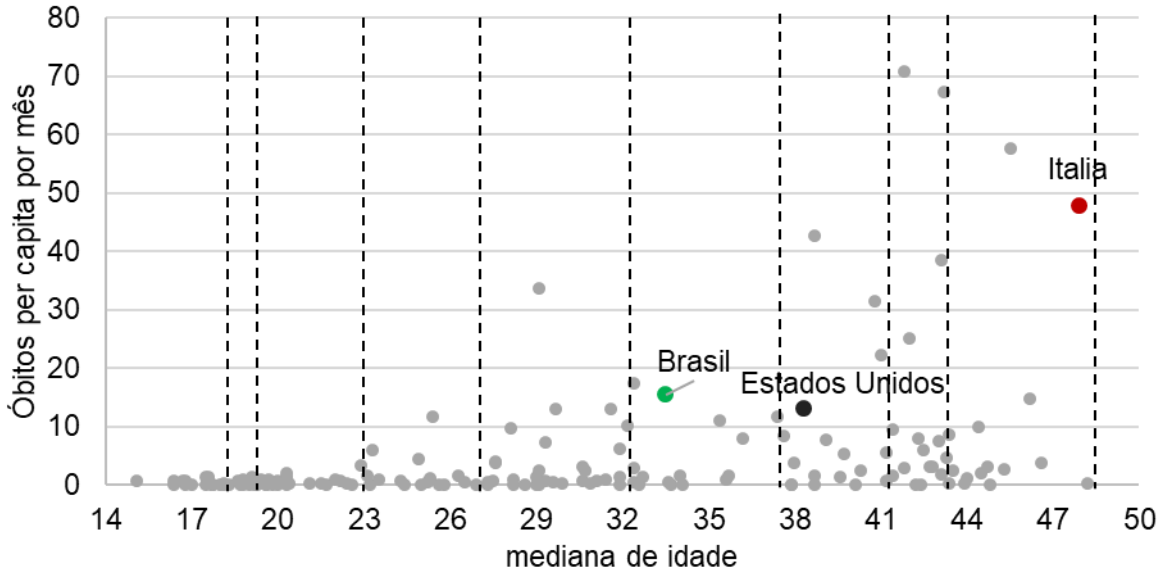


Figura 7: Óbitos per-capita vs mediana da idade da população com o limite das 10 faixas. Brasil, Estados Unidos e Itália foram destacados. Fonte: Própria.

Observa-se que, mesmo com um alto grau de dispersão, é visível que os países com maior quantidade de óbitos per-capita encontram-se com maior concentração do lado dos países com maior idade. Por outro lado, como mostrado na tabela 1 a mediana de idade possui uma correlação de 0,49 com o número de óbitos per capita por mês, assim, é possível validar a idade como variável de separação.

5.7. CLASSIFICAÇÃO DE PAÍSES POR FAIXAS

O algoritmo foi desenvolvido em linguagem Python pela facilidade de implementação de técnicas estatísticas dada pela grande quantidade de documentação e bibliotecas disponíveis. A metodologia escolhida para avaliar a capacidade de uma variável separar os países em grupos com diferentes valores de óbitos per-capita foi inspirada em técnicas utilizadas na área de engenharia da qualidade.

A área de qualidade procura, com base em ferramentas da estatística, criar ferramentas que garantam o bom desempenho de processos industriais mesmo quando os mesmos possuem um alto componente estocástico.

Devido à variabilidade inerente de processos industriais, as medidas individuais de uma característica de qualidade são todas diferentes entre si, mas quando agrupadas elas tendem a formar um certo padrão. Quando o processo é estável, esse padrão pode ser descrito por uma distribuição de probabilidade, Ribeiro e Shwengber [18].

O algoritmo mostrado na Figura 8 pode ser exemplificado a partir da distribuição de países por óbitos por milhão na primeira onda do COVID-19 (normalizado pelo número de

meses de cada período) versus mediana da idade mostrada na Figura 9. Organizando os países de menor a maior mediana de idade podemos criar faixas de mediana, na Figura 9 temos 10 faixas de mediana (15 a 16 países por faixa) com as médias de óbitos por faixa. na Figura 7 temos os limites das faixas gerada.

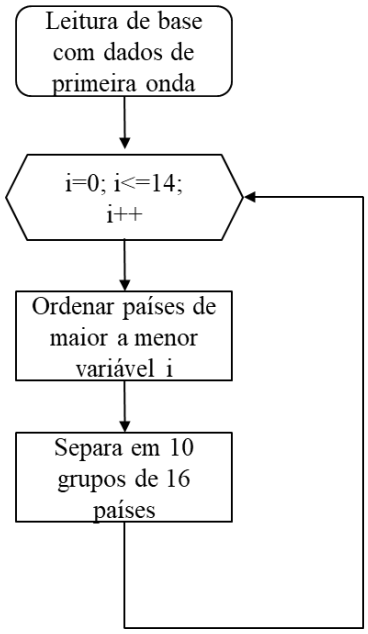


Figura 8: Algoritmo para classificar os países por faixa usando as 14 variáveis escolhidas. Fonte: Própria.

Cada ponto da Figura 9 representa a média de óbitos por milhão para os países da faixa. A partir do desvio padrão da média de óbitos da faixa podemos sinalizar aqueles que estão acima de um desvio padrão da média. A área da curva limite da Figura 4 representa os valores por faixa abaixo da média mais o desvio padrão. A partir do desvio de óbitos por milhão de cada faixa de mediana da idade podemos dividir os países entre aqueles que estão dentro de 1 desvio (dentro da área limite) e aqueles que estão fora de 1 desvio (acima da área).

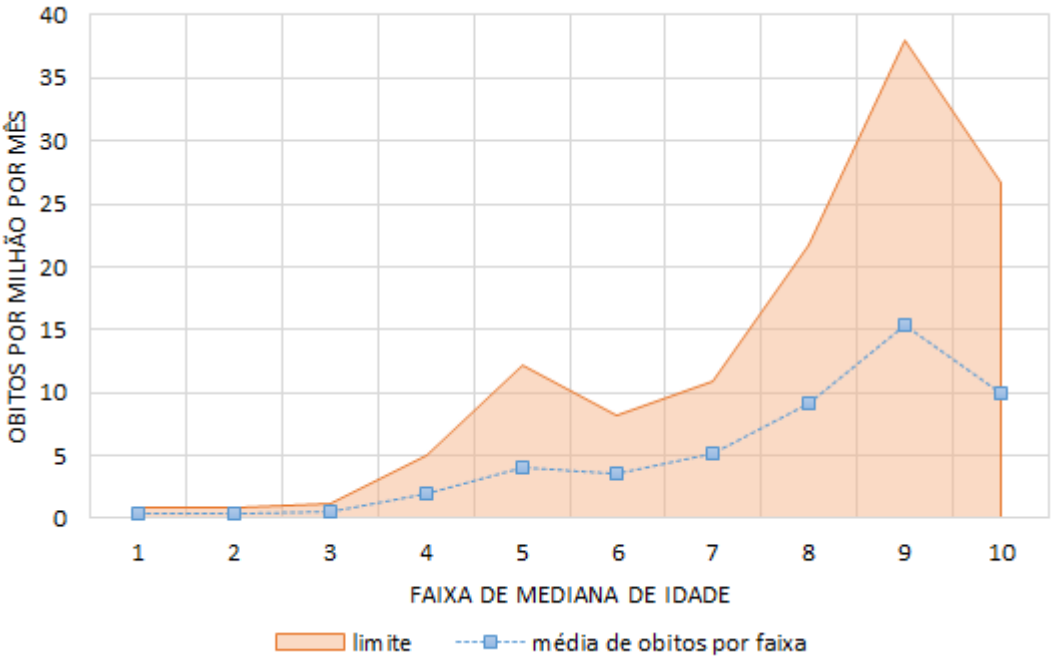


Figura 9: Óbitos per capita por mês vs faixa mediana. Fonte: Própria.

Na Tabela 2 temos os países que compõem a faixa 9, a qual possui óbitos por milhão médios de 15,41 (fazendo a média não ponderada dos países) e um desvio padrão de 37.98..

Tabela 2: Países da faixa 9 de mediana de idade. Fonte: Própria.

País	Mediana idade	Óbitos por milhão	Dias	Faixa 9 mediana idade	Média de óbitos faixa 9	Limite
Dinamarca	42,3	7,96	68	9	15,41	37,98
Singapore	42,4	0,00	38	9	15,41	37,98
Bósnia e Herzegovina	42,5	6,08	47	9	15,41	37,98
Estônia	42,7	3,15	64	9	15,41	37,98
Finlândia	42,8	3,25	84	9	15,41	37,98
România	43	7,58	58	9	15,41	37,98
Suíça	43,1	38,60	38	9	15,41	37,98
Cuba	43,1	1,80	42	9	15,41	37,98
Holanda	43,2	67,37	48	9	15,41	37,98
Czechia	43,3	4,79	37	9	15,41	37,98
Hungria	43,4	8,69	53	9	15,41	37,98
Coreia do Sul	43,4	0,30	45	9	15,41	37,98
Lituânia	43,5	2,39	39	9	15,41	37,98
Latvia	43,9	0,27	42	9	15,41	37,98
Croácia	44	1,15	39	9	15,41	37,98
Áustria	44,4	9,87	36	9	15,41	37,98

Dos 16 países da faixa, Suíça e Holanda possuem média de óbitos por milhão por mês acima da média +1 desvio, podendo ser categorizados como países de risco a partir do critério de mediana da idade. Na Tabela 3 temos todos os países cuja letalidade encontra-se um desvio padrão acima da letalidade média da faixa de mediana de idade.

Tabela 3: Países acima dos óbitos por milhão da faixa de mediana de idade. Fonte: Própria.

País	Óbitos por milhão	Faixa mediana idade	Limite	Acima do limite
Gambia	1,3	1	1,2	1
Burkina Faso	1,3	1	1,2	1
Sierra Leone	1,5	2	1,4	1
Guatemala	3,4	3	2,4	1
Bolivia	11,7	4	6,6	1
Peru	33,7	5	17,5	1
Turquia	13,1	6	12,8	1
Iran	17,5	6	12,8	1
Brasil	15,5	7	15,0	1
Bélgica	70,8	8	46,0	1
Suiça	38,6	9	37,6	1
Holanda	67,4	9	37,6	1
Espanha	57,6	10	35,1	1
Itália	47,9	10	35,1	1

5.8. ROBUSTEZ DA METODOLOGIA DE CLASSIFICAÇÃO POR FAIXAS

Num primeiro momento a ideia de separar os países em especificamente 10 faixas foi uma decisão arbitrária.

Para testar a robustez da separação em 10 faixas foram feitas outras partições em menor ou maior número de faixas. Começando em 5 faxias até 14 passando por todos as possibilidades intermedias: {5.6. ...10. ...14}.

Na matriz que se mostra na Figura 10 foram comparadas a classificação totais dos países em função do número de faixas, a porcentagem de célula da matriz corresponde à porcentagem de concordância das faixas. Por exemplo, com cinco (5) faixas cada país foi comparado com a média de óbitos por faixa de mediana de idade num grupo de 31 países, já com 14 faixas cada país foi comparado com um grupo de 11 países, a matriz mostra que nesse caso 89% dos países tiverem a mesma marcação de risco.

Na figura 11 temos o algoritmo usado para gerar a matriz.

Quando separamos os países em mais ou menos faixas, estamos gerando grupos de comparação diferentes, mesmo assim, vemos que a metodologia aponta na mesma direção independentemente do número de faixas utilizadas.

faixas	5	6	7	8	9	10	11	12	13	14
5	100%									
6	95%	100%								
7	94%	95%	100%							
8	93%	95%	94%	100%						
9	91%	94%	94%	95%	100%					
10	91%	92%	91%	92%	94%	100%				
11	91%	94%	92%	94%	94%	97%	100%			
12	89%	91%	89%	93%	91%	92%	94%	100%		
13	90%	90%	92%	94%	93%	96%	94%	94%	100%	
14	89%	90%	92%	93%	92%	94%	94%	96%	97%	100%

Figura 10: Matriz de faixas de mediana de idade. Fonte: Própria.

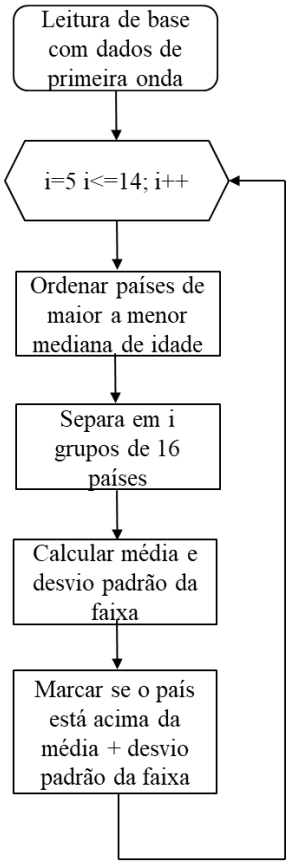


Figura 11: Algoritmo para determinar a matriz de faixas de mediana de idade. Fonte: Própria.

5.9. ANÁLISE DE COLINEARIDADE

Um dos problemas mais frequentes de análises multivariáveis é a existência de correlação entre as diversas variáveis observadas. O cenário ideal é aquele onde todas as variáveis são independentes, o que permite que as contribuições das mesmas para o resultado final de uma regressão sejam independentes. Porém quando variáveis dependentes são utilizadas para uma modelagem estatística, as mesmas podem gerar interpretações erradas sobre a influência das variáveis sobre o resultado final, Mason and Perreault [19].

Uma das técnicas mais utilizadas para revelar a presença de correlações em um conjunto de diferentes variáveis é a elaboração de uma matriz de correlação, onde conste todas as combinações possíveis de correlação entre cada variável. Na Figura 12 se observa a matriz de correlação das variáveis escolhidas, na qual consta o valor absoluto da correlação, isto para ressaltar por meio da paleta de cores quais variáveis possuem alta correlação e quais não.

	Mediana de Idade	Parcela acima de 65 anos	Parcela acima de 70 anos	GDP por habitante	IDH	Fumantes femeninos	% mortes por doenças cardiovasculares
Mediana de Idade	1,0	0,9	0,9	0,6	0,9	0,6	0,3
Parcela acima de 65 anos	0,9	1,0	1,0	0,5	0,8	0,7	0,3
Parcela acima de 70 anos	0,9	1,0	1,0	0,5	0,8	0,7	0,3
GDP por habitante	0,6	0,5	0,5	1,0	0,8	0,4	0,4
IDH	0,9	0,8	0,8	0,8	1,0	0,5	0,4
Fumantes femeninos	0,6	0,7	0,7	0,4	0,5	1,0	0,2
% mortes por doenças cardiovasculares	0,3	0,3	0,3	0,4	0,4	0,2	1,0

Figura 12: Matriz de correlação das variáveis escolhidas. Fonte: Própria.

A matriz nos permite observar que as sete (7) variáveis escolhidas estão fortemente correlacionadas umas com as outras.

5.10. IMPLEMENTAÇÃO DE PCA

Como discutido anteriormente, grande parte das variáveis possuem alta correlação entre elas. A técnica de análise de componente principal nos permite gerar novas variáveis

independentes a partir de combinações lineares das variáveis originais. O primeiro passo para implementar a PCA é a construção de uma matriz de covariância das variáveis sete variáveis selecionadas como relevantes:

- Mediana de idade.
- Parcela da população acima de 65 anos.
- Parcela da população acima de 70 anos.
- GDP por habitante.
- Índice de desenvolvimento humano
- Fumantes femininos
- Porcentagem de óbitos por doenças cardiovasculares.

A primeira etapa para a criação das novas componentes é a elaboração de uma matriz de covariância, ver Figura 13, nela consta a covariância da combinação de cada uma das variáveis escolhidas.

	Mediana de Idade	Parcela acima de 65 anos	Parcela acima de 70 anos	GDP por habitante	Indice de desenvolvimento	Fumantes femeninos	% mortes por doenças
Mediana de Idade	0,10	0,21	0,22	0,21	0,06	0,17	-0,05
Parcela acima de 65 anos	0,21	0,53	0,57	0,39	0,12	0,47	-0,11
Parcela acima de 70 anos	0,22	0,57	0,63	0,41	0,13	0,50	-0,12
GDP por habitante	0,21	0,39	0,41	1,07	0,17	0,32	-0,21
Indice de desenvolvimento Humano	0,06	0,12	0,13	0,17	0,05	0,10	-0,04
Fumantes femeninos	0,17	0,47	0,50	0,32	0,10	0,79	-0,07
% mortes por doenças cardiovasculares	-0,05	-0,11	-0,12	-0,21	-0,04	-0,07	0,21

Figura 13: Matriz de covariância. Fonte: Própria.

A covariância de uma variável com ela mesma é a própria variância da variável, assim, a diagonal da nossa matriz é a variância das variáveis escolhidas. A covariância é comutativa, assim, a matriz é espelhada em relação à diagonal principal.

A próxima etapa é o cálculo dos autovetores e autovalores da matriz de covariância, ver Figura 14. A literatura explica que os autovetores da matriz de covariância representam geometricamente a direção onde é explicada a maior quantidade de variância do sistema, ou

seja, a direção que captura a maior quantidade de informação do conjunto de dados, Maćkiewicz and Ratajczak [20].

	1	2	3	4	5	6	7		
Mediana de Idade	-0,19	-0,01	-0,16	0,12	0,69	-0,66	-0,07		2,24
Parcela acima de 65 anos	-0,45	-0,22	-0,38	0,10	-0,02	0,13	0,76		0,69
Parcela acima de 70 anos	-0,48	-0,24	-0,46	0,09	-0,33	-0,01	-0,61		0,26
GDP por habitante	-0,53	0,79	0,23	0,16	-0,13	-0,02	0,02	autovalores	0,16
Indice de desenvolvimento Humano	-0,12	0,05	-0,07	0,05	0,63	0,73	-0,21		0,02
Fumantes femeninos	-0,47	-0,48	0,72	-0,16	0,04	-0,01	-0,03		0,00
% mortes por doenças cardiovasculares	0,13	-0,17	0,19	0,96	-0,06	0,04	-0,02		0,01

Figura 14: Autovetores e autovalores da matriz de covariância. Fonte: Própria.

Isto é feito partindo da premissa que, quanto maior a variância contida na direção do autovetor, maior a informação contida nessa direção. Por outro lado, os autovalores da matriz de correlação indicam a quantidade de variância contida no respectivo autovalor, assim, como podemos ver na Figura 15 nos primeiros quatro autovetores é contida 99,2% de toda a variância do sistema.

Finalmente para gerar as novas variáveis multiplicamos a matriz transposta dos autovetores (no nosso caso os primeiros 4) pela transposta dos dados originais. Dando como resultado 4 novas variáveis para cada um dos países analisados, ver Equação 5:

$$Componentes\ principal = Matriz\ Autovetores^T * Dados\ Originais^T \quad (5)$$

As novas variáveis foram nomeadas de PC1 (relativo ao primeiro autovetor) PC2, PC3 e PC4. Para entender como variáveis originais contribuíram para a geração de cada uma das componentes podemos retomar os valores dos autovetores da matriz de covariância.

autovetor	autovalores	pesos	ordem de relevância
1	2,24	66,3%	1
2	0,69	20,5%	2
3	0,26	7,7%	3
4	0,16	4,7%	4
5	0,02	0,5%	5
6	0,00	0,1%	7
7	0,01	0,2%	6

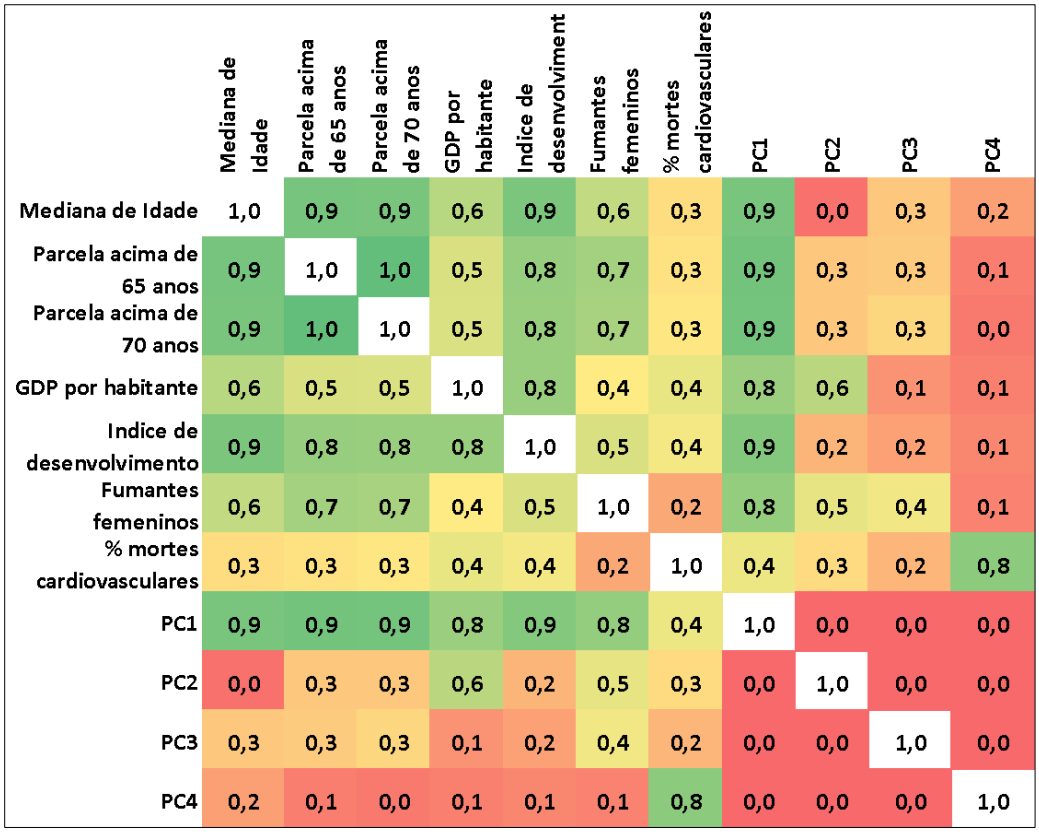
Figura 15: Autovalores da matriz de covariância ordenados na ordem decrescente. Fonte: Própria.

Na Figura 16 se indica do lado esquerdo os autovetores das componentes escolhidas para a análise de componente principal e do lado direito a contribuição de cada variável para o valor da componente. Assim podemos observar que a primeira componente PC1 é formada principalmente pelos valores de Parcela da população acima de 65 anos, parcela acima de 70, GDP por habitante, IDH e Fumantes. Já a segunda componente principal possui uma contribuição de 40% do GDP per-capita e 25% de fumantes femininos, o PC3 é 33% fumantes femininos, 21% parcela acima de 70 anos e 17% parcela acima de 65 e no PC4 a principal contribuição é do % de mortes por doenças cardiovasculares com 58%.

	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
Mediana de Idade	-0,19	-0,01	-0,16	0,12	8%	0%	7%	7%
Parcela acima de 65 anos	-0,45	-0,22	-0,38	0,10	19%	11%	17%	6%
Parcela acima de 70 anos	-0,48	-0,24	-0,46	0,09	20%	12%	21%	5%
GDP por habitante	-0,53	0,79	0,23	0,16	22%	40%	11%	10%
Indice de desenvolvimento Humano	-0,12	0,05	-0,07	0,05	5%	3%	3%	3%
Fumantes femininos	-0,47	-0,48	0,72	-0,16	20%	25%	33%	10%
% mortes por doenças cardiovasculares	0,13	-0,17	0,19	0,96	6%	9%	9%	58%

Figura 16: Do lado esquerdo se tem os autovetores das componentes escolhidas para a análise de componente principal e do lado direito a contribuição de cada variável para o valor da componente. Fonte: Própria.

Na matriz da Figura 17 se observa novamente a correlação das variáveis, dessa vez com as novas componentes junto.



	Mediana de Idade	Parcela acima de 65 anos	Parcela acima de 70 anos	GDP por habitante	Indice de desenvolvimento	Fumantes femeninos	% mortes cardiovasculares	PC1	PC2	PC3	PC4
Mediana de Idade	1,0	0,9	0,9	0,6	0,9	0,6	0,3	0,9	0,0	0,3	0,2
Parcela acima de 65 anos	0,9	1,0	1,0	0,5	0,8	0,7	0,3	0,9	0,3	0,3	0,1
Parcela acima de 70 anos	0,9	1,0	1,0	0,5	0,8	0,7	0,3	0,9	0,3	0,3	0,0
GDP por habitante	0,6	0,5	0,5	1,0	0,8	0,4	0,4	0,8	0,6	0,1	0,1
Indice de desenvolvimento	0,9	0,8	0,8	0,8	1,0	0,5	0,4	0,9	0,2	0,2	0,1
Fumantes femeninos	0,6	0,7	0,7	0,4	0,5	1,0	0,2	0,8	0,5	0,4	0,1
% mortes cardiovasculares	0,3	0,3	0,3	0,4	0,4	0,2	1,0	0,4	0,3	0,2	0,8
PC1	0,9	0,9	0,9	0,8	0,9	0,8	0,4	1,0	0,0	0,0	0,0
PC2	0,0	0,3	0,3	0,6	0,2	0,5	0,3	0,0	1,0	0,0	0,0
PC3	0,3	0,3	0,3	0,1	0,2	0,4	0,2	0,0	0,0	1,0	0,0
PC4	0,2	0,1	0,0	0,1	0,1	0,1	0,8	0,0	0,0	0,0	1,0

Figura 17: Matriz de correlação das variáveis junto com as novas componentes. Fonte: Própria.

Como as componentes principais não são mais que os autovetores da matriz de covariância, eles são todos perpendiculares entre si, pelo qual eles têm correlação de 0% uns com os outros. Na Figura 18 observa-se o algoritmo criado para gerar as componentes principais.

A variável PC1 ou Primeira Componente Principal, possui uma forte correlação com as variáveis de distribuição etária (mediana idade, acima de 65 e acima de 70), fumantes femininos, GDP e IDH, mesmo que a contribuição destas duas últimas seja pequena na combinação linear que gera a componente. A variável PC2 é uma componente onde a maior contribuição vem do GDP per capita, porém, diferente do próprio GPD, não possui correlação forte com as variáveis relativas à distribuição etária. O mesmo acontece com a PC3 e a PC4 onde existe uma correlação forte entre a componente e a variável de maior contribuição (fumantes femininos e % de mortes por doenças cardiovasculares) mas sem ter uma correlação com o resto de variáveis do sistema.

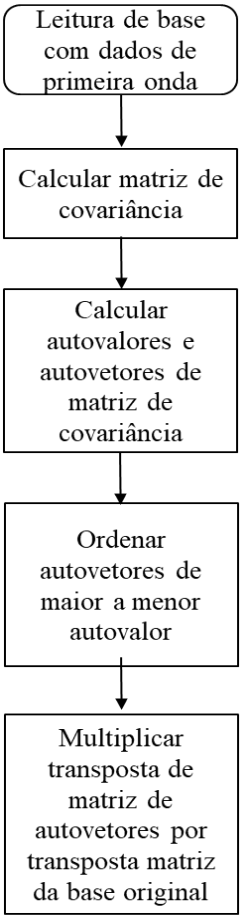


Figura 18: Algoritmo criado para determinar as componentes principais. Fonte: Própria.

6. DISCUSSÃO

Num primeiro momento foi avaliada individualmente a capacidade de cada variável de separar os países em grupos com baixo e alto número de óbitos per capita. Como mostrado nas Figuras 19 até 20, as variáveis relativas à distribuição etária conseguem separar os países em grupos de diferentes magnitudes de óbitos per capita onde países com população mais nova possuem valores de óbitos per capita menores que os países com população mais velha.

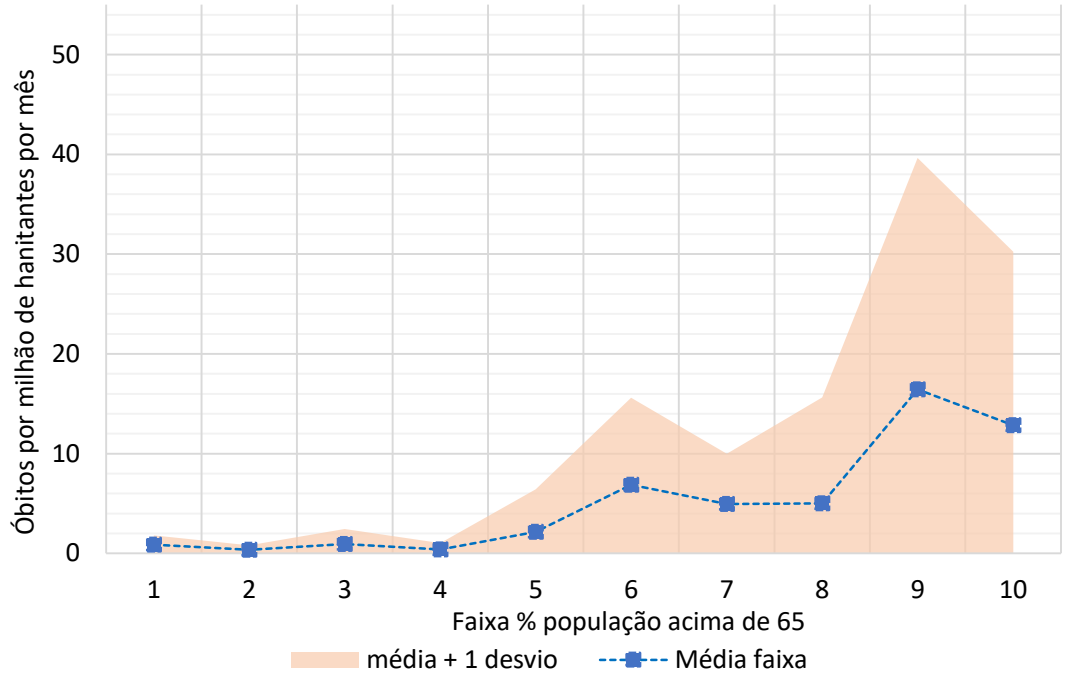


Figura 19: Curva de faixa de idade superior a 65 vs. média mensal de óbitos per capita no período de primeira onda. Fonte: Própria.

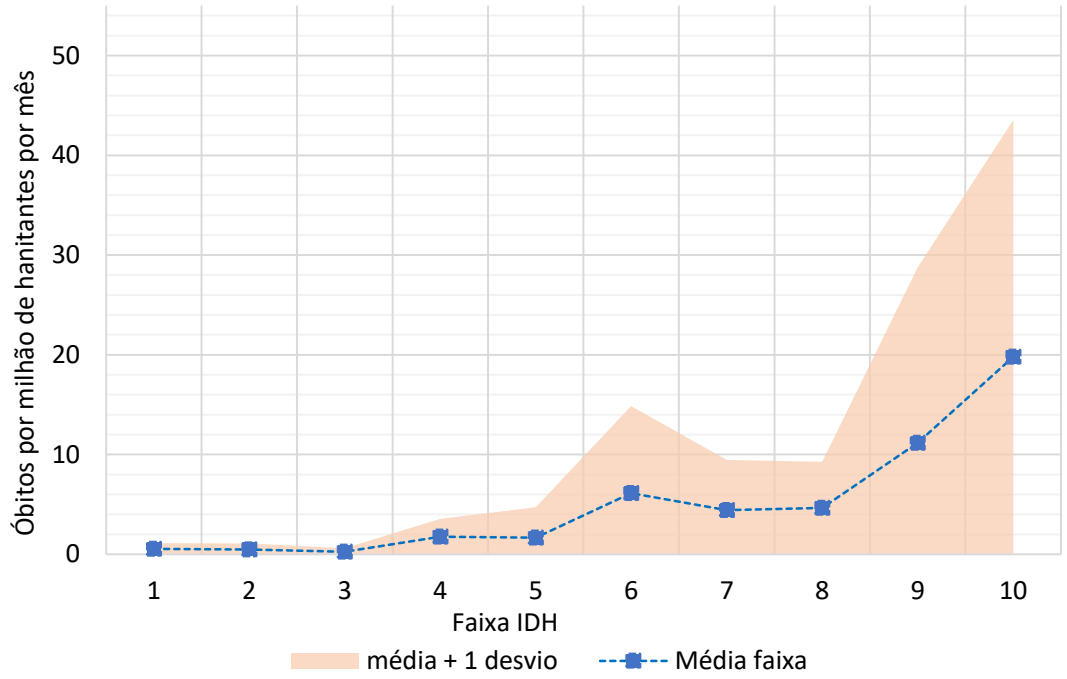


Figura 20: Curva de faixa de idade superior a 70 vs. média mensal de óbitos per capita no período de primeira onda. Fonte: Própria.

Por outro lado, testes com variáveis relativas à economia do país trouxeram resultados contraditórios. Por exemplo, nas Figuras 21 e 22 podemos evidenciar uma correlação positiva entre o produto interno bruto (GDP pelas siglas em inglês) e o total de mortes por

milhão de habitantes, o que é contraditório visto que a expectativa é que países com maior produto interno possuem uma vantagem de recursos à hora de combater a pandemia. A mesma conclusão acontece quando olhamos para os óbitos por milhão contra a faixa de índice de desenvolvimento humano (enquanto maior o número da faixa maior o índice) onde a expectativa, na mesma ideia do dito em relação ao produto interno bruto, vai contra do resultado.

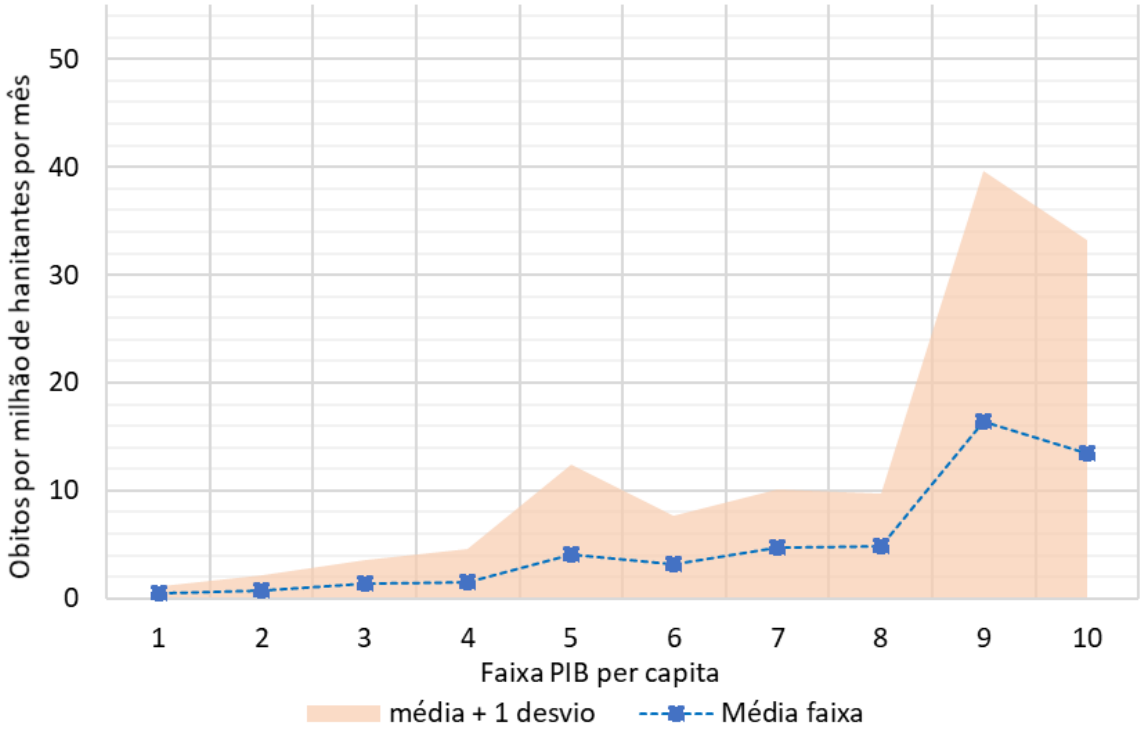


Figura 21: Curva de faixa de PIB per capita vs. média mensal de óbitos per capita no período de primeira onda. Fonte: Própria.

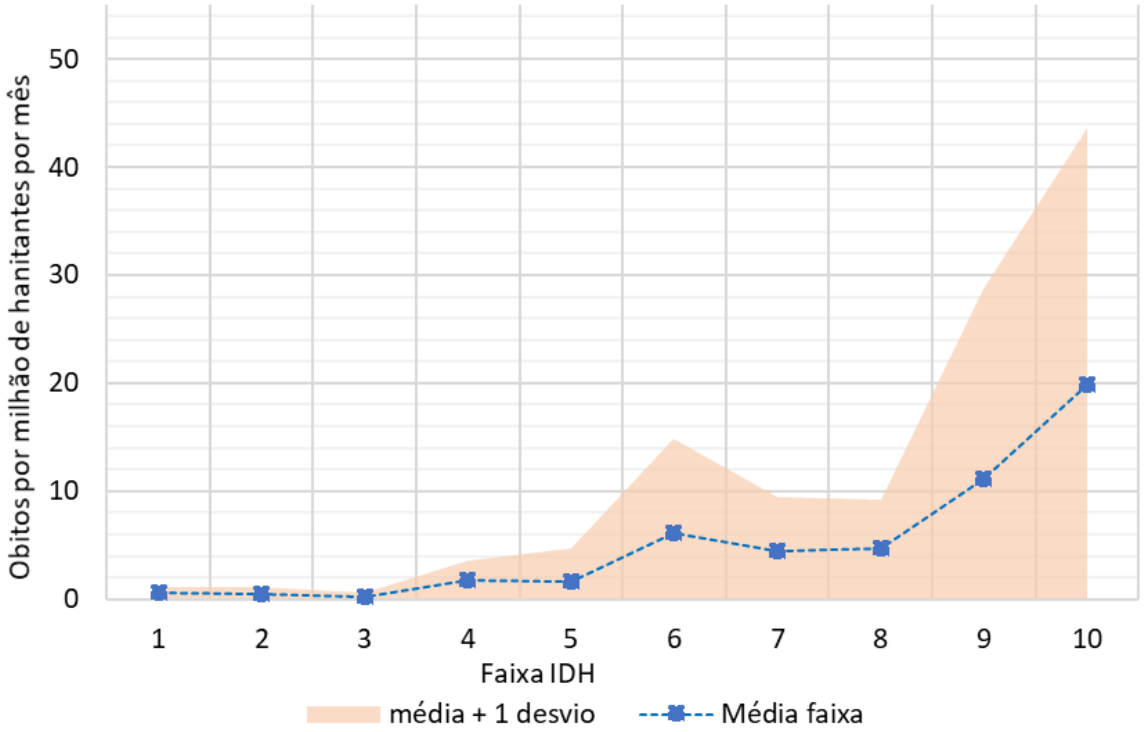


Figura 22: Curva de faixa de IDH per capita vs. média mensal de óbitos per capita no período de primeira onda. Fonte: Própria.

A explicação para esse fenômeno é a forte correlação já evidenciada entre os indicadores econômicos e os indicadores demográficos. Assim, é necessário recorrer a ferramentas como a análise de componente principal para conseguir um conjunto de variáveis independentes a serem trabalhadas.

Partindo do desenvolvimento feito no item 5.10**,** no qual foram geradas as componentes principais do total de variáveis pré-selecionadas para a análise, é necessário avaliar a correlação das novas variáveis com os óbitos per capita. Na Figura 23 temos a matriz de correlação entre as quatro componentes principais geradas junto com os óbitos por milhão. Podemos observar que a primeira componente possui uma clara correlação com a variável de óbitos, já as componentes 2 e 3 não possuem correlação nenhuma e a quarta componente possui uma fraca correlação.

	PC1	PC2	PC3	PC4	Obitos por milhão
PC1	1,00	0,00	0,00	0,00	0,49
PC2	0,00	1,00	0,00	0,00	0,01
PC3	0,00	0,00	1,00	0,00	0,02
PC4	0,00	0,00	0,00	1,00	0,19
Obitos por milhão	0,49	0,01	0,02	0,19	1,00

Figura 23: Matriz de correlação das variáveis principais com os óbitos por milhão de habitantes. Fonte: Própria.

Retomando o discutido do desenvolvimento das componentes principais, a combinação linear da primeira componente possui uma forte contribuição das variáveis relativas à distribuição etária, a segunda componente possui uma maior contribuição do Produto interno bruto, a terceira da porcentagem de fumantes femininos e a quarta da porcentagem de mortes por doenças cardiovasculares.

Assim, das quatro componentes geradas só a primeira componente possui a capacidade de separar os países em grupos com diferentes óbitos per capita. Na Figura 24 temos a curva de faixas de países agrupados pela primeira componente principal contra os óbitos per-capita no período da primeira onda.

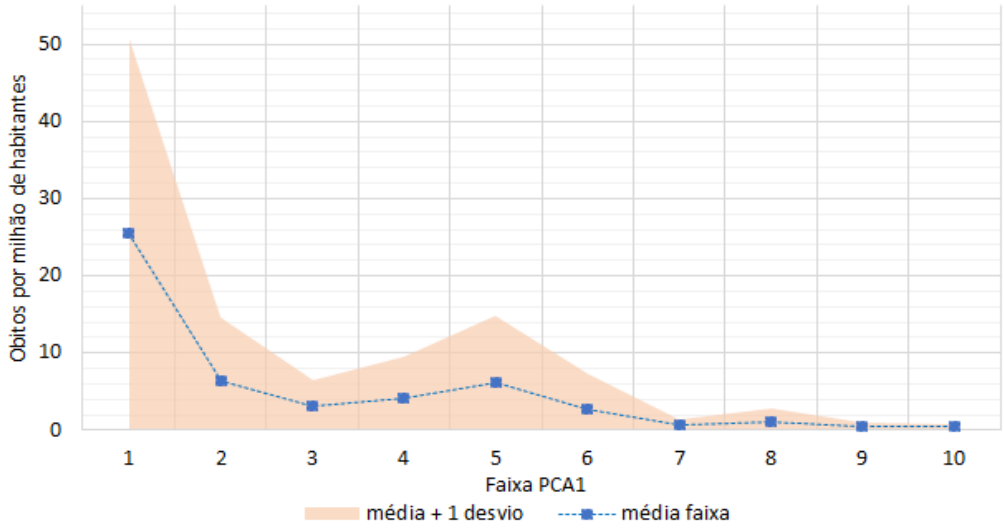


Figura 24: Curva de faixa de PC1 vs. média mensal de óbitos per capita no período de primeira onda. Fonte: Própria.

Nesta análise, reduzimos a um as variáveis a analisar, sendo essa a primeira componente principal da PCA. Na Tabela 4 se mostra o cálculo da quantidade de desvios padrão acima da média para o Brasil, Alemanha, Haiti e Estados Unidos.

Tabela 4: Análise do desvio frente à média da faixa, Brasil, Alemanha, Haiti e Estados Unidos. Fonte: Própria.

Países	PC1	Obitos por mês por milhão	Número da Faixa	Média PC1 faixa	Média + Desvio Faixa	Diferença	**Diferença/ desvio**
Brasil	0,10	16	4	4,24	9,51	6,02	**1,14**
Alemanha	-3,37	4	1	25,52	50,58	-46,73	**-1,86**
Haiti	1,40	1	9	0,51	1,02	-0,26	**-0,51**
EUA	-2,31	13	2	6,40	14,58	-1,52	**-0,19**

O Brasil tem um PC1 de 0,1 o que o deixa na faixa 4, a média de óbitos da faixa é de 4,24 óbitos mensais por milhão de habitantes, já o desvio padrão da faixa 4 é de 5,27. Brasil, no período que foi definido como primeira onde, possui uma média de 16 óbitos mensais por milhão de habitantes, 6,02 acima da média mais um desvio padrão da faixa 4, assim, podemos dizer que o Brasil está a 1,14 desvios padrão da média da faixa 4. Na tabela 5 temos o resumo das médias de óbitos por faixa de PCA.

Tabela 5: Media de: PCA; Òbitos por milhão por mês; desvio padrão. Por faixa de PCA.

FAIXA	Média PCA	Média Óbitos	desvio	# Países
1	-2,83	25,52	25,06	16
2	-2,15	6,40	8,18	16
3	-1,16	3,14	3,26	16
4	-0,11	4,24	5,27	15
5	0,38	6,22	8,58	16
6	0,68	2,65	4,65	16
7	0,97	0,63	0,74	15
8	1,23	1,08	1,74	16
9	1,42	0,51	0,51	16
10	1,62	0,42	0,47	16

Na figura 24 as maiores faixas de PCA possuem uma dispersão muito menor que as primeiras faixas o que dificulta avaliar o limite e a média do grupo para tais faixas. Na figura 26 temos as médias de óbitos por faixas de PCA com o eixo vertical em escala logarítmica.

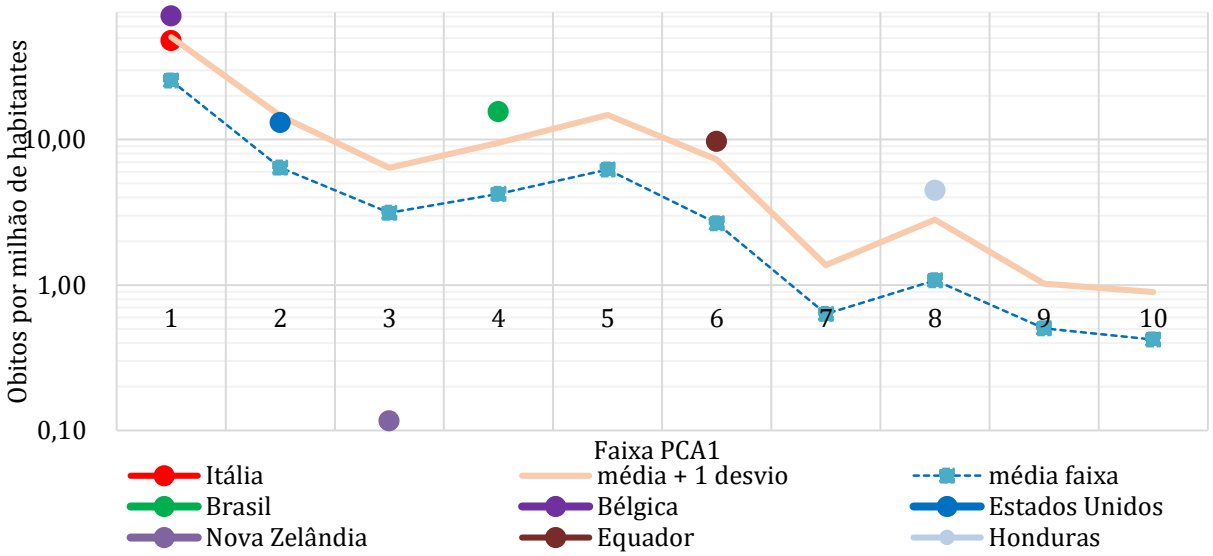


Figura 25: Curva de faixa de PC1 vs. Log média mensal de óbitos per capita no período de primeira onda, constam pontos para Itália, Brasil, Estados Unidos, Nova Zelândia, Bélgica, Equador e Honduras. Fonte: Própria.

Na figura 25 foram colocados os pontos de óbitos por milhão por mês no período de primeira noda para Itália, Brasil, Estados Unidos, Nova Zelândia, Índia e Honduras nas suas respectivas faixas. Estados Unidos possui óbitos por milhão por mês acima de Honduras, contudo a análise de PCA sugere que no caso de Estados Unidos é "justificável" em função da média de países comparáveis, enquanto Honduras possui um valor de óbitos por milhão acima da média do grupo.

Por outro lado, na figura 1 foi apresentado que no início da pandemia Brasil possuía um acumulado de óbitos por milhão abaixo de Itália, Bélgica e em linha com Estados Unidos. A figura 26 mostra que o Brasil, tendo uma distribuição etária menor e consequentemente uma PCA menor possui valores de óbitos por milhão acima do grupo de países comparáveis.

Finalmente, na tabela 6 temos a lista de 23 países com valores acima do limite de óbitos da faixa correspondente. A metodologia desenvolvida indica que esses 23 países possuem o pior desempenho no combate ao COVID-19 no período de primeira onda proposto.

Tabela 6: Lista de 23 países acima do limite da faixa de PCA1.

País	PCA1	Óbitos per capita	Faixa PCA1	Limite
Algeria	0,84	2,38	7	1,37
Bélgica	-2,72	70,78	1	50,58
Brasil	0,10	15,53	4	9,51
Burkina Faso	1,69	1,30	10	0,90
Chile	-1,59	10,97	3	6,40
Equador	0,72	9,71	6	7,30
Espanha	-2,74	57,59	1	50,58
Gambia	1,76	1,29	10	0,90
Holanda	-2,80	67,37	1	50,58
Honduras	1,25	4,49	8	2,82
Iran	0,76	17,48	6	7,30
Liberia	1,64	1,04	10	0,90
Mauritania	1,11	2,11	7	1,37
Mauritius	0,03	11,70	4	9,51
Paquistão	1,33	1,02	9	1,02
Peru	0,50	33,75	5	14,79
Portugal	-2,24	14,77	2	14,58
Reino Unido	-2,36	31,61	2	14,58
Sierra Leone	1,39	1,46	9	1,02
Tajikistan	1,14	6,08	8	2,82
Trinidad e Tobago	-0,47	7,97	3	6,40
Turquia	-0,38	13,07	4	9,51
Yemen	1,47	1,47	9	1,02

CONCLUSÕES

Foi desenvolvida uma metodologia que permitiu avaliar 158 países diferentes em relação ao número de óbitos per capita na primeira onda do covid-19. A partir do modelo SIR foi possível estimar o número médio de reprodução com o qual foi definida uma janela de tempo específica para cada país correspondente à primeira onda. Como consequência de definir diferentes janelas de tempos, foi necessário normalizar os óbitos por milhão pela janela de tempo de cada país.

A partir do conjunto de variáveis disponíveis, analisamos a capacidade das mesmas de separar os países em grupos com mais ou menos óbitos per capita, isto é, por meio de uma metodologia de agregação por faixa. A robustez da metodologia foi revisada por meio de uma matriz onde foram comparadas o risco de cada país relativo a faixas com número de países diferentes, a matriz mostrou que o número de faixas escolhidas para a análise não muda significativamente a marcação final de cada país.

A análise individual levou a conclusões contraditórias no caso de variáveis como PIB e IDH, que foram atribuídas à alta correlação com variáveis demográficas. A técnica de PCA foi proposta para gerar componentes independentes. Por um lado, a implementação de técnica de componente principal foi positiva pois permitiu a criação de uma única variável composta por todas as variáveis escolhidas inicialmente. Por outro lado, o autovetor correspondente à primeira componente principal ainda considera que há uma correlação positiva entre o IDH e o PIB com os óbitos por milhão, enquanto as outras componentes principais (PC2 e PC3) que consideram uma correlação negativa entre IDH e PIB, não possuem nenhuma correlação com número de óbitos. Assim, o resultado final da metodologia desenvolvida permite comparar países em função da sua distribuição etária e identificar países de risco comparando países com características semelhantes.

REFERÊNCIAS

[1] Timeline: WHO's COVID-19 response. World Health Organizations.
https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline

[2] Candido D. et al. Evolutions and epidemic spread of SARS-CoV-2 in Brazil. Journal Science. Vol. 369, No. 6508, 2020. DOI: 10.1126/science.abd2161.

[3] BBC News. Mortos por Coronavirus: por que a Bélgica tem a maior taxa de mortalidade por COVID-19 no mun do. <https://www.bbc.com/portuguese/internacional-52497090>.

[4] Rutger A. Middelburg, Frits R. Rosendaal, COVID-19: How to make between-country comparisons, *International Journal of Infectious Diseases*,Volume 96, 2020.

[5] Cohen, Israel; Huang, Yiteng; Chen, Jingdong; Benesty, Jacob. (2009). Pearson Correlation Coefficient. Noise Reduction in Speech Processing. Volume 2.

[6] Jolliffe I.T. and Cadima J. Principal component analysis: a review and recent development. Philosophical Transactions. Royal Society. A374: 20150202.
https://doi.org/10.1098/rsta.2015.0202

[7] Andrzej Maćkiewicz and Waldemar Ratajczak, Principal components analysis (PCA). Journal Computers & Geosciences. Vol. 19, Issue 3, 1993.

[8] Weber A., Iannelli F., Gonçalves S. Trend analysis of the COVID-19 pandemic in China and the rest of the world. MedRxiv Journal, 2020.
doi: https://doi.org/10.1101/2020.03.19.20037192,2 020.

[9] Our World in Data. <https://ourworldindata.org/coronavirus>

[10] WHO Coronavirus Disease (COVID-19) Dashboard. https://covid19.who.int/

[11] Johns Hopkins University https://dataservices.library.jhu.edu/

[12] Nações Unidas https://population.un.org/wpp/Publications/

[13] World Bank Group < https://data.worldbank.org/ >

[14] Human Mortality Database < https://mortality.org>

[15] Arroyo M. F. and Bullano F. Tracking R of COVID-19: A new real-time estimation using the Kalman filter. Plos One Journal, 2021.
https://doi.org/10.1371/journal.pone.0244474 J

[16] Universal scaling law for human-to-human transmission diseases Ben-Hur Francisco Cardoso, Sebastián Gonçalves Europhysics Letters, 133 (2021) 58001

[17] Dimitar Valev, Relationships of total COVID-19 cases and deaths with ten demographic, economic and social indicators. MedRxiv Journal. ID: ppmedrxiv-20188953.
https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/pt/ppmedrxiv-20188953

[18] José Luís Duarte Ribeiro e Carla Shwengber ten Caten. Porto Alegre: FEENG/UFRGS, 2012. 172p. (Série Monográfica Qualidade) ISBN 85-88085-10-0 Universidade Federal do Rio Grande do Sul. Escola de Engenharia. Programa de Pós-Graduação em Engenharia de Produção. III. Título. IV. Série CDU519.2

[19] Mason, C. H. and Perreault, W. D. Collinearity, Power, and Interpretation of Multiple Regression Analysis. Journal of Marketing Research, 1991.

ANEXOS

ANEXO 1: Lista de 158 países escolhidos para a análise.

1	Afghanistan	41	El Salvador	81	Liberia	121	Senegal
2	Albania	42	Equatorial Guinea	82	Libya	122	Serbia
3	Algeria	43	Eritrea	83	Lithuania	123	Sierra Leone
4	Angola	44	Estonia	84	Madagascar	124	Singapore
5	Argentina	45	Eswatini	85	Malawi	125	Slovakia
6	Armenia	46	Ethiopia	86	Malaysia	126	Slovenia
7	Australia	47	Finland	87	Mali	127	Somalia
8	Austria	48	France	88	Mauritania	128	South Africa
9	Azerbaijan	49	Gabon	89	Mauritius	129	South Korea
10	Bahrain	50	Gambia	90	Mexico	130	South Sudan
11	Bangladesh	51	Georgia	91	Moldova	131	Spain
12	Belarus	52	Germany	92	Mongolia	132	Sri Lanka
13	Belgium	53	Ghana	93	Morocco	133	Sudan
14	Benin	54	Greece	94	Mozambique	134	Sweden
15	Bolivia	55	Guatemala	95	Myanmar	135	Switzerland
16	Bosnia e Herzegovina	56	Guinea	96	Namibia	136	Syria
17	Botswana	57	Guinea-Bissau	97	Nepal	137	Taiwan
18	Brazil	58	Haiti	98	Netherlands	138	Tajikistan
19	Bulgaria	59	Honduras	99	New Zealand	139	Tanzania
20	Burkina Faso	60	Hong Kong	100	Nicaragua	140	Thailand
21	Burundi	61	Hungary	101	Niger	141	Timor
22	Cambodia	62	India	102	Nigeria	142	Togo
23	Cameroon	63	Indonesia	103	North America	143	Trinidad e Tobago
24	Canada	64	Iran	104	North Macedonia	144	Tunisia
25	Rep Africa Central	65	Iraq	105	Norway	145	Turkey
26	Chad	66	Ireland	106	Oman	146	Turkmenistan
27	Chile	67	Israel	107	Pakistan	147	Uganda
28	China	68	Italy	108	Palestine	148	Ukraine
29	Colombia	69	Jamaica	109	Panama	149	Emiratos Arabes
30	Congo	70	Japan	110	Papua New Guinea	150	United Kingdom
31	Costa Rica	71	Jordan	111	Paraguay	151	United States
32	Cote d'Ivoire	72	Kazakhstan	112	Peru	152	Uruguay
33	Croatia	73	Kenya	113	Philippines	153	Uzbekistan
34	Cuba	74	Kosovo	114	Poland	154	Venezuela
35	Czechia	75	Kuwait	115	Portugal	155	Vietnam
36	Congo	76	Kyrgyzstan	116	Qatar	156	Yemen
37	Denmark	77	Laos	117	Romania	157	Zambia
38	Rep Dominicana	78	Latvia	118	Russia	158	Zimbabwe
39	Ecuador	79	Lebanon	119	Rwanda		
40	Egypt	80	Lesotho	120	Saudi Arabia		