What are the main factors affecting case-fatality and mortality rates of COVID-19 in different countries?

Barış Engin
Introduction to Data Analysis
Galatasaray University
Istanbul, Turkey

***Abstract*—This document consists of the description and examination of data concerning COVID-19 cases and outcomes in relation with the development rate of a country. It tries to link the existing mortality rates with numerous factors defining a country's development rate.**

***Index Terms*—covid, coronavirus, development rate, country, data**

I. Introduction

COVID-19 is a contagious disease that was first identified in Wuhan, China in late-2019. It has spread globally ever since, causing a worldwide pandemic. In this paper, we will look into numerous datasets regarding factors of development in countries and analyze them alongside case-fatality and mortality rates of COVID-19 in different countries.

First, we will introduce the dataset. Next, we'll ask a high-level research question and low-level technical questions in order to reach our answer. Finally, we will explain the methods to use in order to answer the question asked.

II. The Dataset

Firstly, we need to introduce the dataset that we are going to work with in detail. I created the dataset on Excel by combining various datasets that include factors that we predict will be useful in our analysis. The dataset consists of nine columns and puts 191 countries together in order to analyze them and find an answer to our question (Fig. 1).

- Country: Consists of 191 countries that we will use to analyze the dataset. Any column that comes after this is related to each given country from this column.
- Country Codes: ISO Alpha-2 country code for every 191 country.
- Confirmed Cases: The amount of confirmed cases of COVID-19 as of December 1st, 2022. Numeric.
- Deaths: The amount of deaths due to COVID-19 as of December 1st, 2022. Numeric.
- Case-Fatality: The percentage of outcomes of COVID-19 cases resulting in death as of December 1st, 2022. Numeric.
- Mortality: The rate of deaths per 100.000 people as of December 1st, 2022. Numeric.
- Elderly Population: The percentage of elderly population (65 years old or more) as of 2021. Numeric.
- Healthcare Index: The estimation of the overall quality of the health care system as of 2022. Numeric.
- Human Development Index: The development rate of a country calculated by taking into account four different factors (mean years of schooling, expected years of schooling, life expectancy at birth, and gross national income (GNI) per capita. Numeric.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 191 entries, 0 to 190
Data columns (total 9 columns):
 #   Column                                  Non-Null Count  Dtype
---  ------                                  --------------  -----
 0   COUNTRY                                 191 non-null    object
 1   COUNTRY CODES                           190 non-null    object
 2   CONFIRMED CASES                         191 non-null    int64
 3   DEATHS                                  191 non-null    int64
 4   CASE-FATALITY                           191 non-null    float64
 5   MORTALITY                               191 non-null    float64
 6   ELDERLY POPULATION (% OF TOTAL IN 2021)  182 non-null    float64
 7   HEALTHCARE INDEX (2022)                 94 non-null     float64
 8   HUMAN DEVELOPMENT INDEX (2021)          191 non-null    float64
dtypes: float64(5), int64(2), object(2)
memory usage: 13.6+ KB
```

Fig. 1. The output of the input "df.info()" for our dataset.

This dataset has numerous null values at the Healthcare Index and Elderly Population columns. For the sake of our data's consistency, while taking into account these columns, we will ignore all countries that don't have this piece of information.

A. Countries

We selected 191 different countries from which mortality and fatality rate, alongside with numerous other factors could be extracted. Considering the fact that United Nations has 193 official member states and that the excluded states have a minor impact on the whole, the amount of countries we selected can very well be considered as representative of the general population.

Figures below give us the numerical values in our dataset (Fig. 2) and the case-fatality to mortality ratio in different countries (Fig. 3).

In Fig. 3. we notice certain extremes like an insanely high case-fatality rate for Yemen (18.1%) and quite a high mortality rate for Peru (659 deaths per 100.000 people).

```
df.describe().apply(lambda s: s.apply('{0:.5f}'.format))
```

	CONFIRMED CASES	DEATHS	CASE-FATALITY	MORTALITY	ELDERLY POPULATION (% OF TOTAL IN 2021)	HEALTHCARE INDEX (2022)	HUMAN DEVELOPMENT INDEX (2021)
count	191.00000	191.00000	191.00000	191.00000	182.00000	94.00000	191.00000
mean	3367621.39267	34737.69110	1.43560	123.15958	9.24725	63.51660	0.72010
std	9802878.18108	110447.19039	1.66878	131.47529	6.69455	10.57255	0.15275
min	3430.00000	7.00000	0.00000	0.32000	1.00000	38.24000	0.36000
25%	50543.00000	676.50000	0.60000	14.41500	3.25000	55.65000	0.60000
50%	344710.00000	4019.00000	1.10000	75.31000	7.00000	63.85000	0.74000
75%	1878670.00000	19270.50000	1.95000	198.30000	15.00000	72.07250	0.83500
max	98788140.00000	1080444.00000	18.10000	659.33000	29.00000	86.04000	0.96000

Fig. 2. The output of the input ”df.describe()” for our dataset.

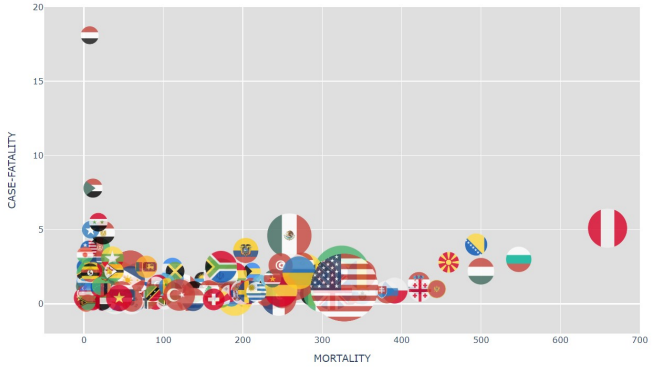

Fig. 3. Case-fatality to Mortality ratio of different countries. Size indicates death toll.

B. Country Codes

These are the ISO Alpha-2 country codes that we used to import flags for our bubble charts.

```
In [45]: print(df[['COUNTRY CODES']].to_string(index=False))
```

```
COUNTRY CODES
           AF
           AL
           DZ
           AD
           AO
           AG
           AR
           AM
           AU
           AT
           AZ
           BS
           BH
           BD
           BB
           BY
           BE
           BZ
```

Fig. 4. Examples of ISO Alpha-2 country codes.

Subsequent graphs will utilize code below with parameters modified accordingly.

C. Confirmed Cases and Deaths

These columns store integer values showing the amount of confirmed COVID-19 cases and deaths due to COVID-19 as of December 1st, 2022 in each country. Case-fatality is calculated using this data. Total population is needed in addition to calculate mortality. Data from these columns will be utilized to size our bubbles in bubble charts.

```
In [63]: from PIL import Image
import plotly.express as px

fig = px.scatter(
    df,
    x="MORTALITY",
    y="CASE-FATALITY",
    hover_name="COUNTRY",
    hover_data=["MORTALITY", "CASE-FATALITY", "DEATHS"]
)
fig.update_traces(marker_color="rgba(0,0,0,0)")

maxDim = df[["MORTALITY", "CASE-FATALITY"]].max().idxmax()
maxi = df[maxDim].max()
for i, row in df.iterrows():
    country_iso = row["COUNTRY CODES"]
    fig.add_layout_image(
        dict(
            source=f"https://raw.githubusercontent.com/matahombres/CSS-Country-Flags-Rounded/master/flags/{country_iso}.png",
            xref="x",
            yref="y",
            xanchor="center",
            yanchor="middle",
            x=row["MORTALITY"],
            y=row["CASE-FATALITY"],
            sizing="contain",
            sizex=np.sqrt(row["DEATHS"] / df["DEATHS"].max()) * maxi * 0.1 + maxi * 0.03,
            sizey=np.sqrt(row["DEATHS"] / df["DEATHS"].max()) * maxi * 0.1+ maxi * 0.03,
            opacity=0.8,
            layer="above"
        )
    )

fig.update_layout(height=600, width=1000, yaxis_range=[-2, 20], xaxis_range=[-50, 700], plot_bgcolor="#dfdfdf")

fig.show()
```

Fig. 5. Piece of code used to display Fig. 3.

D. Case-fatality and Mortality

These columns will be compared to other columns subsequently in order to find an answer to our question. Each following column will have a graph accompanying it, comparing it relative to the columns case-fatality and mortality.

E. Elderly Population

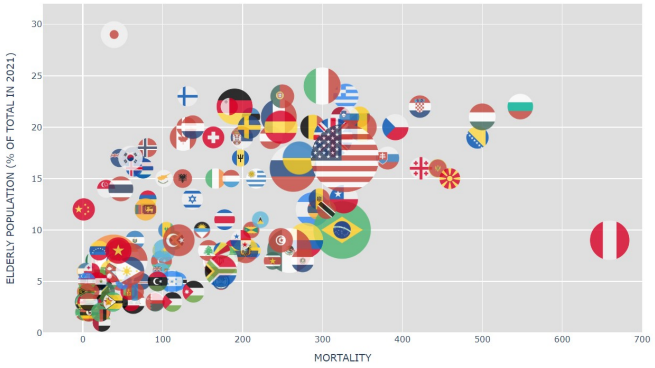

Fig. 6. Elderly Population to Mortality ratio of different countries.

In Fig. 6. we can see that as the elderly population ratio of a country increases, its mortality rate also increases. There are some exceptions to this like Japan with a low mortality rate despite having an older population and Peru with a high mortality rate despite having a younger population.

We cannot see the relation that we saw in Fig. 6. in Fig. 7. This time, visually speaking, it seems like the ratio of elderly population is not affecting the case-fatality rate as strongly as it affects mortality.

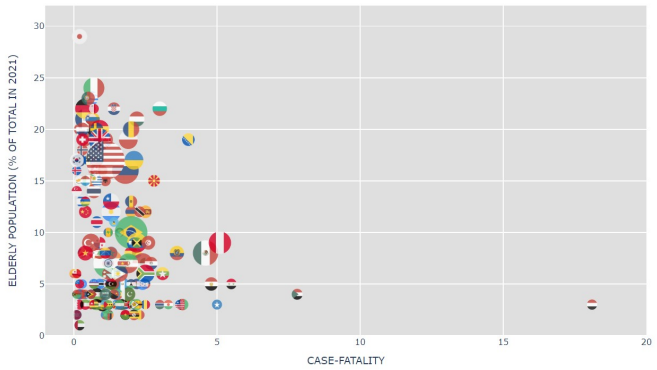

Fig. 7. Elderly Population to Case-fatality ratio of different countries.

F. Healthcare Index

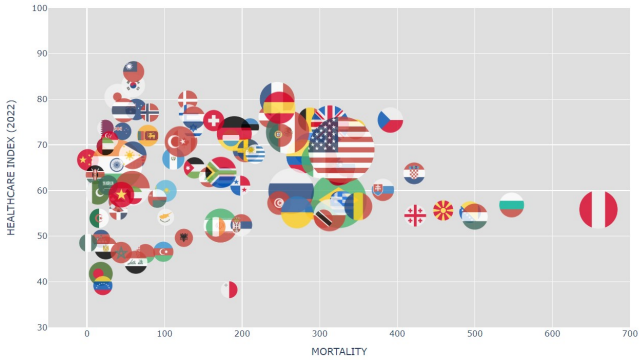

Fig. 8. Healthcare Index to Mortality ratio of different countries.

In Fig. 8. we can see that countries that have above a certain level of mortality do not have a healthcare index exceeding 60. This may be an indicator that the quality of healthcare given in a country affects COVID-19 deaths, however this will be more clear in the case-fatality graph since mortality is based on total population and not only sick people.

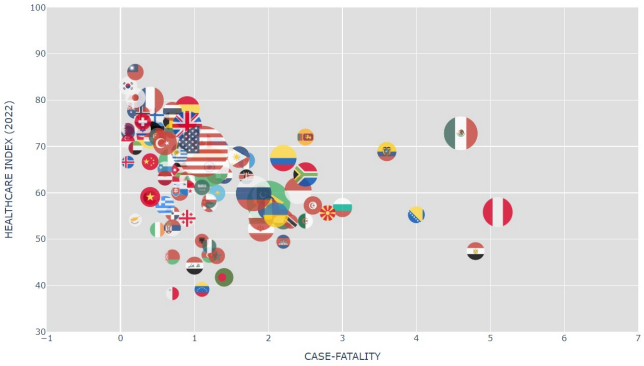

Fig. 9. Healthcare Index to Case-fatality ratio of different countries.

In Fig. 9. we can see clearly that as healthcare index decreases, case-fatality rate increases. There are minor exceptions such as Mexico with a high healthcare index but also a high case-fatality rate.

G. Human Development Index

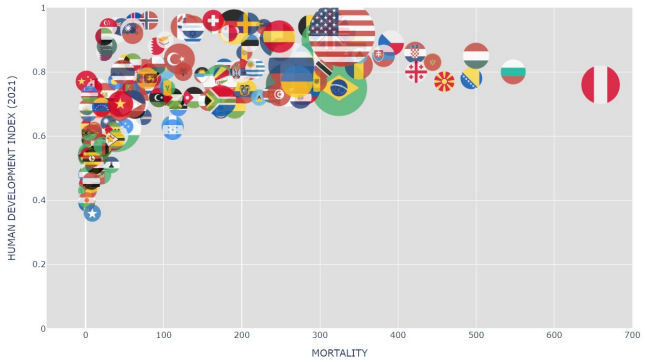

Fig. 10. Human Development Index to Mortality ratio of different countries.

In Fig. 10. we see that as HDI increases, mortality also gradually increases. This may be due to the fact that advanced countries with a higher HDI have more clear and reliable data with higher detection of COVID-19, resulting in increased death tolls these countries.

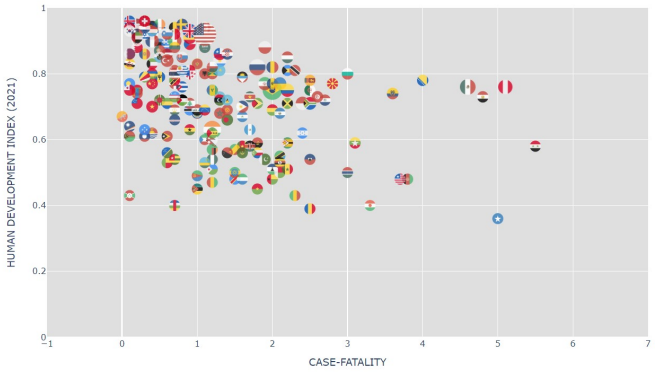

Fig. 11. Human Development Index to Fatality ratio of different countries.

Fig. 11. on the other hand compares HDI with case-fatality. In this graph, we can see that as HDI decreases, case-fatality increases. This is also an indication that HDI may also be an important factor when considering COVID-19 deaths.

III. Research Question and Methods

A. High Level Research Question

For this paper, we wish to see what factors come into play when measuring mortality and case-fatalities. In the future, another pandemic may devastate the world and we need to be more prepared for that. So for that reason we ask: What are the main factors affecting case-fatality and mortality rates of COVID-19 in different countries?

B. Low Level Technical Questions

In order to find an answer to the high level question, we will define certain low level questions:

- Which parameters aid to separate high mortality and case-fatality rates from low ones?
- Are there external factors that need to be taken into account when attempting to verify the validity of this data? If yes, what?

C. Methods

We will attempt to make Z-Tests or T-Tests in order to verify certain hypotheses that we may have while analyzing our data. We will then construct tables and models based on methods that we will learn in the data analysis class during the second half of the semester.