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

II. The Dataset

III. Research Question and Methods

IV. The Construction of Hypothesis Tests and Methods

1) H_0: A higher human development index in a country does not mean lower COVID-19 mortality rates.
   H_a: A higher human development index in a country means lower COVID-19 mortality rates.

2) H_0: A higher human development index in a country does not mean lower COVID-19 case-fatality rates.
   H_a: A higher human development index in a country means lower COVID-19 case-fatality rates.

3) H_0: The ratio of elderly population in a country does not affect COVID-19 mortality rates.
   H_a: The ratio of elderly population in a country does affect COVID-19 mortality rates.

4) H_0: The ratio of elderly population in a country does not affect COVID-19 case-fatality rates.
   H_a: The ratio of elderly population in a country does affect COVID-19 case-fatality rates.

5) H_0: The healthcare index of a country does not affect COVID-19 mortality rates.
   H_a: The healthcare index of a country does affect COVID-19 mortality rates.

6) H_0: The healthcare index of a country does not affect COVID-19 case-fatality rates.
   H_a: The healthcare index of a country does affect COVID-19 case-fatality rates.

Methods to utilize:

- For hypothesis tests number 1 and 2, we will compare the human development index ratings considered high with the ones that aren't, and since the population standard deviation is known and the sample sizes are greater than 30, we will use the Z-test.
- For hypothesis tests number 3, 4, 5 and 6, we will compute the covariance and the Pearson correlation coefficient of the two datasets that we will compare, and see if there is a relation between them.

V. Tests

A. Does higher HDI mean lower mortality/fatality from COVID-19?

A higher human development index may be the one of the factors that reduce COVID-19 related casualties. We will compare the values considered high from this factor with the corresponding countries' COVID-19 case-fatality rates and mortality rates.

Hypothesis test 1:

We will do a one-tailed Z-test between the population and the sample of countries that have a very high HDI ($>$ 0.8). Here is the formula for applying a Z-test between the population and a sample:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Where:

$\bar{x}$ = Mean mortality rate of all countries

μ = Mean mortality rate of countries with a HDI higher than 0.8

σ = Mortality standard deviation for all countries

n = Sample size of countries with a HDI higher than 0.8

```
df["MORTALITY"].describe()
```

```
count    191.000000
mean     123.159581
std      131.475294
min        0.320000
25%       14.415000
50%       75.310000
75%      198.300000
max      659.330000
Name: MORTALITY, dtype: float64
```

Fig. 1. The descriptive statistics of the mortality column

From Fig. 1. we have H_0 = $\mu \geq 123.16$ and H_a = $\mu < 123.16$

```
df_mh = df[df["HUMAN DEVELOPMENT INDEX (2021)"] >= 0.8]
df_mh["MORTALITY"].describe()
```

```
count     69.000000
mean     203.343188
std      127.610643
min       23.740000
25%       90.630000
50%      197.600000
75%      304.900000
max      547.450000
Name: MORTALITY, dtype: float64
```

Fig. 2. The descriptive statistics of the mortality column for countries with a HDI higher than 0.8

Immediately from Fig. 2. we see that the sample mean μ is in fact higher than the population mean $\bar{x}$ so we can almost be certain that our alternative hypothesis will be rejected, but let's proceed with the test anyway.

Let's choose a confidence level of 95% which corresponds to a Z-score of 1.645. If our Z-score exceeds this value we will reject the null hypothesis, else we will accept it.

Our Z-score is:

$$z = \frac{123.16 - 203.34}{\frac{131.48}{\sqrt{69}}} \tag{1}$$

$$= \frac{-80.18}{15.83} \approx -5.07 \tag{2}$$

Our Z-score -5.07 $<$ 1.645, therefore we accept the null hypothesis and we can say with a level of 95% confidence that a higher human development index in a country does not mean lower COVID-19 mortality rates.

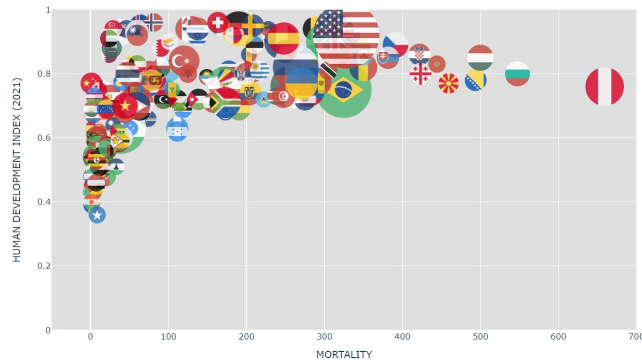

Fig. 3. Human Development Index to Mortality ratio of different countries.

We can see in Fig. 3. that the graph does correspond to our conclusion, for the mortality rates do not go lower as the HDI increases.

Hypothesis test 2:

We will do the exact same analysis that we did on first hypothesis test, but instead of the mortality column we will compare the HDI with the case-fatalities.

Here:

$\bar{x}$ = Mean case-fatality rate of all countries

μ = Mean case-fatality rate of countries with a HDI higher than 0.8

σ = Case-fatality standard deviation for all countries

n = Sample size of countries with a HDI higher than 0.8

```
df["CASE-FATALITY"].describe()
```

```
count    191.000000
mean       1.435602
std        1.668781
min        0.000000
25%        0.600000
50%        1.100000
75%        1.950000
max       18.100000
Name: CASE-FATALITY, dtype: float64
```

Fig. 4. The descriptive statistics of the case-fatality column

From Fig. 4. we have H_0 = $\mu \geq 1.436$ and H_a = $\mu < 1.436$

```
df_mh["CASE-FATALITY"].describe()
```

```
count    69.000000
mean      0.768116
std       0.606216
min       0.100000
25%       0.300000
50%       0.700000
75%       1.000000
max       3.000000
Name: CASE-FATALITY, dtype: float64
```

Fig. 5. The descriptive statistics of the case-fatality column for countries with a HDI higher than 0.8

Like in hypothesis 1, we choose a confidence level of 95% which corresponds to a Z-score of 1.645. If our Z-score exceeds this value we will reject the null hypothesis, else we will accept it.

Our Z-score is:

$$z = \frac{1.436 - 0.768}{\frac{1.669}{\sqrt{69}}} \tag{3}$$

$$= \frac{0.668}{0.201} \approx 3.323 \tag{4}$$

Our Z-score 3.323 $>$ 1.645, therefore we reject the null hypothesis and we can say with a level of 95% confidence that a higher human development index in a country does indicate lower COVID-19 case-fatality rates.

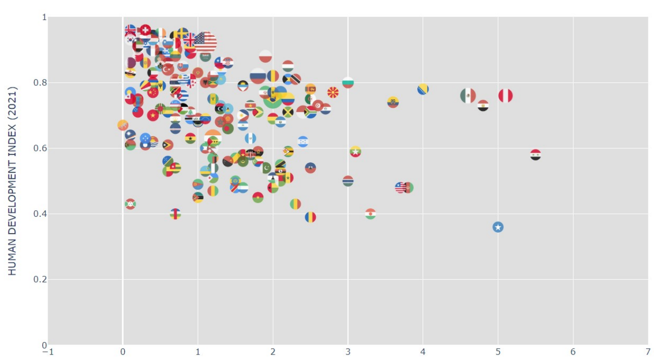

Fig. 6. Human Development Index to Case-Fatality ratio of different countries.

We can see in Fig. 6. that the graph does correspond to our conclusion, for the case-fatality rates go lower as the HDI increases.

B. Does the ratio of elderly population in a country affect mortality/fatality from COVID-19?

In order to answer this question, we will compute the covariance and the Pearson correlation coefficient of the elderly population ratio in function of the COVID-19 mortality and case-fatality rates.

Hypothesis test 3:

In this hypothesis, we are trying to figure out if there is a relation between COVID-19 mortality rates and the ratio of elderly population in a given country. Here is the formula for covariance:

$$s_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})$$

Where:

N = Sample size of countries with an existing data of elderly population ratio

X_i = Elderly population ratio of ith country

$\bar{X}$ = Mean elderly population ratio of all countries

Y_i = Mortality rate of ith country

$\bar{Y}$ = Mean mortality rate of all countries

And here is the formula for correlation coefficient:

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

```
EPR = df["ELDERLY POPULATION (% OF TOTAL IN 2021)"]
M = df["MORTALITY"]
rowind = (~np.isnan(EPR)) & (~np.isnan(M))
np.cov(df.loc[rowind, "ELDERLY POPULATION (% OF TOTAL IN 2021)"], df.loc[rowind, "MORTALITY"])
```

```
array([[   44.81698136,   595.0713882 ],
       [  595.0713882 , 17626.5433922 ]])
```

Fig. 7. Covariance matrix of elderly population ratio and mortality rates

```
rowind = (~np.isnan(EPR)) & (~np.isnan(M))
np.corrcoef(df.loc[rowind, "ELDERLY POPULATION (% OF TOTAL IN 2021)"], df.loc[rowind, "MORTALITY"])
```

```
array([[1.        , 0.66952098],
       [0.66952098, 1.        ]])
```

Fig. 8. Correlation matrix of elderly population ratio and mortality rates

According to Fig. 7. we have s_{xy} = 595.08 and according to Fig. 8. we have r_{xy} = 0.669. We can infer from these values that there is a positive relation between the two variables and as the elderly population ratio increases, mortality rates from COVID-19 also increase.

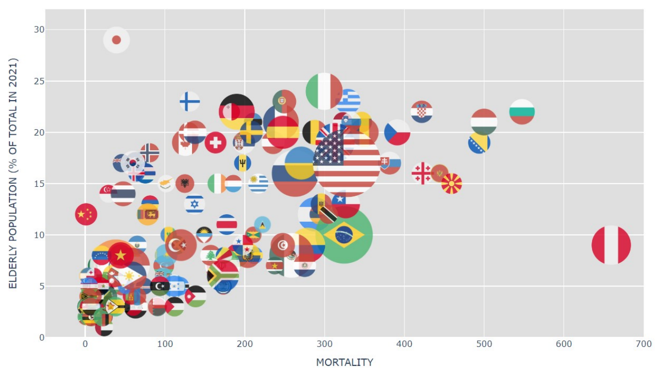

Fig. 9. Elderly Population to Mortality ratio of different countries.

We can clearly see from Fig. 9. that our mathematical conclusion can be visually validated as true.

Hypothesis test 4:

In this hypothesis, we are trying to figure out if there is a relation between COVID-19 case-fatality rates and the ratio of elderly population in a given country. We will use covariance just like in the previous hypothesis.

Here:

N = Sample size of countries with an existing data of elderly population ratio

X_i = Elderly population ratio of ith country

$\bar{X}$ = Mean elderly population ratio of all countries

Y_i = Case-fatality rate of ith country

$\bar{Y}$ = Mean case-fatality rate of all countries

```
EPR = df["ELDERLY POPULATION (% OF TOTAL IN 2021)"]
F = df["CASE-FATALITY"]
rowind = (~np.isnan(EPR)) & (~np.isnan(F))
np.cov(df.loc[rowind, "ELDERLY POPULATION (% OF TOTAL IN 2021)"], df.loc[rowind, "CASE-FATALITY"])

array([[44.81698136, -2.43263918],
       [-2.43263918,  2.86672333]])
```

Fig. 10. Covariance matrix of elderly population ratio and case-fatality rates

```
rowind = (~np.isnan(EPR)) & (~np.isnan(F))
np.corrcoef(df.loc[rowind, "ELDERLY POPULATION (% OF TOTAL IN 2021)"], df.loc[rowind, "CASE-FATALITY"])

array([[ 1.       , -0.2146167],
       [-0.2146167,  1.       ]])
```

Fig. 11. Correlation matrix of elderly population ratio and case-fatality rates

In Fig. 10. we see that s_{xy} = -2.433 and in Fig. 11. it is shown that r_{xy} = -0.215. According to these values, because they are too little to indicate a real relation between the two variables, we cannot draw any conclusions regarding the relation between the elderly population ratio and COVID-19 case-fatality rates.

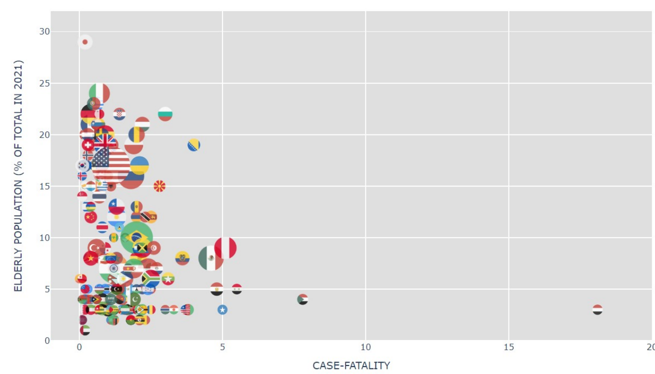

Fig. 12. Elderly Population to Case-fatality ratio of different countries.

In Fig. 12. we can see that there seems to be no apparent covariance between these two variables. Therefore, our conclusion seems correct.

C. Does the healthcare index of a country affect its mortality/fatality from COVID-19?

Similarly to the previous hypotheses, we will use covariance and correlation coefficient in order to draw conclusions from comparing the healthcare index and mortality/fatality rates of a country.

Hypothesis test 5:

Let's compute the covariance and correlation between the healthcare index and mortality from COVID-19.

Here:

N = Sample size of countries with an existing data of healthcare index

X_i = Healthcare index of ith country

$\bar{X}$ = Mean healthcare index of all countries

Y_i = Mortality rate of ith country

$\bar{Y}$ = Mean mortality rate of all countries

```
rowind = (~np.isnan(HI)) & (~np.isnan(M))
np.cov(df.loc[rowind, "HEALTHCARE INDEX (2022)"], df.loc[rowind, "MORTALITY"])

array([[   111.77873452,    -34.30744592],
       [   -34.30744592, 20301.47324165]])
```

Fig. 13. Covariance matrix of healthcare index and mortality rates

```
np.corrcoef(df.loc[rowind, "HEALTHCARE INDEX (2022)"], df.loc[rowind, "MORTALITY"])

array([[ 1.       , -0.0227743],
       [-0.0227743,  1.       ]])
```

Fig. 14. Correlation matrix of healthcare index and mortality rates

From Fig. 13. we see that s_{xy} = -34.307 and from Fig. 14. we see that r_{xy} = -0.023. The correlation value is very close to 0, therefore we can conclude that there are no linear dependencies between the healthcare index and mortality rates and the healthcare index of a country does not affect its mortality rate from COVID-19.

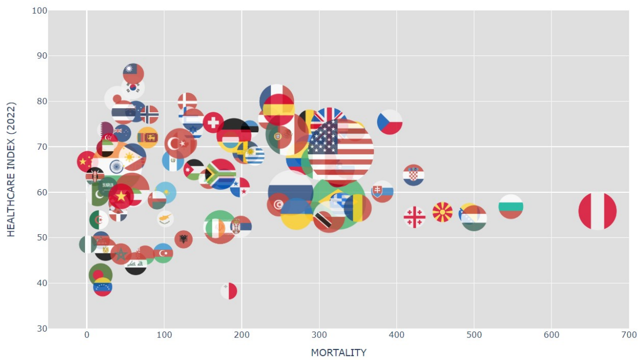

Fig. 15. Healthcare index to mortality ratio of different countries.

In Fig. 15. we can see graphically that there are no linear dependencies between these two variables.

Hypothesis test 6:

For this hypothesis, we will do the same operations for the relation between healthcare index and case-fatality rates.

Here:

N = Sample size of countries with an existing data of healthcare index

X_i = Healthcare index of ith country

$\bar{X}$ = Mean healthcare index of all countries

Y_i = Case-fatality rate of ith country

$\bar{Y}$ = Mean case-fatality rate of all countries

```
F = df["CASE-FATALITY"]
rowind = (~np.isnan(HI)) & (~np.isnan(F))
np.cov(df.loc[rowind, "HEALTHCARE INDEX (2022)"], df.loc[rowind, "CASE-FATALITY"])

array([[111.77873452,  -4.17755296],
       [ -4.17755296,   1.11695722]])
```

Fig. 16. Covariance matrix of healthcare index and case-fatality rates

```
np.corrcoef(df.loc[rowind, "HEALTHCARE INDEX (2022)"], df.loc[rowind, "CASE-FATALITY"])

array([[ 1.       , -0.373873],
       [-0.373873,  1.       ]])
```

Fig. 17. Correlation matrix of healthcare index and case-fatality rates

From Fig. 16. we see that s_{xy} = -4.178 and from Fig. 17. we see that r_{xy} = -0.374. The correlation value is closer to 0 than it is to 1, we can say that there is some relation between these two variables but no strong linear dependencies.

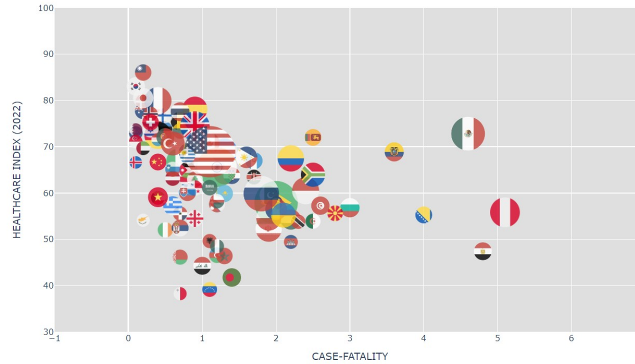

Fig. 18. Healthcare index to case-fatality ratio of different countries.

Finally in Fig. 18. we can observe a linear pattern of case-fatalities increasing as healthcare index goes lower. However, some exceptions such as Mexico, cause the correlation value to be closer to 0 than to 1. In conclusion, we can say that the healthcare index of a country does affect its case-fatality rate to an extent.

VI. Conclusion

In conclusion, we can say that the three factors that we have evaluated affect either case-fatality or mortality rates of COVID-19 in different countries. For instance, we have seen that a higher human development index and healthcare index means lower case-fatality rates, while a lower elderly population ratio means lower mortality rates.

VII. Sources of Data

https://coronavirus.jhu.edu/data/mortality

https://www.numbeo.com/health-care/rankings_by_country.jsp

https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS?name_desc=false