Covid-19 Deaths in Developing vs Developing Countries

OSCARITO
isaho, mbaker6, msemlani, rsalkham

Hypothesis:

COVID-19, an infectious disease caused by the coronavirus, is affecting over 200 nations around the world. Both developed and developing countries were affected by the novel coronavirus; therefore, we sought out to investigate the following hypothesis: developed countries will witness a lower number of cases per million and a lower death rate than developing countries as a result of COVID-19.

Data:

Our data was collected from multiple sources including: World Bank for development indicators, Our World In Data for COVID-19 specific data, The Economist Intelligence Unit for democracy level and regime type data, OpenFlights.org for flight related data, John Hopkins School of Engineering, and more.

Dependent variables: total number of cases, total number of deaths, death rate.

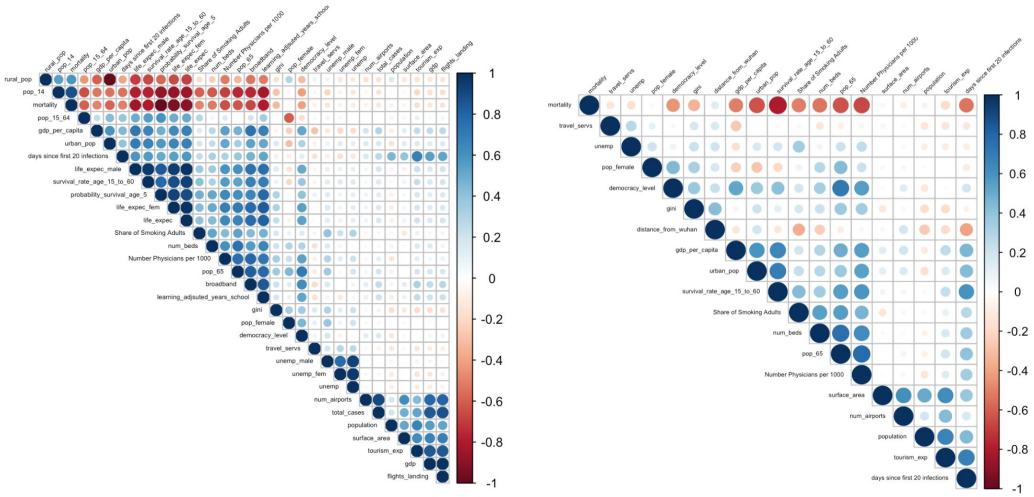
Independent variables: GDP per capita, GDP, total population, urban and rural shares of the population, shares of population under 14, between 15 and 65, and over 65 years of age, female share of population, mortality rate, surface area, life expectancy for males and females, unemployment for females and males, broadband, the gini index, expected years of schooling, learning adjusted schooling, probability of survival under 5, the survival rate between 15 to 60, and the democracy level, share of smoking adult population, number of physicians per 1000, travel services, expenditure on tourism, number of beds per 1000, number of flights landing in a country, number of airports in a country, and regime type.

COVID-19 specific data: number of tests, distance from Wuhan, number of days since first 20 infections.

Findings:

Claim #1: Development indicators and COVID-19 data for each country are correlated with each other.

Support for Claim #1: Although we removed the correlated variables we still found that there are variables that are dependent on each other (i.e. survival rate for people ages 15 to 60 and mortality). The graph shows the correlation level between pairs of variables in our dataset. In the graph to the right, highly correlated variables (threshold of 0.8) have been omitted. Yet, many remain with high correlations. This has made it difficult to distinguish between the effects of different factors on the death rate as they all affect each other.



Claim #2: There is no relationship between the total number of coronavirus cases/deaths in a country and how developed the country is.

Support for Claim #2: Indicators such as GDP/capita and mortality were not significant in our regression, which means that, as opposed to our hypothesis, such development indicators do not affect the number of cases and deaths in a country. Therefore, other factors enter into the equation in determining the number of coronavirus cases/deaths per country

Claim #3: There seems to be a strong relationship between the total number of coronavirus cases/deaths in a country and the movement of people.

Support for Claim #3: Variables such as tourism expenditure and number of airports are the most significant in affecting the number of coronavirus cases/deaths. This travel data shows that the higher number of airports and the higher a country's tourism expenditure, the higher number of coronavirus cases/deaths it will witness.

Other findings: We expected some variables to have an effect on total coronavirus deaths/cases (e.g. overall population and population over 65, number of tests, number of hospital beds), and they indeed were significant. However, as opposed to our expectations, other variables such as number of physicians, share of adults who smoke, distance from Wuhan, and number of days since the 20th infection were not significant. This is important because it shows that the number of cases or deaths in a certain country is not dependent on the number of doctors in the country, the distance of the country from the original pandemic epicenter, or how long the country had the coronavirus for.

Regression on total cases

```
Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                    -1.819e+04  8.481e+04  -0.215   0.8323
gdp_per_capita                 -3.130e-01  3.373e-01  -0.928   0.3644
population                     -3.382e-05  1.236e-05  -2.736   0.0127 *
urban_pop                       2.122e+02  1.985e+02   1.069   0.2977
pop_65                          1.432e+03  9.155e+02   1.564   0.1335
pop_female                      7.185e-01  1.046e+03   0.001   0.9995
mortality                       3.210e+02  2.742e+02   1.171   0.2555
surface_area                   -9.864e-03  1.165e-03  -8.467 4.80e-08 ***
num_beds                       -2.063e+03  1.742e+03  -1.184   0.2504
tourism_exp                     1.508e-06  2.435e-07   6.192 4.76e-06 ***
travel_servs                   -8.933e+01  1.410e+02  -0.633   0.5336
unemp                          -9.125e+01  7.532e+02  -0.121   0.9048
gini                            3.779e+02  1.744e+02   2.167   0.0425 *
survival_rate_age_15_to_60      4.757e+04  8.880e+04   0.536   0.5981
`Number Physicians per 1000`   -2.549e+03  2.582e+03  -0.987   0.3354
`Share of Smoking Adults`      -1.964e+02  3.068e+02  -0.640   0.5294
distance_from_wuhan            -1.249e-03  9.765e-04  -1.279   0.2157
`days since first 20 infections` -4.630e+02  2.958e+02  -1.565   0.1332
num_airports                    2.972e+01  3.215e+00   9.244 1.17e-08 ***
total_tests                     9.146e-02  6.728e-03  13.595 1.45e-11 ***
regime_typeFlawed democracy    -2.224e+04  1.501e+04  -1.482   0.1539
regime_typeFull democracy      -1.721e+04  1.808e+04  -0.952   0.3525
regime_typeHybrid regime       -8.030e+03  1.419e+04  -0.566   0.5777
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Regression on total deaths

```
Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                    -4.080e+03  1.492e+04  -0.274   0.7873
gdp_per_capita                 -5.701e-02  5.933e-02  -0.961   0.3480
population                     -5.149e-06  2.174e-06  -2.369   0.0280 *
urban_pop                      -2.372e+01  3.491e+01  -0.680   0.5046
pop_65                          3.916e+02  1.610e+02   2.432   0.0245 *
pop_female                      9.116e+01  1.840e+02   0.496   0.6256
mortality                       6.853e+01  4.823e+01   1.421   0.1708
surface_area                   -1.086e-03  2.049e-04  -5.302 3.44e-05 ***
num_beds                       -7.024e+02  3.064e+02  -2.292   0.0329 *
tourism_exp                     3.777e-07  4.283e-08   8.820 2.50e-08 ***
travel_servs                    7.566e+00  2.480e+01   0.305   0.7635
unemp                          -5.517e+01  1.325e+02  -0.416   0.6815
gini                           -1.364e+00  3.068e+01  -0.044   0.9650
survival_rate_age_15_to_60      3.621e+03  1.562e+04   0.232   0.8190
`Number Physicians per 1000`    4.374e+01  4.542e+02   0.096   0.9242
`Share of Smoking Adults`       2.639e+01  5.396e+01   0.489   0.6301
distance_from_wuhan             2.541e-04  1.718e-04   1.480   0.1546
`days since first 20 infections` -4.470e+01  5.202e+01  -0.859   0.4003
num_airports                   -1.299e+00  5.655e-01  -2.296   0.0326 *
total_tests                     6.420e-03  1.183e-03   5.426 2.60e-05 ***
regime_typeFlawed democracy    -6.357e+03  2.639e+03  -2.409   0.0258 *
regime_typeFull democracy      -7.859e+03  3.180e+03  -2.471   0.0226 *
regime_typeHybrid regime       -6.200e+03  2.496e+03  -2.484   0.0220 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Regression on death rate

```
Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                    -1.216e-01  2.316e-01  -0.525 0.605367
gdp_per_capita                 -9.452e-07  9.212e-07  -1.026 0.317124
population                     -1.653e-11  3.375e-11  -0.490 0.629657
urban_pop                      -5.569e-04  5.420e-04  -1.027 0.316473
pop_65                          2.355e-03  2.500e-03   0.942 0.357458
pop_female                     -1.026e-03  2.857e-03  -0.359 0.723285
mortality                       4.962e-04  7.488e-04   0.663 0.515175
surface_area                    4.141e-10  3.182e-09   0.130 0.897755
num_beds                       -8.527e-04  4.758e-03  -0.179 0.859588
tourism_exp                     2.710e-12  6.650e-13   4.075 0.000591 ***
travel_servs                    5.031e-04  3.851e-04   1.306 0.206299
unemp                          -2.427e-03  2.057e-03  -1.180 0.251805
gini                           -1.207e-04  4.764e-04  -0.253 0.802512
survival_rate_age_15_to_60      2.733e-01  2.425e-01   1.127 0.273076
`Number Physicians per 1000`    3.460e-03  7.052e-03   0.491 0.629050
`Share of Smoking Adults`       8.461e-04  8.379e-04   1.010 0.324696
distance_from_wuhan             2.408e-09  2.667e-09   0.903 0.377349
`days since first 20 infections` -1.355e-03  8.077e-04  -1.678 0.108908
num_airports                   -2.147e-05  8.780e-06  -2.445 0.023869 *
total_tests                    -1.528e-09  1.837e-08  -0.083 0.934548
regime_typeFlawed democracy     8.663e-03  4.098e-02   0.211 0.834710
regime_typeFull democracy      -2.410e-02  4.937e-02  -0.488 0.630781
regime_typeHybrid regime       -4.387e-03  3.875e-02  -0.113 0.911001
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```