Impacts of Different Factors of G20 Countries on the Covid-19 Vaccine Rates Based on Empirical Analysis

Yihan Lei(✉)

Shanghai Starriver Bilingual School, SSBS, Shanghai, China
yuhongyan0816@163.com

Abstract. COVID-19 vaccine is critical to terminate the global pandemic. Intended to examine the impacts of various factors of a country on its COVID-19 vaccine rate, a data set of new cases per day, hospital beds per person, GDP, population density, and stringency index is collected. We do correlation analysis on the variables. We applied OLS linear regression model to the data set. We find the coefficient of different variables, observe p-value of each variable to examine its significance. Then, we applied lasso variable selection to the variables to find more important variables and put them in the linear regression models in order to solve the problems of variables being insignificant and multicollinearity.

Keywords: COVID-19 · Vaccination rate · Regression analysis · factor · country

1 Introduction

Coronavirus has become one of the most commonly used words in 2020, and it is still having a significant impact on people's lives even in 2021. The impact of the coronavirus's spread is definitely enormous, ranging from countless deaths to the inevitability of global economic collapse and international cooperation. As a result, the most pressing concern is to stop the infection from spreading. Although ensuring effective and fair distribution of COVID-19 vaccinations is a top governmental goal, acceptability is also crucial. The effectiveness of any vaccination program is dependent on the public's trust in vaccinations and the organizations that deliver them [2]. Given that vaccines always play an important role in stopping diseases in human history, the vaccine will be the key to the termination of the coronavirus pandemic. While some countries widely spread vaccination and successfully relieve the pandemic, some other countries are struggling with the vaccination. Indonesia: 21.5% last vaccination per hundred; South Africa 37.53% last vaccination per hundred; South Africa 8.98% last vaccination per hundred; Australia: 40.37% [2]. What kinds of factors influence the COVID-19 vaccination rate in different countries? To address this issue, we found ten factors that can reflect a country's economic strength, medical level, age distribution, education level, and strictness of the epidemic prevention and control policies. To conduct quantitative analysis, we need timely, comparable data across countries to assess the scope and rate of vaccination

Z. Zhan et al. (Eds.): EIMSS 2022, AHCS 7, pp. 14–29, 2023.
https://doi.org/10.2991/978-94-6463-024-4_3

implementation. The COVID-19 vaccinations dataset from Our World in Data provides a public global aggregated dataset on administered vaccinations. Our COVID-19 vaccination dataset is widely used by scientists because of its authentication. For example, the World Health Organization relies on this dataset for its official COVID -19 dashboard. We gathered 10 factors from the dataset and then divided them into two datasets which are developed countries including Canada, America, Britain, France, Germany, Italy Japan, Korea, and Australia, and developing countries including Brazil, Russia, China, India, South Africa, Mexico, Turkey, Indonesia, Argentina, and Saudi Arabia. Applying python's data visualization library (Pandas, Numpy, Matplotlib) to visualize the dataset into box plot, pair plot, and histogram to observe the general difference between developing countries and developed countries. Moreover, we examine the correlation between different factors, and we find different patterns for developed countries and developing countries. For example, the correlations between university rates and vaccination rates are opposite. Although there are 19 countries and 11 variables in total, as we divided the data into 9 developed countries and 11 developing countries, the number of countries is smaller than the number of variables, which can cause insignificance of variables and multicollinearity. Therefore, we introduce Lasso (least absolute shrinkage and selection operator) and Lars (Least angle regression) to select more important variables.

2 Methods and Results

The methods we used includes:

Correlation analysis:

We use Pearson's correlation coefficients to show the correlations between different variables.

The correlation between x and y [9].

μ_X is the expected value.

σ_X is the standard deviation (Fig. 1)

$$\rho_{X,Y} = \mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \quad (1)$$

Fig. 1. Function of Correlation analysis

OLS regression:

First, calculate RSS (residual sum of squares) [10] (Figs. 2 and 3)

$$RSS = \sum_{i=1}^{n} e_i^2 \quad (2)$$

Fig. 2. Function of Residual Sum of Squares

$$\hat{\sigma}_{\beta_1} = \hat{\sigma}_\varepsilon \sqrt{\frac{1}{\sum(x_i - \bar{x})^2}}$$

$$\hat{\sigma}_{\beta_0} = \hat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}} = \hat{\sigma}_{\beta_1} \sqrt{\frac{\sum x_i^2}{n}} \tag{3}$$

Fig. 3. Functions of OLS regression

And then estimate the predicted values.

Lasso variable selection (Fig. 4):

$$\frac{1}{N} \sum_{i=1}^{N} f(x_i, y_i, \alpha, \beta) \tag{4}$$

Fig. 4. Functions of LASSO regression

Lasso variable selection is applied to select more important variables. The general form of lasso is shown above [5].

Descriptive statistical analysis:

After collecting the data of each country and classifying the data of the countries according to developed and developing countries, I conducted a descriptive statistical analysis of the data. In the process of descriptive statistical analysis, I selected a total of three types of graphs.

1. Boxplot

A boxplot is a graphical representation of numerical data groups based on their quartiles.

In this scenario, two data groups, developing countries and developed countries, are graphed based on the data of the 10 factors. We intended to observe the relationship and difference on the 10 factors of these two groups of countries (Fig. 5).

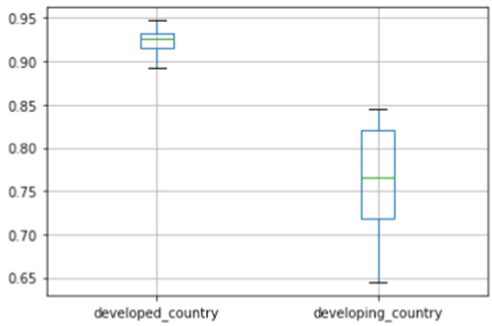

Fig. 5. Box plots of human development index of developed countries and developing countries respectively

This graph is the box plot of the Human Development Index. We can see that the average value of the Human Development Index in developed countries is 0.2 higher than that in developing countries. The highest human development index of developed countries reaches 0.95, while the highest value of developing countries' human development index is 0.85. At the same time, we can see that the lowest value of the Human Development Index of developed countries is 0.05 higher than the highest value of developed countries. We can see that the human development index of developed and developing countries is quite different. The ratio is generally higher in developed countries, while the ratio in developing countries is very low without exception (Fig. 6).

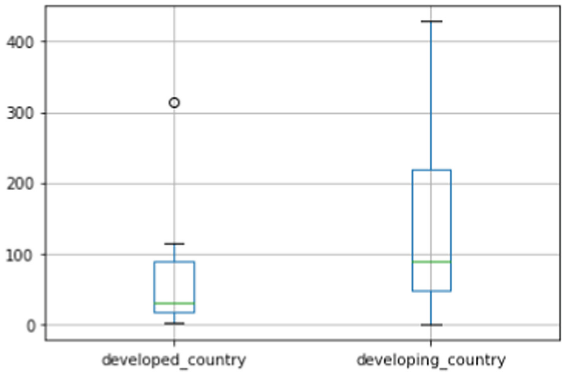

Fig. 6. Box plots of new cases average per million of developed countries and developing countries respectively

The next graph is the box plot of new cases average per million. The median of developed countries is slightly smaller than the median of developing countries. However, the range of developing countries is much larger than the range of developed countries. The new cases average per million of both groups is skewed to the left. It is worth noting that there is an outlier in developed countries which is the United Kingdom. The new case average per million of the UK is 315 (Fig. 7).

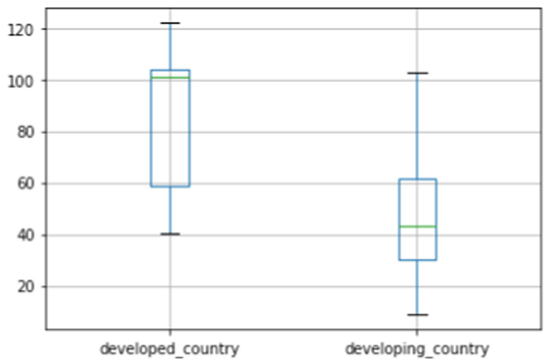

Fig. 7. Box plots of COVID-19 vaccine rates of developed countries and developing countries respectively

In terms of COVID-19 vaccine rate, although the range of data of developing countries and developed countries are the same, the average value of developed countries is 60% higher than that of developing countries. At the same time, South Africa has a vaccination rate of 8% and Indonesia has a vaccination rate of 21%. Canada and the United Kingdom have reached 120%, while only China among developing countries has exceeded 100% (Fig. 8).

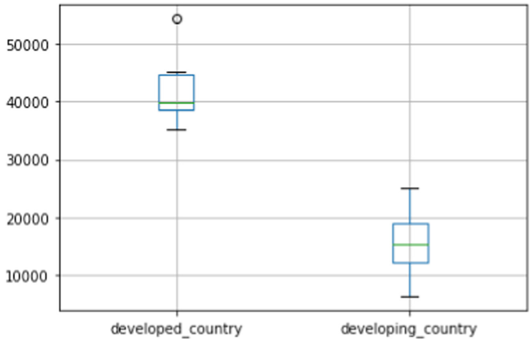

Fig. 8. Box plots of GDP per capita of developed countries and developing countries respectively

GDP per capita reflects the level of productivity and quality of life of a country. The average per capita GDP in developed countries is 20,000 US dollars higher than that in developing countries. At the same time, one of the developed countries has reached more than US$50,000, while the per capita GDP of developing countries is only US$25,000. The lowest per capita GDP in developed countries is also the highest than that in developing countries, with a per capita GDP higher than US$10,000. From this picture, we can see that the per capita GDP, that is, the standard of living, in developed countries is much higher than that in developed countries (Fig. 9).

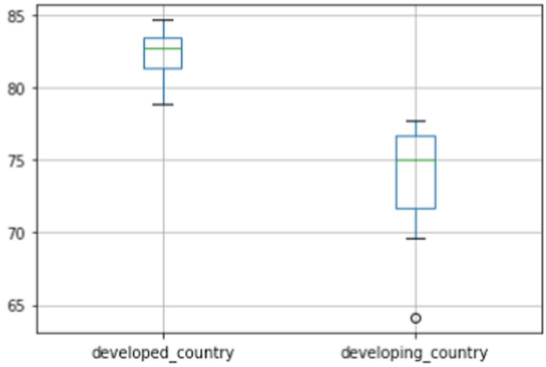

Fig. 9. Box plots of life expectancy of developed countries and developing countries respectively

In terms of life expectancy, developed countries as a whole are higher than developing countries. The highest value of life expectancy in developed countries has reached 85, while that in developing countries is only around 77. The lowest developing country is even lower than 65. From this data, we can also reflect that the general medical level in developing countries is lower than that in developed countries (Fig. 10).

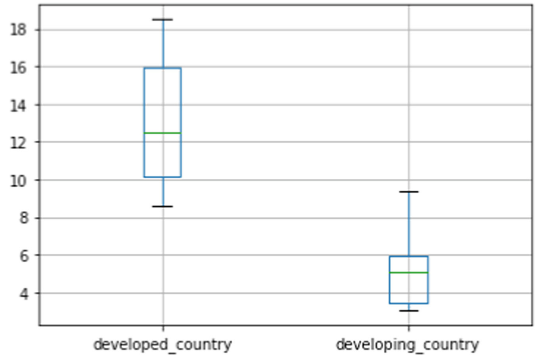

Fig. 10. Box plots of proportion of people who are 70 years old or older of developed countries and developing countries respectively

Taking into account the proportion of the population, the highest population over 70 in developed countries is 18%, and the lowest is 8%, while the majority of developing countries are young and middle-aged. The population over 70 years old is the highest. The general average is around five. Large numbers of old people also lead to higher mortality and transmission rates. At the same time, the vaccine cannot be well implemented in the elderly (Fig. 11).

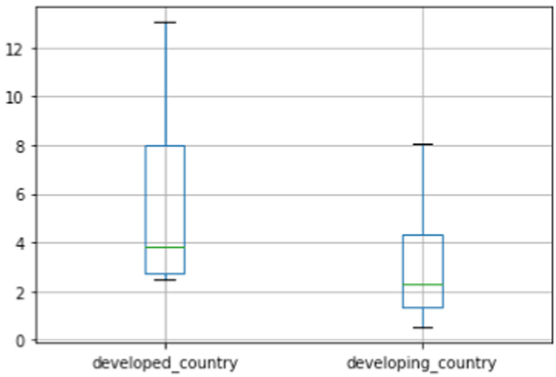

Fig. 11. Box plots of hospital beds per thousand of developed countries and developing countries respectively

The medians and means of hospital beds per thousand for both developed countries and developing countries are between 3 and 4. Yet, the developed countries have a maximum of 13 hospital beds per thousand. Japan and Korea have 13.05 hospital beds per thousand and 12.27 hospital beds per thousand respectively. Meanwhile, the maximum of developing countries is eight. At the same time, many developing countries have fewer than two hospital beds per thousand. The number of hospital beds per thousand can better reflect the medical resources of a country. It has a greater impact on the vaccination rate of a country (Fig. 12).

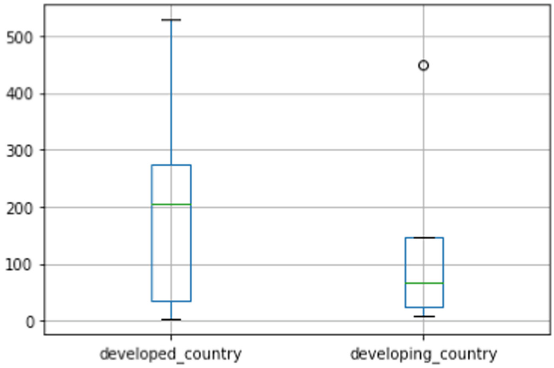

Fig. 12. Box plots of population density of developed countries and developing countries respectively

This graph reflects the population density of developed and developing countries. We can see that the population density distributions of developed and developing countries are the same, and both are left-leaning. However, the average value in developed countries is around 200, while the average value in developing countries is around 100 (Fig. 13).

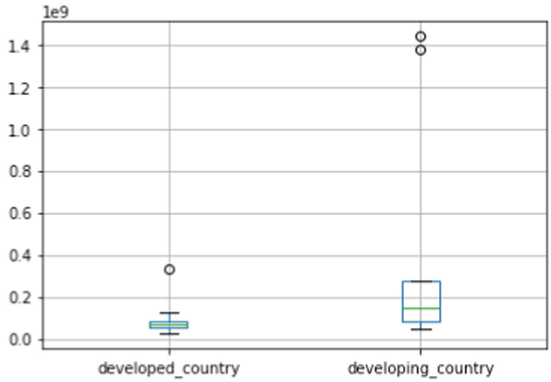

Fig. 13. Box plots of population of developed countries and developing countries respectively

Although in terms of population density, there is not much difference between developed and developing countries. However, in terms of population, the extremes of developing countries are significantly ahead of developed countries. Among the developing countries, India and China have populations of 1.3 billion and 1.4 billion respectively.

The population difference between developed countries is not that big, only the United States has a population of 300 million (Fig. 14).

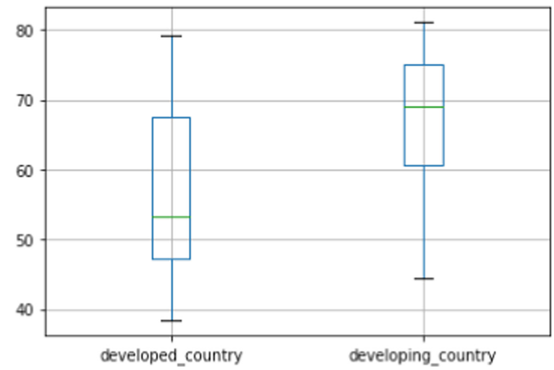

Fig. 14. Box plots of stringency index of developed countries and developing countries respectively

The stringency index reflects the epidemic control strictness of a country in the epidemic. From this chart, we can see the average degree of epidemic prevention in developed countries. It is 15 lower than that in developing countries. Although the highest value for both developed and developing countries is around 80, the strictness of epidemic prevention in developed countries is generally lower than that in developing countries (Fig. 15).

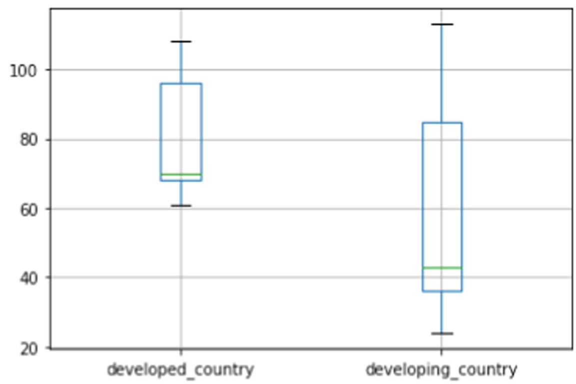

Fig. 15. Box plots of university rates of developed countries and developing countries respectively

This graph reflects a country's university enrollment rate and reflects the quality of education in a country. In developing countries, the university enrollment rate varies greatly among countries. Some countries have only 20%, while some other countries exceed the highest value of developed countries, reaching 120%. At the same time, university enrollment rates in developed countries are relatively similar, and the average is 30% higher than in developing countries.

2. Pairplot (Fig. 16)

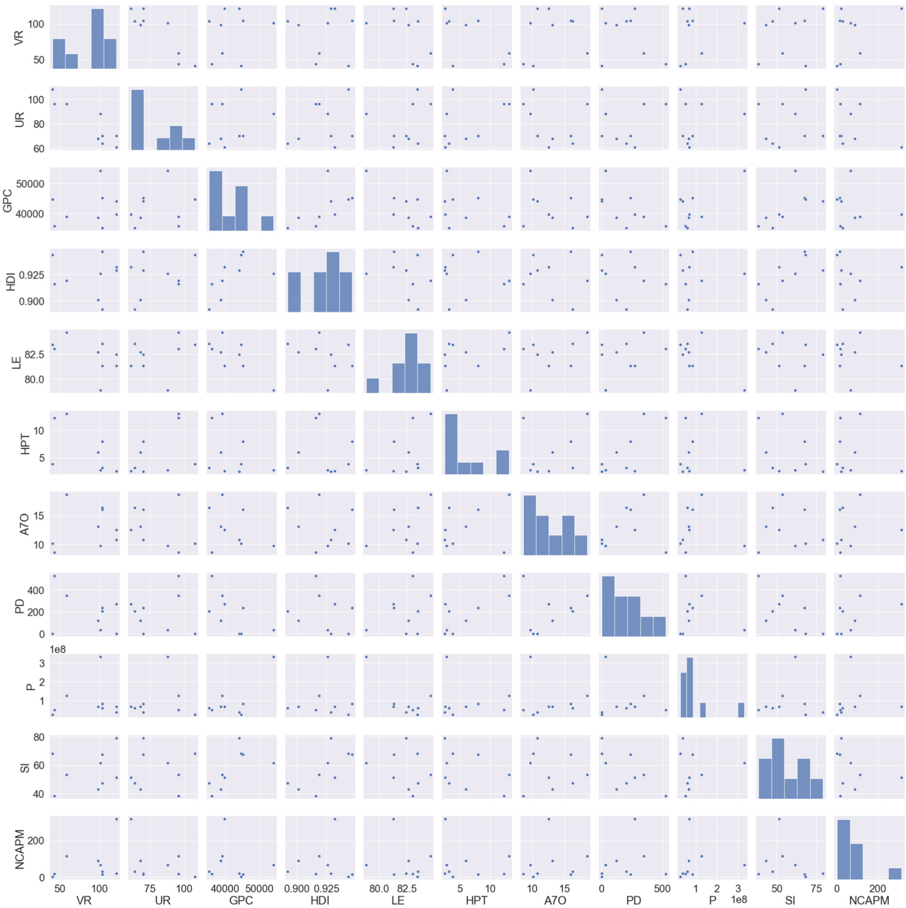

Fig. 16. Pair plot of the dataset of developed countries. VR stands for vaccine rate; UR stands for university rate; GPC stands for GDP per capita; HDI stands for human development index; LE stands for life expectancy; HPT stands for hospital beds per thousand; A7O stands for proportion of people who are 70 years old or older; PD stands for population density; P stands for population; SI stands for stringency index; NCAPM stands for new case average per million.

This pair plot of developed countries' datasets combines scatter plots and histograms. We can observe the relationship between different factors from the scatter plot by examining the patterns while seeing the frequency from histograms (Fig. 17).

From the scatter plots of developed country, I observed only one factor which is university rate has a obvious negative relationship where the points on the scatter plot forms a line downward,

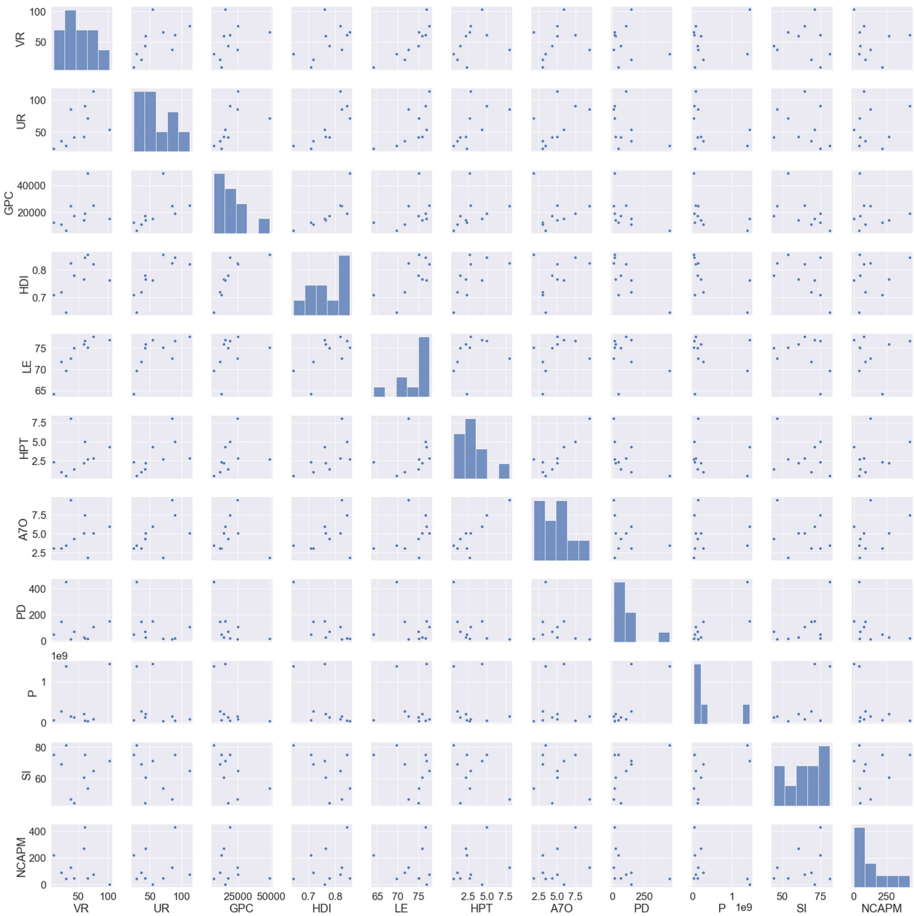

Fig. 17. Pair plot of the dataset of developed countries. VR stands for vaccine rate; UR stands for university rate; GPC stands for GDP per capita; HDI stands for human development index; LE stands for life expectancy; HPT stands for hospital beds per thousand; A7O stands for proportion of people who are 70 years old or older; PD stands for population density; P stands for population; SI stands for stringency index; NCAPM stands for new case average per million.

2) Correlation analysis

By comparing the correlation map of developed countries and developing countries, we observed different patterns.

Figures 18 and 19 are the correlation analysis between various factors in developed countries and developing countries respectively. By comparing the two graphs, there is some significant difference between the two correlation analyses.

Firstly, in developed countries, the correlation coefficient of university rate and vaccination rate is −0.9. Meanwhile, the correlation coefficient of university rate and vaccination rate is 0.54 in developing countries. That is to say, in developed countries, the higher the university rate, the lower the country's vaccination rate, which is a surprising

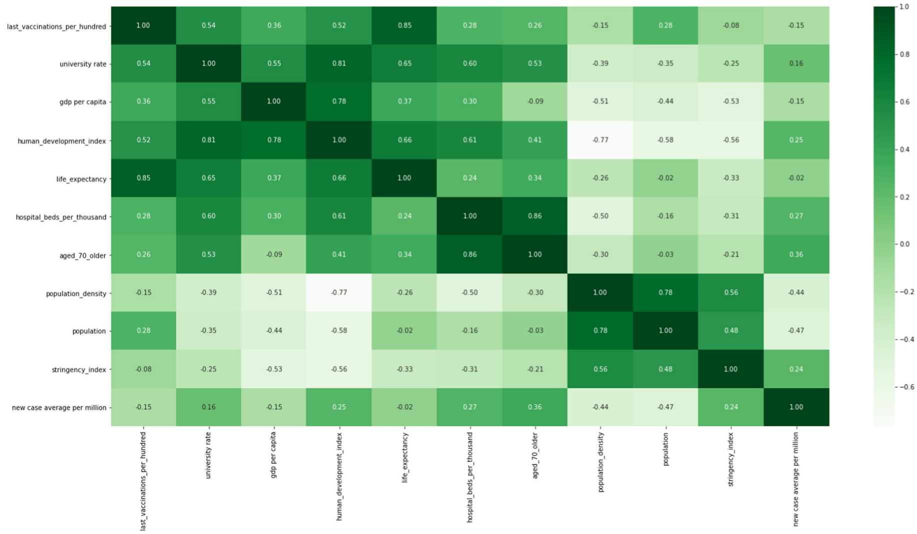

Fig. 18. Correlation map of developing countries

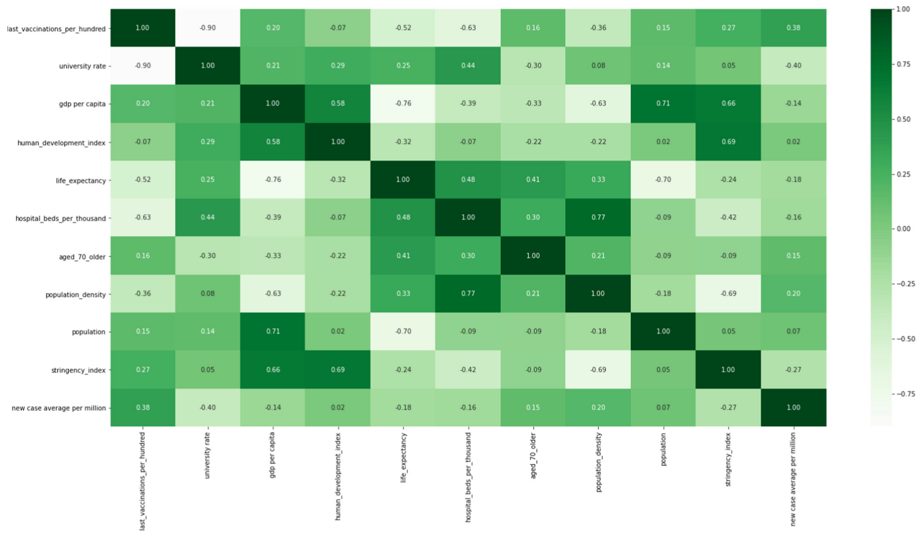

Fig. 19. Correlation map of developed countries

result since the university rate may lead to a higher acceptance of the vaccine. After studying specific data, I think that the two countries that led to this conclusion are Japan and South Korea. Their proportion of going to university is Reached 98% and 97% respectively, and their vaccination rate was only 50%.

Due to this, the correlation between university rate and the vaccination rate in developing countries is in the opposite direction. The correlation has reached a positive correlation of 0.54, which is also in line with our expectations. At the same time, we hope to study why developed countries have such extreme negative correlations in the following research.

In addition, according to the correlation analysis, life expectancy has different impacts on vaccination rates for developed countries and developing countries. The correlation coefficient of life expectancy and vaccination rate is −0.52 and 0.85 for developed countries and developing countries respectively. Higher life expectancy usually means a better medical system that can lead to a higher vaccination rate. Still, Japan and Korea mainly caused this negative correlation. They have a life expectancy of 84.63 and 83.03 with vaccination rates of 59.49% and 42.94% respectively.

3) Linear regression

We performed an OLS (Ordinary Least Squares) regression of the data to study the relationship between these independent variables and vaccination rates. OLS regression is a linear modeling technique that can be used to model multiple explanatory variables. Y can be predicted to some extent by representing the relationship between a continuous response variable (Y) and continuous explanatory variables (X) with a line of best fit [3].

In linear regression, we mainly observe two indicators, p-value, and Adjusted R-squared.

P-value:

The p-value is the probability of obtaining test results at least as extreme as the results observed, under the assumption that the null hypothesis is correct. A very small p-value means that such an extreme observed outcome would be very unlikely under the null hypothesis. Reporting p-values of statistical tests is common practice in academic publications of many quantitative fields. Since the precise meaning of p-value is hard to grasp, misuse is widespread and has been a major topic in metascience [7].

Adjusted r-square:

the coefficient of determination denoted R2 or r2 and pronounced "R squared", is the proportion of the variation in the dependent variable that is predictable from the independent variable(s) [8] (Figs. 20 and 21).

However, these models cannot explain the assumption well for all the p-values of the factors are above 0.4. (A factor is considered as significant when the p-value is lower than 0.05). At the same time, the adjusted r-square of this model is low.

It is because the dimension of the dataset is higher than the number of the countries. Therefore, it will cause redundancy of variables and multicollinearity issues.

OLS Regression Result of developed countries before Lasso variables selections		
variables	regression coefficients	p-values
intercept	2963.7817	0.404
human_development_index	-1563.3771	0.594
hospital_beds_per_thougsand	0.0348	0.996
aged_70_older	5.6956	0.404
population	-4.54E-07	0.552
gdp_per_capita	0.0058	0.693
life_expectancy	-20.8161	0.398
new_case_average_per_million	0.1161	0.675
R-squared	0.867	
Adj. r-squared	-0.064	0

Fig. 20. OLS Regression Result of developed countries before Lasso variables selections

OLS Regression Result of developing countries before Lasso variables selections		
variables	regression coefficients	p-values
intercept	-396.8692	0.02
population_density	0.0951	0.204
population	4.02E-08	0.046
strngency_index	0.3956	0.364
life_expectancy	3.8997	0.079
asged_70_older	-2.3759	0.298
human_development_index	182.805	0.273
R-squared	0.948	
Adj. r-squared	0.844	

Fig. 21. OLS Regression Result of developing countries before Lasso variables selections

3 Lasso Variable Selection

The normal procedure is not well suited for high-dimension data. MLE is not guaranteed usually if the number of observations is smaller than the number of nodes [4].

In order to enhance the prediction accuracy and interpretability of the models, we use lasso (least absolute shrinkage and selection operator) to select more important variables.

Lasso was originally formulated for linear regression models. This simple case reveals a substantial amount about the estimator. These include its relationship to ridge regression and best subset selection and the connections between lasso coefficient estimates and so-called soft thresholding. It also reveals that (like standard linear regression) the coefficient estimates do not need to be unique if covariates are collinear [5] (Figs. 22 and 23).

As a result of LASSO, we selected 7 variables for developed countries and 6 variables for developing countries and get good results on OLS regression (Fig. 24).

	new case average per million	gdp per capita	human_development_index	life_expectancy	hospital_beds_per_thousand	aged_70_older	population_density	population	stringency_index	university rate
0	0.000000	0	0.000000	0.0000000	0.000000	0.000000	0.00000000	0.000000	0.000000	0.00000
1	0.000000	0	0.000000	0.0000000	0.000000	0.000000	0.00000000	0.000000	0.000000	-44.02206
2	0.000000	0	0.000000	0.0000000	-2.485642	0.000000	0.00000000	0.000000	0.000000	-46.50771
3	0.000000	0	0.000000	-7.9552746	-7.326204	0.000000	0.00000000	0.000000	0.000000	-54.72344
4	0.000000	0	0.000000	-13.5376494	-6.274497	0.000000	0.00000000	0.000000	9.098306	-63.81897
5	0.000000	0	0.000000	-8.1385880	-7.078014	0.000000	0.00000000	9.773398	15.255571	-71.77858
6	4.481742	0	0.000000	-3.1548362	-7.249120	0.000000	0.00000000	17.836098	21.991234	-77.17501
7	5.045557	0	0.000000	-0.5062929	-6.457667	-2.051227	0.00000000	20.083426	23.240604	-79.27357
8	5.165444	0	0.000000	0.0000000	-6.259519	-2.456382	-0.05200246	20.510551	23.468259	-79.69146
9	6.212783	0	0.000000	0.0000000	-3.789211	-4.055334	-1.73209503	21.070866	24.188625	-81.47398
10	10.082370	0	-6.641101	0.0000000	0.000000	-7.552531	-1.34459213	21.557189	32.189878	-82.08355
11	11.380660	0	-9.355443	0.0000000	0.000000	-8.538397	0.00000000	21.805941	35.561251	-81.73731
12	11.975809	0	-10.378461	0.0000000	0.000000	-8.890689	0.00000000	21.864161	36.570168	-81.55944
13	13.074126	0	-12.193937	0.0000000	1.835798	-10.095069	0.00000000	22.006824	38.966975	-82.08097
14	13.810110	0	-13.038759	0.0000000	3.650310	-10.921532	-0.97129390	22.025686	39.897830	-82.67077

AIC+LASSO:AIC=-12.8347

Fig. 22. Lasso variables selection of developed countries

	new case average per million	gdp per capita	human_development_index	life_expectancy	hospital_beds_per_thousand	aged_70_older	population_density	population	stringency_index	university rate
0	0.000000	0.000000	0.000000	0.000000	0.0000000	0.000000	0.0000000	0.00000	0.000000	0.000000
1	0.000000	0.000000	0.000000	46.659714	0.0000000	0.000000	0.0000000	0.00000	0.000000	0.000000
2	0.000000	0.000000	0.000000	57.363761	0.0000000	0.000000	0.0000000	10.70405	0.000000	0.000000
3	0.000000	5.090239	0.000000	58.804370	0.0000000	0.000000	0.0000000	16.19154	0.000000	0.000000
4	0.000000	6.009395	0.000000	59.014797	0.5093409	0.000000	0.0000000	17.33078	0.000000	0.000000
5	0.000000	5.596153	1.927211	58.233114	0.0000000	0.000000	0.0000000	18.35886	0.000000	0.000000
6	0.000000	5.569331	9.203744	55.528041	0.0000000	0.000000	0.0000000	24.45379	0.000000	0.000000
7	0.000000	5.714698	9.115419	55.520790	0.0000000	0.000000	-0.4772069	24.94000	0.000000	0.000000
8	0.000000	8.725811	6.487691	56.610962	0.0000000	0.000000	-6.9949234	29.45289	3.037818	0.000000
9	0.000000	11.490972	0.000000	57.607794	0.0000000	0.000000	-12.9738667	32.49801	4.390425	2.774049
10	0.000000	15.728920	0.000000	56.303106	0.0000000	0.000000	-26.8201901	45.34785	10.973353	4.753431
11	-1.820470	15.618771	0.000000	56.423439	0.0000000	0.000000	-29.7208043	46.19617	13.075778	5.100626
12	0.000000	13.670996	0.000000	54.746487	0.0000000	-3.301525	-31.6351312	50.14523	10.547012	8.809540
13	0.000000	0.000000	0.000000	46.158072	0.0000000	-22.592318	-54.3891346	73.18375	5.498256	31.740318
14	0.000000	0.000000	0.000000	45.924592	0.0000000	-23.101912	-55.2546645	74.07243	5.607195	32.267285
15	0.000000	0.000000	0.000000	43.991710	-3.2169218	-21.850950	-58.1156696	76.90137	5.162485	34.615773
16	0.000000	11.046536	0.000000	32.628827	-30.3951506	0.000000	-71.2135220	89.69394	6.405245	40.380425
17	0.000000	10.958471	0.000000	32.304595	-30.8203470	0.000000	-71.7608185	90.24588	6.306293	40.868271
18	5.246465	10.587427	0.000000	24.536312	-40.0583517	0.000000	-79.2941480	103.50669	0.000000	51.038919
19	6.081164	10.705780	0.000000	22.175526	-43.5740682	0.000000	-83.8015669	108.29617	0.000000	54.592879
20	17.546364	9.786802	0.000000	5.885038	-62.5148913	0.000000	-98.2472975	135.64179	-14.392854	75.638963
21	27.564546	16.606997	-20.385971	0.868960	-72.2035935	0.000000	-112.5170111	152.40170	-24.907343	93.613035

AIC+LASSO：AIC=10.15429

Fig. 23. Lasso variables selection of developing countries

After applying lasso variables selection, we believe our regression model can interpret this problem well. From the perspective of adjusted r-squared in developing countries, this value reached 0.914. Among them, the p-value of the two parameters of population and population density is 0.05. It can be said that population and population density have a greater impact on the vaccination rate in developing countries. At the same time, the population's coefficient is 1.3, and the population density's coefficient

OLS Regression Result of developing countries after Lasso variables selections		
variables	regression coefficients	p-values
intercept	8.33E-17	1
stringency_index	0.0791	0.536
aged_70_older	-0.3327	0.068
population	0.9835	0.013
population_density	-0.7563	0.021
life_expectancy	0.522	0.034
university_rate	0.4441	0.067
R-squared	0.977	
Adj. r-squared	0.932	

Fig. 24. OLS Regression Result of developing countries after Lasso variables selections

is negative 1.08. That is to say, when the population is larger, the vaccination rate will increase, and when the population density is larger, the vaccination rate will decrease (Fig. 25).

OLS Regression Result of developed countries after Lasso variables selections		
variables	regression coefficients	p-values
intercept	-1.11E-16	1
stringency_index	0.4443	0.013
hospital_beds_per_thougsand	0.0322	0.139
aged_70_older	-0.1189	0.03
population	0.243	0.011
new_case asverage_per_million	0.151	0.025
human_development_index	-0.146	0.035
university_rate	-0.9064	0.004
R-squared	1	
Adj. r-squared	1	

Fig. 25. OLS Regression Result of developed countries after Lasso variables selections

The vaccination rates in developed countries can be better interpreted by its linear regression model with the adjusted r-squared of 1.000. Among all the variables, the p-value of the university rate is 0.004. This reflects that university rate has a very significant impact on a country's vaccination rate, and it also matches the negative 0.9 we obtained in the correlation analysis.

4 Conclusion

In conclusion, the relationships between the covid-19 vaccine rates and the 10 factors of developed countries and developing countries are largely different. Based on the correlation analysis, linear regression, and lasso regression, we successfully predict the COVID-19 vaccine rates in one country with its variables. This research can help to increase the COVID-19 vaccine rates by using the models we built and help to ease the COVID-19 pandemic.

References

1. Our World in Data covid-19 dataset. https://github.com/owid/covid-19-data/tree/master/public/data
2. de Figueiredo A, Simas C, Karafillakis E, Paterson P, Larson HJ (2020) Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study. Lancet 396:898–908. K. Elissa, "Title of paper if known," unpublished
3. Craven BD, Islam SMN (2011) Ordinary least-squares regression. In The SAGE dictionary of quantitative management research, pp 224–228
4. Meinshausen N, Bühlmann P (2006) High-dimensional graphs and variable selection with the lasso. Ann Stat 34(3):1436–1462
5. Lasso (statistics) Wikipedia. https://en.wikipedia.org/wiki/Lasso_ (statistics)
6. Least-angle regression Wikipedia. https://en.wikipedia.org/wiki/Least-angle_regression
7. P-value Wikipedia. https://en.wikipedia.org/wiki/P-value
8. Coefficient of determination Wikipedia. https://en.wikipedia.org/wiki/Coefficient_of_determination
9. Correlation Wikipedia. https://en.wikipedia.org/wiki/Correlation
10. OLS regression Wikipedia. https://en.wikipedia.org/wiki/Ordinary_least_square