Are Wealthy Countries Healthier?

An IFP mathematics coursework by Valeriy Proklov

09/02/2023

Contents:

- **Hypothesis 1:** The residents of countries with higher GDP / Capita have longer life expectancies.
- **Hypothesis 2**: Countries with a higher GDP / Capita have lower covid deaths / million.
- **Hypothesis 3:** Countries with higher GDP / Capita have more hospital beds / thousand.

Introduction:

The goal of the paper is to answer the question of “Are Wealthy Countries Healthier”. Because of the complex nature of the question, and a wide range of possible answers, the paper aims to answer the question by individually testing the veracity of 3 hypotheses outlined in the contents. Hypotheses that may benefit from subdividing the data into further categories to more accurately answer the question are split, and the hypothesis is investigated independently for the two subdivisions.

Hypothesis 1:

The residents of countries with higher GDP / Capita have longer life expectancies.

Introduction:

This hypothesis was chosen because of its high relevancy to answering the question. The relationship between GDP per capita and life expectancy is a very good indicator of the relationship between wealth and health in countries.

Method of investigation:

I will test the hypothesis for “poor” countries and “rich” countries independently. A “poor” country I will define as having a GDP/Capita that is lower than the median GDP/Capita. A “rich” country would be the reverse; They all have a GDP/Capita that is higher than the median. By subdividing the countries into “poor” and “rich” I can individually verify the accuracy of the conclusion in both cases.

To visually compare the distribution of life expectancy between “poor” and “rich” countries I will plot two box plots, one for “poor” countries and one for “rich” countries.

Subsequently, I will plot scatter graphs of life expectancy over GDP per Capita for both “poor” and “rich” countries, so that I can compare the regression equations in both cases. I will conclude by testing my hypothesis using the PMCC value of both cases.

Investigation:

Cleaning:

I will first clean the data by removing any entries with missing information for the GDP / Capita and life expectancy columns.

As a result of this operation, I removed the entries of 8 countries: Kosovo, Andorra, Liechtenstein, Monaco, Somalia, Syria, Taiwan and Vatican.

The data is likely missing because the countries do not have the resources to conduct a representative sample of their population or are intentionally withholding information.

Subdivision:

To subdivide the countries as described in the method I first find the median GDP/Capita.

Median GDP/Capita = $11691.49

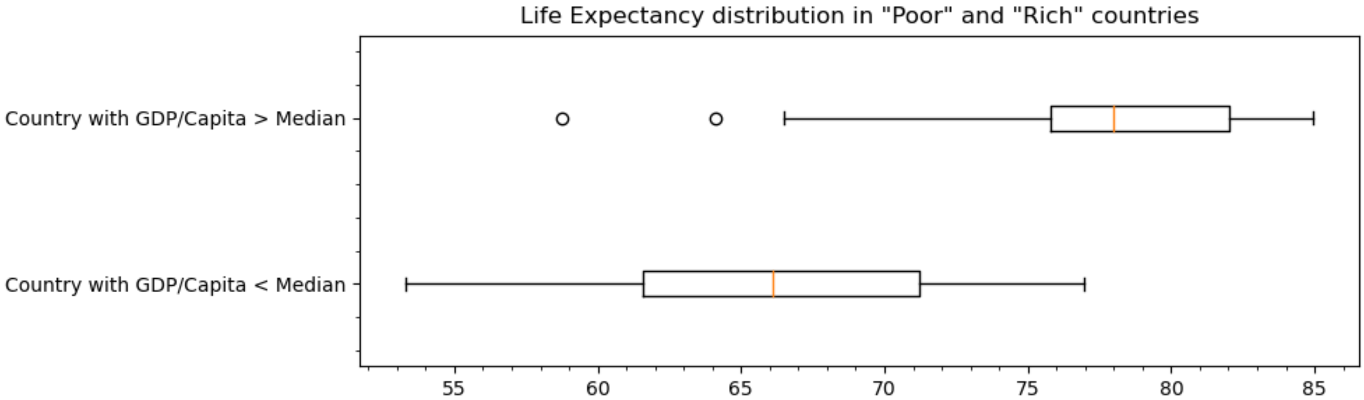
Next, I subdivide the countries into above median, and below median. There are 69 countries in each subdivision.

Below median: Afghanistan, Angola, Armenia, Bangladesh, Benin, Bhutan, Burkina Faso, Burundi, Cambodia, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Congo, Cote d'Ivoire, Democratic Republic of Congo, Djibouti, Egypt, Eritrea, Eswatini, Ethiopia, Gambia, Georgia, Ghana, Guinea, Guinea-Bissau, India, Indonesia, Jordan, Kenya, Kyrgyzstan, Laos, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Moldova, Morocco, Mozambique, Myanmar, Namibia, Nepal, Niger, Nigeria, Pakistan, Palestine, Philippines, Rwanda, Sao Tome and Principe, Senegal, Sierra Leone, South Sudan, Sri Lanka, Sudan, Tajikistan, Tanzania, Timor, Togo, Tunisia, Uganda, Ukraine, Uzbekistan, Vietnam, Yemen, Zambia and Zimbabwe.

Above median: Albania, Algeria, Austria, Azerbaijan, Bahrain, Belarus, Belgium, Bosnia and Herzegovina, Botswana, Brunei, Bulgaria, China, Croatia, Cyprus, Czechia, Denmark, Equatorial Guinea, Estonia, Finland, France, Gabon, Germany, Greece, Hungary, Iceland, Iran, Iraq, Ireland, Israel, Italy, Japan, Kazakhstan, Kuwait, Latvia, Lebanon, Libya, Lithuania, Luxembourg, Malaysia, Maldives, Malta, Mauritius, Mongolia, Montenegro, Netherlands, North Macedonia, Norway, Oman, Poland, Portugal, Qatar, Romania, Russia, San Marino, Saudi Arabia, Serbia, Seychelles, Singapore, Slovakia, Slovenia, South Africa, South Korea, Spain, Sweden, Switzerland, Thailand, Turkey, United Arab Emirates and United Kingdom.

Box plots:

Next, I will plot the two boxplots of life expectancy (for “poor” and “rich” countries). Two outliers were found using the “1.5 IQR” method.



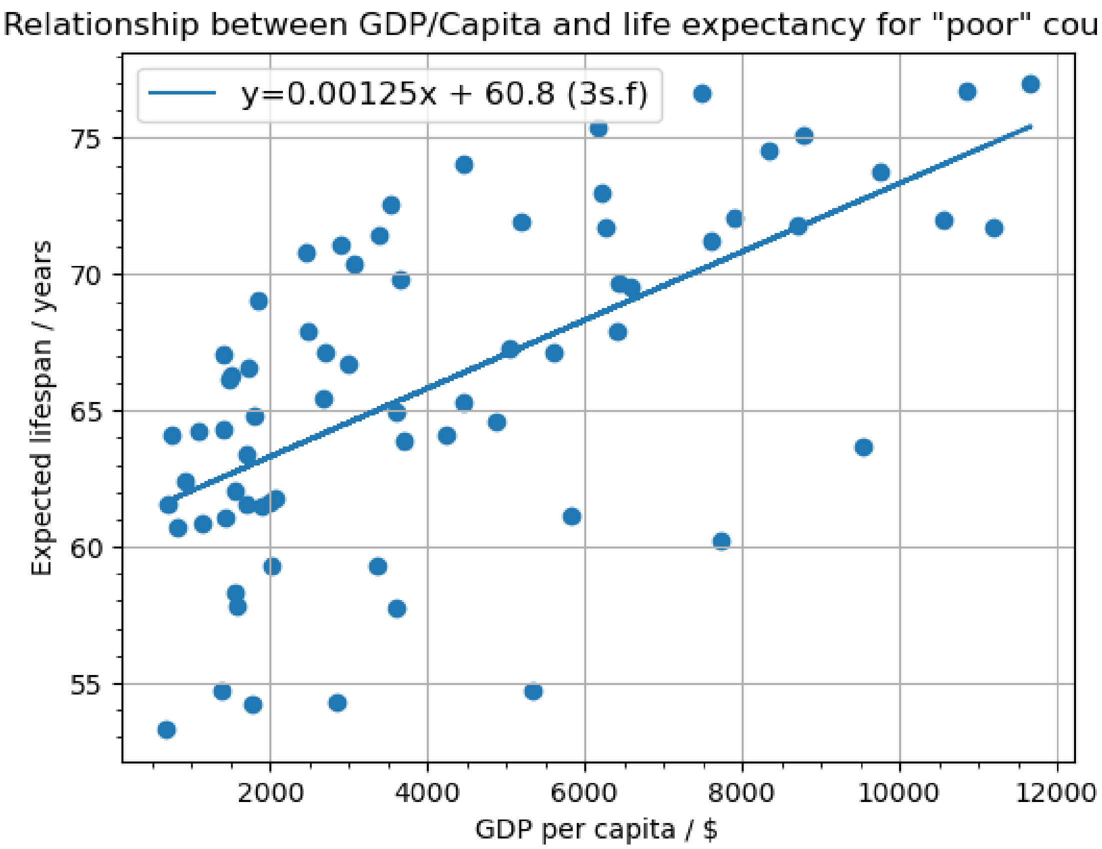
Analysis:

The median life expectancy for “rich” countries (GDP/Capita > Median) is 77.97, compared to 66.12 for “poor” (GDP/Capita < Median) countries. The IQR of “poor” countries was significantly higher than the IQR for rich countries, with a value of 9.76 compared to just 6.5 for “rich” countries.

This suggests that countries with a higher GDP/Capita have (on average) higher expected lifespans, but “poor” countries are less consistent in life expectancy. Additionally, the skew for the “rich” countries is significantly positive, indicating that the data is not evenly distributed, and is concentrated closer to the first quartile. This is not the case for “poor” countries, suggesting that their life expectancies are evenly distributed.

Scatter graphs:

Next, I will plot a scatter graph of life expectancy on GDP / Capita for both categories and calculate the PMCC.

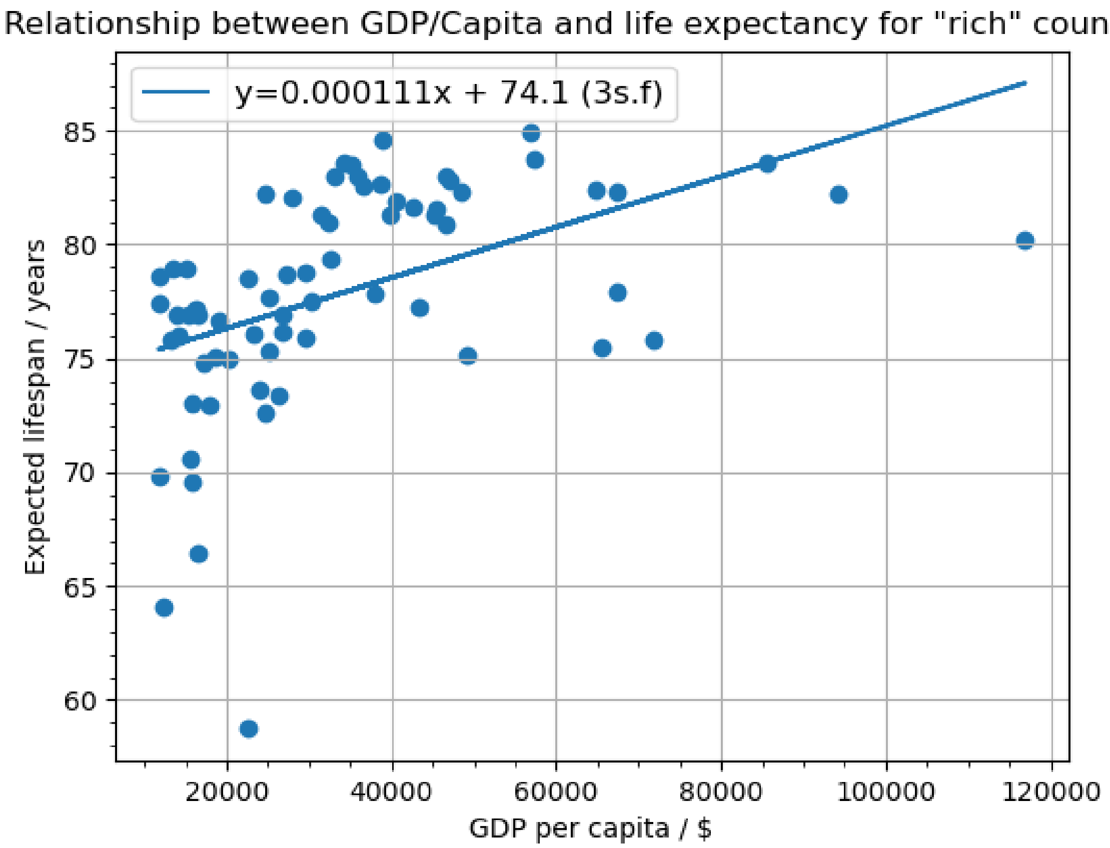


Analysis:

This is a scatter graph of life expectancy on GDP per capita for “poor” countries. PMCC = 0.623 for this data, indicating a moderately strong positive linear correlation. Therefore, I can interpret the linear regression equation.

The slope of 0.00125 indicates that for every 1000$ increase in a “poor” countries GDP, there is a 1.25 year (1 year + 91 days) increase in life expectancy. This is likely because for every added 1000$ the country would have better equipped health services.

Additionally, the Y intercept can be interpreted. A theoretical country with a GDP / Capita of 0 would have a life expectancy of 60.8 Years. This estimate, however, is unreliable as it involves extrapolating the regression line.



Analysis:

This is a scatter graph of life expectancy on GDP per capita for “rich” countries. PMCC = 0.471 for this data, suggesting that there is a lower correlation between GDP / Capita and expected lifespan for “rich” countries. A hypothesis test later in the paper reveals that there is correlation between GDP per capita and life expectancy, so the regression line can be interpreted.

The slope of 0.000111 indicates that for every 1000$ increase in GDP per Capita, there is an increase of 0.111 years (40.5 days) in life expectancy. Interestingly, the gradient of the regression line for “rich” countries is significantly lower than for “poor” countries, suggesting that GDP / Capita is not as important for life expectancy in rich countries as it is for poor countries.

The Y intercept of the regression line suggests that a theoretical “rich” country with a GDP / Capita of 0 would have a life expectancy of 74.1 years. This is likely inaccurate because it involves extrapolating the regression line, and a “rich” country by definition cannot have a GDP / Capita close to or equal to 0.

Hypothesis testing:

For “rich” countries:

H_0: r = 0 H_1: r > 0 PMCC=0.471 Significance level: 5% n=69

For this data, critical value≈0.2 (interpolated). This is less than the PMCC, so there is sufficient evidence to reject H_0 and accept the alternative hypothesis.

Despite the relatively low PMCC of 0.471 there is evidence to suggest that there is a positive correlation between GDP / Capita and expected lifespan in “rich” countries.

For “poor” countries:

H_0: r = 0 H_1: r > 0 PMCC=0.623 Significance level: 5% n=69

For this data, critical value≈0.2 (interpolated). This is less than the PMCC, so there is sufficient evidence to reject H_0 and accept the alternative hypothesis.

This suggests that there is a positive correlation between GDP / Capita and expected lifespan in “poor” countries.

Conclusion:

In conclusion, in order to test the hypothesis of “The residents of countries with higher GDP / Capita have longer life expectancies.” I have subdivided the data to test the hypothesis for “poor” countries and “rich” countries, analysed the distribution of life expectancy in both cases by plotting box plots and scatter graphs, and conducted hypothesis tests for both cases.

The location of the “poor” box and the “rich” box indicated that, at least in the scope of quartiles, there was a positive correlation between GDP / Capita and life expectancy. The scatter graphs also support the positive correlation in both cases. The subdivided hypothesis tests further support the hypothesis in both categories of countries, indicating that even though the PMCC values are not very high for “rich” countries there is still positive correlation.

Interestingly, the PMCC value was significantly lower for the “rich” subdivision than the “poor” subdivision, suggesting that factors other than GDP / Capita may be more impactful towards life expectancy for “rich” countries compared to “poor” countries.

The hypothesis is supported unanimously (but to varying degrees) by all actions described in the method. The hypothesis is therefore declared as true, demonstrating that wealthy countries are healthier, however, the importance of wealth in health is maximized in poor countries as indicated by the difference in regression gradients.

Hypothesis 2:

Countries with a higher GDP / Capita have lower covid deaths / million.

Introduction:

This hypothesis was chosen because it permits exploration of the question in two continents with significantly different approaches to healthcare in the particular case of covid. It allows for trends created by the different policies to be easily compared.

Method of investigation:

In order to investigate the hypothesis, I will subdivide the data into Asian countries and European countries. I do this because the two continents have experienced covid at a slightly different time and implemented substantially different policies to combat it. Investigating the hypothesis in this manner allows for isolating factors unrelated to the hypothesis. (Differences in the covid response policy will be isolated to their respective subdivisions, and not impact the data as a whole).

I will first plot scatter graphs to interpret regression equations and PMCC values for both continents. This will allow easy visualization of trends for both continents and allow for the regression line to be interpreted. Subsequently, I will perform a correlation hypothesis test on both cases.

Note: covid deaths per million is an abbreviation of total covid deaths per million in 2020.

Investigation:

Cleaning:

I will first clean the data by removing any entries with missing information for the GDP / Capita and covid deaths per million columns.

As a result of this operation, I removed the entries of 14 countries:

Bhutan, Cambodia, Laos, Mongolia, Seychelles, Timor, Vatican, Andorra, Liechtenstein, Monaco, Somalia, Syria, Taiwan and Vatican

The cleaning is done before filtering the countries to keep the structure consistent between hypotheses. Like in the first hypothesis, these countries likely did not conduct trustworthy samples of their population or choose to withhold the information.

Subdivision:

To subdivide the countries as described in the method I need to create a subdivision with Asian countries and another of North American countries. To do this I will filter the countries by continent for Asia and North America.

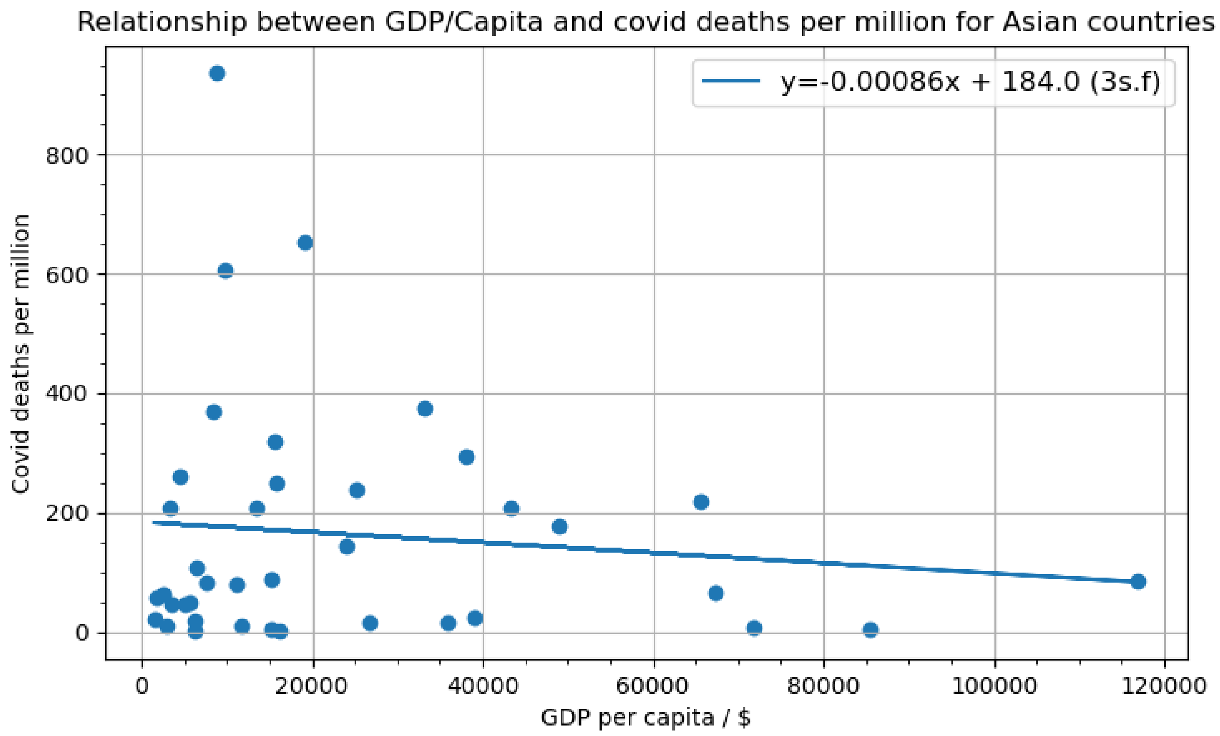
Asian countries (39):

Afghanistan, Armenia, Azerbaijan, Bahrain, Bangladesh, Brunei, China, Georgia, India, Indonesia, Iran, Iraq, Israel, Japan, Jordan, Kazakhstan, Kuwait, Kyrgyzstan, Lebanon, Malaysia, Maldives, Myanmar, Nepal, Oman, Pakistan, Palestine, Philippines, Qatar, Saudi Arabia, Singapore, South Korea, Sri Lanka, Tajikistan, Thailand, Turkey, United Arab Emirates, Uzbekistan, Vietnam and Yemen

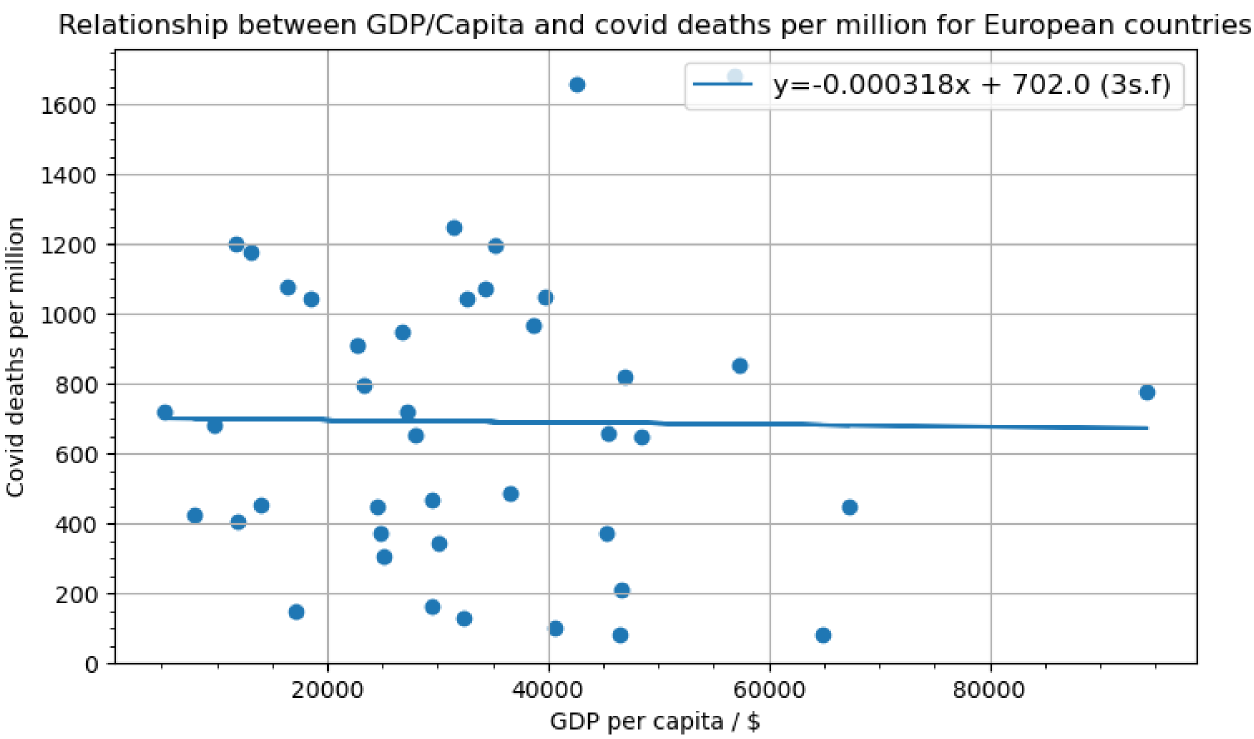
European countries (42):

Albania, Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Kosovo, Latvia, Lithuania, Luxembourg, Malta, Moldova, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Russia, San Marino, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine and United Kingdom

Scatter graphs:



Analysis: PMCC = -0.1115 for this data, signifying that there is hardly any correlation between GDP / Capita and covid deaths per million for Asian countries. A 5% significance level hypothesis test later in the paper determines that there is sufficient evidence to suggest that there is no correlation between GDP / Capita and covid deaths per million for Asian countries. The regression line cannot be interpreted because it does not accurately represent the data.



Analysis: PMCC = -0.0138 for this data, signifying that there is hardly any correlation between GDP per capita and covid deaths per million for European countries. The regression line cannot be interpreted because later in the paper a 5% significance level hypothesis test determines that there is sufficient evidence to suggest that there is no correlation.

Hypothesis testing:

For Asian countries:

H_0: r = 0 H_1: r < 0 PMCC=-0.112 Significance level: 5% n=39

For this data, critical value = 0.268 (interpolated). The absolute value of PMCC is less than the critical value, so there is sufficient evidence to accept H_0 and reject the alternative hypothesis.

There is sufficient evidence to suggest that there is no correlation between GDP / Capita and covid deaths per million in Asian countries. It is therefore unreasonable to interpret the regression equation as it does not accurately represent the data.

For North American countries:

H_0: r = 0 H_1: r < 0 PMCC=-0.0138 Significance level: 5% n=42

For this data, the critical value= 0.258 (interpolated). The PMCC value is less than the critical value, so there is sufficient evidence to accept H_0 and reject the alternative hypothesis.

This suggests that there is no correlation between GDP / Capita and covid deaths per million in North American countries. The regression equation therefore cannot be interpreted as it does not accurately represent the data.

Conclusion:

To investigate the hypothesis “Countries with a higher GDP / Capita have lower covid deaths / million” the data has been subdivided into Asian and European countries. This was done to ensure that different covid response strategies do not falsely affect the conclusion.

Both the scatter diagrams and the hypothesis testing show that there is no correlation between GDP per Capita and covid deaths per million for both continents. This hypothesis answers the question of “Are Wealthy Countries Healthier” with a definite no.

Hypothesis 3:

Countries with higher GDP / Capita have more hospital beds / thousand.

Introduction:

This hypothesis was chosen because it is also highly relevant to answering the question, as the availability of hospital beds is a reasonable indicator of availability of healthcare in a country. Analysing the relationship between GDP per Capita and hospital beds per thousand is highly pertinent to the question.

Method of investigation:

In order to investigate the hypothesis of "Countries with higher GDP / Capita have more hospital beds / thousand", I will plot a scatter graph of hospital beds / thousand over GDP per capita. I will then calculate the PMCC value for the data and perform a one-tailed hypothesis test with a 5% significance level.

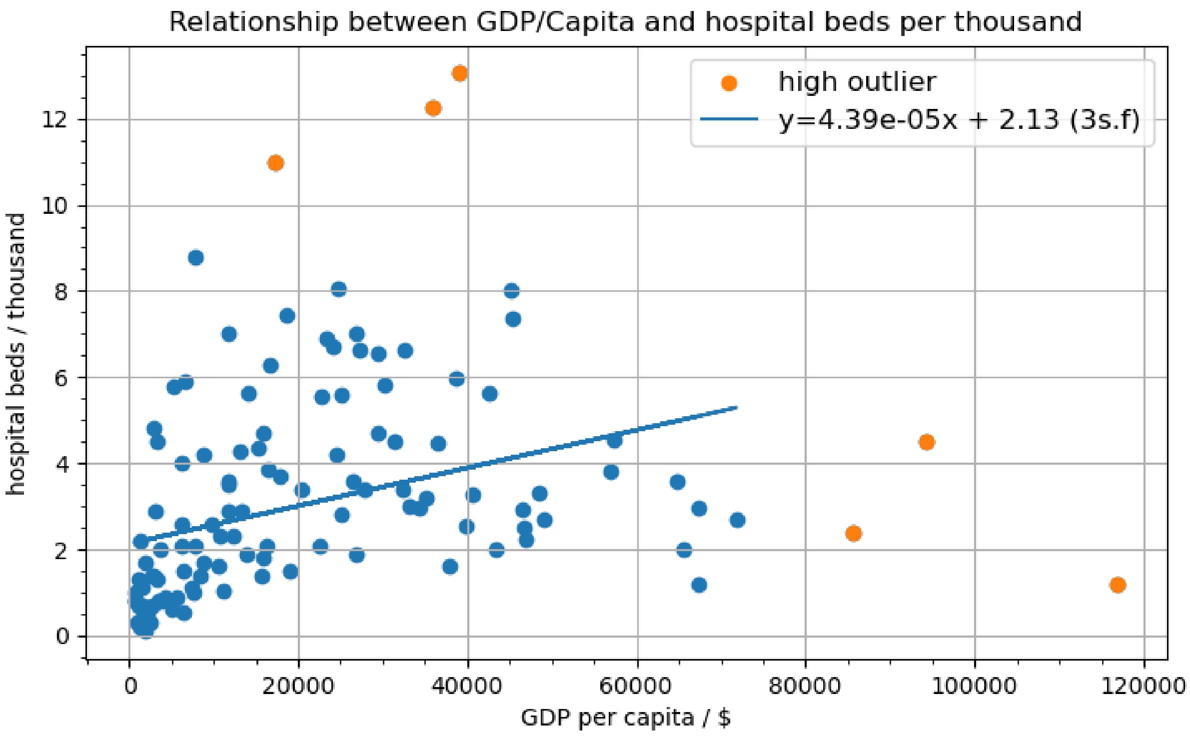
Investigation:

Cleaning:

First, the data is cleaned to remove any entries with missing data. 27 countries are removed: Andorra, Angola, Chad, Congo, Cote d'Ivoire, Democratic Republic of Congo, Guinea-Bissau, Kosovo, Lesotho, Maldives, Mauritania, Namibia, Nigeria, Palestine, Rwanda, Senegal, Sierra Leone, South Sudan, Taiwan, Vatican, Andorra, Liechtenstein, Monaco, Somalia, Syria, Taiwan and Vatican

These countries possibly did not conduct trustworthy samples of their population or choose to withhold the information.

Scatter graph:



Outliers:

Before plotting the graph the following outliers were found using the 1.5 IQR method. No low outliers were found. The outliers are plotted on the graph but ignored for the regression calculation.

High GDP per Capita: Luxembourg, Qatar, Singapore

High hospital beds / thousand: Belarus, Japan, South Korea

Analysis:

A hypothesis test at a 5% significance level later in the investigation suggests that there is a positive correlation, so the regression line can be considered as a good representation of the data.

Therefore, it is possible to interpret that for every 10,000$ increase in GDP per capita there is a 0.439 increase in the number of beds per thousand. Additionally, a theoretical country with a GDP per capita of 0 would have 2.13 hospital beds per 1000 people. That is not reliable, however, as it involves extrapolating the regression line.

Hypothesis testing:

H_0: r = 0 H_1: r > 0 PMCC=0.371 Significance level: 5% n=116

Because the n value is greater than the maximum given in the formula booklet (see references), the closest n value will be taken. The critical value for the test = 0.1654, with n=100. The PMCC > critical value, so there is sufficient evidence to reject the null hypothesis, and accept the alternate.

There is therefore evidence to suggest that there is a positive correlation between GDP per capita and hospital beds per thousand.

By performing the hypothesis test using the p-value (probability that randomly distributed data produces a |PMCC| equal to the actual PMCC), it is possible to avoid the issue of the table being out of range. The p-value = 2.04x10^-5. That is less than the significance level, so there is sufficient evidence to reject the null hypothesis and accept the alternate.

The n=100 estimation concurs with the more accurate method.

Conclusion:

To investigate the hypothesis “Countries with higher GDP / Capita have more hospital beds / thousand” the data was cleaned, and then a scatter diagram was plotted with outliers removed. The PMCC value with the outliers removed was used for a one-tailed hypothesis test at a 5% significance level. Because interpolating the values from the table was not possible, n was taken as the maximum value in the table, and the calculation was then verified against an accurate method. The hypothesis test determined that there is a positive correlation between GDP per Capita and hospital beds per thousand. The correlation is weak, however, as is indicated by the somewhat low PMCC value even with the outliers removed.

It should be noted that hospital beds per thousand does not directly translate into health. A country may have many hospital beds but insufficient medicine, trained personnel, or equipment to provide adequate treatment. Despite this, the hypothesis was included because the availability of hospital beds is a good indicator of the country’s general medical infrastructure (availability of hospitals), and by a rough extension the health of its residents.

The hypothesis test suggests that there is a positive correlation between GDPs per capita and hospital beds per thousand. The above-mentioned hypothesis is therefore valid. If the availability of hospital beds is considered as a reasonable measure of a country’s health, then there is some indication that wealthy countries are healthier from this hypothesis.

Overall conclusion:

To summarize, this paper aims to answer the question of “Are Wealthy Countries healthier?” To achieve this, the structure was split into the following 3 hypotheses that explore the question in a particular context:

- Hypothesis 1: The residents of countries with higher GDP / Capita have longer life expectancies.
- Hypothesis 2: Countries with a higher GDP / Capita have lower covid deaths / million.
- Hypothesis 3: Countries with higher GDP / Capita have more hospital beds / thousand.

The first hypothesis suggests that wealthy countries are healthier as it unanimously shows a positive correlation between GDP / Capita and life expectancy across a wide range of GDP / Capitas.

The investigation of the second hypothesis returns results that are both of low certainty (PMCC) and inconsistent between the continents that it was investigating. It concludes with a low certainty that wealthy countries are not necessarily healthier.

The third investigation was somewhat more productive than the second. Its results showed some indication that wealthy countries are healthier, although the data in the investigation also had low PMCC values.

In conclusion, the question of “Are Wealthy Countries healthier?” is answered somewhat positively by two of the three hypotheses. Hence, the final ruling is that wealthy countries are indeed healthier.

Notes and references:

All the calculations in this paper were performed by a quickly prototyped Python library that acts as a thin wrapper over common statistical libraries such as pandas, Matplotlib and SciPy. Each hypothesis has a corresponding Python script that uses the library to reproduce all the calculations near instantly, given the same Excel file input. The scripts and the library also implement some conditional logic to work with a wide range of inputs (for instance, outliers are plotted automatically, if found). This functionality took far more time than the paper itself.

The library, affectionately called “Shortcourse”, **is published on Github.**

The hypothesis testing calculations in this paper were done by interpolating between different n values in the following table, for ease of marking:

15

Critical Values for Correlation Coefficients

These tables concern tests of the hypothesis that a population correlation coefficient ρ is 0. The values in the tables are the minimum values which need to be reached by a sample correlation coefficient in order to be significant at the level shown, on a one-tailed test.

Product Moment Coefficient					
Level					Sample size, n
0.10	0.05	0.025	0.01	0.005	
0.8000	0.9000	0.9500	0.9800	0.9900	4
0.6870	0.8054	0.8783	0.9343	0.9587	5
0.6084	0.7293	0.8114	0.8822	0.9172	6
0.5509	0.6694	0.7545	0.8329	0.8745	7
0.5067	0.6215	0.7067	0.7887	0.8343	8
0.4716	0.5822	0.6664	0.7498	0.7977	9
0.4428	0.5494	0.6319	0.7155	0.7646	10
0.4187	0.5214	0.6021	0.6851	0.7348	11
0.3981	0.4973	0.5760	0.6581	0.7079	12
0.3802	0.4762	0.5529	0.6339	0.6835	13
0.3646	0.4575	0.5324	0.6120	0.6614	14
0.3507	0.4409	0.5140	0.5923	0.6411	15
0.3383	0.4259	0.4973	0.5742	0.6226	16
0.3271	0.4124	0.4821	0.5577	0.6055	17
0.3170	0.4000	0.4683	0.5425	0.5897	18
0.3077	0.3887	0.4555	0.5285	0.5751	19
0.2992	0.3783	0.4438	0.5155	0.5614	20
0.2914	0.3687	0.4329	0.5034	0.5487	21
0.2841	0.3598	0.4227	0.4921	0.5368	22
0.2774	0.3515	0.4133	0.4815	0.5256	23
0.2711	0.3438	0.4044	0.4716	0.5151	24
0.2653	0.3365	0.3961	0.4622	0.5052	25
0.2598	0.3297	0.3882	0.4534	0.4958	26
0.2546	0.3233	0.3809	0.4451	0.4869	27
0.2497	0.3172	0.3739	0.4372	0.4785	28
0.2451	0.3115	0.3673	0.4297	0.4705	29
0.2407	0.3061	0.3610	0.4226	0.4629	30
0.2070	0.2638	0.3120	0.3665	0.4026	40
0.1843	0.2353	0.2787	0.3281	0.3610	50
0.1678	0.2144	0.2542	0.2997	0.3301	60
0.1550	0.1982	0.2352	0.2776	0.3060	70
0.1448	0.1852	0.2199	0.2597	0.2864	80
0.1364	0.1745	0.2072	0.2449	0.2702	90
0.1292	0.1654	0.1966	0.2324	0.2565	100

International Foundation Programme *Bio-Medical Science and Engineering Pathway*