Predicting Tuberculosis Fatality Rate

Stat Squad

Pledege

Please type your names in the appropriate space below. Failing to do so will result in a 0 on this assignment.

"We have neither given nor received unauthorized help on this assignment"

- Member 1: Evan Zimmerman
- Member 2: Paige Crawford
- Member 3: Drew Sternberg
- Member 4: Shiloh Gelnett

Background

Tuberculosis is a deadly disease that can have extreme consequences if not identified and treated properly. According to an article by the National Institute of Health, "A hospital-based TB diagnosis is a critical opportunity to identify those at high risk of early and overall mortality" (Osman, 2021, pg. 1). In addition, healthcare differs around the world, which can make it difficult for TB to be correctly diagnosed or beaten. It is important to understand the variables that affect the TB fatality rate, so that it can be prevented in the future. According to a journal article by Nature Microbiology, "for the first time in more than a decade, both the estimated incidence of TB and mortality owing to TB increased" (Pai, 2023, pg. 1). This increasingly deadly disease has motivated us to consider what specific factors affect the fatality rate for TB cases.

Research Questions:

- Does a higher HIV rate in a country make the tuberculosis death rate higher?
- Does a country having a low GDP make the tuberculosis death rate higher?
- Do the countries with the least amount of government spending on healthcare have the highest tuberculosis death rate?

Some factors like HIV have a strong relationship with TB. For example, "Recent data estimates show that 3-7 million HIV patients develop TB per year and up to 5 million people develop acute pulmonary TB" (Obeagu, 2023, pg. 128). Another factor, GDP, per capita "was highest for the type IV countries (high-income countries), which indicates that a lower TB incidence is accompanied by increasing affluence" (Lei, 2023, pg. 5). Overall, questions considering HIV rate, GDP, and amount of spending on healthcare are relevant to TB death rate. They explain and emphasize how countries value their healthcare systems and how prosperous their economy is, which can affect how TB is handled and diagnosed.

Data Description

GDP per capita

GDP per capita data was collected by the World Bank with the International Comparison Program (ICP) in 2023. It was collected globally with the help of the United Nations Statistical Commission (UNSC). Data collection involved countries estimating their own GDP, submitting an economic plausibility assessment, and then providing this information to the ICP. The purpose of this was to produce purchasing power parties (PPPs) and comparable price level indices (PLIs) for participating economies. Furthermore, it was created to convert volume and per capita measures of gross domestic product (GDP) and its expenditure components into a common currency using PPPs. However, there are some potential issues. For example, some countries might not be honest about their GDP

with inflation, there could be concerns with the accuracy of the methods of data collection, some countries might have larger GDPs than others which could cause a skew in the data.

Healthcare expenditure per capita

Healthcare expenditure per capita was collected and tracked by the SHA (System of Health Accounts) in a given country over a defined period of time regardless of the entity or institution that financed and managed that spending. In 2023, the WHO converted the expenditure data using PPP time series extracted from WDI (based on ICP 2017) and OECD data. Where WDI/OECD data were not available, IMF or WHO estimates were utilized. The data was presented by the World Bank globally under the framework of the System of Health Accounts in 2011. It was collected to discover relationships between healthcare spending and economic status. Nonetheless, there are some potential issues with this data collection. For example, here could be some issues with accuracy converting the currency units, rounding errors, and mislabeling of country names . For example, "Preparing this original data involves several processing steps. Depending on the data, this can include standardizing country names and world region definitions, converting units, calculating derived indicators such as per capita measures, as well as adding or adapting metadata such as the name or the description given to an indicator." Overall, there could be some rounding or calculation errors.

Number of people living in a country with HIV and Multidrug resistant tuberculosis

Both of the data sets corresponding to the number of people living with HIV and the number of people living with multidrug resistant tuberculosis were collected through the 2019 Global Burden of Disease Study (GBD) conducted by the Institute of Health Metrics and Evaluation. The data was collected globally by the GBD research team through data from hospitals, governments, surveys, and other databases to create a comprehensive healthcare study. This data was collected to support the study's mission of improving healthcare systems and decreasing global health care disparities. A potential issue of data from this study is that researchers created projections and used modeling tools for various locations because they did not have access to data from some locations. The projections that the researchers made, while they are well informed, may not accurately predict the true data for those locations.

Tuberculosis fatality rate and rate of new cases

The data sets that include the rate of new cases and the tuberculosis fatality rate were collected by the World Health Organization (WHO) through their annual Global TB report in 2019. The data is reported by countries to the WHO and then the data is compiled by the WHO. The data is collected by the WHO with the end goal of eradicating tuberculosis cases across the world and lessening the impacts of the illness. A potential issue with this data is various countries may not have as accurate reporting as others. The true statistics may potentially be greater than what countries are reporting to the WHO.

Hemisphere

The geography of the northern and southern hemisphere was collected by Science Direct in 2024 and presented by the World Population Review. This was collected globally to understand the geographical locations of countries and differences in size and population between the northern and southern hemispheres. There are no potential issues because it is all factual, geographical information.
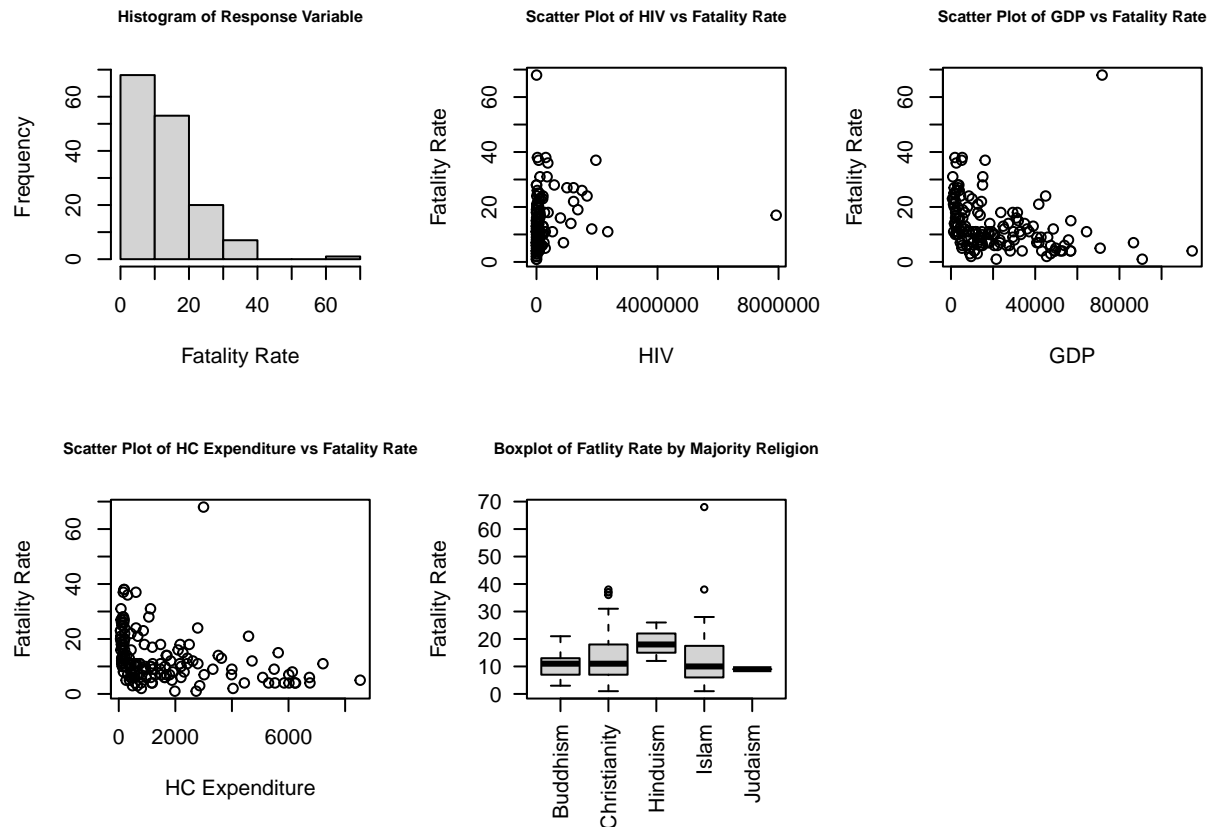
Religious affiliation

Religion by country was collected by Pew Research Center in 2022 and presented by Wise Voter. It was collected globally to provide demographic projections related to religion and to understand human history and culture as it is a source of guidance and inspiration for people around the world. A potential issue is that the size of every dominant religion varies from country to country depending on the population/size and how dedicated/affiliated a country is to religion. This could cause some smaller countries to have a strong majority, while a larger country might be more diverse and have less of a majority.

Political Regime

The political regimes of the world data set was collected by Varieties of Democracy in 2023 for the year 2019 (V-Dem). This data set classifies countries as either a closed autocracy, electoral autocracies, electoral democracies, and liberal democracies. This data was created by V-Dem consulting with over 3,500 political country experts, typically around 25 experts per country. V-Dem then averages the data provided by the experts and supplements the data with their own experts data to assign countries with what political regime they fall into based on a classification system made by V-Dem. This data set was compiled by V-Dem so that people can accurately understand democratization and why it happens. Some flaws in this data include that the four categories of political regime are very broad and can be misleading. Along with this, the data is understood from liberal and electoral viewpoints of democracy which does not account for egalitarian political systems. This can lead to certain countries being inaccurately represented in the data.

Exploratory Data Analysis

Histogram of Response Variable

Scatter Plot of HIV vs Fatality Rate

Scatter Plot of GDP vs Fatality Rate

Scatter Plot of HC Expenditure vs Fatality Rate

Boxplot of Fatlity Rate by Majority Religion

Conclusion

The histogram of the response variable, Fatality Rate, shows the distribution of Fatality Rate to be asymmetric, unimodal, and right-skewed. The skewness of the Fatality Rate is 1.97, which evinces the strong right skew. The mean and standard deviation of the Fatality Rate are 13.34 and 9.44 respectively. The strong right skew indicates that lower Fatality Rates are more common, but there are a few countries who suffer from higher rates. Fatality Rate is suitable for regression because it is continuous with few outliers, however, while a right-skewed response variable like Fatality Rate can be used in regression, the skewness suggests we should check the regression assumptions carefully, and consider transforming the fatality rate if necessary.

The first two graphs we will look at are the scatter plots of GDP vs the fatality rate, and HC Expenditure vs fatality rate. As you can clearly see in the graphs there is a negative exponential relationship between GDP and fatality rate, and HC Expenditure and fatality rate. This makes perfect sense because you would assume the lower the GDP, and HC Expenditure of a country the higher the death rate. The correlation of these variables are -0.2730207, and -0.3130875 which shows there is a moderate negative correlation between the two. This is a low number because the

majority of the countries in our data have low GDP, and low HC Expenditure.

The HIV graph differs from our other plots in that it is virtually a vertical line. There is no obvious visual trend in the data. This is backed by the correlation of 0.1737426, this means that there isn't any linear relationship between the two variables.

For the boxplot of Majority religions, all the fatality rates seem pretty similar across all religions, except for Hinduism. All of the countries have a median between 9 and 11 except for Hinduism, which has a median of 18. Its first and third quartiles are also substantially larger than the quartiles of the other religions. Islam and Christianity have much wider ranges of their IQR than the others, but this is likely due to the fact that Islam and Christianity are the majority religion for 90% of the countries in our dataset, and there is greater variation amongst the countries that practice those religions.

We will continue to examine these predictors by manipulating and transforming the variables to discover new relationships among them. For example, we will use the log transformation to make the exponential graphs more suitable for linear regression. To conclude our preliminary findings, it seems that the HIV rate does not have a relationship with the tuberculosis death rate, a country having a low GDP does correspond with a higher tuberculosis death rate, and the countries with the least amount of government spending on healthcare have the highest tuberculosis death rate.

Appendix A: Data Dictionary

Variable Name	Abbreviated Name	Description	Units	Levels (if Qualitative)
Fatility Rate	Fatality Rate	Percentage of deaths among diagnosed tuberculosis cases	Fatality percentage	
GDP per capita	GDP	The gross domestic product that measures a country's economic well-being	Currency in international-$	
Healthcare expenditure per capita	HC expenditure	The amount spent on healthcare services divided by a country's population	Currency in international-$	
Number of people living in a country with HIV	HIV	The number of people living with diagnosed HIV (human immunodeficiency virus) in each country	Number of people living with diagnosed HIV	
Multidrug resistant Tuberculosis	Multidrug Resistant TB	A type of TB that is resistant to at least 2 different types of anti-TB drugs	Number of people with diagnosed multidrug resistant TB	
Rate of New Cases	Rate of New TB	The proportion of people recently diagnosed with TB per 100,000 people in each country	Proportion of people with recently diagnosed TB	

Variable Name	Abbreviated Name	Description	Units	Levels (if Qualitative)
Political Regime	Political Regime	A set of rules, protocols, and cultural norms that regulate how a government functions	N/A	Closed autocracy, Electoral autocracy, Electoral democracy,Liberal democracy
Hemisphere	Hemisphere	Hemisphere that a country belongs to	N/A	Northern, Southern
Dominant Religious Affiliation	Majority Religion	Dominant religion of a country	N/A	Christianity, Islam, Buddhism, Hinduism, Judaism

Appendix B: Data Rows

```
      Country Year        GDP Fatality Rate HC expenditure         HIV
1 Afghanistan 2019   2079.922            14       285.5581    5125.301
2      Algeria 2019 11627.280            11       750.4487    9485.271
3       Angola 2019   6602.424           18       178.0261 383909.750
4    Argentina 2019 22071.748             6      2198.8804 188657.250
5      Armenia 2019 14317.553             6      1616.1779    1389.422
6    Australia 2019 49379.094             4      5294.4630   17305.775
  Multidrug Resistant TB Rate of New TB    Political Regime Hemisphere
1             1762.45100          189.0 electoral autocracy      North
2              221.38165           61.0 electoral autocracy      North
3             3405.68300          351.0 electoral autocracy      South
4              120.33472           29.0 electoral democracy      South
5              214.82237           26.0 electoral democracy      North
6               37.55193            6.8   liberal democracy      South
  Majority Religion
1             Islam
2             Islam
3      Christianity
4      Christianity
5      Christianity
6      Christianity
```

Appendix C: References

Background

1. Lei, Y., Wang, J., Wang, Y., & Xu, C. (2023). Geographical evolutionary pathway of global tuberculosis incidence trends. BMC Public Health, 23(1), 755.

2. Obeagu, E. I., & Onuoha, E. C. (2023). Tuberculosis among HIV Patients: A review of Prevalence and Associated Factors. Int. J. Adv. Res. Biol. Sci, 10(9), 128-134.

3. Osman, M., Meehan, S. A., von Delft, A., Du Preez, K., Dunbar, R., Marx, F. M., Boulle, A., Welte, A., Naidoo, P., & Hesseling, A. C. (2021). Early mortality in tuberculosis patients initially lost to follow up following diagnosis in provincial hospitals and primary health care facilities in Western Cape, South Africa. PloS one, 16(6), e0252084. https://doi.org/10.1371/journal.pone.0252084

4. Pai, M., Dewan, P. K., & Swaminathan, S. (2023). Transforming tuberculosis diagnosis. Nature Microbiology, 8(5), 756-759.

Data

1. Healthcare expenditure vs. GDP per capita. Our World in Data. (n.d.). https://ourworldindata.org/grapher/healthcare-expenditure-vs-gdp?tab=table

2. Northern Hemisphere countries 2024. (n.d.). https://worldpopulationreview.com/country-rankings/northern-hemisphere-countries

3. Number of people living with HIV. Our World in Data. (n.d.-b). https://ourworldindata.org/grapher/number-of-people-living-with-hiv?tab=table Pew Research Center. (2022, December 21). Religious composition by country, 2010-2050. Pew Research Center's 4. Religion & Public Life Project. https://www.pewresearch.org/religion/interactives/religious-composition-by-country-2010-2050/

4. Saloni Dattani, Fiona Spooner, Hannah Ritchie and Max Roser (2023) - "Tuberculosis" Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/tuberculosis' [Online Resource]

5. Southern Hemisphere countries 2024. (n.d.). https://worldpopulationreview.com/country-rankings/southern-hemisphere-countries V-Dem (2023) – with major processing by Our World in Data. "Political regime – Regimes of the World" [dataset]. V-Dem, "Democracy and Human rights, OWID based on Varieties of Democracy (v13) and Regimes of the World v13" [original data]. Retrieved March 23, 2024 from https://ourworldindata.org/grapher/political-regime

6. WHO (2023); Population based on various sources (2023) – with minor processing by Our World in Data. "Rate of new tuberculosis cases" [dataset]. WHO, "Global Tuberculosis Report"; Various sources, "Population" [original data]. Retrieved March 23, 2024 from https://ourworldindata.org/grapher/incidence-of-tuberculosis-sdgs

7. World Bank (2023) – with minor processing by Our World in Data. "GDP per capita – World Bank" [dataset]. World Bank, "World Bank World Development Indicators" [original data]. Retrieved March 23, 2024 from https://ourworldindata.org/grapher/gdp-per-capita-worldbank