Fiona Sternberg 22795532

Funding Health, Extending Life – Final Report

Introduction

The life expectancy at birth in the LAC region (Latin America and the Caribbean) has continuously changed in the last few years. This project analyzes the influence on health care expenses on the life expectancy in the LAC region to answer the question: Do higher health expenses per capita increase the life expectancy at birth in Latin American and Caribbean countries?

Used Data

The analysis is performed using an SQLite database containing two tables life_expectancy and health_expenditure. Both tables contain data for the years 2000 to 2021 for the same 33 countries in the LAC region. If there were any missing values in the original datasets, they were interpolated so there are no null values.

Analysis

The analysis was performed using python, specifically the pandas, scipy and statsmodel modules. Pandas was used to utilize its dataframe handling. The scipy module provided the needed functions to compute Pearson's correlation coefficient and the statsmodel module was used to perform a linear regression with an OLS model (Ordinary Least Squares).

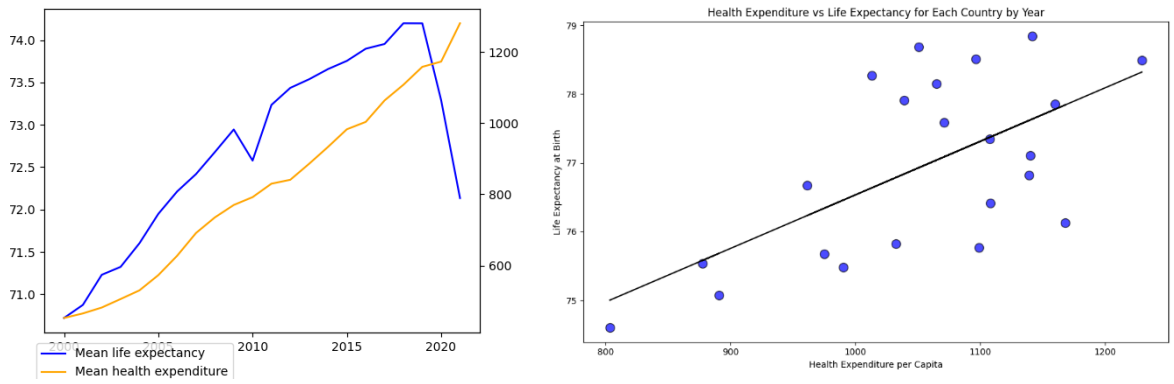


Figure 1: Left: mean life expectancy and health expenditure over the years, right: example scatter plot (Argentina)

Descriptive analysis

First, a descriptive analysis was performed to get acquainted with the data. A graph was plotted containing the mean values for life expectancy and health expenditure for each year (see Figure 1). This figure indicated that for most years, higher health expenditure and higher life expectancy occurred together. Additionally, scatter plots for each country were produced to identify a trend in the data (see Figure 1). The descriptive analysis indicated that for most countries (n=29), higher health expenditure led to a higher life expectancy.

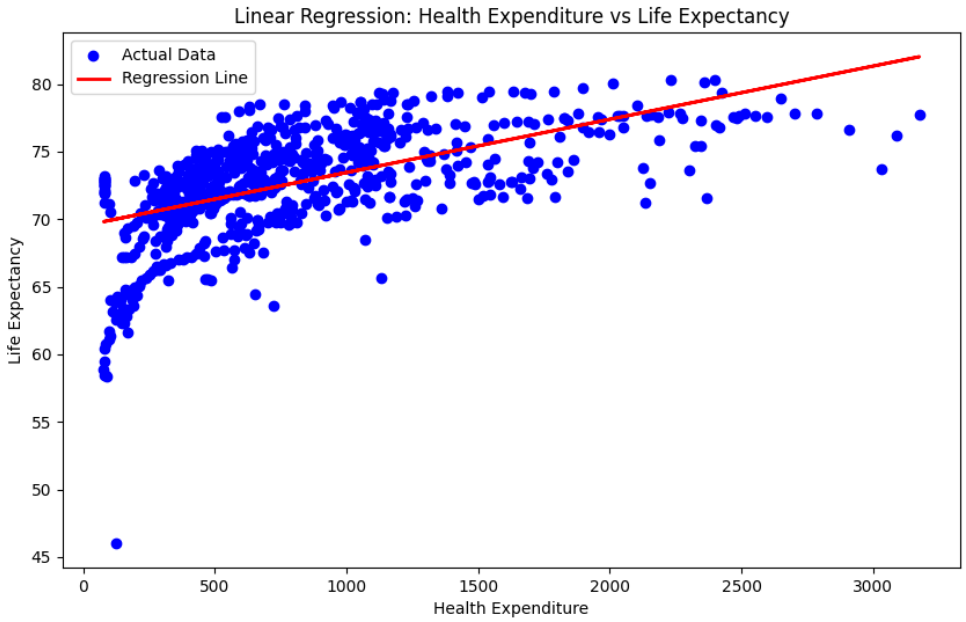
Correlation analysis

Second, a correlation analysis was performed during which Pearson's correlation coefficient was computed for each country. This analysis was used to identify a possible correlation between life expectancy and health expenditure. There were positive and statistically significant (p-value $\leqslant$ 0,05) correlations for 25 of the 33 analyzed countries and only one negative statistically significant correlation for the country Venezuela which had very few recorded data points concerning health expenditure.

The mean correlation was 0.515 if all statistically significant samples were included. Without the negative sample, the mean increases to 0.709. This shows that there tends to be a positive correlation between life expectancy and health expenditure.

Regression analysis

To determine whether higher health expenditures caused higher life expectancy, a linear regression was performed using an OLS model. The results of the model are statistically significant with a p-value of 0.000. The slope of the regression was 0.0039 which means that for each unit increase (1$) in health expenditure, the life expectancy increases by 0.0039 years (approximately 1.4 days). The intercept was 69.5184 which is the estimated life expectancy without any health expenditures. The model predicts an increase in life expectancy with higher health expenditures (see Figure 2).



.

Figure 2: Scatter plot of linear regression

However, the R-squared value was relatively low (0.334) which means that only about 33% of the variance in the life expectancy could be explained by the health expenditure. This indicates that other factors (e.g. GDP) have an influence on the life expectancy in the LAC region.

Conclusions

The analysis shows that higher health expenses per capita do increase the life expectancy at birth in Latin American and Caribbean countries but there are other influences like the GDP.

However, there were several limitations in the analysis. Firstly, only 33 of 42 countries in the LAC region could be included in the analysis due to missing data. If sufficient data about these countries were known, the outcomes might differ. Secondly, only the analysis only included measures of linear correlation, which assume a straight-line relationship between variables. However, in cases where the relationship between the variables is non-linear, linear methods may not capture the true association.