Factors influence health expenditure

Yihang Luo

Research Question

Health expenditure has always been a large part of household expenses. With the advance of technology in recent years, families have been required to spend more and more on health. Beside the basic health insurance covered by part of the salary and the expenses given to the government for medical care, additional costs may be incurred to pay clinicians or for unexpected illnesses. As a student from a middle-class background with high personal health conscious, I am more than aware of how the high cost of a serious illness or accident can tear a family with a normal income apart. Thus, studying the factors that influence spending on health can improve the awareness of people related to health, may to some extent prevent the excessive health expenditure by families who are not financially well off.

Background/Literature

A search on "factors influence health expenditure" yields 2370000 journals and articles, and I choose three of them which are helpful to my topic. The first journal from WHO, the determinants of health expenditure: a country-level panel data analysis (Ke, X., Saksena, P., & Holly, A., 2011), conducted static and dynamic panel models of government health expenditures, out-of-pocket costs, and total health expenditures, showing that factors besides income contribute to this variation, from population age structure, technology progress to health system characteristics. In some level, it contributes to the factors I choose other than GDP, while I focus the factors more on personal health conditions, educations, or society factors like mortality or number of passed abortion laws.

The second article, Household catastrophic health expenditure: a multicountry analysis (Xu, K., Evans, D. B., Kawabata, K., Zeramdini, R., Klavus, J., & Murray, C. J., 2003), explored variables associated with catastrophic health expenditure in 59 countries by regression analysis. They discovered that the proportion of households facing catastrophic payments from out-of-pocket health expenses varied widely between countries, which is why the topic is a great concern to lots of families around the world.

The third journal, the determinants and effects of health expenditure in developed countries (Hitiris, T., & Posnett, J., 1992), re-examined the results of previous work using a larger sample of 560 pooled time-series and cross-section observations. They confirmed the importance of GDP as a determinant of health spending and the importance of some non-income variables such as demographic structure, which let me reassure the inclusion of GDP variables and the try to add some demographic structure factors such as unemployment rate, number of passed abortion laws and education level.

Exploratory Data Analysis

Data Source and Description

There are 9 variables I would like to study for this project:
- response variable: percentage health expenditure of total expenditure

- predictor 1: GDP
- predictor 2: life expectancy
- predictor 3: adult mortality
- predictor 4: alcohol
- predictor 5: BMI
- predictor 6: schooling
- predictor 7: number of passed abortion laws
- predictor 8: unemployment rate

Based on the background research conducted, these variables all appear to be related to the response variable, so I include them to be consistent with previous results. They come from Kaggle data for studying life expectancy among 16 years from 2000 to 2015 and unemployment rate in 2007, deriving from WHO data, UNdata and World Bank data, including 134 countries as observations. I calculate the mean value for the 16 years to find an approximate number for the variables in recent years. The predictors come from both personal health conditions and demographic structure among varies countries so people and government can focus more on these factors in order to decrease the expenditure on health.

Website: 1. https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who
2. https://www.who.int/data
3. https://www.kaggle.com/datasets/pantanjali/unemployment-dataset
4. https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS
5. http://data.un.org/Data.aspx?q=abortion&d=GenderStat&f=inID%3a11

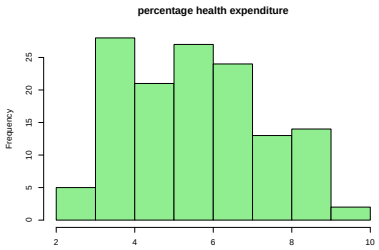
Summaries of the Variables

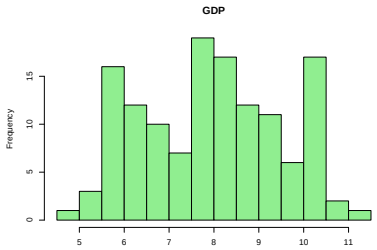
Table 1 presents numerical summaries of the variables, that show a bunch of really interesting properties, which I will discuss below.

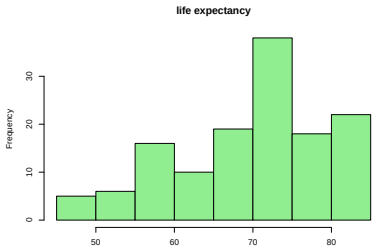
Variable	Mean	Min	Max
Response(percentage health expenditure)	5.593	2.150	9.255
Predictor 1(GDP)	7.948	4.975	11.012
Predictor 2(life expectancy)	69.55	45.79	82.53
Predictor 3(adult mortality)	4.903	3.808	6.154
Predictor 4(alcohol)	4.724	0.010	13.497
Predictor 5(BMI)	12.51	8.28	65.95
Predictor 6(schooling)	12.159	3.927	20.013
Predictor 7(number of passed abortion laws)	0.5756	0	1
Predictor 8(unemployment rate)	6.921	0.580	19.441

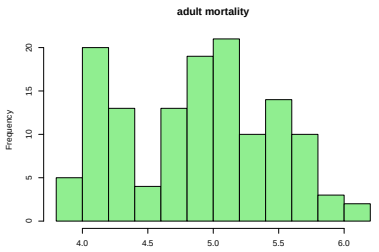
Table 1: All variables summary

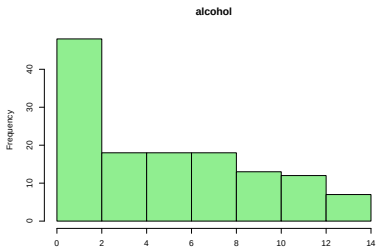
We can also notice some ingenuity in the shape of some variables with the histogram and boxplot presented in Figure 1.

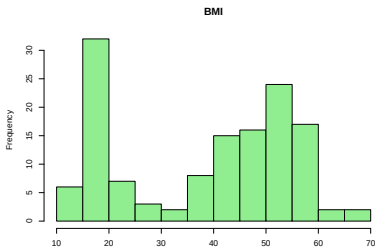

percentage health expenditure
Frequency

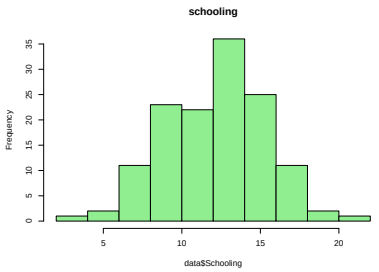

GDP
Frequency

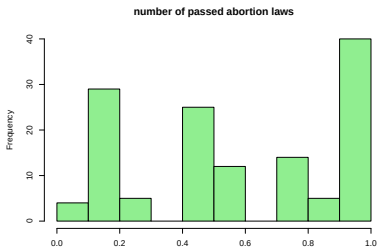

life expectancy
Frequency

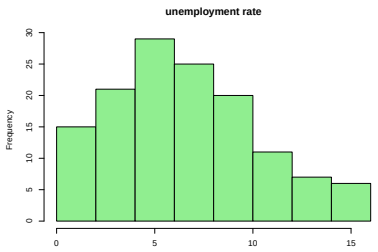

adult mortality
Frequency

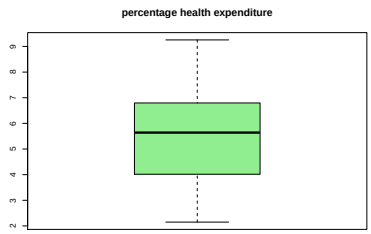

alcohol
Frequency

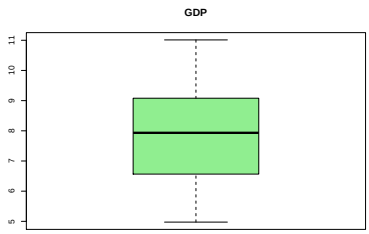

BMI
Frequency

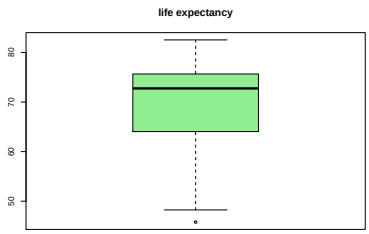

schooling
Frequency
data$Schooling

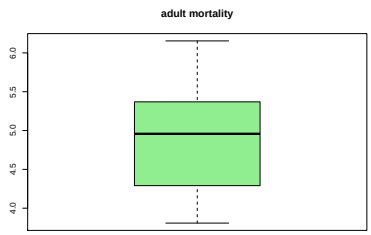

number of passed abortion laws
Frequency

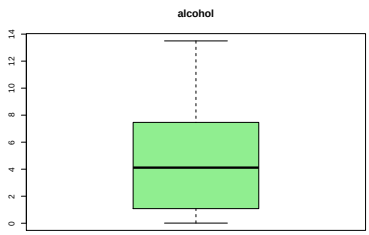

unemployment rate
Frequency

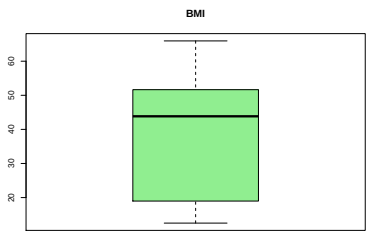

percentage health expenditure

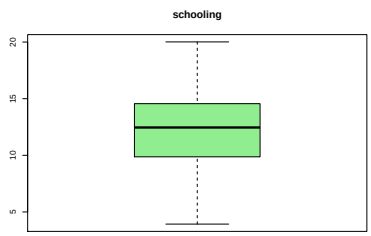

GDP

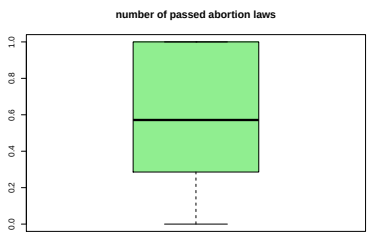

life expectancy


adult mortality


alcohol


BMI


schooling


number of passed abortion laws

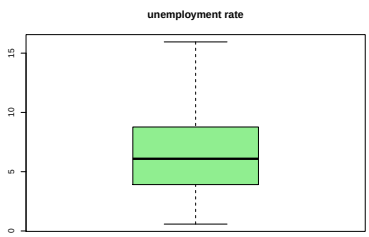

unemployment rate

Figure 1: Histogram and boxplot of each variables

Then, we can make the scatter plots in figure 2 that represent the linear relationship between response variables and the predictors.

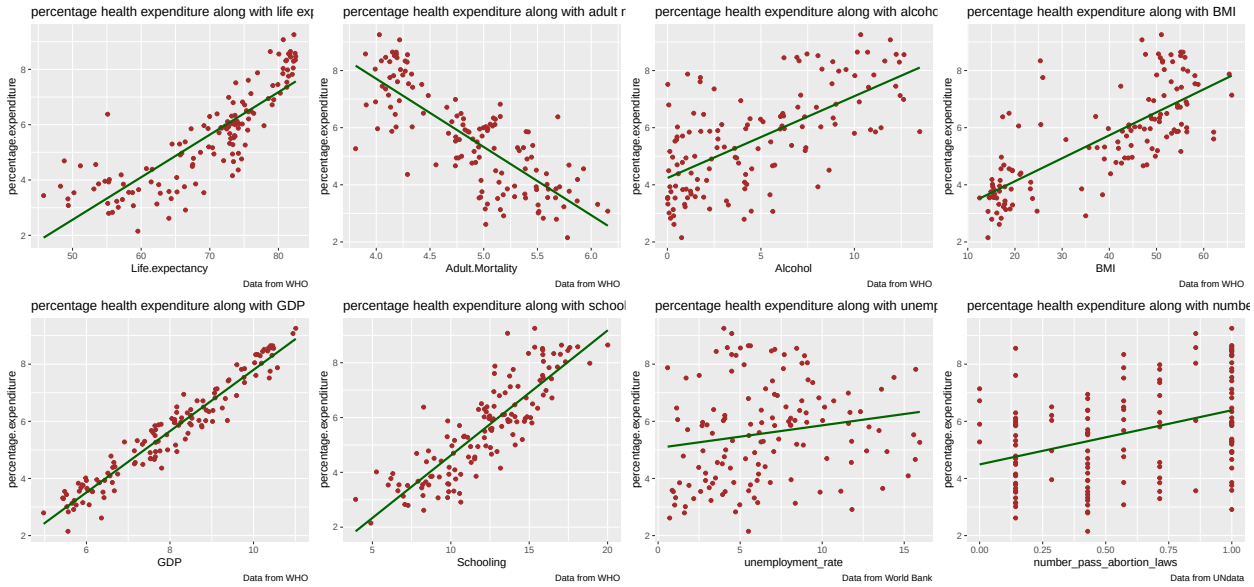


Figure 2: Scatter plot and linear relationship between percentage health expenditure and other factors

Discussion of Linear Regression Models

The response variable of the project is Y which is the percentage health expenditure, and predictor 1, GDP, would be the predictor of interest. All of other predictors are added as the additional predictors related to the response variable in order to ensure the unbiased estimation of the relationship. Since most of the predictors are numerical data and all predictors showed the significant fitted linear regression lines with the response variable, it can be nice to make a multi-linear regression model based on these predictors.

Assumptions of linear model:

- Assumption 1: Linearity
  As we can see from figure 2, all of the predictors are showing significant or somewhat significant fitted linear regression line with the response variable.
- Assumption 2: Uncorrelated
  All the observations are data from different countries, which means all of them are perfectly uncorrelated.
- Assumption 3: Constant variance
  From linear regression between response variable and predictors in figure 2, the dispersion of the distribution is relatively uniform instead of Trumpet-shaped, which means each of the variables have a relatively constant variance.
- Assumption 4: Normal
  From the histogram of percentage expenditure in figure 1, we can see the data of percentage expenditure is symmetric distributed with no few skewness, which means data as well as the residual of response variables shows normal distribution.

References

Hitiris, T., & Posnett, J. (1992). The determinants and effects of health expenditure in developed countries. *Journal of health economics*, 11(2), 173-181.

Xu, K., Evans, D. B., Kawabata, K., Zeramdini, R., Klavus, J., & Murray, C. J. (2003). Household catastrophic health expenditure: a multicountry analysis. *The lancet*, 362(9378), 111-117.

Ke, X., Saksena, P., & Holly, A. (2011). The determinants of health expenditure: a country-level panel data analysis. *Geneva: World Health Organization*, 26(1-28).