Avoiding a coup

An analysis of social factors and their effect on healthy life expectancy at birth

# Abstract

When attempting to predict life expectancy, the current focus is on the effect factors directly linked to health and wellness have. However, to accurately estimate life expectancy, factors beyond these need to be considered. This paper analyzes the impact societal factors have on healthy life expectancy at birth on a country wide level. The data used within this analysis comes from an aggregation of 12 years of surveying the happiness of citizens of countries around the world. Models were created and compared with an emphasis on simplicity of understanding. Factors were split into 2 categories, those that increased estimated healthy life expectancy at birth, and those that decreased it. The values that increased estimated healthy life expectancy at birth in descending order are as follows: Negative affect, Positive affect, Freedom to make life choices, Social support, Log GDP per capita, Life ladder, and Delivery quality. The values that decreased estimated healthy life expectancy at birth in descending order are as follows: GINI world bank estimate, Generosity, Confidence in national government.

# Introduction

The average life expectancy of a country’s residents is an important statistic to both its citizens and governments. Citizens may use this information to determine whether to stay in the country, or if possible, move to a country with a higher life expectancy. Governments need this statistic to help manage their countries population. When attempting to estimate this statistic, an emphasis is placed on factors directly linked to health and wellness. Factors that fall under this category include existing medical conditions, diet, exercise habits, etc. In many cases, the impact that a countries societal factors may have been not considered. This report attempts to analyze what societal factors have a significant statistical impact on country wide healthy life expectancy as well as their level of affect.

Data

Data used in this analysis comes from the World Happiness report of 2020. Values come from country wide averages of survey responses from the years of 2005 to 2017. Variables with more than 50% of the values missing were removed. Variables that created too granular an analysis, such as year or country, were not included in the analysis. Remaining variables containing missing values were imputed using the median values of said variables. These remaining variables are comprised of average values of survey responses to many social aspects of citizen life as well as the average value for the GINI index of the country for that year. Descriptions of variables used in this analysis can be found in the following table.

|  |  |
| --- | --- |
| Data Dictionary | |
| Negative affect | The average of binary responses to survey questions asking about the feelings worry, sadness, and anger. |
| Positive affect | The average of binary responses to survey questions asking about laughter, happiness, and enjoyment. |
| Freedom to make life choices | The average of responses to the question “Are you satisfied or dissatisfied with your freedom to choose what you do with your life?” A yes is equivalent to 1, and 0 to no. |
| Social support | The average of responses to the question “If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?” A yes is equivalent to 1, and 0 to no. |
| Log GDP per capita | This statistic is relative to purchasing power parity at 2011 international dollar prices. |
| Life ladder | The average of responses to the question “Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?” |
| Delivery quality | This refers to the governments exercise of power in the following aspects: rule of law, control of corruption, regulatory quality, and government effectiveness |
| GINI world bank estimate | A numerical representation of distribution of income. A value closer to 0 represents a more income equality distribution |
| Generosity | The average of responses to the question “Have you donated money to a charity in the past month?” A yes is equivalent to 1, and 0 to no. |
| Confidence in national government | The average of responses to a question asking about citizens confidence in their government. A yes is equivalent to 1, and 0 to no. |

Table 1. Description of variables used

# Methods

After data preprocessing, the data was split into two categories, training and validation. The training data set comprised 70 percent of the original data and was used to create all the models further discussed in this section. The remaining 30 percent was used to evaluate the models using techniques outlined in the results sections. A total of 6 linear regression models were created. The predictor variables within these models were selected using stepwise regression. Akaike information criterion (AIC) was used for comparison by the stepwise regression for variable selection. 3 of the models were created using a formula containing all predictor variables remaining after cleaning, and allowed for interaction between them. The models created using this formula differed in the direction the stepwise regression traveled in. These directions were forward, backward and both directions. The remaining 3 models were created using a formula containing all predictor variables remaining after cleaning, but did not allow for variable interaction. Again, the models created using these formulas differed in the direction the stepwise regression traveled in: forward, backwards and both directions.

# Results

The 6 models created were evaluated using the criteria of R2, or the proportion of variance of estimated healthy life expectancy at birth that can be explained by the predictor variables,as a representation of the goodness of model fit. Evaluation also consisted of comparison of number of predictor variables comprising the model as a representation of model complexity. The models created with a formula that allowed interactions between predictor variables will be referred to by the following: forward interaction model (FIM), backward interaction model (BWIM) and both interaction model (BIM), for the stepwise regression directions parameter of forwards, backwards, and both respectively. The models created with the formula disallowing interactions between predictor variables will be referred to by the following: forward model (FM), backward model (BWM), and both model (BM). The following table displays the R2 values and the number of variables in each model.

|  |  |  |
| --- | --- | --- |
|  | R2 | Number of Variables |
| FIM | 0.82 | 45 |
| BWIM | 0.82 | 55 |
| BIM | 0.82 | 39 |
| FM | 0.73 | 11 |
| BWM | 0.73 | 11 |
| BM | 0.73 | 11 |

Table 2. Comparison of each model using R2 and number of variables

In looking at figure one, FIM, BWIM, and BIM all have the same R2 values of 0.82. Similarly, FM, BWM, and BM all have R2 values of 0.73. When comparing the number of variables, FM, BWM, and BM are all equivalent. This is because the stepwise regression with no interaction between variables resulted in the same model being created for all 3 directions. With an emphasis on simplicity of model and how it relates to ease of understanding, BM was selected as the model to continue analysis on as the gain in simplicity outweighed the loss of model performance.

Within the selected model, BM, the following predictor variables are included. Delivery quality, Freedom to make life choices, Life Ladder, Log GDP per capita, Negative affect, Positive affect, Social support, Confidence in national government, Generosity, GINI World Bank Estimate. For each of these predictors, the coefficients are interpreted as having their value impact on the healthy life expectancy at birth (HLE) for every 1 unit they are increased while all other variables remain constant. All variables within this model are significant at the a = 0.05 level. The following table displays the predicted variables that are interpreted as having a positive impact on healthy life expectancy at birth as well as the coefficients for each variable.

|  |  |
| --- | --- |
| Variable Name | Increase to HLE for every unit increase |
| Delivery quality | 1.02 |
| Freedom to make life choices | 3.74 |
| Life ladder | 1.35 |
| Log GDP per capita | 2.89 |
| Negative affect | 9.0 |
| Positive affect | 3.9 |
| Social support | 3.38 |

Table 3. Variables and their increase to HLE

The following table displays the predicted variables that are interpreted as having a positive impact on healthy life expectancy at birth as well as the absolute value of the coefficients for each variable.

|  |  |
| --- | --- |
| Variable Name | Decrease to HLE for every unit increase |
| Confidence in national government | 2.16 |
| Generosity | 3.19 |
| GINI world bank estimate | 13.66 |

Table 4. Variables and their decrease to HLE

Discussion

For the purposes of comparison, three example countries have been created. Country A was created using the average value for each predictor. Country B was created using the minimum value for each predictor. Finally, country C was created using the maximum value for each predictor. The values for each predictor variable for each are displayed in the following figure.

Figure 1. Values for min max and average factor values

When the values for each country are inputted into the model, the corresponding estimates for healthy life expectancy at birth are displayed in the following graph.

Figure 2. Estimated HLE for mean, min, and max values

When considering societal factors, to improve the country wide healthy life expectancy at birth Country A needs to focus on both increasing factors that contribute positively and decreasing factors that contribute negatively. Increasing Negative effect up to the maximum value can increase Country A’s life expectancy by 3.94 years. Giving positive affect the same treatment can result in a 2.106 year increase to life expectancy. Decreasing the GINI world bank estimate to the minimum value found within the dataset leads to an increase of 1.841 years of healthy life expectancy at birth.  
 As Country B has the lowest value for each predictor, a focus should be placed on increasing factors that have a positive impact on healthy life expectancy at birth. Increasing Negative effect up to the maximum value can increase Country B’s life expectancy by 5.6 years. Giving positive affect the same treatment can result in a 2.42 year increase to life expectancy. Increasing Freedom to make life choices to the maximum value within the dataset results in a 2.72 year increase to life expectancy.

For Country C to increase its life expectancy, it needs to decrease the 3 factors that reduce life expectancy. Decreasing GINI world bank estimate to the minimum value found within the data results in a 5.123 year increase in life expectancy. Decreasing Generosity to the minimum value results in a 2.1385 year increase to life expectancy. Decreasing Confidence in national government to the minimum value results in a 2.02 year increase to life expectancy.

# Conclusion

Healthy life expectancy at birth is an estimated value that requires the input of many factors to accurately predict. An emphasis on physical factors is common, but social factors also impact this estimate. Without knowledge of what factors have an effect and how much they affect this estimate efficient actions cannot be taken. This report highlighted several factors as well as the impact on the number of years they increase or decrease countrywide healthy life expectancy at birth. Further analysis at the individual level is recommended to allow each citizen to improve their unique healthy life expectancy at birth.   
Limitations

The largest limitation of this analysis is the emphasis on model simplicity. The ability for those without broad analytic knowledge to understand and use the information provided is paramount for this analysis to be useful. Another major limitation was the granularity of the data. Each row of data was collected at a country wide level using the averages of survey responses. For individuals to know what factors are directly important to them, the responses of each individual would need to be accessed.