LIFE EXPECTANCY BY COUNTRY

Michael Wallace and Brandon Berube

Southern New Hampshire University

Abstract

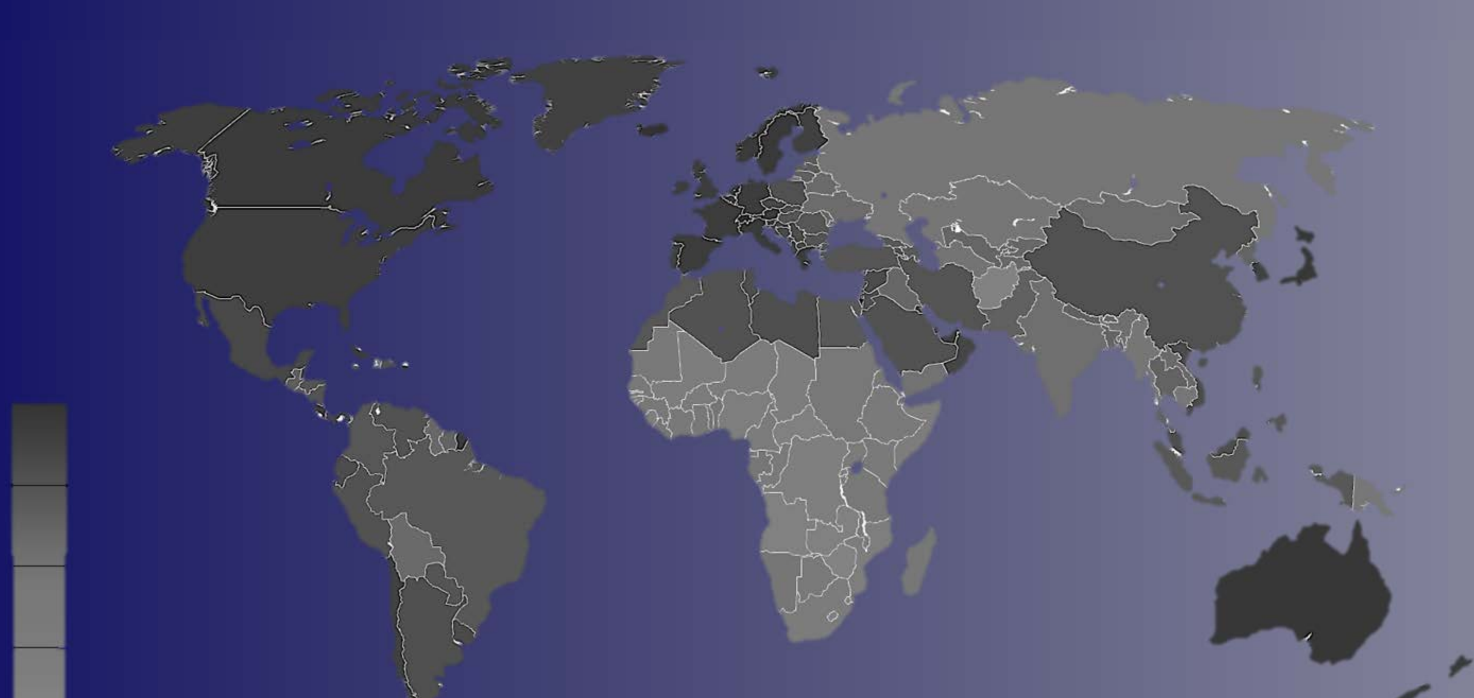
For this project, we questioned if it was possible to use regression analysis to predict the average life expectancy of a country's citizen. The world is concerned about health due to poor air quality, inadequate sanitation, and lack of healthy drinking water. We used regression analysis to analyze these variables to see if they contributed to the overall life expectancy of a nation. We selected effective variables and eliminated ones that were ineffective. Although there are some variables that may seem to be effective, after checking residuals and correlations, we selected only those variables proven to be useful. With the research conducted, we will be able to show the inadequacies of nations that affect life expectancy, and how to increase the average life expectancy of their citizens.

Introduction

We chose a random sample of twenty countries in order to predict an accurate model for life expectancy of any country. It is evident that there are many factors of life expectancy, thus after brainstorming, we identified four main variables:

1) Access to improved sanitation facilities
2) Health expenditure per capita
3) Access to improved water source
4) Food production index

With these variables, we proposed a model to run a hypothesis test, either to reject or fail to reject our null hypothesis. Our model will have utility upon such rejection.



Method

Model 1

H_o: $\beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 = 0$ (Model is not useful)
H_a: $\beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 \neq 0$ (Model has utility)

The regression equation is:
Life Expectancy = 36.2 + 0.233 Sanitation Score + 0.00139 Healthcare + 0.0632 Water Quality + 0.0763 Food

Predictor	Coef	SE Coef	T	P
Constant	36.16	11.79	3.07	0.008
Sanitation Score	0.23250	0.05535	4.20	0.001
Healthcare	0.0013937	0.0006336	2.20	0.044
Water Quality	0.06319	0.08017	0.79	0.443
Food	0.07626	0.07634	1.00	0.334

S = 4.02304 R-Sq = 83.3% R-Sq(adj) = 78.9% P-Value 0.000

Model 2

H_o: $\beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5 X_3X_4 + \beta_6X_5 = 0$ (Model is not useful)
H_a: $\beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5 X_3X_4 + \beta_6X_5 \neq 0$ (Model has utility)

The regression equation is
Life Expectancy = 49.0 + 0.228 Sanitation Score + 0.00131 Healthcare - 0.082 Water Quality - 0.033 Food + 0.00128 Food*Water + 0.47 Air Quality

Predictor	Coef	SE Coef	T	P
Constant	49.00	53.12	0.92	0.373
Sanitation Score	0.22756	0.06163	3.69	0.003
Healthcare	0.0013147	0.0009828	1.34	0.204
Water Quality	-0.0821	0.6107	-0.13	0.895
Food	-0.0325	0.4462	-0.07	0.943
Food*Water	0.001275	0.005276	0.24	0.813
Air Quality	0.470	3.673	0.13	0.900

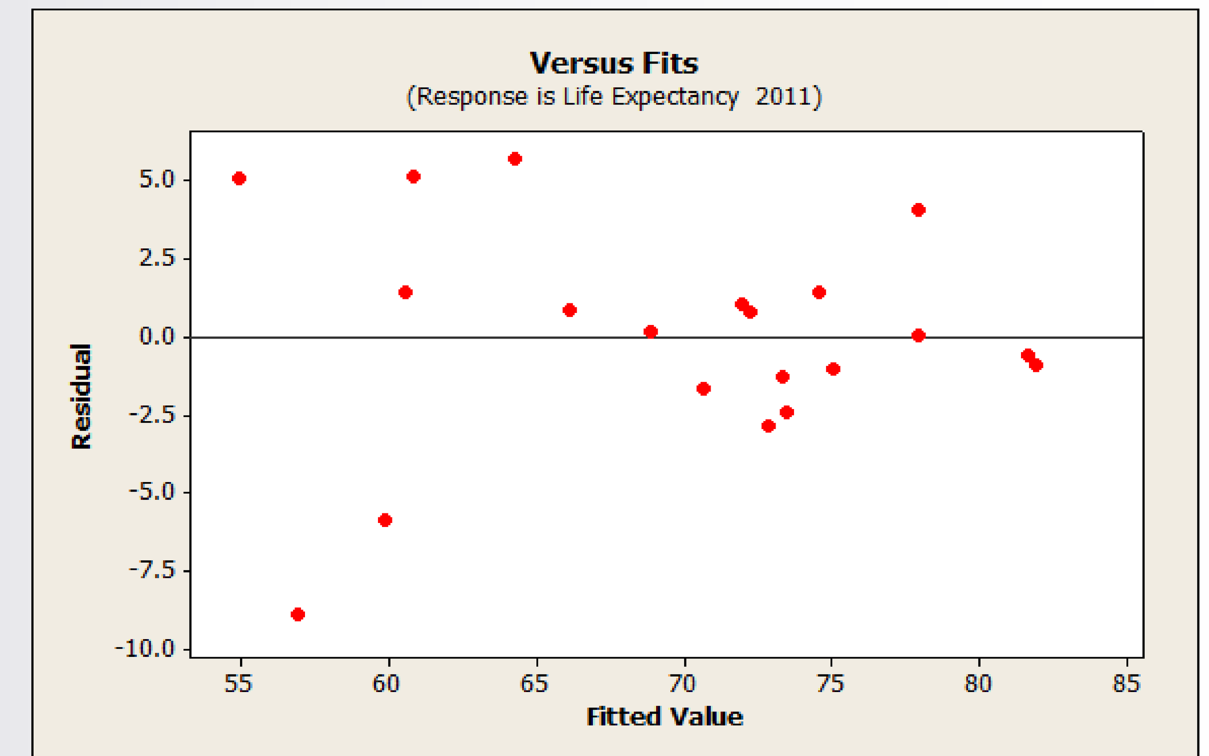
S = 4.30674 R-Sq = 83.4% R-Sq(adj) = 75.8% P-Value 0.000

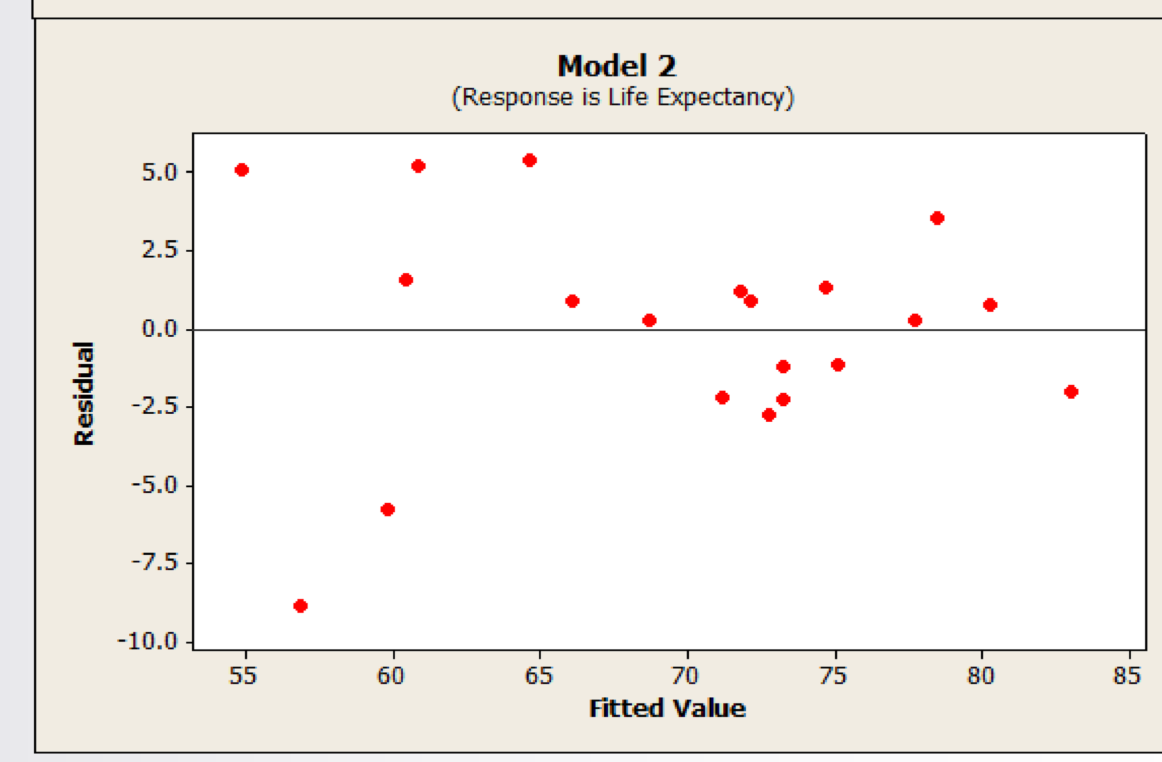
Results

Model 1 is stronger than model 2. Not only is the R^2 (adj) value higher by 3.1 percent, but the two negative coefficients within Model 2 represent multicollinearity.

Important observations of Model 1:
1. Our overall model's p-value is .000
2. Our adjusted R^2a is 78.9%
3. We reject our H_0, which states that our model is useful in predicting life expectancy

Important observations of Model 2:
1. Our overall model's p-value is .000
2. Our adjusted R^2a is 75.8%
3. We reject our H_0, which states that our model is useful in predicting life expectancy





Significance

We successfully created a model that is approximately 80 percent effective for predicting the average life expectancy. This equation can be applied to a variety of areas, for example the weighing of risk management in insurance agencies to government statistics for their census. A country with a lower average life expectancy can use this equation to improve the variables represented in our model, with the potential of increasing their citizen's life span.

References

Mendenhall, W. & Sincich, T. (2012). *A second course in statistics : regression analysis*. Boston: Prentice Hall

The World Bank, (n.d.). *Life expectancy at birth, total (years)*. Retrieved from website: http://data.worldbank.org/indicator/SP.DYN.LE00.IN

The World Bank, (n.d.). *Improved sanitation facilities, urban (% of urban population with access)*. Retrieved from website: http://data.worldbank.org/indicator/SH.STA.ACSN.UR/countries/1W?display=default

The World Bank, (n.d.). *Health expenditure per capita (current us$)*. Retrieved from website: http://data.worldbank.org/indicator/SH.XPD.PCAP

The World Bank, (n.d.). *Improved water source (% of population with access)*. Retrieved from website: http://data.worldbank.org/indicator/SH.H2O.SAFE.ZS

The World Bank, (n.d.). *Food production index (2004-2006 = 100)*. Retrieved from website: http://data.worldbank.org/indicator/AG.PRD.FOOD.XD

The World Bank, (n.d.). *Co2 emissions (kt)*. Retrieved from website: http://data.worldbank.org/indicator/EN.ATM.CO2E.KT/countries