Does Money Buy Happiness?

Abstract

In this paper, we build a multiple regression model to investigate what factors influence national happiness. After finding significant variables such as money, health, inequality, and human freedom, we create an initial model and use a variant of purposeful selection to systematically trim insignificant interactions. The resulting model is not only significant but also predicts national happiness well, with an adjusted R^2 value of .7891. It turns out money has a huge effect on national happiness.

1 Introduction

People say "money can't buy happiness." Of course, this oft-quoted adage contains truth for individuals, since human needs such as community and a sense of purpose are what make people truly happy. But the question naturally arises: what contributes to an entire nation's happiness? In this project, we will explore data on several factors such as money, health, inequality, and human freedom in order to understand what promotes national happiness.

2 Data

This project uses data from reputable private research institutes along with government agencies such as the Central Intelligence Agency. Let's explore each variable:

1. **Happiness Index:** Every year, the World Happiness Report gathers data from the Gallup World Poll. Respondents across the world rate their quality of life on a scale from 0 to 10, with 0 being the worst possible life and 10 being the best possible life. Then, researchers rank each country according to its happiness level. In this project, the happiness index is the single response, and is continuous with possible values from $0 \leq y \leq 10$.
2. **GDP per capita:** GDP per capita is "the per person market value of all final goods and services produced within a country in a given period of time" (Mankiw). Economists regard GDP as the best measure of productivity and wealth of a nation. This continuous predictor variable has possible values from $0 \leq x < \infty$.
3. **Life expectancy:** There are many ways to measure the health of a nation. Perhaps the best way is also the simplest–life expectancy at birth. Average life expectancy varies across countries, but the predictor variable has continuous possible values from $0 \leq x < \infty$.
4. **Human Freedom:** The Cato Institute calculates the Human Freedom Index in order to understand human freedom around the world. Measuring personal, civil, and economic freedom for every country on a scale from 0 to 10, with 10 representing more freedom, results in a continuous predictor variable that has possible values from $0 \leq x \leq 10$.
5. **Inequality:** One can quantify overall inequality in a nation by examining the distribution, or lack of distribution, of wealth in a population. To do so, economists use the Gini coefficient. The following figure illustrates how to calculate the statistic (Bourne):

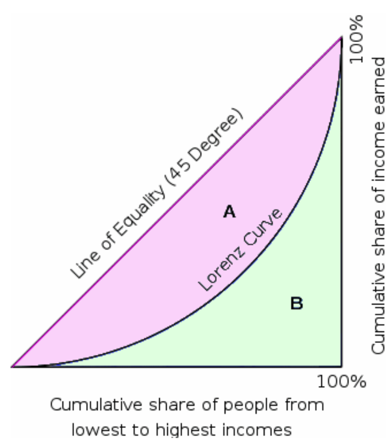


Figure 1: Distribution of wealth in a hypothetical country

Mathematically, the Gini coefficient is $G = {}^{A}\!/_{A+B}$ where *A* and *B* are the areas in Figure 1. As a result, a Gini coefficient of zero implies perfect equality where all income is the same, while a Gini coefficient of one expresses complete inequality. Therefore, this continuous predictor variable has possible values from $0 \leq x \leq 1$.

We considered using other explanatory variables such as taxation levels around the world. However, these interactions seemed to have weaker correlations with national happiness. As a result, the following table shows a few rows of the final data. Consider the differences between Syria, China, and the United States:

	Happiness Index	GDP per Capita	Life Expectancy	Human Freedom Index	Gini Coefficient
United States	6.886	\$59500	80.0 years	8.39	.408
China	5.246	\$16700	75.7 years	6.01	.421
Syria	3.462	\$2900	75.1 years	4.04	.358

The next section builds a model in order to predict happiness around the world.

3 Exploratory Data Analysis

Collected data on the five variables discussed in the previous section will allow us to understand the relationship between per capita GDP, life expectancy, human freedom, inequality, and happiness. However, before building a model, it would be beneficial to examine the individual data sets. Appendix A contains a pairs plot that shows the connection between every variable. Notice most of the data is fairly linear with somewhat strong associations. There is, however, one glaring exception between GDP per capita and the happiness index. Fortunately, although the plot clearly curves, the relationship is intrinsically linear since a log transformation fixes the curved pattern. With this linear data, we can consider how each individual variable affects the predictor variable. In order to get a general understanding of the data, we created simple linear regression models between each individual variable and the happiness index. The following table gives the resulting P-values of the slope coefficient hypothesis test for each predictor variable:

Predictor Variable	P-Value
GDP per capita	$< 2.2 \times 10^{-16}$
Life expectancy	$< 2.2 \times 10^{-16}$
Human freedom	$< 2.2 \times 10^{-16}$
Inequality	.105

Table 1: P-values for each explanatory variable

According to the table, the first three explanatory variables are highly significant when used to predict happiness. Additionally, inequality displays adequate significance at P-value $= 0.105$.

4 Model Building

Since all of our data is significant, we can build a full model. For this dataset, we will start with a full model and systematically remove predictors. The complete model with interactions is:

$$\hat{y} = \alpha + \beta_1 a + \beta_2 b + \beta_3 c + \beta_4 d + \beta_5 ab + \beta_6 ac + \beta_7 ad + \beta_8 bc + \beta_9 bd + \beta_{10} cd \tag{1}$$

α = constant; a = GDP per capita; b = Life expectancy; c = Human freedom; d = Inequality

In order to build a significant model, we will follow a variant of purposeful selection and systematically remove variables with the least significance. Additionally, we will not remove any variables involved in a significant interaction, even if the variable has a high P-value for the slope coefficient hypothesis test. Using this logic results in eliminating the GDP/Human freedom interaction, followed by the GDP/Life expectancy interaction, followed by the GDP/Inequality interaction. As shown in the table below, in this trimmed model, each of the surviving variables is significant ($\alpha \leq .05$) or involved in a significant interaction:

	Estimate	P-Value
Intercept	13.2139	.0070
GDP per capita	1.1296	3.11×10^{-9}
Life expectancy	-.2912	8.92×10^{-6}
Human freedom	-.8202	.2808
Inequality	-.0744	.2287
Life expectancy : Human freedom	.0265	.0006
Life expectancy : Inequality	.0036	.0003
Human freedom : Inequality	-.0229	.0171

Table 2: Estimates and P-Values for Trimmed Model

Before accepting this model, we have to verify assumptions. Primarily, it is safe to say the data is independent since each country is isolated from one another, so individual data points should not affect one another. To address other assumptions, the Q-Q plot and standardized residuals plots in Appendix B demonstrate there are no unacceptable patterns in our data. As a result, the model fulfills assumptions in normality, homoscedasticity, and linearity.

5 Results

According to Table 2, here is the full equation for our model:

$$\hat{y} = 13.21 + 1.13log(a) - .29b - .82c - .07d + .03bc + .004bd - .02cd$$

a = GDP per capita; b = Life expectancy; c = Human freedom; d = Inequality

This equation gives the predicted happiness index for a country based on its GDP per capita, life expectancy, human freedom, and inequality. Additionally, the model has a very good adjusted R^2 value of .7891. At first glance, the model looks nonsensical since there are negative coefficients in front of life expectancy and human freedom, implying a reduction will increase happiness. This clearly contradicts our data. However, the model still makes sense when you take into account all of the interactions. Additionally, another confusing part of the model is its initial intercept of 13.21. Since the happiness index has a maximum value of 10, it is impossible for a country to have a happiness index of 13.21. Consequently, it seems as if we should be careful when applying the model at extreme prediction values. Finally, notice the size of the coefficient in front of GDP per capita. Even though the coefficient represents the increase in happiness from a one unit increase in the logarithm of GDP, it appears wealth has an enormous effect on happiness.

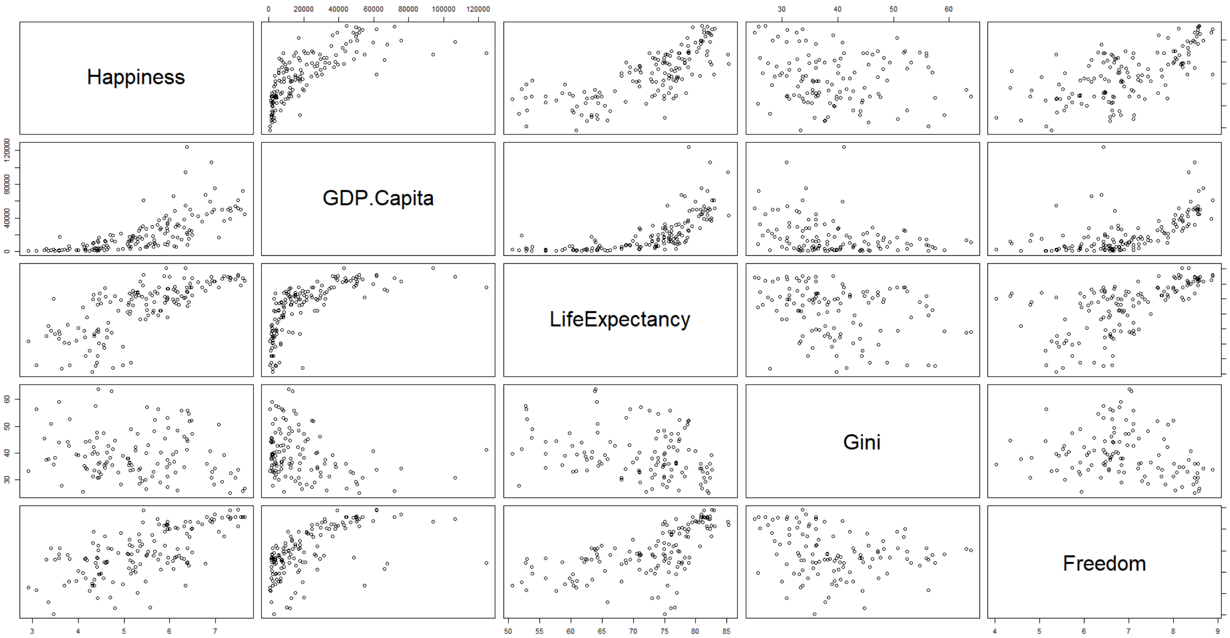
Yet, there is one more fundamental limitation in the model. Throughout this project, we have assumed the happiness index perfectly measures happiness, GDP perfectly measures wealth, and the Gini coefficient perfectly measures inequality. This is a big assumption. In the real world, it is much harder to understand such subjective variables, especially across cultures. As a result, we could improve the model in the future by finding better ways to measure these variables.

6 Conclusion

This project demonstrates the relationship between money, health, inequality, human freedom, and happiness. Through multiple regression techniques, we were able to build a functioning model in order to predict the overall happiness of a nation. It appears as if money might not buy happiness, but it is a good down payment.

Appendix A

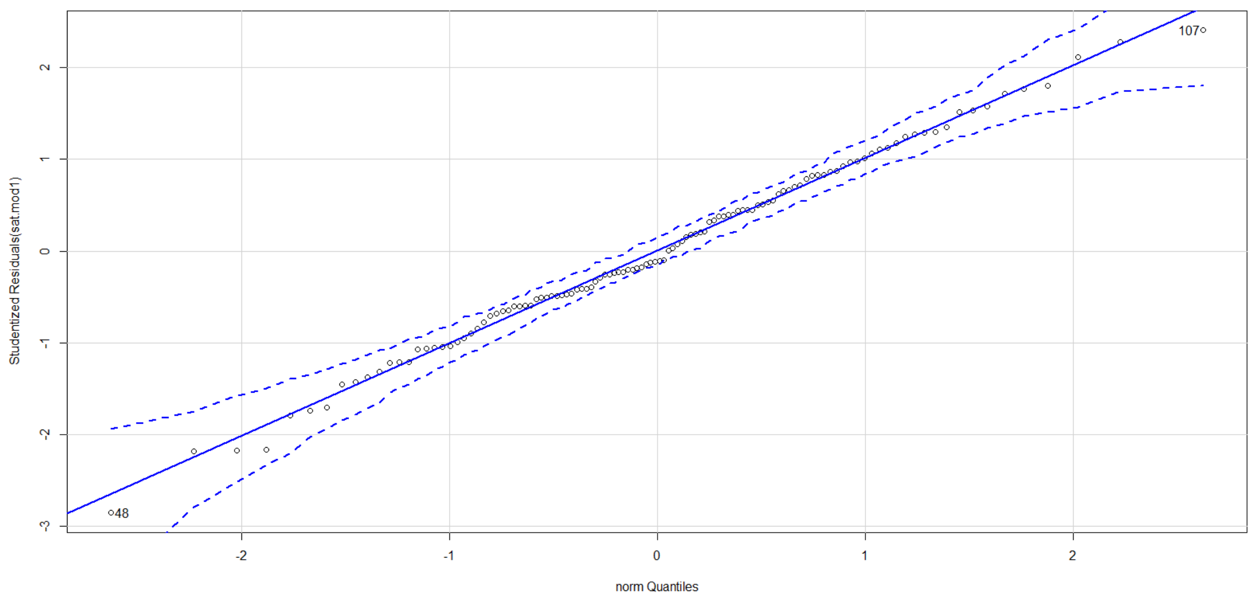
This appendix contains a pairs plot that shows the connection between every variable. Since the happiness index is the predictor variable, focus on the top row to get a general understanding of the data:

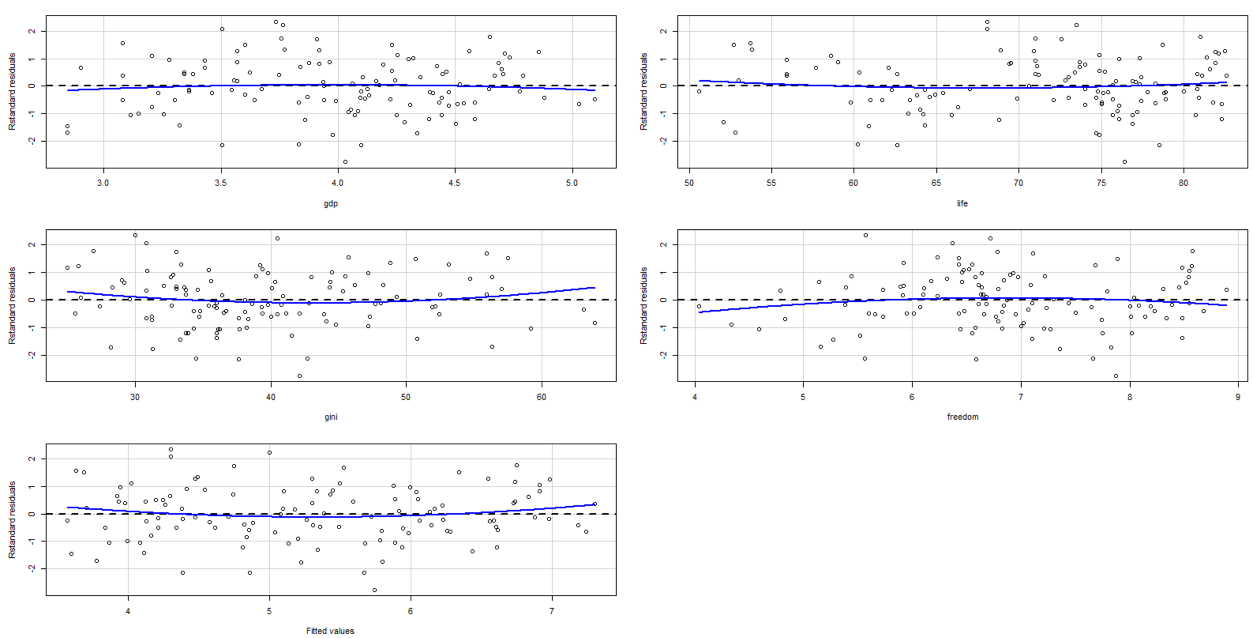


The relationship between GDP per capita and the happiness index is not linear. However, a log transformation on GDP per capita fixes this problem.

Appendix B

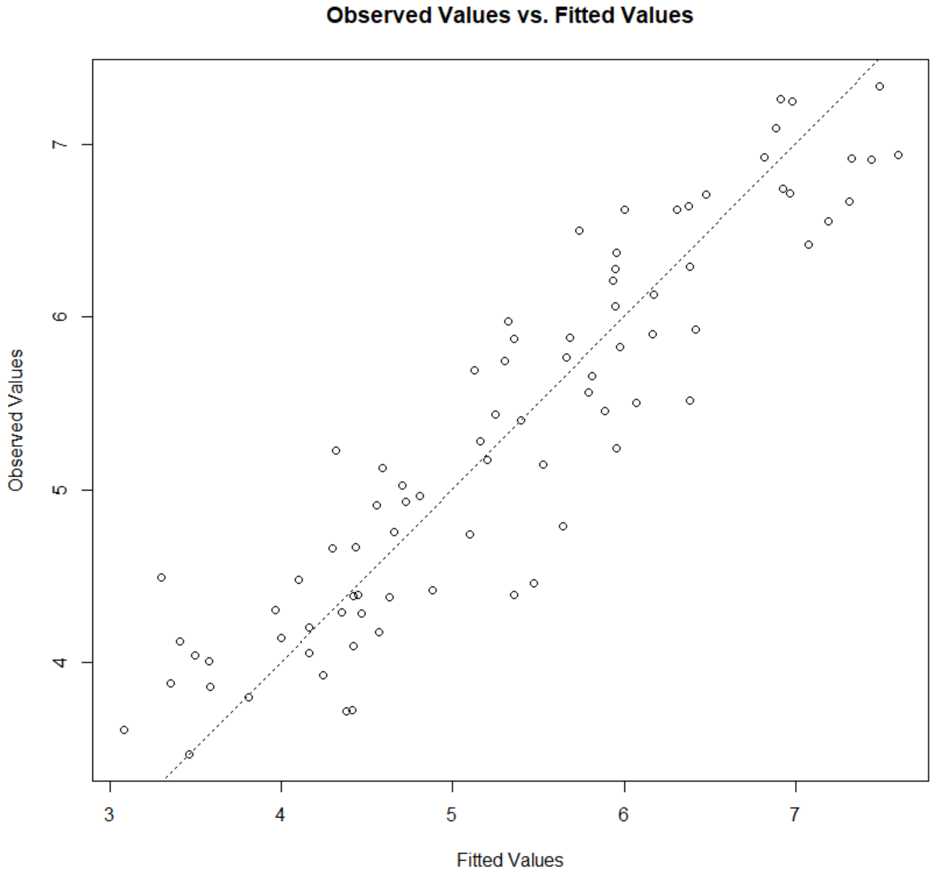
This appendix contains plots that allow us to address assumptions in the model:





As seen in the Q-Q plot and standardized residuals plots, there are no unacceptable patterns in our data. As a result, the model fulfills assumptions in normality, homoscedasticity, and linearity.

Additionally, the following y versus $\hat{y}$ plot demonstrates random scatter for our model:



Works Cited

[1] Bourne, Murray. “The Gini Coefficient of Wealth Distribution.” *INTMath*, SquareCirclez, 24 Feb. 2010, www.intmath.com/blog/mathematics/the-gini-coefficient-of-wealth-distribution-4187.

[2] Helliwell, John F., et al. *World Happiness Report 2018*. Sustainable Development Solutions Network, 2018.

[3] Mankiw, Nicholas. *Principles of Economics*. 8th ed. Singapore: Cengage Learning, 2017.

[4] United States. *Country Comparison: GDP Per Capita (PPP)*. Central Intelligence Agency, 2017, www.cia.gov/library/publications/the-world-factbook/rankorder/2004rank.html.

[5] —. *Country Comparison: Life Expectancy At Birth*. Central Intelligence Agency, 2017, www.cia.gov/library/publications/the-world-factbook/rankorder/2102rank.html.

[6] Vasquez, Ian, and Tanja Porcnik. *The Human Freedom Index 2017: A Global Measurement of Personal, Civil, and Economic Freedom*. Cato Institute, 2017.

[7] World Bank. *World Development Indicators 2013*. World Bank, 2013, http://data.worldbank.org.