An Examination of Factors Shaping Countries’ Happiness Scores

Nina Austria

2024-03-13

## Abstract

A 150-200 word abstract. (One paragraph)

We explore different models for assessing the relationship between countries’ happiness scores and economic wealth and available resources using data on countries from 2018.

### Introduction

Type background and significance section here. Present your research question(s) and provide relevant background info or lit survey.

How does a country’s happiness score vary in relation to its economic wealth and available resources? In this project, I …

QUESTION: Should I talk at all about background on the random forest var imp measures here? Or should I save that for the methodology section and talk more about why my research question is important here (happiness scores of countries and how they relate to economic factors).

Does a smaller student-to-faculty ratio help students succeed? Do lower tuition costs have an effect on a student wanting to stay for the full degree at one school? Does institutional financial assistance play a role in a higher completion rate at a specific college? In this study, we examine a multilevel dataset on completion rates at US colleges. We are given one response variable of completion rate and seven explanatory variables. We will attempt to narrow down the strongest contributing factors to determine what really affects college completion rates.

Prior research on this topic has yielded consistent results across the board. Sheilynda Stewart, Doo Hun Lim and JoHyun Kim (2015) found unsurprisingly that more assistance to students both academically and financially led to a greater amount of persistence for students to continue at their respective college beyond the first year. Their study took place only across public institutions but was a longitudinal study covering a variety of explanatory variables and using persistence beyond the first year as their response variable. Although we did not examine persistence beyond the first year, if a student chooses to stay at an institution beyond one year, they are more likely to complete their degree at said institution for the same reasons. In another cross-institutional study, Cullen F. Goenner and Sean M. Snaith (2004) similarly concluded that student-faculty ratios, total expenditures, and cost play a significant role in completion rates at the colleges they examined.

In this paper, we continue the ongoing investigation of what factors have the strongest influence on completion rates as US colleges. Any institution looking to boost their completion rates could take a look at our results and get a better idea of what to change in order to retain more of their students through completion of their degrees. This is important, for even though college enrollment has been trending upwards, completion rates have remained mostly stagnant. We expect our conclusions to be similar to the above research in that faculty will likely play a role in determining completion rate, as well as tuition and typical amount of an institutional grant among recipients (in $1000s) or percentage of students who receive an institutional grant. Both of these last two variables are related to the institution giving back to the student and that seemed to be a common factor in prior research.

### Data

##   
## Table 2: Summary Statistics

##   
##   
## | | GDP| Happiness\_score| Health| Internet| Military| Unemployment|  
## |:-------|---------:|---------------:|---------:|--------:|---------:|------------:|  
## |Mean | 15980.00| 5.423219| 11.255373| 53.75735| 6.181550| 6.780882|  
## |Std.Dev | 21861.07| 1.143796| 5.860822| 28.72397| 4.924869| 5.217286|  
## |Min | 275.00| 2.905000| 0.000000| 2.00000| 0.000000| 0.100000|  
## |Q1 | 1704.50| 4.447000| 7.030000| 27.40000| 3.030000| 3.400000|  
## |Median | 6120.50| 5.472000| 10.415000| 57.60000| 4.760000| 5.250000|  
## |Q3 | 21626.00| 6.310000| 14.510000| 78.80000| 8.450000| 9.050000|  
## |Max | 114340.00| 7.632000| 39.460000| 98.30000| 31.900000| 27.000000|

The data in this research involved the merging of three datasets. The first dataset is the 2018 World Happiness Report obtained from Kaggle, using data collected from the Gallup World Poll. From this dataset, the variables of interest are the countries or regions and our response variable, the happiness scores in each location. These scores were derived from the Cantril ladder, a measure in happiness research which involves asking individuals to rate their current life satisfaction on a scale from 0 to 10, with 0 representing the worst possible life and 10 representing the best possible life.

The second dataset is the AllCountries dataset sourced from the ‘Lock5Data’ package in R, which has data on the countries of the world collected for 2018 or the most recently available year. The data were gathered online from the World Bank (<http://data.worldbank.org/>). The variables of interest extracted from this dataset are Gross Domestic Product (GDP) per capita in US dollars, percentage of government expenditures directed toward the military, percentage of government expenditures directed toward healthcare, percentage of the population with access to the internet, and percent of the labor force unemployed.

Finally, the third dataset used is the Countries of the World Dataset, also sourced from Kaggle and comprising data from the US government. This dataset links country names to their respective regions across the globe.

Merging the 2018 World Happiness Report with the AllCountries dataset resulted in a dataframe containing data on happiness scores from 137 countries or regions. Merging this \_\_\_\_ Since I am solely focused on the year 2018, in total there are 137 observations and 8 variables. Table 1 presents the first five rows of the data and the main variables important to this research.

### Methods

Our main purpose is to explore how a country’s happiness score varies in relation to its economic wealth and available resources. Additionally, we aim to compare the rankings of feature importance across various variable importance measures.

We started off our model-building process with an Ordinary Least Squares (OLS) regression model. Since we are interested in how happiness scores vary with economic wealth, I included GDP per capita in US dollars as an explanatory variable. I also included percentage of government expenditures directed toward the military, percentage of government expenditures directed toward healthcare, percentage of the population with access to the internet, and percent of the labor force unemployed. (maybe talk about correlation plot here)

Figure 2 is a correlation plot of the numerical variables in our dataset, with darker red

Here’s our OLS model, in which i represents region and j represents country:

Note, however, that in an OLS regression model, we assume that all entries in the data are independent. Given that we are working with country-level data, it is likely that the happiness scores of countries in the same region are likely to be more highly correlated.

Linear mixed effects models contain both fixed effects and random effects, variables that control for the random variability within each level in our model (Roback and Legler).

I will use terms to represent the fixed effects. These are the variables that are constant over the country it represents.

Here’s our final LMER model, in which i represents region and j represents country:

where and .

Simulation: The variable importance, from least to most important, should be: x4, x3, x2, x1/x5. We are testing to see whether random forest variable importance measures overemphasize the importance of correlated features. x5 represents unemployment, and x2 represents GDP. Does random forest variable importance put x2/x3/x4 higher than x5 because it is highly correlated with x1?

### Results

Write your results section here. Include the fixed and random effects tables, as below.

Intercept: The intercept is 1.63, representing the estimated average happiness score when all predictor variables are zero. log(GDP): A one-unit increase in the log(GDP) is associated with an estimated increase of 0.362 in the happiness score. Internet: A one-unit increase in the Internet variable is associated with an estimated increase of 0.0122 in the happiness score. Health: A one-unit increase in the Health variable is associated with an estimated increase of 0.0207 in the happiness score. Military: The Military variable does not seem to have a significant effect on happiness, as the p-value is high (0.9815). Unemployment: A one-unit increase in the Unemployment variable is associated with an estimated decrease of 0.0428 in the happiness score.

### Discussion and Conclusions

Type discussion section.

### References

References in APA format. For example:

Sadler, M. E., & Miller, C. J. (2010). Performance anxiety: A longitudinal study of the roles of personality and experience in musicians. Social Psychological and Personality Science, 1(3), 280-287.

Lin, M. C. (2019). An Investigation Of Music Performance Anxiety In Taiwanese Pianists, Vocalists, String And Wind Instrumentalists At The College Level.

Stoeber, J., & Eismann, U. (2007). Perfectionism in young musicians: Relations with motivation, effort, achievement, and distress. Personality and Individual Differences, 43(8), 2182-2192.

Roback, P., & Legler, J. (2021). Beyond Multiple Linear Regression: Applied Generalized Linear Models And Multilevel Models in R. CRC Press.

# mine are below hehe

Kaushik, S., et al. (2024). World Bank Open Data. Retrieved from data.worldbank.org/.

Lasso, F. (2018). Countries of the World. Kaggle. Retrieved from www.kaggle.com/datasets/fernandol/countries-of-the-world.

Lower infant mortality and access to contraception reduce fertility in low- and middle-income nations. Scientific Figure on ResearchGate. Retrieved from <https://www.researchgate.net/figure/World-map-showing-four-regional-classes-used-as-a-random-effect-in-the-general-linear_fig1_357138541>

Pramanick, S. Happiness-Index-2018. Kaggle. Retrieved from www.kaggle.com/datasets/sougatapramanick/happiness2018

Random Forest. Retrieved from <https://anasbrital98.github.io/blog/2021/Random-Forest/>

Roback, P., & Legler, J. (2021). Beyond Multiple Linear Regression: Applied Generalized Linear Models And Multilevel Models in R. CRC Press.

Fig-1. This is how a decision tree looks. Analytics Vidhya, medium.com. Retrieved from <https://medium.com/analytics-vidhya/how-exactly-decision-trees-are-built-with-complete-example-dbda4a34cf1d>

## Appendix (optional)