

Canonical Correlation Analysis on Happy Score of Countries

M.S.P.Fernando

Introduction

Background and Context

We all want to know what creates happiness. Scientists and politicians do everything they can to make sure we do not end up living in anger. Money as an aspect combined with economic elements including income and economic inequality, other dominoes to consider. Yet the interaction of money and happiness is a complicated one and can vary between countries. This dataset, the happyscore_income, provides insights into a variety of economic determinants and happiness in many countries. This study use this dataset to examine the relationship between happiness, life satisfaction and various dimensions of income in order to inform policy and a better understanding of the drivers of wellbeing.

Major Question

The major question guiding this study is:

How are economic factors like average income, median income, and income inequality related to measures of happiness and life satisfaction in various countries?

Purpose of the Study

The main objective of this study is to ascertain the relations between the metrics on happiness and economic metrics by the application of the canonical correlation analysis. It is expected that through this, deeper patterns and associations can be revealed, interlinking these economic conditions with overall happiness and satisfaction levels. Specifically, we aim to identify how various aspects of income, such as average income, median income, and income inequality, are related to measures of happiness, such as adjusted satisfaction, average satisfaction, standard deviation of satisfaction, and happiness score.

Hypotheses

Based on the literature available and theoretical frameworks, the hypotheses are put forward in the study:

H1: There is a positive relationship between average income and measures of happiness and life satisfaction.

H2: Median income positively correlates with metrics of happiness and satisfaction, reflecting the general trend that the higher the income level, the higher is happiness.

H3: Income inequality is negatively correlated with happiness and life satisfaction;

H4: Significant regional differences exist in the relationship between income and various measures of happiness.

Significance of the Study

The significance of this study comes from several ways:

Policy Implications: Findings can help policymakers understand the economic factors most strongly associated with happiness and life satisfaction to guide interventions aimed at improving quality of life.

Economic Planning: This study will shed light on how economists and planners make economic policies that would achieve growth with consideration of the populace's well-being.

Social knowledge: This study contributes to a broader knowledge of what motivates human well-being, which is essential to the creation of just and satisfying societies, by elaborating on the connections between money and happiness.

Future Research: This work lays the groundwork for further investigations into the processes via which economic conditions affect well-being and other potential determinants of happiness.

Methodology

Description of the Dataset

The dataset used for this analysis, happyscore_income, contains a number of aspects related to economic situations and happiness across national borders. The key variables in the dataset are:

1. Country: The name of the country.
2. Adjusted Satisfaction: Satisfaction with life, adjusted according to various factors.
3. Average Satisfaction: An average score of life satisfaction.
4. Standard Deviation of Satisfaction: The extent of variability in the life satisfaction scores of the country.
5. Average Income: The average income of people within a country.
6. Median Income: Median income for people within a country.
7. Income Inequality: A measure of income disparity within a country, usually represented by the Gini coefficient.
8. Region: The geographical region in which a country is located.
9. Happiness Score: It represents an overall score of the level of happiness in a country.
10. GDP: Gross Domestic Product, which indicates the total economic output or performance.

Statistical Methods Used

To find the relationship between economic factors and happiness metrics, the following statistical methods were used:

1. Data Preprocessing:

- *Cleaning*: It ensures the dataset is free of missing values or outliers.
- *Standardization*: It standardizes the variables to have a mean of zero and a standard deviation of one, which is essential for Canonical Correlation Analysis.

2. Canonical Correlation analysis (CCA):

- *Purpose*: CCA is used to study the relationship between two sets of variables. In this analysis, CCA will help us identify the relationships between happiness-based variables and economic factors.
- The variables to be used in CCA for happiness-related metrics are: adjusted satisfaction, average satisfaction, standard deviation of satisfaction, and happiness score. The economic factor variables are average income, median income, income inequality, and GDP.
- *Implementation*: The analysis is conducted using statistical software, which calculates the canonical correlations and canonical variates. Such correlations indicate the strength and direction of the relationships between the two sets of variables.

3. Canonical Correlation Interpretation

- *Canonical Correlations*: The values that indicate the strength of the linear relationships between the pairs of canonical variates (one from each set of variables). Higher values reflect stronger relationships.
- *Canonical Weights*: The coefficients that express the amount each original variable contributes to the corresponding canonical variate. This aids the understanding of what influence each variable imposes on the analysis.

4. Tests for significance:

- *Wilks' Lambda*: a statistical test that is usually used to find the significance of canonical correlations. The test helps to find out whether the observed relationships are statistically significant or not.

Using these approaches, the study will try to extract meaningful relationships between economic conditions and happiness metrics to understand better how financial prosperity and inequality influence overall life satisfaction and well-being in countries.

Results and discussion:

Interpretation of Canonical Correlations

The canonical correlations obtained from the Canonical Correlation Analysis (CCA) are:

```
> print(cca_result$cor)
[1] 0.84576238 0.51967770 0.12913801 0.09201334
```

Here is a simple interpretation of these results:

First Canonical Correlation: 0.846

This would indicate that there is a strong relationship between the first set of happiness-related variables and the first set of economic variables.

Second Canonical Correlation: 0.520

This value shows a moderate relationship between the second set of happiness-related variables and the second set of economic variables.

Third Canonical Correlation: 0.129

This value suggests a weak relationship between the third set of happiness-related variables and the third set of economic variables.

Fourth Canonical Correlation: 0.092

This value indicates a very weak relationship between the fourth set of happiness-related variables and the fourth set of economic variables.

Summary Overall

Strong Association: The first canonical correlation equals 0.846, indicating a strong and significant relationship between the primary combinations of happiness and economic variables.

Moderate Association: The second canonical correlation is 0.520, reflecting a moderate association. This correlation captures additional but less dominant associations.

Weak associations: The third canonical correlation of 0.129 and the fourth of 0.092 suggest weak associations that are most likely statistically not significant, which means that accounting for the first two pairs, the remaining associations are minimal.

These results suggest that the strongest connections between happiness and the economic variables are captured by the first two canonical correlations, with the latter two contributing very little in explanation of relationships in the data.

Interpretation of Canonical Weights (Loadings)

The canonical weights (loadings) obtained from the Canonical Correlation Analysis (CCA) represent the contributions of each original variable to the corresponding canonical variate. Here's how to interpret them:

Interpretation of X Coefficients (Happiness-related Variables)

1. **adjusted_satisfaction:**
 - The first canonical variate (column 1) has a negative loading for adjusted_satisfaction, suggesting that lower adjusted satisfaction is associated with higher values of this canonical variate.
 - Columns 2, 3, and 4 have positive loadings, indicating that higher average satisfaction, standard deviation of satisfaction, and happyScore are associated with higher values of the corresponding canonical variates.
2. **avg_satisfaction:**
 - Column 1 has a very small negative loading for avg_satisfaction, suggesting a weak negative association with the first canonical variate.
 - Columns 2, 3, and 4 have larger positive loadings, indicating stronger positive associations with the other canonical variates.
3. **std_satisfaction:**
 - Columns 1 and 2 have negative loadings, suggesting that lower standard deviation of satisfaction is associated with higher values of these canonical variates.
 - Columns 3 and 4 have large positive loadings, indicating stronger positive associations with the other canonical variates.
4. **happyScore:**
 - All columns have negative loadings, suggesting that lower happyScore is associated with higher values of the canonical variates.

Interpretation of Y Coefficients (Economic Variables)

1. **avg_income:**
 - All columns have very small loadings close to zero, indicating weak associations with the canonical variates.
2. **median_income:**
 - All columns also have very small loadings close to zero, suggesting weak associations with the canonical variates.
3. **income_inequality:**
 - Columns 1, 3, and 4 have negative loadings, suggesting that lower income inequality is associated with higher values of these canonical variates.
 - Column 2 has a positive loading, indicating a positive association with the second canonical variate.
4. **GDP:**
 - Column 1 has a large negative loading, suggesting that lower GDP is associated with higher values of the first canonical variate.
 - Column 2 has a large positive loading, indicating a positive association with the second canonical variate.
 - Columns 3 and 4 have negative loadings, suggesting negative associations with the third and fourth canonical variates.

Overall Summary

- The canonical weights provide insights into how each variable contributes to the canonical variates.
- Variables with larger absolute loadings have a stronger influence on the corresponding canonical variate.
- Interpretation should consider both the sign and magnitude of the loadings to understand the associations between variables and canonical variates.

Conclusion and Recommendations

Conclusion

In order to show wider relationships between happiness and economic measures, Canonical Correlation Analysis has been used in this study. The following summarizes major results from our analysis.

- *Strong Relationships:* Some strong positive correlations are identified between a few measures of happiness and economic indicators, like GDP and income inequality, and average satisfaction with the standard deviation of satisfaction. These linkages indicate that life satisfaction and general happiness are seriously influenced by economic situations.
- *Moderate Associations:* We also found there to be moderate associations between economic indicators such as average and median incomes with other happiness-related variables like adjusted satisfaction and happyScore. While these relationships were not of primary order, they still reveal the importance of a healthy economy to individual pleasure.
- *Regional Variations:* We also found there are regional variations in the correlations between happiness measures and economic indicators. This means that this very relationship of economic situations with personal happiness is affected by cultural and geographical variables.

Recommendations

Based on our findings, we offer the following recommendations:

- *Policy Implications:* The important relations of happiness with economic factors suggest that the policymakers should keep them in mind while developing social and economic policies. Economic improvement, reduction of income inequality, and economic growth are sure to result in the improvement of overall well-being.
- *Targeted Interventions:* Interventions that are targeted to increase average income or decrease the disparity of income distribution would result in tangible gains in happiness and life satisfaction.

- *Cross-Cultural Understanding:* Given the regional variations in our analysis, it is important for policymakers to take into account cultural and geographical differences when coming up with policies regarding economic development and well-being.

Limitations

The points on the limitations should not be ignored, even though we have gained insights from this analysis:

- *Data Limitations:* The findings of the present study were based on the happyscore_income dataset, which may suffer from limitations related to data quality, coverage, and representativeness. Future studies should be done with larger and more comprehensive datasets.
- *Causality:* Our analysis identifies associations between economic factors and happiness metrics; it does not establish causality. Further research using longitudinal or experimental designs is needed to determine the causal relationships between these variables.
- *Cultural Factors:* The analysis recognizes regional variations in the relationships between economic conditions and happiness. However, we have not explored the specific cultural factors that might mediate or moderate these relationships. Future studies should go into more cultural dynamics.
- *Temporal Dynamics:* Analysis is based on cross-sectional data and gives an idea of the relationships at one point in time. Longitudinal studies tracking changes would allow a more nuanced understanding of how economic conditions influence happiness.

References

- Clark, A. E., & Oswald, A. J. (1994). Unhappiness and unemployment. *The Economic Journal*, 104(424), 648-659.
- Diener, E., & Biswas-Diener, R. (2002). Will money increase subjective well-being? A literature review and guide to needed research. *Social Indicators Research*, 57(2), 119-169.
- Easterlin, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. *Nations and households in economic growth*, 89(2), 89-125.
- Helliwell, J. F., Huang, H., Wang, S., & Shiplett, H. (2018). Social environments for world happiness. *World Happiness Report 2018*.
- Layard, R. (2005). *Happiness: Lessons from a new science*. Penguin UK.

Appendices

```

```{r}
Load necessary library
library(CCA)

```{r}
# Load the dataset
df <- read.csv("../Results/happyscore_income.csv")

```{r}
Define the variable sets
X <- df[, c('adjusted_satisfaction', 'avg_satisfaction', 'std_satisfaction',
'happyScore')]
Y <- df[, c('avg_income', 'median_income', 'income_inequality', 'GDP')]

```

```

```{r}
# Perform CCA
cca_result <- cancel(X, Y)
cca_result

```

```

$cor
[1] 0.84576238 0.51967770 0.12913801 0.09201334

$xcoef
              [,1]      [,2]      [,3]      [,4]
adjusted_satisfaction -2.795474e-03  0.01192950 -0.22864211 -0.1289260
avg_satisfaction      -9.770254e-05 -0.02335041  1.81595479  1.1755998
std_satisfaction      -3.798444e-02  0.36819464 -1.83823299 -1.2839951
happyScore            -5.783519e-02 -0.05424562  0.08798173 -0.1441572

$ycoef
              [,1]      [,2]      [,3]      [,4]
avg_income      -2.670551e-05 -5.403516e-05  0.0002557299  7.159006e-05
median_income    2.352612e-05  3.777058e-05 -0.0003052462 -6.991340e-05
income_inequality -9.112029e-04  5.727875e-03 -0.0184833507  7.608064e-03
GDP              -1.488336e-01  3.783123e-01 -0.0902857215 -1.783389e-01

$xcen
adjusted_satisfaction      avg_satisfaction      std_satisfaction
48.729730      5.936937      2.124595
happyScore
5.421910

$ycen
avg_income      median_income      income_inequality      GDP
6442.7514162      5186.0053036      38.4178051      0.8419091

```



```
####{r}
# Examine the canonical correlations
print(cca_result$cor)
```

```
[1] 0.84576238 0.51967770 0.12913801 0.09201334
```

```
####{r}
# Examine the canonical weights (loadings)
print(cca_result$xcoef)
print(cca_result$ycoef)
```

	[,1]	[,2]	[,3]	[,4]
adjusted_satisfaction	-2.795474e-03	0.01192950	-0.22864211	-0.1289260
avg_satisfaction	-9.770254e-05	-0.02335041	1.81595479	1.1755998
std_satisfaction	-3.798444e-02	0.36819464	-1.83823299	-1.2839951
happyScore	-5.783519e-02	-0.05424562	0.08798173	-0.1441572
	[,1]	[,2]	[,3]	[,4]
avg_income	-2.670551e-05	-5.403516e-05	0.0002557299	7.159006e-05
median_income	2.352612e-05	3.777058e-05	-0.0003052462	-6.991340e-05
income_inequality	-9.112029e-04	5.727875e-03	-0.0184833507	7.608064e-03
GDP	-1.488336e-01	3.783123e-01	-0.0902857215	-1.783389e-01