The Surprising Factor about GDP

Jiaqin Li*

Suzhou Ulink, Suzhou, 215215, China

* Corresponding author. Email: 1512196946@qq.com

Abstract. This paper involves a search about GDP (gross domestic product) factors, such as social support and life ladder. Some studies about how to analyze are also included in this article. Four different methods have been implemented in this paper: factors analysis, Linear regression, Multiple regression, and clustering. The figures reveal that these factors are helpful in GDP but to various extent. Gross domestic product (GDP) significantly influences each country in many fields. Each country they have different GDP. This paper needs to analyze the relationship within it. Factor Analyzing is the most significant factor, so it does the most research on it.

Keywords： GDP, health life, generosity, positive and negative effects

1   Introduction

Gross domestic product (GDP) is the total monetary or market value of all the goods and services produced in a specific period. It measures overall domestic production and functions as an all-around performance in a given country. GDP is the essential economic statistic in the macroeconomy because it is considered the most critical indicator of the development of the national economy. It is increasing people's concern about economics. It is not only because GDP ranking for countries and this rank can show the countries' strength and power, but it is also an indicator of the economic growth in recent times and reflects whether or not a problem is involved in the country gradually. The consequence of the GDP can be influenced by many variables: net export, migration, and so on.

In this paper, a search is wroten about whether or not some other factors can change the result. A novel range of factors is listed in the paper to analyze: Life Ladder, Social support, healthy life, freedom to have choices, generosity, positive effects, and adverse effects. All standard paper components have been specified for four ways of analysis.
(1) factor analyzing,
(2) Linear regression
(3) Multiple regression,
(4) clustering.
There are 1949 countries involved in this paper. The figures reveal that all of these factors are useful to GDP, but in different extent.

2   Factor analyzing

Table 1. Communalities

	Initial	Extraction
Life Ladder	1.000	0.974
Log GDP per capita	1.000	0.637
Social Support	1.000	0.956
Healthy life expectancy	1.000	0.942
Freedom to make life choices	1.000	0.963
Generosity	1.000	0.817
Perception of corruption	1.000	0.747
Positive affect	1.000	0.925
Negative affect	1.000	0.833

2.1   Extraction Method: Principal Component Analysis

As the figures shown in the Table 1, it is clear to see that each factors have varying extraction which are in different extent, even these data have same initial value. The highest value is about life ladder, 0.974 while the least one is Log GDP per capita. Extraction is defined as how many independent variables the factors we extracted can explain, the closer the value is, the better (maximum value is 1.00).

Then, by illustrating the first diagram which is about communalities, the work knows that the extraction changes by different extent, all of the initial is 1.000. the highest extraction is Life Ladder. It means that Life Ladder is the most principles value among these figures. The work describes the history and features of the Ladder of Life Detection, a tool intended to guide the design of investigations to detect microbial life within the practical constraints of robotic space missions. The Ladder of Life Detection is not intended to endorse specific biosignatures or instruments for life-detection measurements, and is by no means a definitive, final product. It is intended as a starting point to stimulate discussion, debate, and further research on the characteristics of life, what constitutes a biosignature, and the means to measure them.[1] The least figure is 0.747 in perceptions of corruption, demonstrating that it is less likely to build a relationship to the GDP.

Table 2. KMO and Bartlett's Test

Kaiser Meyer Olkin Measure of Sampling Adequacy		0.892
	Approx. Chi-square	29370.451

Bartlett's Test of Sphericity	df	36
	Sig.	0.000

Factor analysis is the most essential method in searching the relationship in this paper. 'Applied Factor Analysis was written to help others apply factor analysis throughout the sciences with the conviction that factor analysis is a calculus of the social sciences. The book developed from research undertaken to do a 236-variable cross-national analysis.' [2] And factor analysis is one category in the dimensional reduction. By using that, we can analyse the data, especially big database easily.

Table 2 shows that some factors are involved in the process (Life Ladder, Social support, healthy life, freedom to have choices, generosity, positive effects and negative effects). It also integrates variables with complex relationships into several factors to reproduce the relationship between the original variables and factors. At the same time, variables can be classified according to different factors.

First of all, the paper classifies the data to two groups according to the median (9.24850), to these values that under the median, marked to 0 and above it to be 1. And then as shown in table 2 can be produced in the proceed. The chi-square is surprisingly large, 29370.451, whereas the sig. is only 0.000. So, it demonstrates that the data is suitable for the analysis. Also, totally, the variance refers to the proportion of selected factors that can be explained by all item variables. And the cumulative rate of the first component is 73.843%, and the second component is 86.602%. The other cumulative rate is all larger than 90.000%. The fugures above is calculated by the program which is called SPSS. After searching some certain figures online, I categorized into two sections with the median value, also with the SPSS, it is easier for me to find the relationship between the populations and the factors. There are many functions in that, like factor analying and clustering. Eventually, I achieved the figures mentioned above.

Also, it is necessary to use KMO and the Bartlett sphericity for the factor analysis. The value of KMO is 0.892, means that the fitness of the factors. And because that the data in the Bartlett is less than 0.05, so that the data is appropriate to the relationship.

3   Linear regression

Linear regression is also an important research technique. According to some researchers, healthy exactly effects GDP. World Health Organization (WHO) had a new report "World Health Statistics 2018" released in Geneva about Healthy life expectancy at birth. "Per ca-pita life expectancy" is the average number of years of surviving after a certain age mortality level of X years. It is an indicator to measure the health level of residents of a country, ethnic group and region. It is calculated on the basis of the current age-specific mortality rate, but in reality, the mortality rate is constantly changing and, therefore, the average life expectancy is a putative indicator. Alison Clements-Hunt (World Health Organization spokesman) said: " A characteristic of healthy life expectancy is that in China, Japan, South Korea and some other, high-income Asian countries, people lose less health than in higher-income Western countries." Some developed countries are shown higher figures in healthy part. Because the individuals live

in the developed countries may have a higher quality of living standard and the countries can provide a much advanced equipment and experienced doctors to the patients. Then the healthy level may rise in a great extent.

According to the saying, there is a relationship between health and the economic growth.' Chronic ill health and the prospect of premature mortality haunt billions of people around the developing world. Few goals can be more worthy than that of ameliorating these burdens. Today, there exists a widespread belief that significant improvements in health are within reach of affordable policy interventions, whether led by governments or by large-scale philanthropies. There also exists a widespread consensus that improving the health of people in poor countries will lead to significant economic gains. The prospect of such economic benefits is often cited as an important secondary justification for health initiatives. For example, the report of the Commission on Macroeconomics and Health (2001), chaired by Jeffrey Sachs, finds evidence that health is one of the most important determinants of a country's economic success. Similarly, the Abuja Declaration of 2000, signed by 53 African heads of state, notes that "malaria has slowed economic growth in African countries by 1.3% per year as a result of which GDP for African countries is now 37% lower than it would have been in the absence of malaria." 1 In this paper we assess the claim that improvements in health lead to increases in economic growth. Our findings are not supportive of the popularly held view. The work finds that large improvements in health lead, in the long run, to modest increases in GDP per capita. Further, these increases in GDP per capita take several decades to arrive. Controlling specific diseases that have a high burden in developing countries would also produce small effects. For example, the effect of eradicating malaria in a typical sub-Saharan country would be to raise GDP per capita by only about 2% in the long run. Our evidence thus suggests that proponents of efforts to improve health in developing countries should rely on humanitarian rather than economic arguments'.[3]

And for the other factor--generosity. Morality is an evolved aspect of human nature, yet is heavily influenced by cultural environment. This developmental study adopted an integrative approach by combining measures of socioeconomic status (SES), executive function, affective sharing, empathic concern, theory of mind, and moral judgment in predicting sharing behavior in children (N = 999) from the age of 5 to 12 in five large-scale societies: Canada, China, Turkey, South Africa, and the USA. Results demonstrate that age, gender, SES, culture, and social cognitive mechanisms explain over 20% of the variance worldwide in children's resource allocation. These findings are discussed in reference to standard cultural comparisons (individualist/collectivist), as well as the degree of market integration, and highlight continuities and discontinuities in children's generosity across urban contexts.[4]

Firstly, the Linear Regression about its healthy in a specific year for these countries. It is noticeable to see that the confidence interval is 0.611 for the factor of year and -42.128 for the cause of healthy. The figures show that both of the year and the healthy has a impact on the economic growth. And the data recording GDP has changed by me to 2 categories which are 0 and 1. those figures under 9.24850 record as 0, and others are recorded to (9.24850 is the median). There are 232 cases in this linear regression

about GDP. And for the expected value is 62.74 in average, max for 70.67, min for 48.63. But for the standard one is -3.46 in minimal and up to 1.95.

Table 3. ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	4253.671	8	531.709	1634.448	0.000^b
Residual	543.275	1670	0.325		
Total	4796.946	1678			

A. Dependent Variable: Log GDP per capita.
B. Predictors: (Constant), negative effects, positive effects.
Perceptions of corruptions, Generosity, Healthy life expectancy at birth, social support, Life Ladder, Freedom to make life choices.

Table 4. Coefficients

Model	Unstandardized B	Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
(constant)	1.730	0.193		8.966	.000
Life Ladder	-0.002	0.000	-0.669	-9.359	.000
Social Support	0.823	0.060	1.386	13.703	.000
Healthy life expectancy at birth	0.124	0.002	1.592	55.716	.000
Freedom to make life choices	0.011	0.008	0.128	1.251	.211
Generosity	-0.205	0.095	-0.036	-2.155	.031
Perceptions of corruptions	-0.335	0.087	-0.060	-3.871	.000
Positive effects	-0.699	0.141	-0.048	-4.948	.000
Negative effects	-0.691	0.178	-0.071	-3.876	.000

According to the figures in table 3, it is clear to see that for all the predictors, the figures are different. In standardized, -0.002 is for life ladder, 0.823 is for social support, 0.124 is for health, 0.011 is for freedom to make choices, -0.205 is for generosity, -0.335 is for corruption, -0.699 and -0.691 is for positive and negative effects respectively. Also. It reveals that the sig. Value is all 0, except freedom to make choices and generosity.

From the data in table 4, there are two tests in it which is t and sig.(p) tests. Most of the data is 0.00 except Freedom to make life choices (0.211) and Generosity (0.031) in the table 4 If the P-value is $0.01<P<0.05$, then representing the presented a significant difference, if P-value<0.01, then the difference is enormous.

4   K-mean Clustering

Clustering is defined as a small group of people or things that are closely together. 'Clustering is a common technique for statistical data analysis, which is used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics. Clustering is the process of grouping similar objects into different groups, or more precisely, the partitioning of a data set into subsets, so that the data in each subset according to some defined distance measure. This paper covers about clustering algorithms, benefits and its applications. Paper concludes by discussing some limitations.' [5] This is an overview introduction of the clustering.

Use of traditional k-mean type algorithm is limited to numeric data. This paper presents a clustering algorithm based on k-mean paradigm that works well for data with mixed numeric and categorical features. It proposes new cost function and distance measure based on co-occurrence of values. The measures also take into account the significance of an attribute towards the clustering process. It presents a modified description of cluster center to overcome the numeric data only limitation of k-mean algorithm and provide a better characterization of clusters. The performance of this algorithm has been studied on real world data sets. Comparisons with other clustering algorithms illustrate the effectiveness of this approach [6-8].

It is necessary for us to cluster the data, especially for the big and complex database. Also, it fits this search as there are 1949 samples involved in the paper. So, K-mean clustering is more suitable. There are 2 number of clusters. According to the final cluster diagram, the most largest figure is 2013.242 in number 2 clustering in Life Ladder. And for the initial cluster diagram, the figure is 2020.000.

5   Conclusion

After analyzing and evaluating the four methods and the factors involved, it can conclude that all of the methods are very suitable for the data. However, more high-tech methods can be used in the paper. Nevertheless, these four are very representative, especially for the factor analysis. After more than a month of effort, a research GDP paper was finally completed. In the whole process, the paper encountered many problems, but with the help of teachers and students solved, the process; the paper is a process of learning, from the initial structure of the paper to the overall problem (GDP factors have a clear cognition).

References

1. Marc Neveu, Lindsay E Hays, Mary A Voytek, Michael H New, Mitchell D Schulte, Astrobiology 18 (11), 1375-1402, 2018
2. Northwestern University Press, 1988, Rudolf J Rummel.
3. Barani., R. , Deepika., S. , Jenitta, M. J. , Madhuladevi., S. , & Surekhaa., V. . SECURED SMART VOTING SYSTEM USING IOT.
4. Jason M Cowell, Kang Lee, Susan Malcolm-Smith, Bilge Selcuk, Xinyue Zhou, Jean Decety, Developmental science 20 (4), e12403, 2017
5. Quamrul H Ashraf, Ashley Lester, David N Weil, from NBER macroeconomics annual 23 (1), 157-204, 2008
6. Ahmad A, Dey L. A k-mean clustering algorithm for mixed numeric and categorical data[J]. Data & Knowledge Engineering, 2007, 63(2):503-527.
7. A M A, B L D. Biological relation extraction and query answering from MEDLINE abstracts using ontology-based text mining - ScienceDirect[C]// Elsevier B.V. Elsevier B.V. 2007:228-262.
8. Abulaish, Muhammad, and L. Dey. "An Ontology-Based Pattern Mining System for Extracting Information from Biological Texts." Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing pt.2; Lecture Notes in Artificial Intelligence; 3641 Department of Mathematics, Jamia Millia Islamia (A Central University), New Delhi-110025, India, 2005.