Cause of Cost of Living Index

Claire, Floyd, Caleb

March 2024

1 Introduction

With a group that carries international experiences, we were curious to understand the price across national factors. We wanted to bring to light how the cost of living across nations differs greatly affecting the price of specific goods. By utilizing global measures of GDP, GNI, and PPP, we found it interesting how the cost of living reflected the different economic measures listed above.

1.1 Guiding question

What's the cause of the Cost of living index under the consideration of macroeconomic data and price of different products?

1.2 Method

For our project, we will use macroeconomic data and a series of price data as explanatory variables to analyze the reasons for the different Cost of living index in different countries. For macroeconomic data, we will use literature research and big data analysis methods to explore the relationships. For price data, we will use data analysis, and we will use typical countries as examples for comparative analysis. We will pick two products that are universal in the way where cultural factors will not be a confounding variable in the situation. For example, rice is a staple grain that is consumed globally in copious amounts. By including rice as a necessary product, we will analyze how the price of rice and consumption of it reflects the cost of living in different countries. We will use Rstudio as the tool for data analysis. We will also use tools such as Directed acyclic graphs to assist research.

1.3 Related Works

Steve Dowrick and John Quiggin's journal International Comparisons of Living Standards and Tastes [1]: A Revealed- Preference Analysis asks similar questions to what we are curious to test in our group project. Dowrick and Quiggin compared the prices and quantities of goods and services consumed across 60 nations by using measures of GDP. They utilized data from the United Nations International Comparison Project and used the revealed-preference approach which compares the welfare of countries based on their consumption choices: "If A could have afforded B's consumption bundle (at A's prices) while B could not afford A's bundle (at B's prices) then we say that A is better off than B, basing this judgment on an assumption of common tastes and optimizing behavior."

In the visual below, the researchers a figure 1 which describes how different countries rank in terms of their GDP per capita. The rankings are based on a revealed-preference analysis, which compares the consumption patterns of different countries to determine which countries have higher levels of well-being.

A "+" sign in row i, column j indicates that i dominates j by revealed-preference criteria: a "$-$" indicates the reverse; NC indicates non comparability by these criteria; "$'!!'$" indicates that the assumption of common tastes is violated. The use of $>$ in column i, row j indicates that consumption of all categories of goods is higher in country i than in country j, and vice versa for $<$.

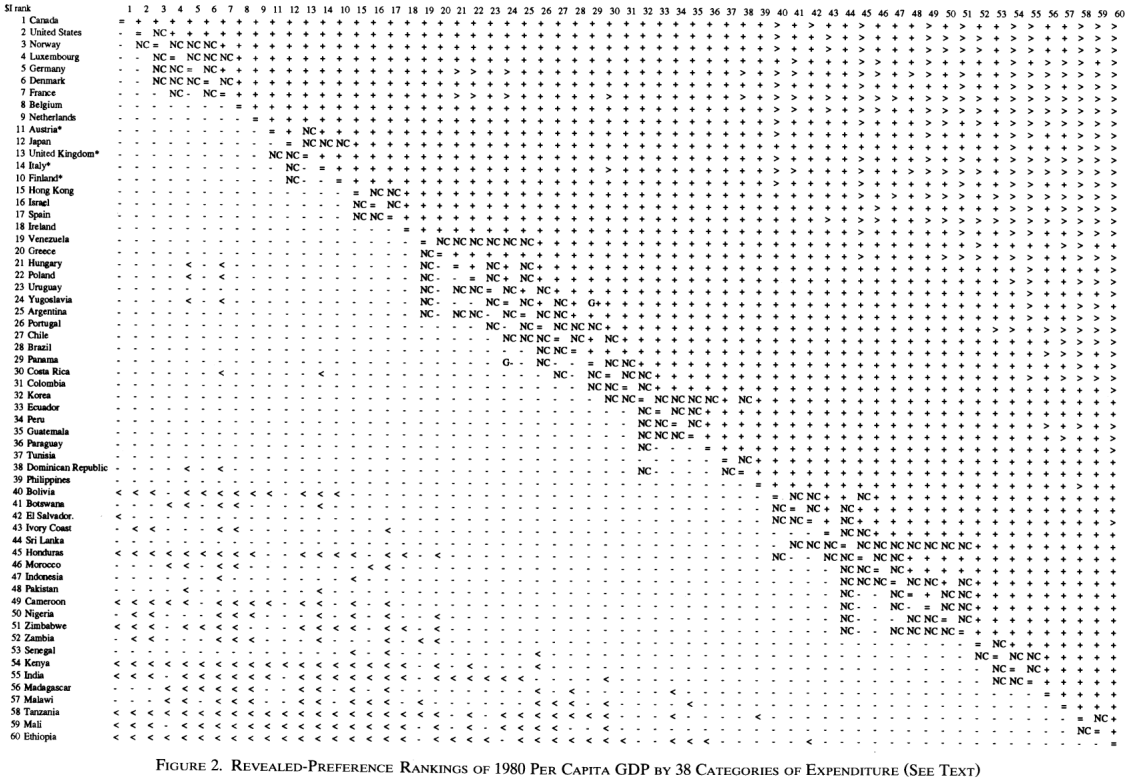


FIGURE 2. REVEALED-PREFERENCE RANKINGS OF 1980 PER CAPITA GDP BY 38 CATEGORIES OF EXPENDITURE (SEE TEXT)

Figure 1: How different countries rank in terms of their GDP per capita

The researchers found that constant price rankings are often not confirmed by revealed-preference criteria, especially when countries are within a similar level of development. However, rankings of groups of countries corresponding to "development groups" are generally confirmed by revealed-preference criteria. This was confirmed under the Generalized Axiom of Revealed Preference which indicates consumer preference by assuming consumer behavior. It suggests that differences in consumption patterns between countries may be due to adjustments to differences in relative prices rather than different tastes.

Dowrick and Quiggin found no evidence to reject the hypothesis of common tastes between countries. The researchers conclude that constant-price measures of "real GDP" should be used with caution for comparisons of countries at a similar level of development. They also question the common practice of citing small movements up or down international league tables as an indicator of policy success or failure. Dowrick and Quiggin's approach used consumption bundles as measures of economic comparison. It taught us what variables we should use with caution such as GDP.

Another related work we found is Currencies, commodities and consumption : measurement and the world economy by Kenneth Clements. It is a book that helps our research and our understanding of what variables work best to compare different countries based on their cost of living. Clements book forms around the fact that currency values, prices, consumption and incomes are intimately linked and are at the heart of the economic performance of all countries. We learned a lot from his method as he utilizes the Big Mac Index (BMI), popularized by The Economist magazine, which uses hamburger prices as the standard of measurement. His book establishes that the BMI contains substantial information regarding economic fundamentals, making it a viable alternative for cross-country income comparisons. Clement also introduces Engel's law, which claims that the proportion of consumption devoted to food (the food share) declines with income. Food shares of a reasonable quality are published regularly by most countries (even the poorest), and as these are pure numbers that are independent of currency values and price levels, they are readily comparable across countries. His book demonstrates that food is a viable approach for making reliable international income comparisons. With confidence from Clement's studies, our project will test various variables that range from rice, bread, cigarettes, and more. By expanding on Clement's research, we will be able to analyze international consumption patterns and the price distributions that vary across countries.

2 Data processing and presentation

According to the guiding question we asked, we collected three sets of data and combined them into one set of total data. Our main data is the cost of living index of 139 countries provided on Kaggle, with only two columns: country and index. Our second set of data is a set of price data, including 95 countries and the prices and rankings of rice, water, bread, cigarettes, and other products in each country. The third data set we collected comes from the World Bank. They include more than 200 countries, along with corresponding GDP, GNI, GDP per capita, GNI per capita and PPP. We combined these three sets of data to obtain a merged dataset of 82 countries that contains all information from the three original datasets. The processing data we will use in our project contains 82 objects and 32 variables. When presenting the data, we sort the data alphabetically by country.

2.1 Price data normalization

To study the impact of the cost of living index, we searched for a set of price data. This set of data is updated in real time, and the download time is March 2024. This set of data includes 8 products from 95 countries. The measurement units of initial data are shown in the table 1 below, and the price unit is US dollars.

Since different products have different prices and measurement units, there is no suitable standard to compare these product data. In order to explore the characteristics of these price data themselves, we normalize these data. We use Rstudio to find the range for each set of price data so that the range of all data is unified to 1. In this way, we can better explore the shape of the data. Here is the boxplot of the processed price data, in figure 2.

Product	Unit
Rice	$1kg$
Imported Beer	0.44 liter
Water	1.5 litter
Cigerattes	20 packs
Cheese	$1kg$
Egg	A dozen
Milk	1.5 liter
White bread	$500g$

Table 1: Table of price and unit in the raw data of the products.

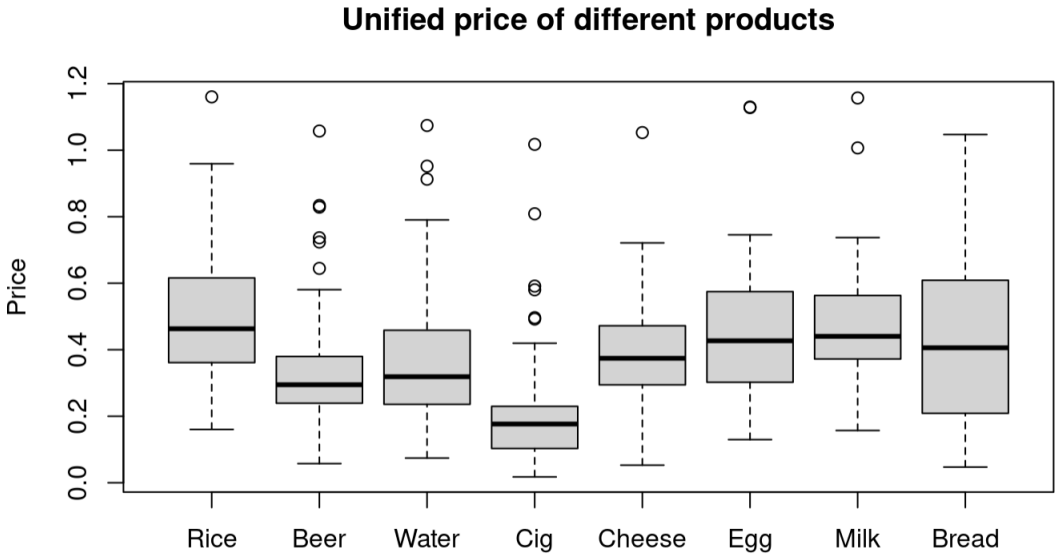
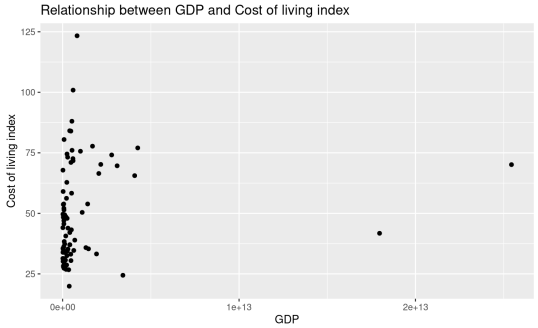


Figure 2: Box-plot of normalized price data of the eight products.

2.2 Macroeconomic data process

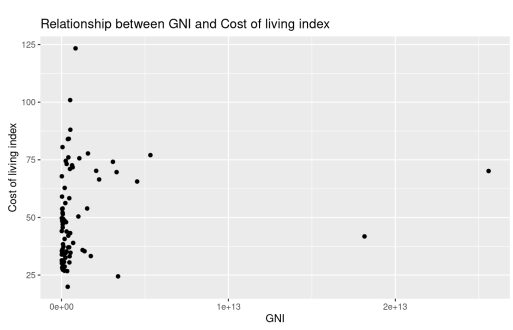
For the world macroeconomic data downloaded from the World Bank, GDP, GDP per capita, GNI, and GNI per capita are four relatively similar sets of data. We will process and screen them to obtain data suitable for research and to formulate hypotheses. For PPP and Import proportion, we will make assumptions after preliminary processing of them.

We use Rstudio to draw scatter plots of GDP, GDP per capita, GNI, and GNI per capita and cost of living index respectively. In the scatter plot, the scatter plots of GDP (GDP per capita) and GNI (GNI per capita) are almost the same. This is because, in economics, the values of GDP and GNI should be very similar. For our study, the difference between these two figures is within a negligible range. Therefore, we will focus on the relationship between the two sets of data, GDP and GDP per capita, and the cost of living index. From the figure 3, we can see that the linear relationship between GDP per capita and cost of living index is more obvious, while the images of GDP and cost of living index are more irregular and have obvious outliers. These outliers should represent several populous countries in the world.

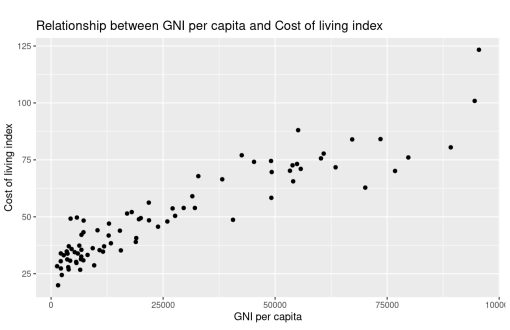


(a) GDP and Cost of living index

(b) GDP per capita and Cost of living index



(c) GNI and Cost of living index



(d) GNI per capita and Cost of living index

Figure 3: Macroeconomic data and Cost of living index

3 Analysis

3.1 Directed acyclic graphs

First, we will make a directed acyclic graphs to sort out the relationship between the three sets of data we studied. This will make it easier for us to design the model. In economics, changes in price data lead to changes in macroeconomic data. Changes in macroeconomic data lead to changes in the cost of living index. Therefore, the directed acyclic graphs of our data are as shown in the figure 4.

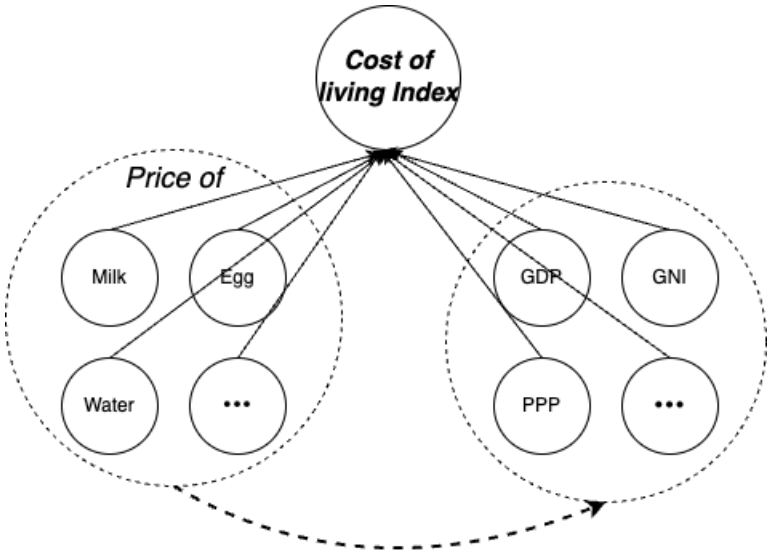


Figure 4: Directed acyclic graphs of our data sets.

3.2 Research on the macroeconomic data sets

Cost of living approximates the total cost that a person needs to spend in one year to live, in most cases this includes an estimate of cost of food, heat, a/c, electricity, rent, taxes, healthcare, and transportation. GDP per capita on the other hand is a measure of the value of the products produced by one country per person in that country. A rise in GDP per capita will most of the time reflect a rise in the cost of living since a rise in GDP per capita often accompanies a rise in gross income which also corresponds to a rise in prices. But this is not always the case, in fact, gross national income per capita correlates better with cost of living than GDP per capita. This makes sense because the labor cost of any product are often the biggest component of overall cost, so when GNI per capita increases, the cost of goods and services should rise as well, and a rise in cost of goods and services means a rise in cost of living as well.

Based on the preliminary data processing and literature research mentioned above, we will only conduct modeling research on GNI per capita data and cost of living index. From the figure 5, there is indeed a linear relationship between the two sets of data. We used Rstudio to draw the image of the linear regression between them, and calculated the R^2 value of this linear regression to detect the strength of the linear connection between them. The R^2 of the linear model between these two data sets is 0.8578091, which shows there is a strong linear relationship in between.

$$R^2 = \frac{\sum(r_0)^2 - \sum(r_{GNI})^2}{\sum(r_0)^2} \tag{1}$$

3.2.1 The situation of countries with different economic levels

Based on our observations of the model, we found that the linear relationship between GNI per capita and cost of living index in low-GNI countries is different from that in high-GNI countries. Judging from the image, the data in the left half of the image, that is, the data for low-GNI countries, has a higher slope. We distinguish between low-GNI and high-GNI by taking the median of the data as a dividing line. The graph is shown in figure 6 below.

Judging from the linear model returned by Rstudio, for the overall GNI data $x = x$, the linear regression has a slope of 7.293×10^{-04}. For low GNI countries $x < median$, the linear regression has a slope 9.764×10^{-04}. For high GNI countries $x < median$, the linear regression has a slope 6.61×10^{-04}.

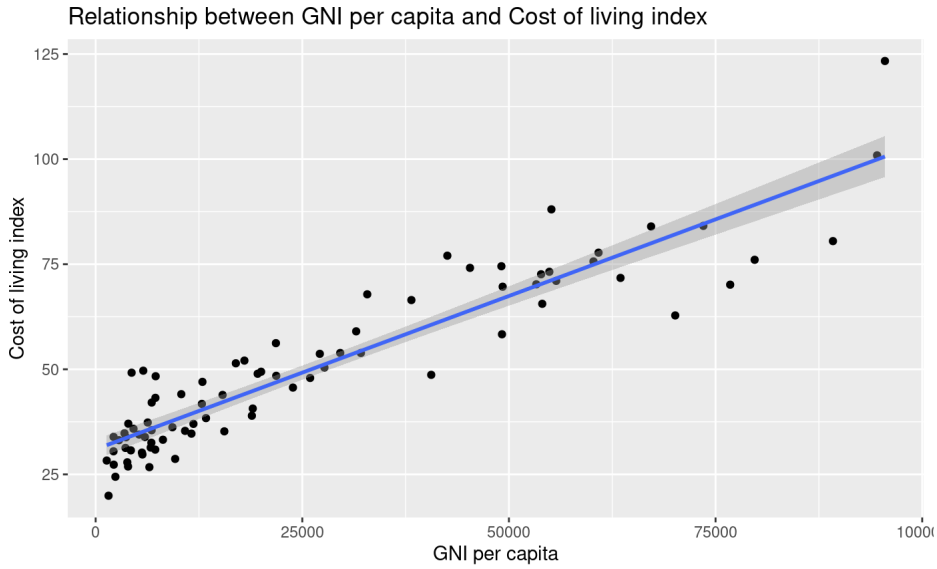


Figure 5: Linear model between Cost of living index and GNI per capita.

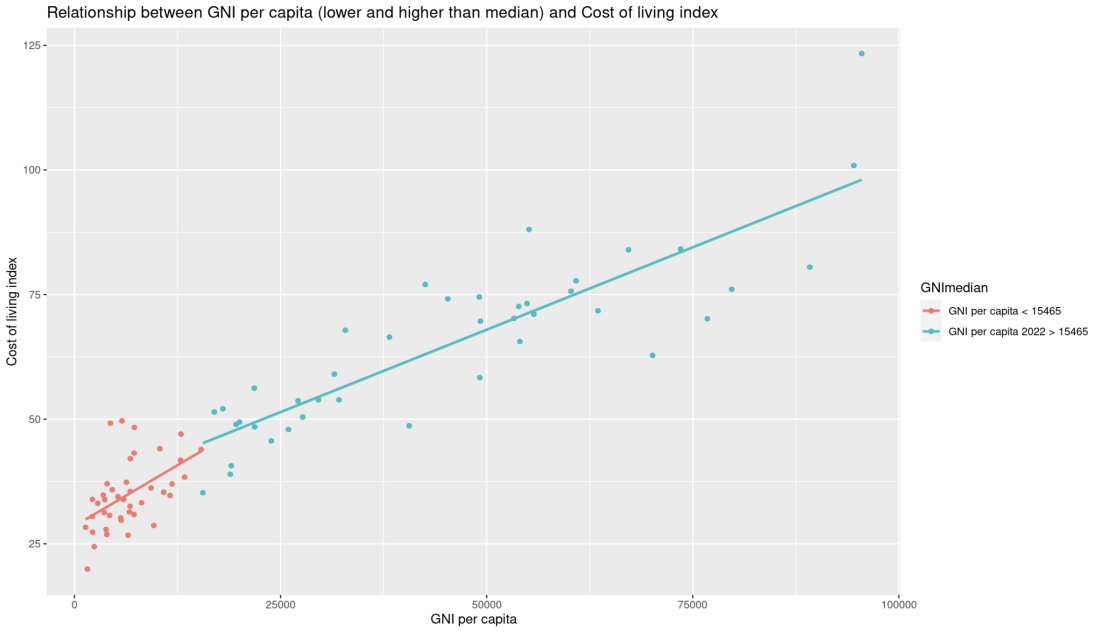


Figure 6: The red part shows the low-GNI countries with a higher gradient in linear model. The blue part shows the high-GNI countries with a lower gradient in linear model.

A higher slope for low-GNI countries means that when the average salary of people in this stage of the country increases, they suffer a greater increase in the cost of living index. This means their marginal cost of living is greater. In contrast, people in countries with higher GNI per capita have lower marginal costs of living. This is very reasonable in real life. Although people in low-income countries have relatively low living costs, their lives are harder and resources are scarce.

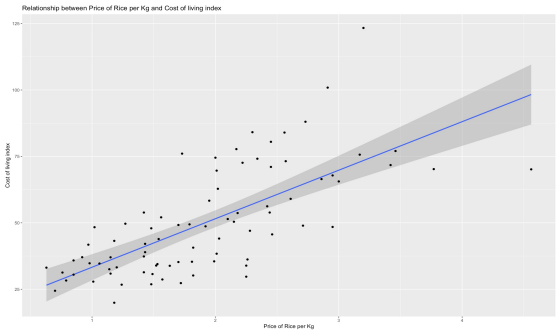
3.2.2

3.3 Research on the price data sets

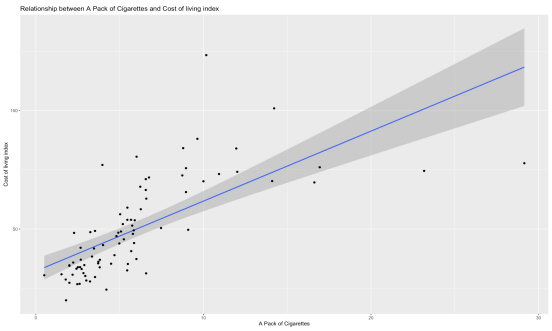
For the price data sets, we first analyzed horizontally between the products. We used Rstudio to create linear models for the eight different price data sets we have.

4 Result presentation

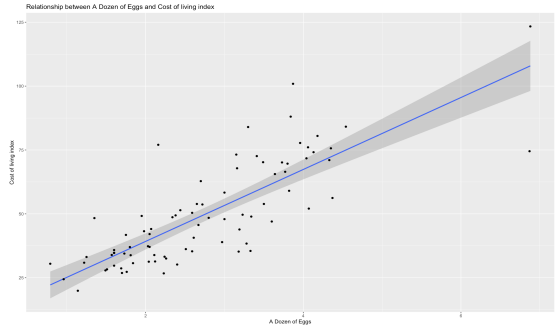
5 Evaluation and conclusion



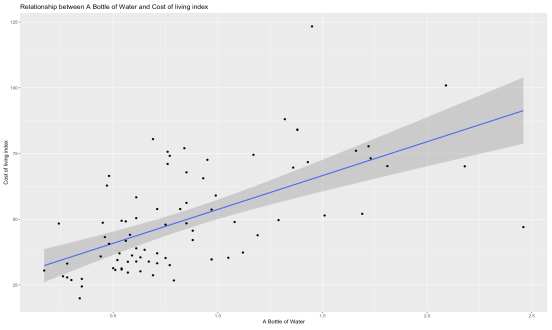
(a) Price of rice and Cost of living index



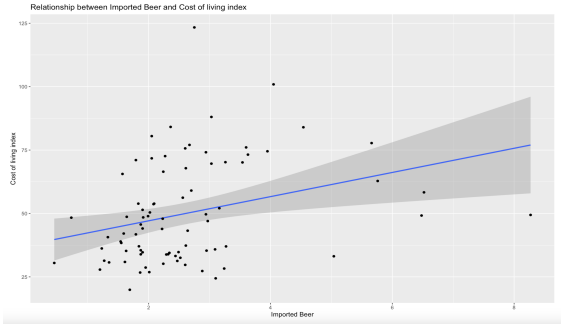
(b) Price of cigeratte and Cost of living index



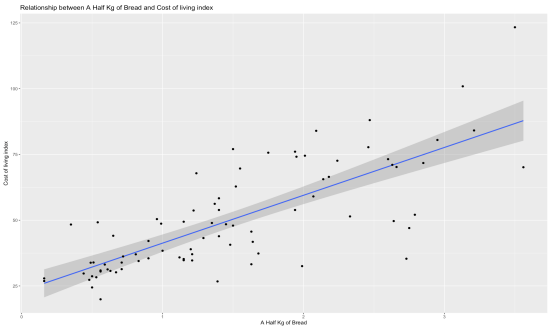
(c) Price of egg and Cost of living index



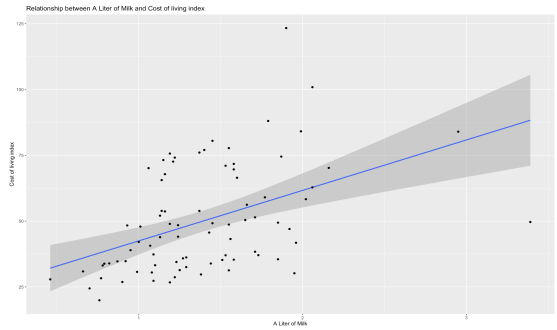
(d) Price of water and Cost of living index



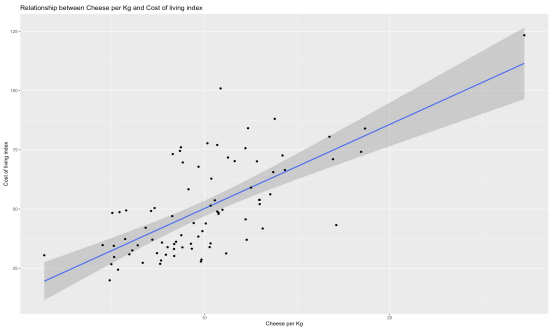
(e) Price of imported beer and Cost of living index



(f) Price of white bread and Cost of living index



(g) Price of milk and Cost of living index



(h) Price of cheese and Cost of living index

Figure 7: The figures of that compare the cost of living of different countries to the price of rice, cigarettes, cheese, eggs, milk, water, bread, beer all show a positive linear correlation. They reflect how as the price of a product gets higher, the cost of living does too. We can prove from this consistent positive linear result, that prices of products indeed explain the cost of living, or development of countries.

References

[1] Steve Dowrick and John Quiggin. International comparisons of living standards and tastes: A revealed-preference analysis. *The American Economic Review*, 84(1):332–341, 1994.