Correlates of the Cost of Living Index

Claire, Floyd, Caleb

May 2024

1 Introduction

We were curious to understand how the price of goods varies across different nations. 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 these different economic measures.

1.1 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 to inform the relationships that we explore. For price data, we will use typical countries as examples for comparative analysis. We will pick two products that are universal in a 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 its consumption 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.

1.2 Related Works

Steve Dowrick and John Quiggin's article International Comparisons of Living Standards and Tastes [1]: A Revealed- Preference Analysis asks similar questions to what we are curious to test in our 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."

Figure 9 in Appendix displays 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.

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.

Kenneth Clements's book, Currencies, commodities, and consumption: measurement and the world economy, 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]

2 Data processing and presentation

According to our guiding question, 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.

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.

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.

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 1.

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.

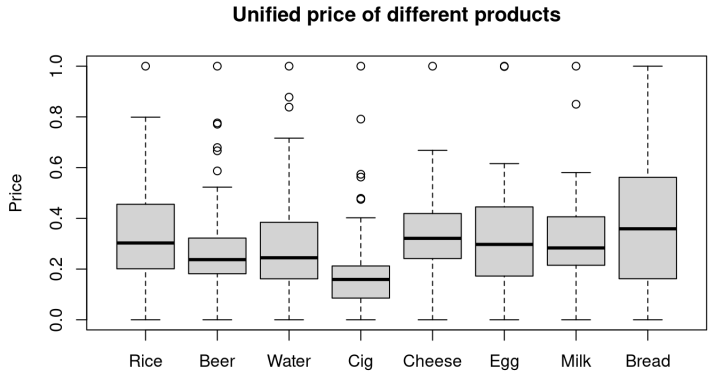


Figure 1: Box-plot of normalized price data of the eight products.

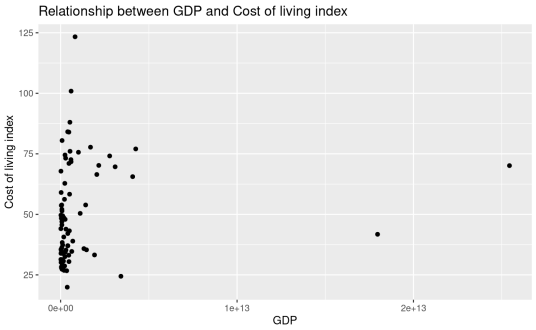
Figure 2 presents scatter plots of GDP, GDP per capita, GNI, and GNI per capita and cost of living index respectively. In the scatter plot, the scatter plots of GDP (GDP per capita) and GNI (GNI per capita) are almost the same. This is because, in economics, the values of GDP and GNI should be very similar. 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 2, 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.

The 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 the 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 is 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 the cost of goods and services means a rise in the cost of living as well.

3 Analysis

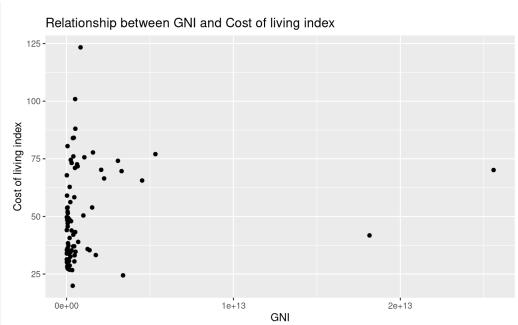
3.1 Directed acyclic graphs

First, we will make a directed acyclic graph to tease out the relationships between the three sets of data we are studying. This will make it easier for us to design the model. The cost of living index is the responsive variable we will study. Obviously, changes in product prices and income data GNI per capita will lead to changes in the Cost of living index. However, it is worth analyzing the relationship between product price data and GNI per capita data. Since the relationship between these two sets of data is complex, we will follow common sense and set the product price data as the explanatory variable of the GNI per capita data. In this way, product price data is the confounder in the relationship between GNI per capita and the Cost of living index.

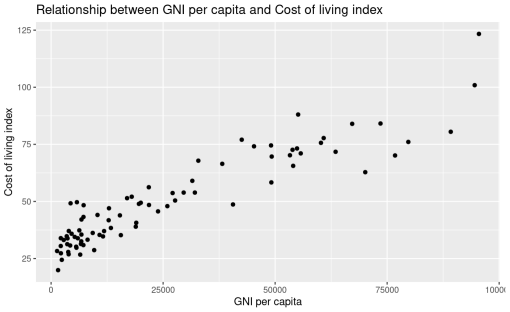


(a) GDP and Cost of living index

(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 2: Macroeconomic data and Cost of living index

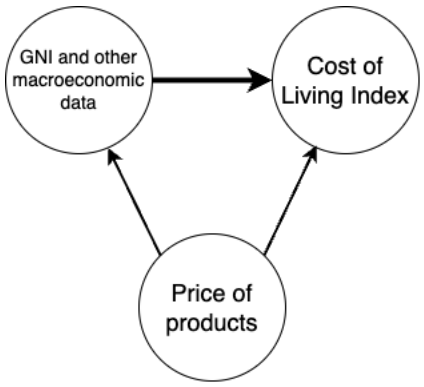


Figure 3: Directed acyclic graphs of our data sets.

3.2 Research on the price data sets

In order to study the relationship between product price data and the Cost of living index, we first analyzed horizontally between the products. We create linear models for the eight different price data sets we have in figure 10 in Appendix.

3.2.1 Price comparison in different areas

By eliminating the possible confounders of cultural practices, we decided to compare two countries with drastically different GNIs, Cost of Living, and Prices per product, but exist in the same global region. For

this experiment, we utilized data from Singapore and Malaysia. Singapore is an affluent country with a GNI of 67200 while Malaysia is still considered a developing country with a GNI of 11830. Singapore's success is attributed to its rapid industrialization in the 1960s which catapulted the nation's development from a low-income to a high-income country dramatically. Its manufacturing and service sectors are the main driver of growth in the country. On the other hand, Malaysia has been diversifying its economy, since gaining independence in 1957, from one that was initially agriculture and commodity-based, to one that focuses on manufacturing and service sectors. These industries have propelled the country to become a big exporter of electrical appliances, parts, and components. These countries have had a history of British colonization and Singapore joined the Federation of Malaysia in 1963 but seceded to become an independent state on August 9, 1965. They are culturally intertwined with their similar histories, but carrying different economic situations. Our figure compares the value of different products in Malaysia and Singapore. We can see a clear pattern of Singapore's prices being extensively higher than Malaysia's in every food category. Figure 6 depicted below shows this with the orange bars in the bar graph showing the price of different products sold in Singapore and the green bars illustrating the price of different products sold in Malaysia. [3] [4]

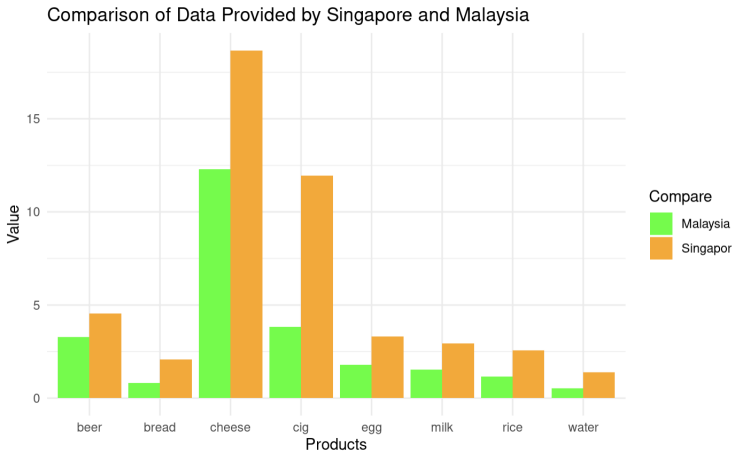


Figure 4: Bar graph of the price comparison between Singapore and Malaysia.

3.2.2 Price comparison of different countries in different regions

In Figure 5, we created a linear model that shows the interaction between the *Cost of Living Index* and *Price of rice per kg* across different countries in similar regions. We created a variable, Cultural Region, that separates the tested countries into the regions of Eastern, Middle Eastern, Latin American, South-East Asian, Sub-Saharan African, and Western. Our previous work comparing the prices of products between Singapore and Malaysia is visible under the South-East Asian Line depicted in bright blue. We can see the same high correlation of Cost of Living being a response to the explanatory *Prices of rice per kg*.

This pattern of high slope is seen in Middle Eastern regions as well. What was interesting was that for Western countries, the slope was less prominent. This reveals how rice is less of a driver of the Cost of Living in Western nations because it is a culturally less relevant product. This pattern in slope per region is a great indicator of understanding how rice as a product is a good variable for some countries' Cost of living Index, while not so good for others. It greatly affects our project as we come to understand that the Price of a Product is not the best overall variable for understanding the cost of living across all nations. Instead, it is a good predictor of nations in a particular region with similar cultural habits.

This contrast illustrates how the price of products relative to the cost of living can vary greatly between regions, influenced by factors such as income levels and economic conditions. The "steepness of this slope" mentioned suggests that the relationship between the cost of living, product prices, and income levels can

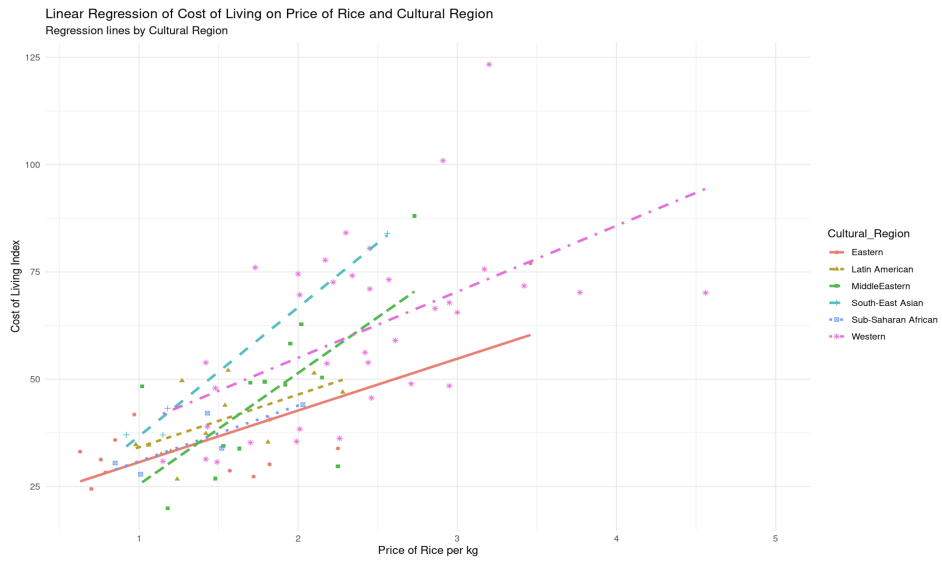


Figure 5: Line graph of the price comparison between different countries

create significant disparities and impact economic dynamics within a region. Another interesting find from this experiment was with the category, Western. There were a lot of scattered points in this comparison with a high residual value for its line. This leads us to believe that the price of rice is not a good predictor of the Cost of Living for this region. The reason for this can be explained by cultural factors of rice not being as in demand in Western countries.

3.2.3 Most correlated product investigation

In order to study the different impacts of different products on the cost of living index, we constructed a linear model between product data and cost of living index data and extracted the coefficients. As shown in Table 2, we extracted the Estimate coefficients and t-value coefficients of eight models. Among them, the larger the Estimate coefficient, the greater the impact of the product on the cost of living index. The larger the t-value coefficient is, the more significant the credibility of the product's impact on the cost of living index.

Product	Estimate	t value
Rice	18.262	8.878
Cigarettes	2.9527	8.463
Cheese	3.5513	8.282
Egg	14.074	12.21
Milk	19.178	4.535
Bread	18.192	10.670
Import Beer	1.648	2.891
Water	25.751	6.796

Table 2: Table of coefficients in linear model between the eight products and Cost of living index.

The coefficient results shown in the table are clearly consistent with common sense. Among them, water has the highest Estimate coefficient (25), which is significantly higher than milk, rice, bread, and eggs (14-

19), and significantly higher than beer, cheese, and cigarettes (¡10). In life, water is the only important basic substance, while milk, rice, bread, and eggs are daily necessities. Beer, cheese, and cigarettes are relatively expensive and only have a greater impact in some areas. In terms of credibility significance, only beer and milk have relatively low t-values. This shows that these two products are relatively affected by other factors. The possible reason is that only richer countries can generally consume milk and imported beer, or some areas do not have the habit of drinking milk and imported beer.

3.3 Research on the macroeconomic data sets

Based on the preliminary data processing and literature research mentioned above, we will only conduct modeling research on GNI per capita data and the cost of living index. From the figure 6, 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 between them.

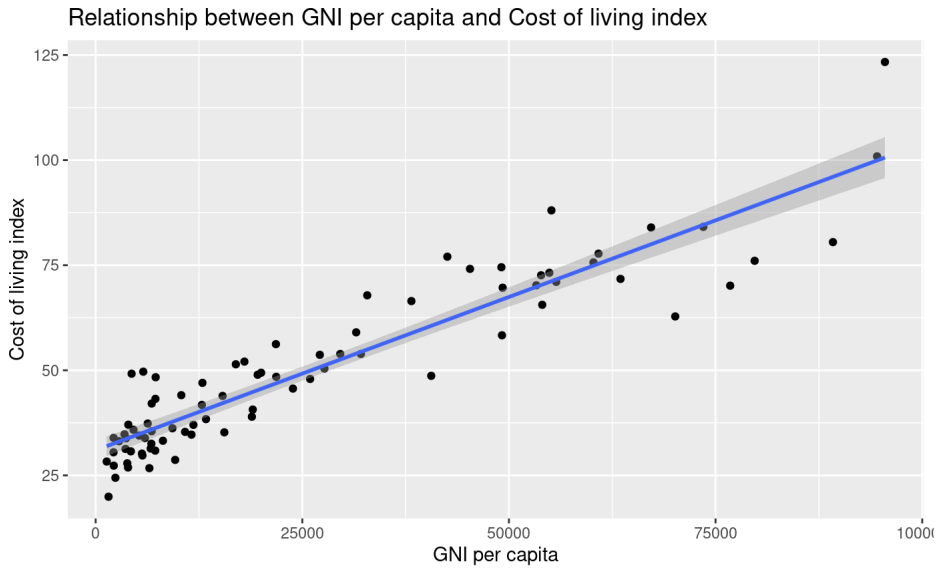


Figure 6: Linear model between Cost of living index and GNI per capita.

3.3.1 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 the 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 7 below.

For the overall GNI data, 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}.

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

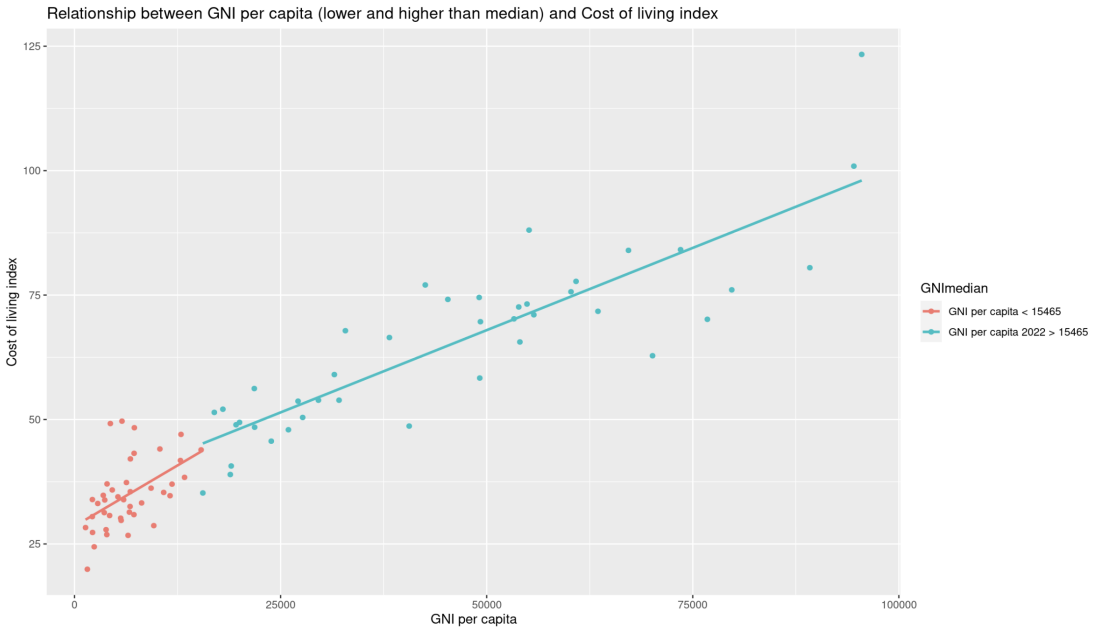


Figure 7: The red part shows the low-GNI countries with a higher gradient in linear model. The blue part shows the high-GNI countries with a lower gradient in linear model.

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.

3.3.2 Outlier extraction and logarithmic simulation

We can generalize this conclusion based on the further observation that countries with lower incomes have higher marginal costs of living. For this observation, we verify it using simulations of the logarithmic function of the Cost of Living Index $\sim$ GNI per capita data. Before the simulation, we removed outliers in the GNI data so as not to affect the accuracy of the fit. The GNI per capita data has $Q1 = 6045, IQR = 29140, Q3 = 44605$, so we removed Luxembourg, Norway, and Switzerland with GNI per capita greater than 88315. As shown in Figure 1, this logarithmic function curve better simulates the data of the Cost of Living Index $\sim$ GNI per capita. Excel provides an R^2 value of 0.9374 for this logarithmic model.

4 Result presentation

4.1 Most impact product of cost of living index

For the eight products we studied, we found that different products have different impacts on the cost of living index and have different significance. In terms of magnitude of impact:

$$price_{water} > price_{milk} > price_{rice} > price_{bread} > price_{eggs} > price_{cheese} > price_{cigarettes} > price_{beer}$$

On the significance of credibility:

$$price_{eggs} > price_{bread} > price_{rice} > price_{cigarettes} > price_{cheese} > price_{water} > price_{milk} > price_{beer}$$

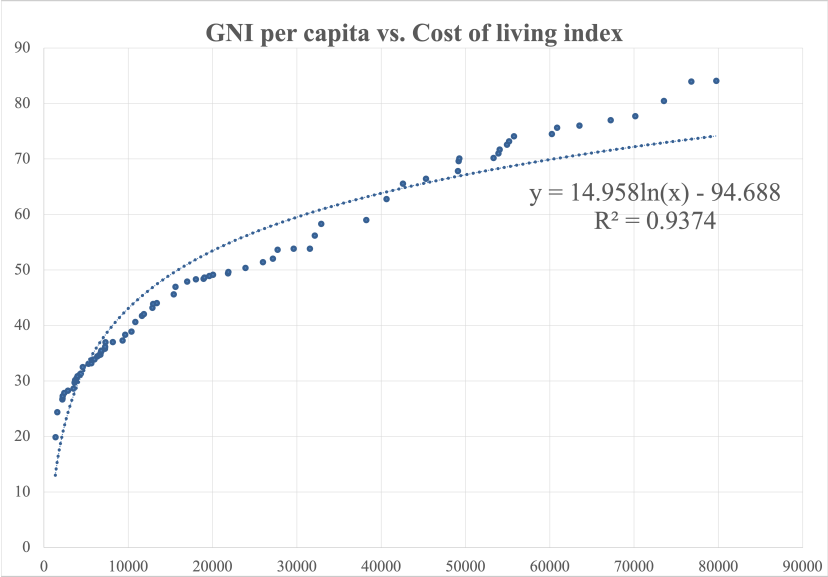


Figure 8: The logarithmic model of Cost of Living Index $\sim$ GNI per capita with labeled formula and R^2 value.

We can see that the more basic a product is to people's needs, the greater its impact on the cost of living index. In terms of significance, the three beverages are less significant. This may indicate that there is a certain relationship between product salience and product functions.

4.2 Income and cost of living

In the research above, we study the relationship between national income and cost of living through GNI per capita and Cost of living index. In a linear relationship, countries with higher GNI per capita have higher Cost of living index. After that, we further used the logarithm model to simulate the logarithm relationship between GNI per capita and Cost of living index. This shows that for countries with lower per capita wages, for every unit increase in the country's GNI index, the Cost of Living Index will increase even more. Relatively speaking, for richer countries, when the GNI index increases by one unit, the Cost of Living Index increases slightly.

4.3 Products and regions

In our research between Singapore and Malaysia example, we compare the products of these two nations taking into account cultural factors that could influence what products are more in demand. We found that their prices are consistently varied with Singapore carrying higher prices of products with their higher cost of living. In our second experiment, we wanted to magnify the scale of our first experiment. We accomplished this by creating a line graph that demonstrated the different correlations between the Price of Products and the Cost of Living with different countries categorized by region. Our countries were separated into their respective Eastern, Latin American, Middle-Eastern, South-East Asian, Sub-Saharan African, and Western regions. We concluded from our line graph that strong correlations for the price of rice per kg exist in regions where rice is a culturally prevalent grain and a weak correlation for regions where rice is not so much of a culturally relevant grain.

5 Evaluation and conclusion

5.1 Advantages and flaws

In this project, we analyzed the relationship between cost of living index data, product price data, and GNI per capita data. We mainly used linear models to conduct a comparative analysis of these variables. We used images of linear and logarithm models as visual representations in our project. We not only applied the most basic model to study how product prices affect the cost of living index, but we also studied the special connection between per capita wage level and GNI, and the impact of cultural regions on the cost of living index $\sim$ product price model. We made extensive use of the models, instructions, and research methods learned in class, including linear models, graph plotting, error measurement, and confounder justification. However, there is still room for improvement and the possibility of continuation of our research. Our study has several limitations and flaws. First, our data only includes 82 countries. This makes the results of the study ignore the situation of many small countries and countries with insufficient data. Secondly, the number of our products is only 8. This makes our product price data overall too microscopic. Finally, our research methods are relatively rudimentary, which results in relatively basic and error-prone research results, and we can only draw qualitative conclusions.

5.2 Improvement and further research

Based on the flaws in our project, further research is necessary and possible. First, the data can include more countries or use more product data. For example, study the relationship between price data for 100 common products and the cost of living index. This will make the conclusions more valuable and closer to the true relationship. Secondly, follow-up research can also continue what we have already studied. For example, products can be divided into economic necessities and luxuries to study the different impacts of different types of products on the cost of living index. Alternatively, a culturally specific product, such as tofu, could be analyzed and compared to a more universal product, such as water. Notably, an important result of our study is that poorer countries have higher marginal costs of living. Poor countries not only cannot enjoy low living costs but also suffer from material scarcity and huge increases in living costs due to development, as GNI per capita increases. This reminds us that the content that should be further researched should go far beyond statistical theory, but to the actual life that needs to be paid attention to.

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.

[2] Kenneth W. Clements. *Currencies, Commodities and Consumption*. Cambridge University Press, 2013.

[3] World Bank. Malaysia overview, 2024.

[4] World Bank. Singapore overview, 2024.

6 Appendix

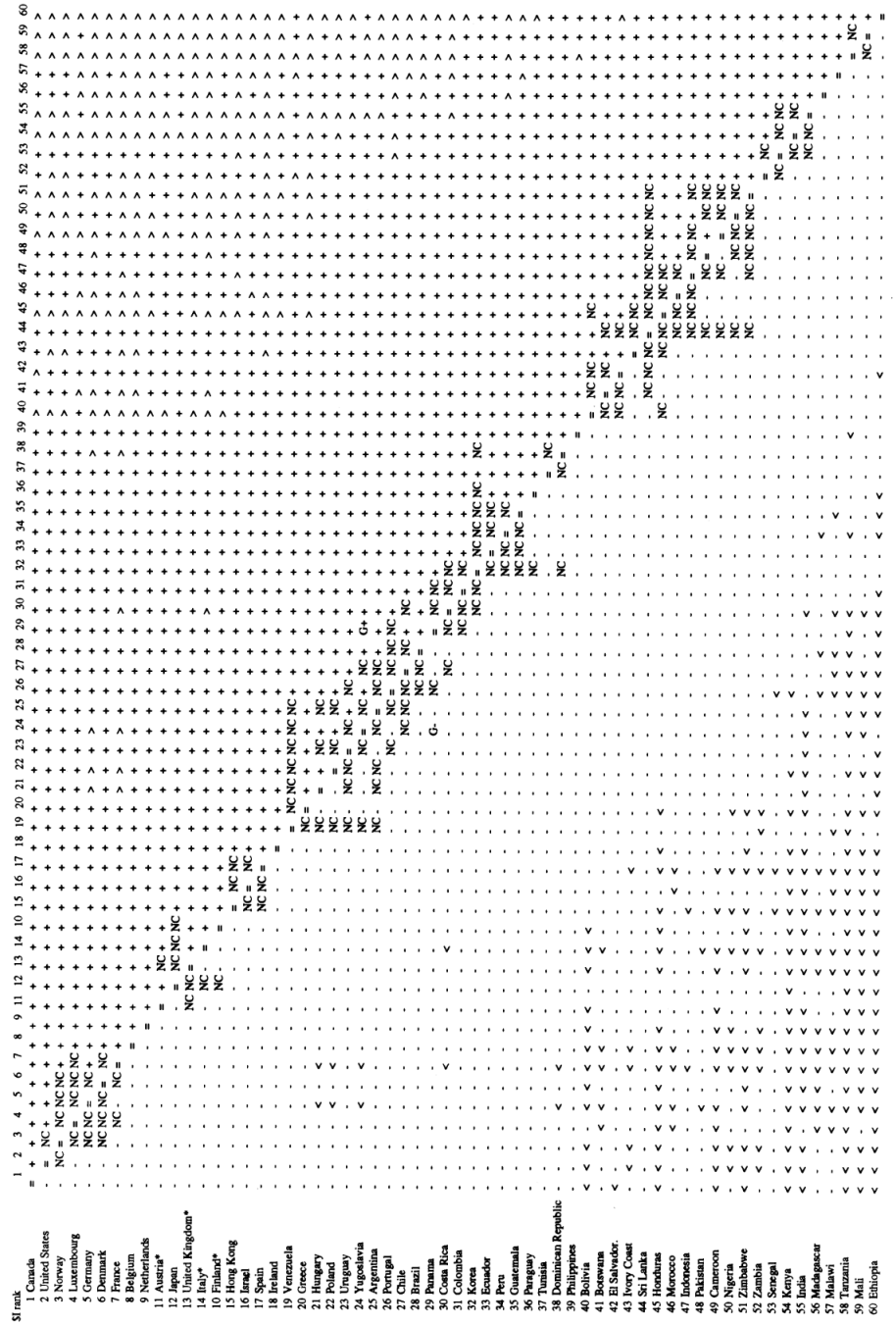
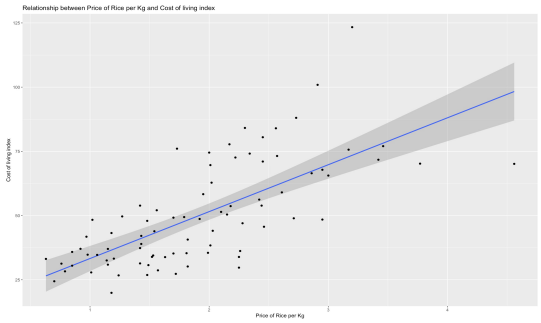
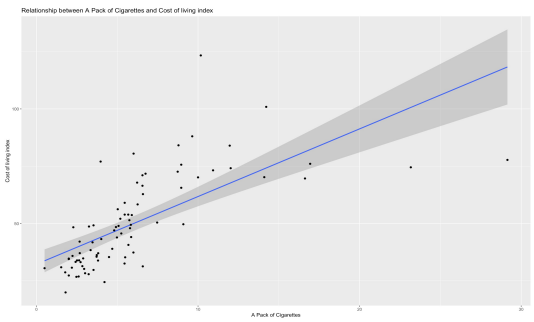


FIGURE 2. REVEALED-PREFERENCE RANKINGS OF 1980 PER CAPITA GDP BY 38 CATEGORIES OF EXPENDITURE (SEE TEXT)

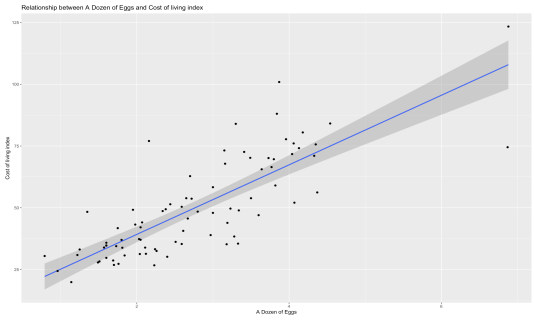
Figure 9: How different countries rank in terms of their GDP per capita [1]



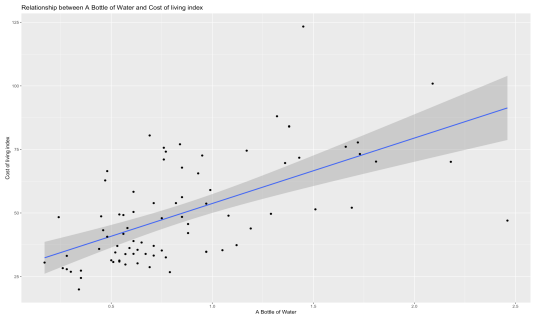
(a) Price of rice and Cost of living index



(b) Price of cigeratte and Cost of living index

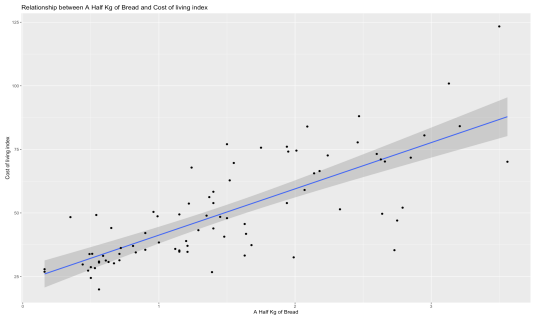


(c) Price of egg and Cost of living index

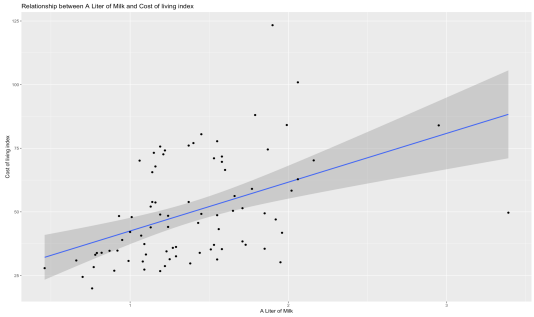


(d) Price of water and Cost of living index

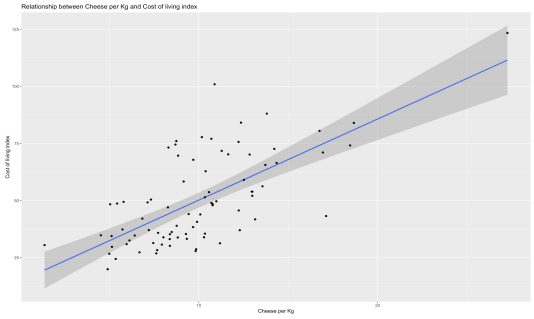
(e) Price of imported beer and Cost of living index



(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 10: The figures that compare the cost of living of different countries to the price of rice, cigarettes, cheese, eggs, milk, water, bread, and 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 the prices of products indeed explain the cost of living, or development of countries.