|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
|  |  |  |  |
| Global food prices  impact analysis on population, mORTALITY, bIRTH- rATE and GDP  Group Assignment – Data Science 2 – Introduction to Statistics | |
|  | Presented by - Group 2  Jobbin Samuel  Ricky Lam  Maxim Smetin | Ramila Mudarth  Avery Lizhong Wang  James Lising |  |

Table of Contents

[iNTRODUCTION 4](#_Toc90061670)

[Data preparation 4](#_Toc90061671)

[data ANALYSIS 6](#_Toc90061672)

[Linear regression Analysis 8](#_Toc90061673)

[Birth Rate analysis 8](#_Toc90061674)

[Child Mortality Analysis 9](#_Toc90061675)

[DEATH Rate Analysis 10](#_Toc90061676)

[Rice Contribution in Afghanistan and Regional LEVEL 11](#_Toc90061677)

[CONCLUSION 12](#_Toc90061678)

[Bibliography 13](#_Toc90061679)

[ApPENDIX 14](#_Toc90061680)

# Objectives

In this project we are going to explore a question about possible impact of food prices into population. Is there correlation between changes in food prices and birth rate, death rate and child mortality?

We choose dataset with food prices for the last twenty years in developing countries, as well as birth rate, death rate and child mortality. Moreover, how GDP impacts population growth in developing countries?

# iNTRODUCTION

The global food prices have always been subjected to external influences like fuel prices, natural disasters as a result of global warming activities. Some countries have been affected more than others depending on their ability to endure the fluctuations and food availability. There are assumptions made that the food prices influence the mortality, the GDP, affordability and income which in turn effects the population trends. Producers benefit from rise in prices where are consumers benefits from lower food prices. Any fluctuation in prices will have an effect especially on the lower income individuals as a result causing food shortages.

The objective of our assignment was to determine the impact of food prices of specific commodity in developing countries and compare their GDP, Mortality Rate to determine any kind of correlation that might exists using last twenty years of data. Additionally, we intend to pair it with restaurant index data to see if there are any insights derived.

We intend to resolve the two hypotheses

* **Hypothesis 1:**How much food price influence Population? Null Hypothesis is food price isn’t a key driver of population. The alternative Hypothesis is that food price somewhat affect population.
* **Hypothesis 2:**How much do food prices impact all parameters? Null Hypothesis is that food prices impact all parameters equally. The alternative hypothesis is that there are some differences between some parameters affected by food prices.

# Data preparation

Preparation of the data set required compiling and sourcing from multiple location. Each data set had to be solved for challenges presented and transformed for the required analysis

A picture containing text, businesscard, screenshot

Description automatically generated

Figure Data Schema

|  |  |  |
| --- | --- | --- |
| The core dataset contains Global Food Prices data from the World Food Programmed covering foods such as maize, rice, beans, fish, and sugar for 98 developing countries and some 1,500 markets. The data goes back as far as 1992 for a few countries, although many countries started reporting from 2003 or thereafter. The Data is collected by [WFP(The World Food Program)](https://www.wfp.org/) and the dataset was distributed by [HDX](https://data.humdata.org/) . Data includes developing countries, locality, market, goods purchased, price & currency used, quantity exchanged, and month/year of purchase.  The unit of measure (UOM) for the commodities were not consistent per observation. We have created a table for normalizing the retail value to kilograms. This step let us to perform aggregation of food prices across the whole dataset per Continent, Country, Commodity and Year. A conversion table was manually created to do the math. No conversion was applied to fuel, and commodity item such as toothbrush, toothpaste, and other liquid that is unmeasurable. UOM in file has not been changed in the file. But those that have been change will be reflect in retail price. *e.g. where the retail was 2500 dollar for 12 kg was converted to 208 dollars per KG.*    Figure 2 The data frame structure  The Food prices were in local currency. The classification of the commodity category was too granular, so we had to aggregate it at a boarder level*. e.g. Rice commodity name has 82 different versions. But we created a category by using the first word in the string.*  The other features merged into the table are   * Regional Classification - Countries were reclassified by the region for higher level analysis and to obtain greater number of observations eg. Northern Africa', 'Eastern Africa', 'Middle Africa', 'Southern Africa', * Gross domestic product (GDP) – Twenty years of GDP information trend per country * Child Mortality Rate - Child Mortality counts since 1967 per country * Birth and Death Rate – Birth and Death count per country since 1960 * Fertility Rate - Fertility Rate since 1950 per country   We added another layer of classification for the commodities   * Raw – Milk, Eggs, Rice * Processed – Bread, Curds * Other - Fuel, Internet  data ANALYSIS By splitting to categorical view, there is no relationship between commodity, & price with any value/ table shown in the matrix below    Figure 3 - Correlation Matrix for Categorical Variables  The prepared data set was further compressed and aggregated to give commodity per country per area. In the prepared data set, we observed that Rice was one commodity that had sufficient data point. We created a subset data frame to run an analysis on. A heat Map was generated using this data set.    Figure Evaluation of Pearson Coefficient across the data frame  Initial high-level observations from the Heat Map shows strong positive correlation between the following parameters   * Population and GDP in USD * Birth Rate and Death Rate * Population and Child Mortality   The negative correlation exists in the following parameters   * GDP versus Birth Rate as well as Death Rate * Price of Rice versus Child Mortality, Birth and Death   The parameters that are more likely to be independent are   * Shelf Price of Rice and GDP is mutually exclusive.    Linear regression Analysis We performed Linear regression modeling on Birth Rate, Death Rate and Mortality at a regional level to see the relationship between the parameters in the data Birth Rate analysis In the Birth Rate OLS Model – A , we observed adj R Squared is 68% making the model reasonable fit.  All the p values are zero as a result we can say, we accept the null hypothesis. All parameters are acceptable in the model.    Figure Birth Rate analysis  We ran Birth Rate OLS Model – B, after removing the outliers from the Influential Plot and the model had an Adjusted R Square of 72.3%.    Figure Influential plot Child Mortality Analysis In the Child Mortality analysis, the Adjusted R-Squared is 15% as a result the null hypothesis is unacceptable.    Figure Child mortality analysis DEATH Rate Analysis The model is at 52%. The price of Rice is not relevant as we need to remove the parameter as the p value is greater than 0.05    Figure Death rate analysis Colliders Birth rate and average price scatter plot shows quite disperse data distribution on both scales. For example Figure 5 shows the same data but with different granularity level – starting from continent Asia, then Regions, and finally by countries. We observed that the correlation between two parameters become stronger if regions and countries being includes.  Linear regression analysis shows no relationship between birth rate and rice on the continent level.    Figure 5. Average rice price and birth rate by continents. Countries in Asia and South America  Pearson squared coefficient is equal to 0.07, p-value for rice coefficient is 20%. We fail to reject null hypothesis – there are no influence of rice price on the continent level. However, on the country level the picture is significantly different. Distributions of birth rates and rice prices becomes narrower. For example, Afghanistan (Figure 6) R squared is equal to 0.62, the coefficient is -45.2 and p-value 1%. Therefore, we reject null hypothesis and conclude that there is an impact of the food price on the birth rate.  Figure 6 Birth rate and rice price in Afghanistan |  |  |

|  |  |
| --- | --- |
|  |  |

|  |
| --- |
| CONCLUSION *What did we infer from our analysis?*  We have studied relations between food prices and different aspects of population changes – birth rate, death rate and child mortality. We explored GDP contribution as well.  Having chosen rice as the most representative commodity we have inferred what it has negative correlation with population.  Overall dataset has small correlation between food price and birth rates. However, if we introduce continents and countries, and apply inferred analysis there, we have observed strong correlation between food price and birth rates.  We have observed that the birth rate in Afghanistan, can be represented through linear regression in the following expression:  *Birthrate = 55.5 – 45 x food price*  inside our dataframe, and therefore reflecting   * **Hypothesis 1:**Is food price influence Population? Null Hypothesis is food price isn’t a key driver of population. The alternative Hypothesis is that food price somewhat affect population.   We are rejecting null hypothesis 1.   * **Hypothesis 2:**How much do food prices impact all parameters? Null Hypothesis is that food prices impact all parameters equally. The alternative hypothesis is that there are some differences between some parameters affected by food prices.   We failed to reject null hypothesis 2, since there one of three population parameters – child mortality did not have **correlate with rice price.** |
|  |

# Bibliography

*Exchange Rates*. (n.d.). Retrieved from https://data.oecd.org/conversion/exchange-rates.htm.

*Global Food Prices*. (n.d.). Retrieved from https://data.humdata.org/dataset/wfp-food-prices.

*Gross Domestic Product*. (n.d.). Retrieved from https://stats.oecd.org/Index.aspx?DatasetCode=SNA\_TABLE1.

*Population Trends*. (n.d.). Retrieved from https://data.worldbank.org/indicator/SP.POP.TOTL.

Ritchie, M. R. ((2013) ). *Food Prices". Published online at OurWorldInData.org*. Retrieved from 'https://ourworldindata.org/food-prices' [Online Resource].

# ApPENDIX