Global Food Prices and population Trends

Data Science 2 – Statistics for Data Science Summary Report

**Submitted by: Group 2**

Jobbin Samuel

Ricky Lam

Maxim Smetin

Ramila Mudarth

Lizhong Wang

James Lising

# Objective

The objective in our assignment was to determine the impact of food prices of specific commodity in developing countries against its GDP and Population Trends – Birth Rate, Death Rate and Child Mortality to determine any kind of correlation that might exists using last twenty years of data.

We predefined the two hypotheses prior to our analysis as follows

**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 population trends and GDP 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 Source

The preparation of the data set for modelling required combination of Global Food Prices Data from [WFP (The World Food Program)](https://www.wfp.org/),  GDP Trends, Unit of Measure Equivalization Table, Exchange Rate, Birth Rate, Death Rate and Child Mortality. The final table had twenty years of Food prices in US dollars as well as the corresponding Birth, Death, Child Mortality and GDP. The US dollar conversion we used the current rate for simplicity.

# Data Analysis

When a categorical correlation matrix was ran for different commodities type, the values were close to zero as a result, the conclusion was that there could be no relationship between the different commodities.

When a correlation matrix was run using one commodity – Rice, 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.

By running a scatterplot distribution for all Rice prices and birth rate, there was no correlation observed. But when the countries and regional parameter was added, the correlations were more defined.

# Modelling

**Linear Regression**

Multi Variate Linear regression modelling was conducted for the Rice commodity Price against

**Boot strapping Results**

# Conclusion