GLOBAL FOOD PRICES AND POPULATION TRENDS

Data Science 2 – Statistics for Data Science
Summary Report

Submitted by: Group 2
Jobbin Samuel
Ricky Lam
Maxim Smetanin
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 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 hypotheses prior to our analysis as follows

Hypothesis 1: Is there a correlation between food prices and population trends? Null Hypothesis is food price isn't a key driver for population trend. The alternative hypothesis is that food price affects population trends.

Data Source

The preparation of the data set for modelling required combination of Global Food Prices Data from WFP (The World Food Program), 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 correlation matrix was ran for the full data set, the values were close to zero. As a result, the conclusion was that there could be definitive correlation between all the parameters.

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

- Birth Rate, Fertility, Death Rate and Child Mortality – suggesting that they are colinear. This is a natural observation, for example, we would expect birth rate to depend on fertility.

The negative correlation exists between the following parameters

- GDP per capita and birth rate, child mortality, fertility, death rate
- Price of Rice versus Child Mortality, Birth, Fertility and Death rates

Modelling

Linear Regression: Analysis of relationship between rice price and birth rate.

Scatter plot of Rice prices and birth rate shows no correlation on global and continent levels (Figure 1A). Linear regression analysis shows no relationship between birth rate and rice on the continent level. 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, when we increase granularity and examine by Asian regions (Figure 1B), countries (Figure 1C) and country level (Figure 1D), the correlations becomes more defined. For example, Afghanistan (Figure 1D) R squared is equal to 0.62, the coefficient is -45.2 and p-value 1%. Therefore, with 99% confidence we reject null hypothesis and conclude that there is an impact of the food price on the birth rate in Afghanistan.

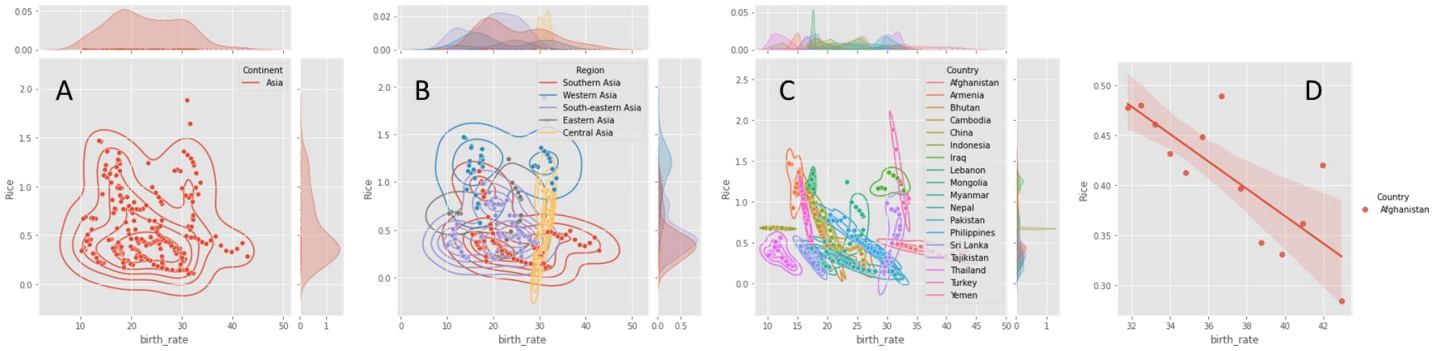

Figure 1. Average rice price and birth rate in Asia (A), Asia regions (B), Asian countries (C) and Afghanistan(D)

Multi-Variable Linear regression: Analysis of relationship between rice price and other parameters

We performed multi-variable Linear regression modeling on Birth Rate, Death Rate and Mortality at a continents level to see the relationship between the parameters in the data

In the Birth Rate OLS Model – A, we observed adj R Squared is 0.725 making the model reasonable fit. All the p values are around zero as a result we can say, we accept the null hypothesis. All parameters are acceptable in the modelling. We ran Birth Rate OLS Model – B, after removing the outliers from the Influential Plot and the model had an Adjusted R Square of 0.726. We can infer that there we no significant outliers.

In the Child Mortality OLS Model, the R Squared was at 0.163. The p values for continent of Europe were greater than 0.05. The null hypothesis was rejected.

In the Death Rate mortality OLS model, R Squared was at 0.52. The price of Rice is not relevant as we need to remove the parameter as the p value is greater than 0.05

Conclusion

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 that 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, 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:

$$Birthrate = 55.5 - 45 \times food\ price$$

Reviewing our hypothesis, previously defined, we can conclude as follows

Hypothesis 1: Is there a correlation between food prices and population trends? Null Hypothesis is food price isn't a key driver for population trend. The alternative hypothesis is that food price affects population trends.

Conclusion: We are rejecting null hypothesis 1. The relationship between food price and birth rate can be expressed with linear regression on the country level. P-values for slope coefficients are less than 1%.