GLOBAL FOOD PRICES

IMPACT ANALYSIS OF FOOD PRICES ON POPULATION TRENDS

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

OBJECTIVES

In this project we are going to explore the impact of food prices on population trends - birth rate, death rate and child mortality. Through the various statistical modelling technique, we will obtain a possible correlation and define a good fit model between changes in food prices and birth rate, death rate and child mortality of the countries in our data set.

We chose a dataset with food prices for the last twenty years in developing countries and their corresponding birth rate, death rate and child mortality rate. We have extended the analysis to include Gross Domestic Product (GDP) per capita.

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 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. Countries that have a higher population of lower income will have a greater impact on the population trends.

We set the purpose of our analysis to determine the impact of food prices of a specific commodity in developing countries and compare their GDP and population trends to determine any kind of correlation that might exists using last twenty years of data.

We have defined and intent 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.

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 the necessary transformation required for analysis.

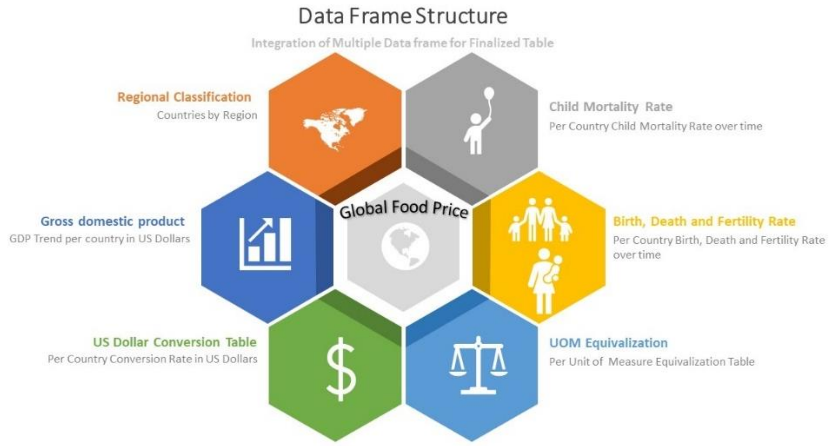

Figure 1 - Data Schema

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) and the dataset was distributed by HDX . 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.*

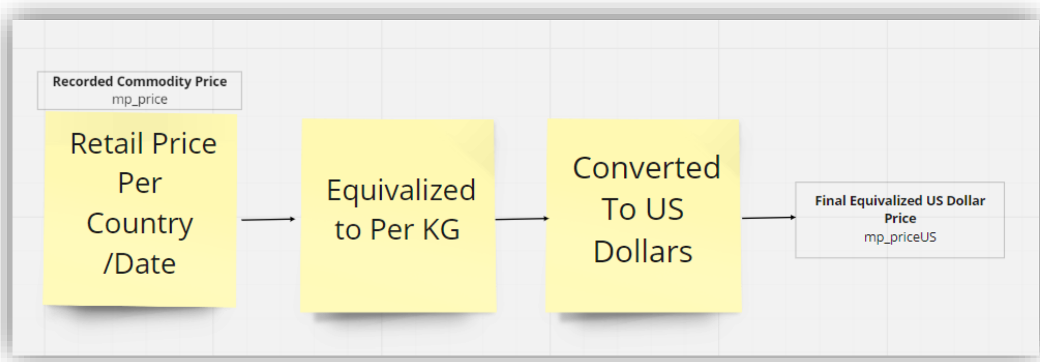

Figure 2 - The data frame structure

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.*

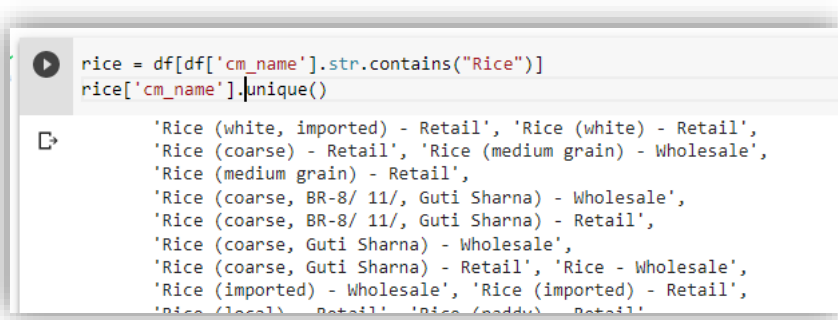

Figure 3 - Different Version of Rice Category

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 e.g. Northern Africa, Eastern Africa, Middle Africa, Southern Africa.
- Gross domestic product (GDP per capita) – Twenty years of GDP information trend per country

- Child Mortality Rate - Child Mortality counts since 1967 per 1,000
- Birth and Death Rate – Birth and Death count per country since 1960 per 1,000
- Fertility Rate - Fertility Rate since 1950, shown per woman

We added another layer of classification for the commodities

- Raw – Milk, Eggs, Rice
- Processed – Bread, Curds
- Other - Fuel, Internet

DATA ANALYSIS

Initial analysis was done for all countries presented in the dataset to see if there was any relationship between all parameters. After running the correlation matrix, it can be inferred that there was no definitive correlation between parameters with any value/ table shown in the matrix below.

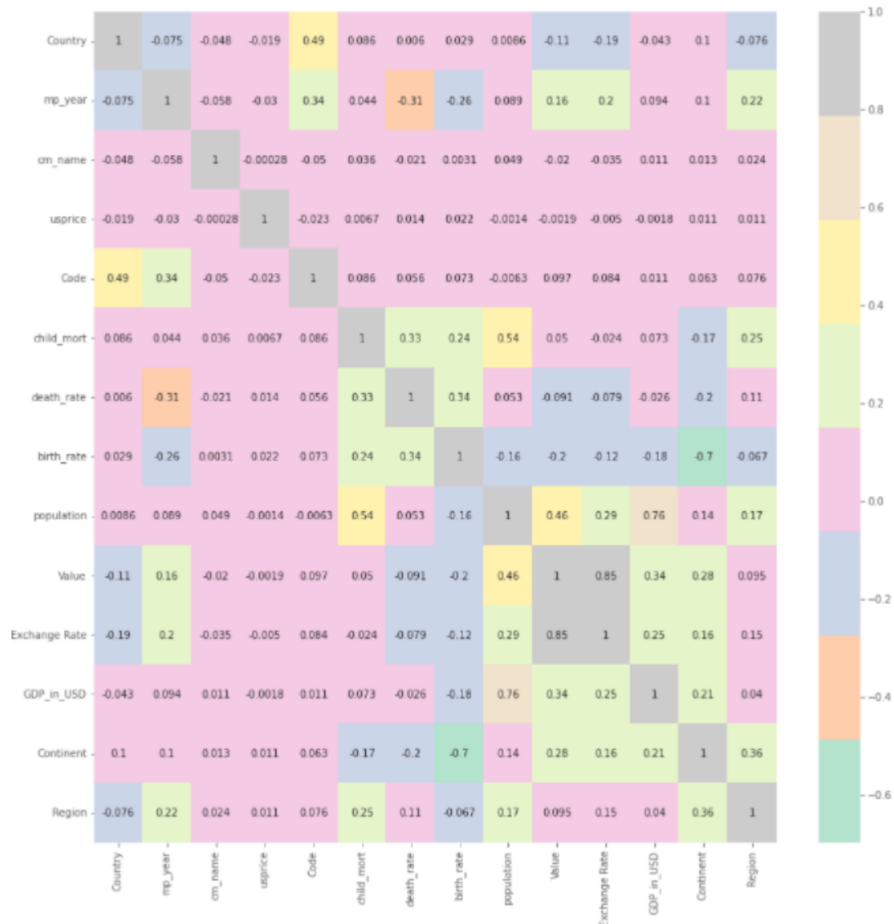

Figure 4 - Correlation Matrix for All Data Set Variables

We observed that Rice had a good set of observations across countries in the dataset. We chose Rice as a most represented commodity and would be a good sample observation for further analysis. We can assume that any changes in food prices will be captured in rice prices as well.

A heat Map was generated using this data set.

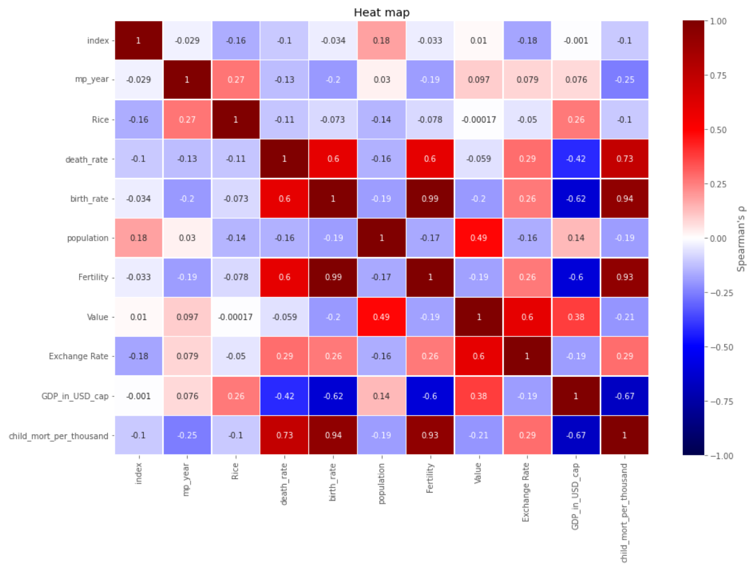

Figure 5 Evaluation of Pearson Coefficient across the data frame

Initial high-level observations from the Heat Map shows positive correlations 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 in the following parameters

- GDP per capita and birth rate, child mortality, fertility, death rate
- Price of Rice versus child mortality, birth, and death.

We took a deeper dive into the relation between rice price and birth rate. Rice being chosen as it is present across our dataset. This would mean that this is a common meal ingredient across most of the countries, and its price fluctuations is colinear with other grocery ingredients. Therefore, the rice price was taken as a representative of the global food prices. Here we want to study how the food price impacts population growth, namely birth rate.

Linear Regression Analysis: Rice and Birth Rate

Once our dataset was built, we ran a check for relation between Rice price and birth rate Figure 6. Overall data was quite scattered. When the linear regression was run for the data set, there was no significant relation between the two parameters. Indeed, heat map (Figure 5), shows Pearson coefficient between them is **-0.071**, and slope has high p-value >0.05. Therefore, we failed to reject null hypothesis and conclude that rice price doesn't impact birth rate.

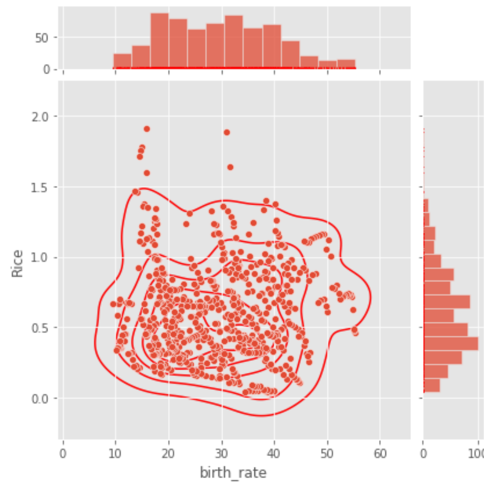

Figure 6 - Scatter plot with rice cost and birth rate distribution across the dataset.

When we extend the dataset to include the developing countries parameter, the observation is quite different. What if we will have a closer look at this dataset and zoom into Asia continent.

CASUAL INFERENCE: ASIA

BIRTH RATE AND RICE PRICE

Whole Asian part of the dataset have birth rate and rice price quite disperse distribution on both scales. For example Figure 7, shows the same data but with different granularity level – starting from continent Asia (Figure 7A), by regions (Figure 7B), by countries (Figure 7C) and one country – Afghanistan (Figure 7d). 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. 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.

Meanwhile, when we examine our relationship on region level (Figure 7B) – the intercept and slope values for each region are not the same. In case of Central Asia – the birth rate is independent of rice price.

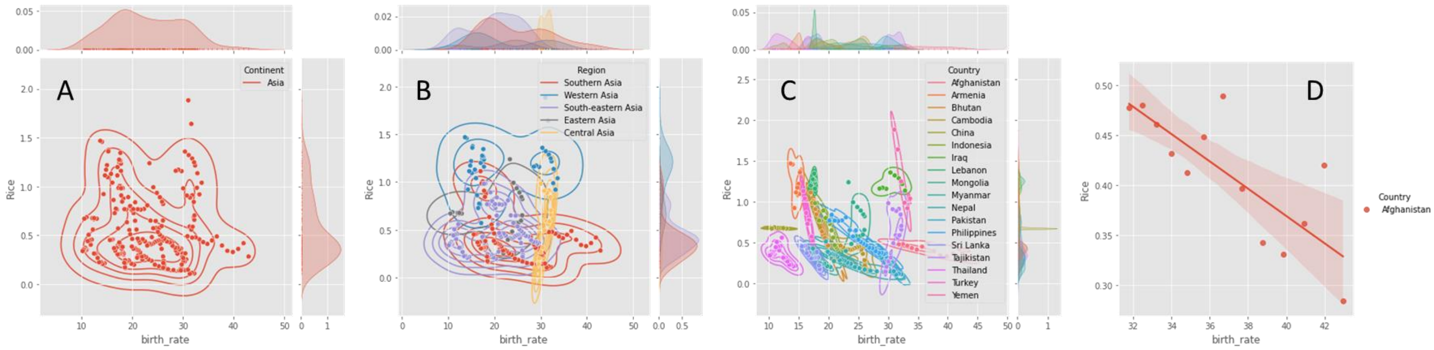

Figure 7 - Average rice price and birth rate in Asia, Asia regions and Asian countries

Once we increase further the granularity and have the distribution by countries – the diversity becomes obvious. Each country will have different intercept – mean values. However, the slope has negative sign – suggesting that with increase of the rice will cause decline in the birth rate.

For example, Afghanistan (Figure 7D) 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.

ASIA: CHILD MORTALITY RATE AND RICE PRICE

Let's examine if there are relationship between food prices and child mortality. On the continent level the data has a strong spread (Figure 8A), and no correlation ($R^2 = 0.115$). Rice slope coefficient has negative 0.95 value. But, even from this level, one can see that in case of small child mortality – between 0 and 2 per thousand – the rice price has no impact. However, as the child mortality becomes more then 2 the rice role becomes stronger.

From region level we can see that rice price impacting Southern Asia countries (Figure 8B). Figure 8C shows the relationship on the country level. For some countries the data points are almost perpendicular to x-axis – showing indeed no dependency from y-axis (rice price). We were happy to see that, as this mean that food accessibility doesn't contribute to the child mortality in those countries. However, in case of Afghanistan R-squared is equal to 0.62, and slope for rice price defined with 1.0% p-value. In this case child mortality can be expressed:

$$Child\ mortality_{Afghanistan} = 6.7 - 9.2 \times Rice\ Price_{Afghanistan}$$

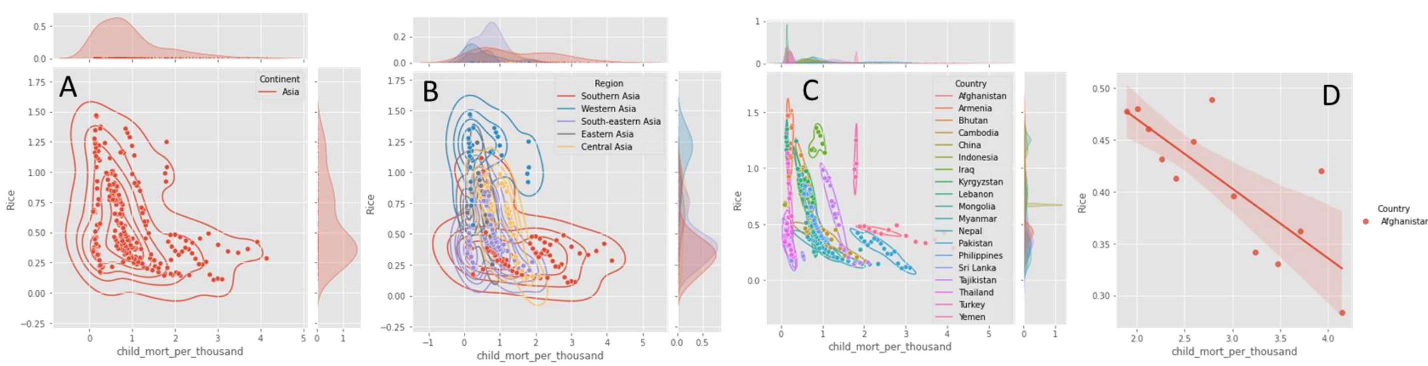

Figure 8 - Average rice price and child mortality in Asia, Asia regions and Asian countries

MULTI-VARIABLE LINEAR REGRESSION MODELLING

We performed multivariable Linear regression modeling on Birth Rate, Death Rate and Mortality at a continents 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 72.5% 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 .

OLS Regression Results

A

Dep. Variable:	birth_rate	**R-squared:**	0.725
Model:	OLS	**Adj. R-squared:**	0.722
Method:	Least Squares	**F-statistic:**	282.2
Date:	Wed, 08 Dec 2021	**Prob (F-statistic):**	2.83e-205
Time:	03:12:03	**Log-Likelihood:**	-2335.2
No. Observations:	758	**AIC:**	4686.
Df Residuals:	750	**BIC:**	4723.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	27.3039	0.940	29.037	0.000	25.458	29.150
Continent[T.Americas]	-8.5213	0.649	-13.128	0.000	-9.795	-7.247
Continent[T.Asia]	-10.1269	0.547	-18.514	0.000	-11.201	-9.053
Continent[T.Europe]	-30.9440	2.705	-11.439	0.000	-36.255	-25.633
GDP_in_USD_cap	-6.476e+05	4.54e+04	-14.257	0.000	-7.37e+05	-5.58e+05
death_rate	1.0433	0.091	11.519	0.000	0.865	1.221
child_mort	5.332e-06	1.47e-06	3.634	0.000	2.45e-06	8.21e-06
Rice	-0.2215	0.064	-3.455	0.001	-0.347	-0.096

Omnibus:	1.745	**Durbin-Watson:**	0.221
Prob(Omnibus):	0.418	**Jarque-Bera (JB):**	1.631
Skew:	0.110	**Prob(JB):**	0.442
Kurtosis:	3.054	**Cond. No.**	3.76e+10

OLS Regression Results

B

Dep. Variable:	birth_rate	**R-squared:**	0.726
Model:	OLS	**Adj. R-squared:**	0.723
Method:	Least Squares	**F-statistic:**	278.5
Date:	Wed, 08 Dec 2021	**Prob (F-statistic):**	4.88e-202
Time:	03:13:41	**Log-Likelihood:**	-2285.2
No. Observations:	744	**AIC:**	4586.
Df Residuals:	736	**BIC:**	4623.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	27.2356	0.932	29.215	0.000	25.405	29.066
Continent[T.Americas]	-8.3556	0.649	-12.869	0.000	-9.630	-7.081
Continent[T.Asia]	-9.8954	0.544	-18.174	0.000	-10.964	-8.827
Continent[T.Europe]	-31.0300	2.681	-11.575	0.000	-36.293	-25.767
GDP_in_USD_cap	-6.412e+05	4.51e+04	-14.216	0.000	-7.3e+05	-5.53e+05
death_rate	1.0535	0.090	11.730	0.000	0.877	1.230
child_mort	5.07e-06	1.46e-06	3.482	0.001	2.21e-06	7.93e-06
Rice	-0.2333	0.064	-3.647	0.000	-0.359	-0.108

Omnibus:	1.767	**Durbin-Watson:**	0.230
Prob(Omnibus):	0.413	**Jarque-Bera (JB):**	1.642
Skew:	0.111	**Prob(JB):**	0.440
Kurtosis:	3.063	**Cond. No.**	3.77e+10

Figure 9 - 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 72.6%.

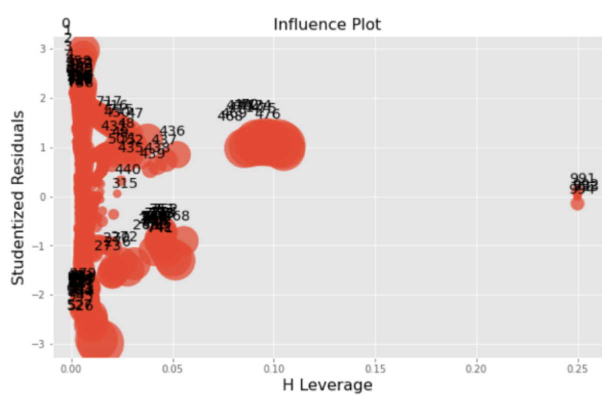

Figure 10 Influential plot

CHILD MORTALITY ANALYSIS

In the Child Mortality analysis, the Adjusted R-Squared is 15% as a result the null hypothesis is unacceptable.

OLS Regression Results

Dep. Variable:	child_mort	**R-squared:**	0.163
Model:	OLS	**Adj. R-squared:**	0.156
Method:	Least Squares	**F-statistic:**	21.31
Date:	Mon, 06 Dec 2021	**Prob (F-statistic):**	2.52e-26
Time:	04:17:23	**Log-Likelihood:**	-10187.
No. Observations:	772	**AIC:**	2.039e+04
Df Residuals:	764	**BIC:**	2.043e+04
Df Model:	7		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	-1.958e+05	3.07e+04	-6.386	0.000	-2.56e+05	-1.36e+05
Continent[T.Americas]	3.384e+04	1.75e+04	1.937	0.053	-448.449	6.81e+04
Continent[T.Asia]	8.24e+04	1.57e+04	5.247	0.000	5.16e+04	1.13e+05
Continent[T.Europe]	7565.5993	7.1e+04	0.107	0.915	-1.32e+05	1.47e+05
GDP_in_USD	10.9276	4.319	2.530	0.012	2.448	19.407
death_rate	9368.6059	2016.469	4.646	0.000	5410.128	1.33e+04
birth_rate	4812.8530	803.995	5.986	0.000	3234.551	6391.155
Rice	7318.5372	1574.478	4.648	0.000	4227.720	1.04e+04

Omnibus:	659.667	**Durbin-Watson:**	0.093
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	12562.181
Skew:	3.956	**Prob(JB):**	0.00
Kurtosis:	21.109	**Cond. No.**	1.81e+04

Figure 11 - 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.

OLS Regression Results

Dep. Variable:	death_rate	**R-squared:**	0.527
Model:	OLS	**Adj. R-squared:**	0.523
Method:	Least Squares	**F-statistic:**	121.6
Date:	Mon, 06 Dec 2021	**Prob (F-statistic):**	1.20e-119
Time:	04:17:35	**Log-Likelihood:**	-1739.8
No. Observations:	772	**AIC:**	3496.
Df Residuals:	764	**BIC:**	3533.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	5.1025	0.525	9.713	0.000	4.071	6.134
Continent[T.Americas]	-2.4887	0.296	-8.397	0.000	-3.071	-1.907
Continent[T.Asia]	-1.8045	0.275	-6.557	0.000	-2.345	-1.264
Continent[T.Europe]	8.1572	1.221	6.681	0.000	5.760	10.554
GDP_in_USD	0.0004	7.56e-05	4.746	0.000	0.000	0.001
birth_rate	0.1377	0.014	10.067	0.000	0.111	0.165
Rice	-0.0382	0.028	-1.353	0.177	-0.094	0.017
child_mort	2.933e-06	6.31e-07	4.646	0.000	1.69e-06	4.17e-06

Omnibus:	168.052	**Durbin-Watson:**	0.236
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	356.695
Skew:	1.200	**Prob(JB):**	3.51e-78
Kurtosis:	5.309	**Cond. No.**	2.36e+06

Figure 12 - Death Rate Analysis

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%.

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].