Child mortality: possible correlation with geographical, economical, religious features of the countries

Massimo Schiappa

Abstract

This analysis is intended to investigate the possible correlation between economical, geographical, religious features of the countries and the child mortality in the same countries. The child mortality is analyzed focusing on the time evolution of the phenomenon taking count that in the same time interval a better economical situation should be a indirect mitigating factor.

Motivation

The goal of this analysis is to describe the terrible phenomenon of child mortality in the world, trying to find any possible correlation with the geographical, economical and religious features of the countries. This can be useful for a better understanding of the problem and could be a starting point for a more accurate investigation on the world population behavior in relation with the place where they live and their cultural features.

Dataset(s)

The data used have been taken from different sources: the World Health Organization, the International Monetary Fund, the on line USA CIA world factbook and the GeoNames on line database. The considered time interval is from 2000 to 2015 for the Child Mortality and the economical behavior of the countries, while the geographical and the religious features of countries has been considered stable. The following links are the data source described above:
http://apps.who.int/gho/data/node.main.ghe1002015-by-cause?lang=en
http://www.imf.org/external/pubs/ft/weo/2017/01/weodata/download.aspx
https://www.cia.gov/library/publications/the-world-factbook/fields/2122.html AND WIKIPEDIA
https://developers.google.com/public-data/docs/canonical/countries_csv AND
http://download.geonames.org/export/dump/

Data Preparation and Cleaning

Each used dataset has been someway managed to be used in the analysis:

World Health Organization (WHO) data set

- the data have been downloaded in CSV format;
- the three header rows of each file have been normalized as one single header row containing all the information needed in compact format according to the following mapping schema:

Data Preparation and Cleaning - WHO

ALRI	Acute lower respiratory infections
BABT	Birth asphyxia and birth trauma
OCPNC	Other communicable, perinatal and nutritional conditions
SOICN	Sepsis and other infectious conditions of the newborn
CA	Congenital anomalies
DD	Diarrhoeal diseases
AIDS	HIV/AIDS
INJ	njuries
MAL	Malaria
MEA	Measles
MEN	Meningitis/encephalitis
OND	Other noncommunicable diseases
PER	Pertussis
PRE	Prematurity

N0-27D	from 0 to 27 days of life
N1-59M	from 1 to 59 months of life
N0-4Y	from 0 to 4 years of life (N0-27D + N1-59M)

Data Preparation and Cleaning – IMF – GEO

International Monetary Fund (IMF) data set

- the data have been downloaded in Microsoft Excel format;
- the files have been opened and saved as CSV files using a free linux software;
- some columns have been deleted so that only the columns useful for the analysis remained in the data set;

GEONAMES (GEO) data set

- the data have been downloaded in CSV format;
- the files have been loaded using the pandas python library dataframe, and each dataset have been cleaned from unwanted columns;
- all the data sets have been joined and saved in a single file using the pandas python library (dataframe merge method and to_csv method);

Data Preparation and Cleaning – REL

CIA Factbook on RELIGIONS in world countries (REL) data set

- the data have been saved from html page into a text file;
- the file have been manipulated with some linux utility (awk, grep, sed) to remove empty lines, to remove all comments on data closed between parenthesis, to remove all notes' rows marked with the beginning words 'note:', to separate each column with commas;
- the informations have been grouped (local specific religions have been put inside the generic column 'Others', all the muslims' different beliefs have been grouped into the column 'Muslim', and the same have been made for all the Christian Protestants into the column 'Protestant');
- Some approximation have been made from original data to make the percentage division coherent (the sum of all the columns contributions must be 100%)
- Some integration to the data have been made using the on line site wikipedia wherever information were missing;

Research Question(s)

The goal of this analysis is to investigate the correlation (if any correlation is there), from a population life behavior (Child mortality) and the economic, geographical and religious feature of that population: a mitigation of the problem (Child mortality) is expected to be observed for the countries getting a better economic situations in the time, while there is no specific reason to suppose that the geographical position, and the religious beliefs of populations could influence it. As an approximation, the geographical and religious data are considered a stable feature of each country, at least for the time interval 2000-2015. The wars during the considered time interval have not been considered, so some outliers is expected.

Methods

The data have been analyzed using a graphical representation by python folium library, and by pandas dataframe corr method.

Findings

As expected the gdp per capita of the countries is somehow correlated with child mortality (negative correlation whose magnitude is about 0.15).

The latitude seems not to be correlated to the child mortality and the longitude of countries seems to be someway positively correlated to the child mortality (correlation value magnitude is about 0.10).

The religious beliefs seems to be correlated to child mortality: for one religion thre is a positive correlation whose magnitude is about 0.40, and for other two religions there are correlation magnitude equal about to 0.11, one of these is a positive correlation, the other is a negative correlation.

Limitations

This analysis doesn’t consider relevant factors (wars, natural disasters, …) which can highly affect the result of correlation calculation.

For the religious beliefs, the spread of a religion could be a strong reason for a wrong correlation calculation.

Conclusions

This analysis has the goal of showing some possible way to study the data of countries and populations to find some unexpected correlation among quite different aspects.

Acknowledgements

The data used have been taken from different sources:
- the World Health Organization;
- the International Monetary Fund;
- the on line USA CIA world factbook and the GeoNames on line database;
- Wikipedia.

References

In this work, I followed a my original idea, without referring to other works.