Assignment

Computational Methods and Principles of Bayesian Inference - DATA.STAT.430-2021-2022-1-TAU

Duy Vu – H292118

1. Dataset

The dataset used for this assignment/coursework is the Maailmanpankin väestötilastoja (tentatively translated in English as World Bank population statistics) 2016 - 2018 from the [Finnish Social Science Data Archive (FSD)](https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvanti/spss/aineistot/). The dataset contains basic information about the countries of the world, other territories and groups formed by countries. In total, there are 30 variables, most of which are from World Bank indicators. Some example variables are birth rate, morality, adult literacy rate by male and female, expenditure on education and military, and GDP per capita.

In this assignment, only two variables are taken into consideration: infant mortality rate (per 1000 live births) and life expectancy (at birth). By definition from the data source, infant mortality rate is the number of infants dying before reaching one year of age, per 1000 live births, and life expectancy is the number of years a new-born infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life. Also, the data used in this assignment is taken specifically from the year 2017.

According to The World Bank, these two variables, along with other variables like mortality rate are among the most important indicators of health status of a country, and high mortality in young age groups of a country tends to significantly lower the life expectancy of that country. Hence, a causal relationship may exist between these two variables.

Since the data has not yet been processed, before analysing the data, some pre-processing steps are needed to be carried out. Since the dataset also consisted of region and groups of countries which makes values of variables in those rows aggregate, those data are not used.

Since there are some countries don’t have data for this attributions, we have to also remove data of this countries from the analysis.

2. Statistical basis of the method applied to the selected data

3. Results