Effect of suicide rates on life expectancy

Filip Zoubek (e11938258@student.tuwien.ac.at)

Abstract

In 2015, approximately 55 million people died worldwide, of which 8 million committed suicide. In the USA, one of the main causes of death is the aforementioned suicide, therefore, this experiment is dealing with the question of how much suicide rates affects the statistics of average life expectancy.

The experiment takes two datasets, one with the number of suicides and life expectancy in the second one and combine data into one dataset. Subsequently, I try to find any patterns and correlations among the variables and perform statistical test using simple regression to confirm my assumptions.

Data

The experiment uses two datasets - WHO Suicide Statistics[1] and WHO Life Expectancy[2], which were firstly appropriately preprocessed. I have started with the dataset Life Expectancy, where I selected the variables Country, Year, Life expectancy, Adult Mortality, Infant deaths, Alcohol, Under-five deaths, HIV/AIDS, GDP, Population, Income composition of resources and Schooling. Then, I selected the following variables country, year, Suicides number and population from the second dataset Suicide Statistics and I had to also group the data by year and country and drop all rows, where the values in the column suicides number were missing. After first preprocessing I have merged these two datasets by index – country and year. Then I have cleaned missing data, I unified population variables, for life expectancy I deleted all rows with missing values and for others I added values using the function fill, which fill the values according to the previous row.

The final input dataset to the experiment has 13 variables, where country and year are used as index: **Country**, **Year**, **Suicides number**, **Life expectancy**, **Adult Mortality**, which is probability of dying between 15 and 60 years per 1000 population, **Infant deaths**, which is number of Infant Deaths per 1000 population, **Alcohol**, which is alcohol, recorded per capita (15+) consumption, **Under-five deaths**, which is number of under-five deaths per 1000 population, **HIV/AIDS**, which is deaths per 1 000 live births HIV/AIDS, **GDP**, which is Gross Domestic Product per capita, **Population**, **Income composition of resources,** which is Human Development Index in terms of income composition of resources, and **Schooling,** which is number of years of schooling.

Experiment

Before I started the experiment, I created a second dataset, which was appropriately scaled by the min-max normalization. Firstly, I displayed a correlation according to which I then filtered more interested variables into the issue: Suicides number, Life expectancy, Population, Income composition of resources, Schooling, and correlated a second time. You can see the result in figure number 1. After correlation, I also displayed a pairplot of filtered variables and looked for the patterns. However, at the first sight, it seemed that there was no effect suicide rate on life expectancy, because I did not find any pattern and the correlation was very weak.

[1] https://www.kaggle.com/szamil/who-suicide-statistics

[2] https://www.kaggle.com/kumarajarshi/life-expectancy-who

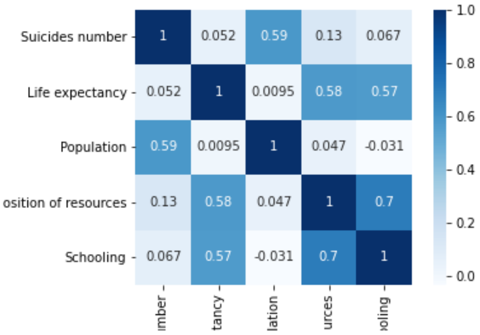
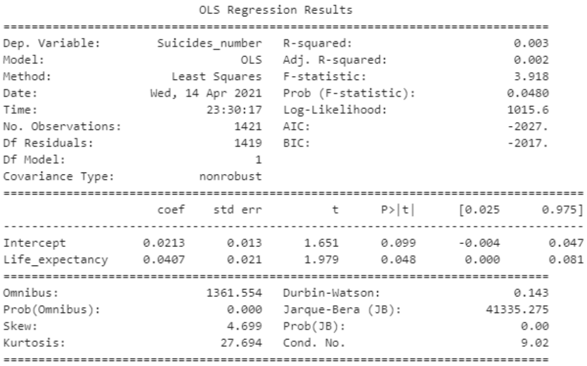


Figure n. 1: Correlation among the filtered variables.



```
                            OLS Regression Results
==============================================================================
Dep. Variable:         Suicides_number   R-squared:                       0.003
Model:                             OLS   Adj. R-squared:                  0.002
Method:                  Least Squares   F-statistic:                     3.918
Date:                 Wed, 14 Apr 2021   Prob (F-statistic):             0.0480
Time:                         23:30:17   Log-Likelihood:                 1015.6
No. Observations:                 1421   AIC:                            -2027.
Df Residuals:                     1419   BIC:                            -2017.
Df Model:                            1
Covariance Type:             nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
Intercept           0.0213      0.013      1.651      0.099      -0.004       0.047
Life_expectancy     0.0407      0.021      1.979      0.048       0.000       0.081
==============================================================================
Omnibus:                      1361.554   Durbin-Watson:                   0.143
Prob(Omnibus):                   0.000   Jarque-Bera (JB):            41335.275
Skew:                            4.699   Prob(JB):                         0.00
Kurtosis:                       27.694   Cond. No.                         9.02
==============================================================================
```

Figure n.2: Results of simple regression.

Therefore, I decided to use simple regression to prove the hypothesis, because the predicate and outcome variable are quantitative. From the results, figure 2, we can see that the difference is statistically significant because the p-value is smaller than the selected significance value of 0.05. Therefore, the null hypothesis was rejected, and I can say that there is effect of suicide rates on life expectancy.

Figure 3 also shows a diagram that visually explains the experiment.

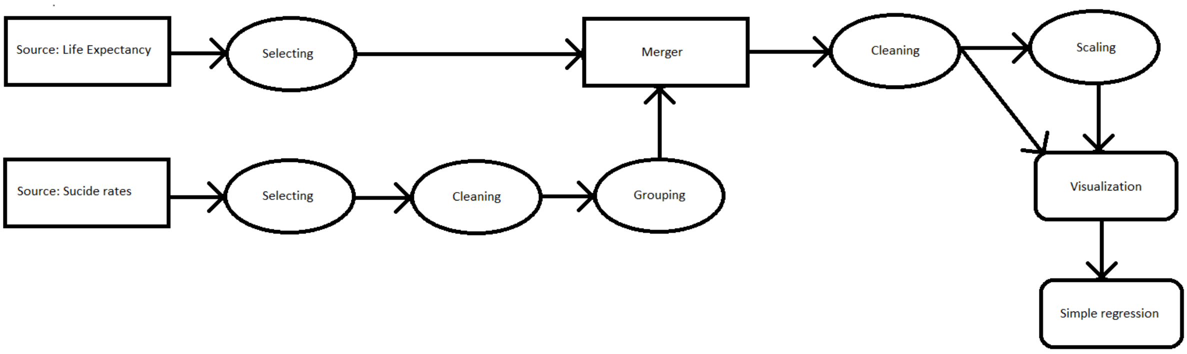


Figure n.3: Diagram explaining the experiment.

Both correlated figures, as well as a pair plot and OLS output, can be found in the project output folder. The experiment was performed on the laptop with parameters: Intel Core I5 9th Gen, 16 GB RAM memory, GTX1650, and used Jupyter Notebook with Python 3.7.9. The source code can be found on the GitHub[3].

Summary

The task of this experiment was to prove the possible effect of suicide rates on life expectancy. Although at first it seemed that there was no effect between these variables, in the end, after a simple regression, it turned out that suicide rates have the effect on life expectancy.

[3] https://github.com/e11938258/effect-of-suicide-rates-on-life-expectancy