INFSCI 2725 Data Analytics

Assignment 5 – Probabilistic Approaches

Date: 11/07/2015

Student name: Ting LI, Hongyuan Cui

pittUserLogin: til42, hoc27

Tutor name: Marek Druzdzel

Executive Summary

Referring to the analysis of Franz H. Messerli, there is a correlation between countries's annual per capita chocolate consumption and the number of Nobel laureates per 10 million populations. This report first this analysis, followed by the analysis of relationship between Nobel laureates with other variables including chocolate consumption, alcohol consumption, meat consumption, tea consumption and computer consumption. The second one is about the relationship between coffee consumption and Nobel laureates. It turned out except for coffee consumption, the correlation coefficients of all the variables with Nobel laureates are all less than 0.5, resulting in our neglecting of these variables considering their relationship with Nobel. As for coffee consumption, this analysis is based on our observed phenomenon that coffee can help people stay clear-minded which in turn can contribute to improving the efficiency of work. Coffee consumption per capita per year for 34 countries was used. This result demonstrates a comparatively strong correlation between coffee consumption and Nobel Laureates. However, as is the case of chocolate consumption, we cannot arrive at their causal relationship. Matlab is used in the analysis and all the code is attached in the appendix.

1. Replicate the Analysis

Dr. Messerli did his analysis utilizing chocolate consumption data and Nobel laureate data of 23 countries. We collect Nobel laureate data from Wikipedia[1] and chocolate consumption data from various sources[2]. Then we use matlab to plot their correlation. Figure 1 is the plot and there is a significant linear correlation between two variables (r= 0.7698 p= 0.0000174<0.0001).

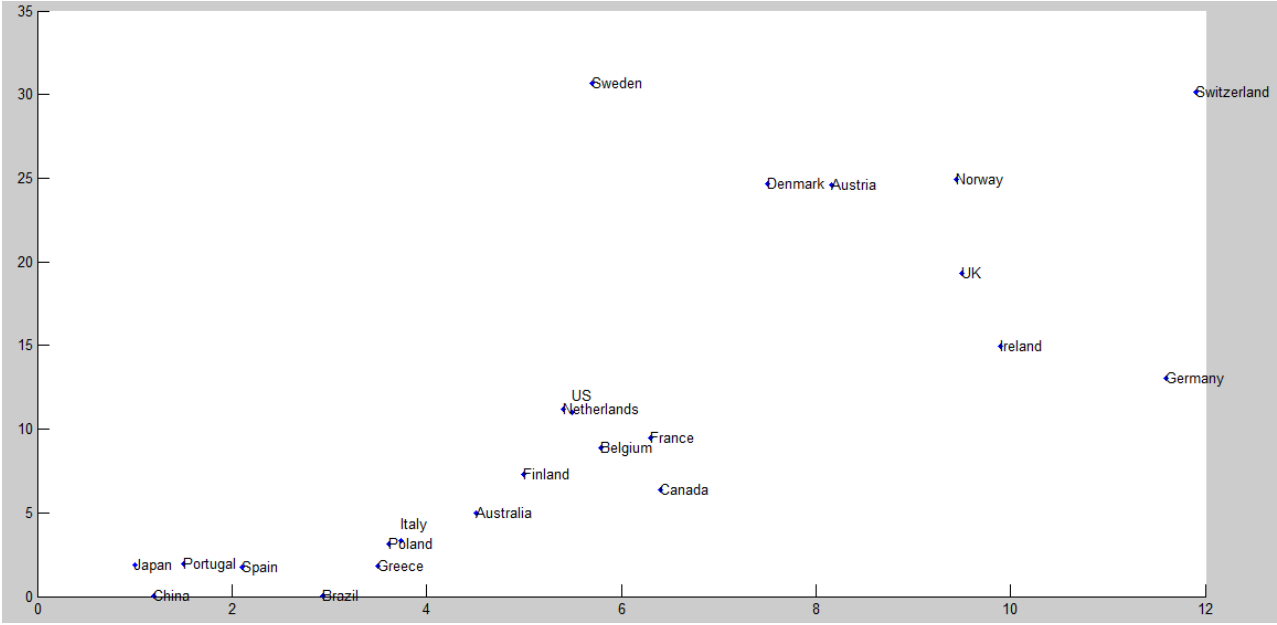

Figure 1. Correlation between Nobel laureate and Chocolate Consumption

2. Analysis of Observed Phenomenon

Coffee is a helpful beverage for people to refresh themselves whenever they feel tired. Because coffee contains caffeine which can excite nerve system and keep person concentrated on their work. Thus we can assume that coffee consumption can improve research efficiency both for individuals and whole populations. We use the number of Nobel laureate as an estimate reflecting a country's research ability. Based on these observed phenomenon, we collect data about coffee consumption per capita per year of 34 countries and conduct analysis between coffee consumption and Nobel

[1] http://en.wikipedia.org/wiki/List_of_countries_by_Nobel_laureates_per_capita

[2] http://www.helgilibrary.com/indicators/index/consumption-of-chocolate-products-per-capita/greece, http://www.statista.com/statistics/262981/amount-of-chocolate-consumed-per-person-by-country/

laureate. Data about Nobel laureate is from 2015 on[3] data about coffee consumption is from 2013 on[4].Figure 2 is a plot about correlation between Nobel laureate and coffee consumption and the linear correlation is significant(r=0.7199 p= 0.0000015919<0.0001).

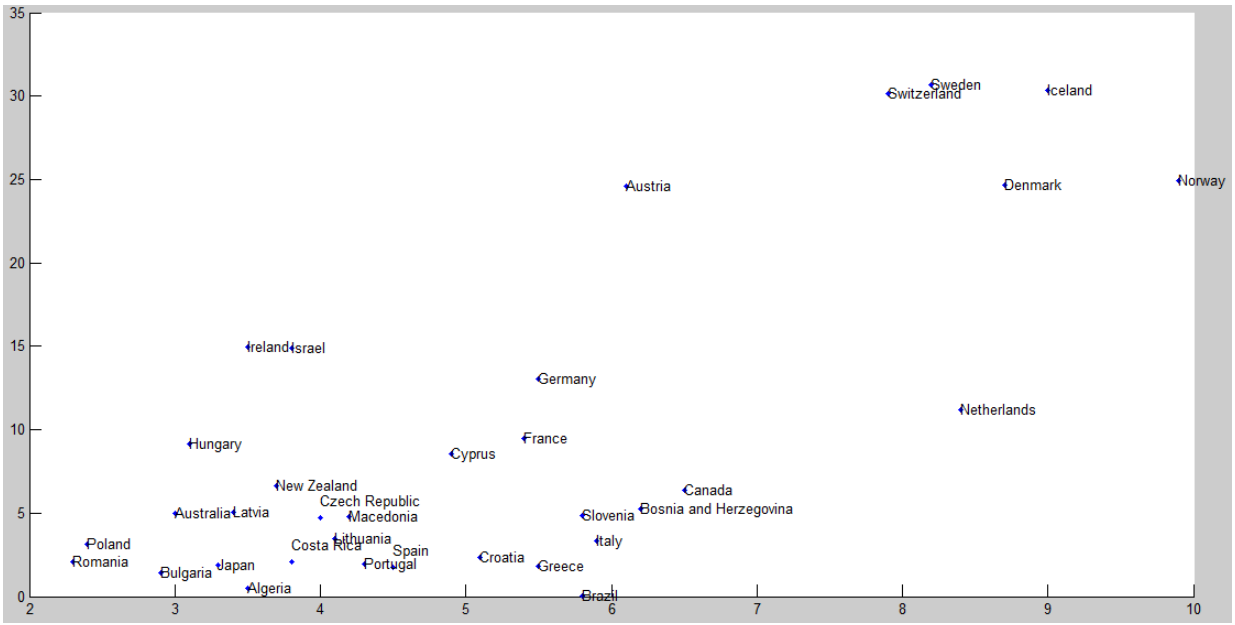

Figure 2. Correlation between Coffee Consumption and Nobel Laureate

Conclusion

The first analysis's outcome is not quite same as that of Dr. Messerli, because the data we collect is not exactly from the same year as Dr. Messerli's and data online varies from websites. However, the r value (correlation coefficient) is almost the same. In the second analysis, we can see the significant linear correlation. We can conclude that coffee consumption is correlated with Nobel laureate. However, we cannot determine whether coffee contribute to Nobel prize.

[3] https://en.wikipedia.org/wiki/List_of_countries_by_Nobel_laureates_per_capita

[4] https://www.google.com/fusiontables/DataSource?docid=1C fn6nSe21acP0xJIO1T1x0wohqfMYCQyJjbqdk#rows:id=1

Appendix

```
clc,clear
load('coffee.mat')
load('raw2010.mat')
y_name=Name_Nobel;
y=Nobel;
y_name(25)=[];
y(25)=[];
x_name=Coffee_Name;
x=Coffee;
cept=intersect(y_name,x_name);
for i=1:length(y_name)
    if(~ismember(y_name(i),cept))
        y_name(i)={''};
        y(i)=nan;
    end
end
for i=1:length(x_name)
    if(~ismember(x_name(i),cept))
        x_name(i)={''};
        x(i)=nan;
    end
end

y_name(isnan(y))=[];
y(isnan(y))=[];
x_name(isnan(x))=[];
x(isnan(x))=[];

[y_name,index] = sort(y_name);
y=y(index);
[x_name,index] = sort(x_name);
x=x(index);
scatter(x,y,'.')
[r,p]=corr([x,y])
y(2)=y(2)+1;
y(13)=y(13)+1;
y(10)=y(10)+0.7;
y(34)=y(34)+1;
y(20)=y(20)+1;
% y(13)=y(13)+1;
% y(23)=y(23)+1;
text(x,y,x_name)
```