A study about different factors influencing life expectancy in Brazil

Student: Wen Han 7879607
Course: STAT 3380, Winter 2021
Instructor: Dr. Zenaida Mateo

Short abstract

Everyone cares about how to live long. From long time ago, longevity has been a classic and enduring topic in research studies. There are many complex factors that could affect living long: genetics, gender, income, health care, personal living habits, social situation, community factors and so on. Some of these factors may be positive for longevity, some of them may be negative.

This study aims to find the relationship between two factors and life expectancy in Brazil, to discuss whether they are doing a positive effect or a negative effect. These two factors are alcohol consumption and level of education (Schooling) in Brazil.

Introduction

There have been many studies in the past about factors affecting life expectancy, this study referenced other previous studies. In this study line chart graph is used to determine the trends of life expectancy changes in the past 10 years. And, to determine, whether drinking alcohol share a positive or negative relationship with life expectancy. Also, if there is any correlation between level of education and life expectancy. In this dataset, life expectancy is predicted by years, alcohol consumption is recorded per capita (over 15 years old) in litres of pure alcohol. The number of years in school, starting at primary school will be referred to level of education.

There are always traditional sayings that drinking alcohol is harmful to health and life expectancy. However, some controversies against it throughout the years, there are other sayings indicate that moderate alcohol consumption contribute to health. In this case, alcohol consumption is doing a positive effect to life expectancy or negative effect, needs to be clarified.

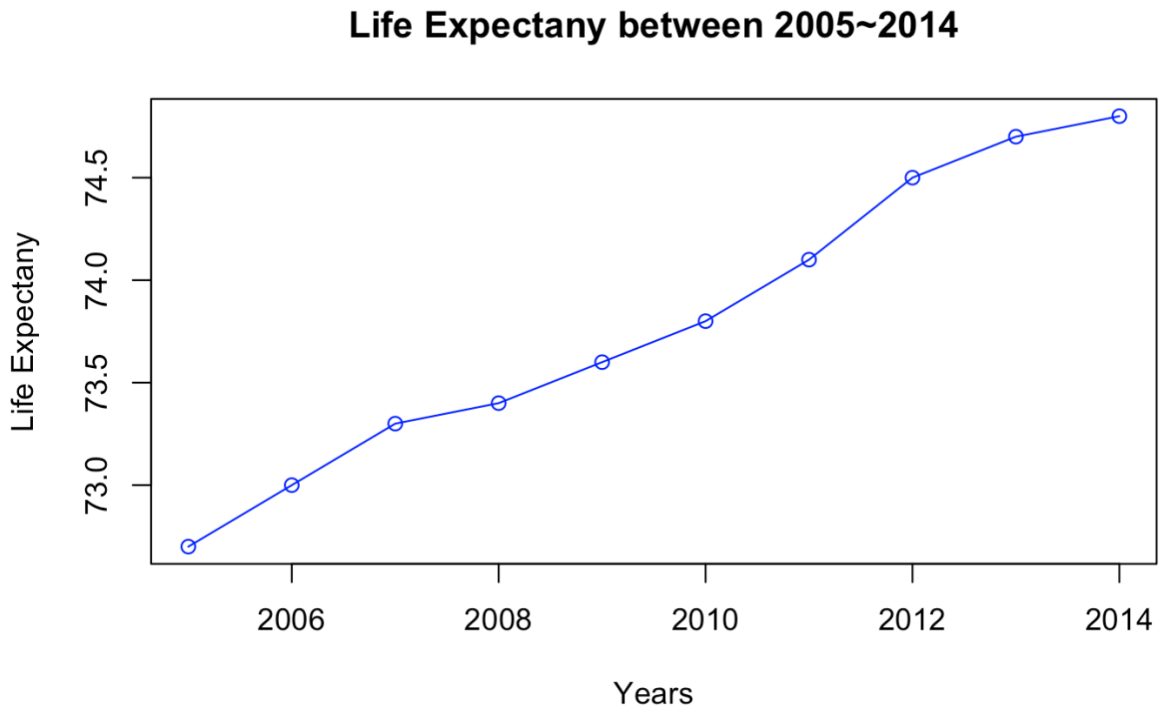
It is known that people with higher level education will gain more awareness about keeping healthy. However, the conclusion has to be based on statistics result. A positive correlation between schooling and life expectancy is assumed, with alternative hypothesis that they are actually negative correlation.

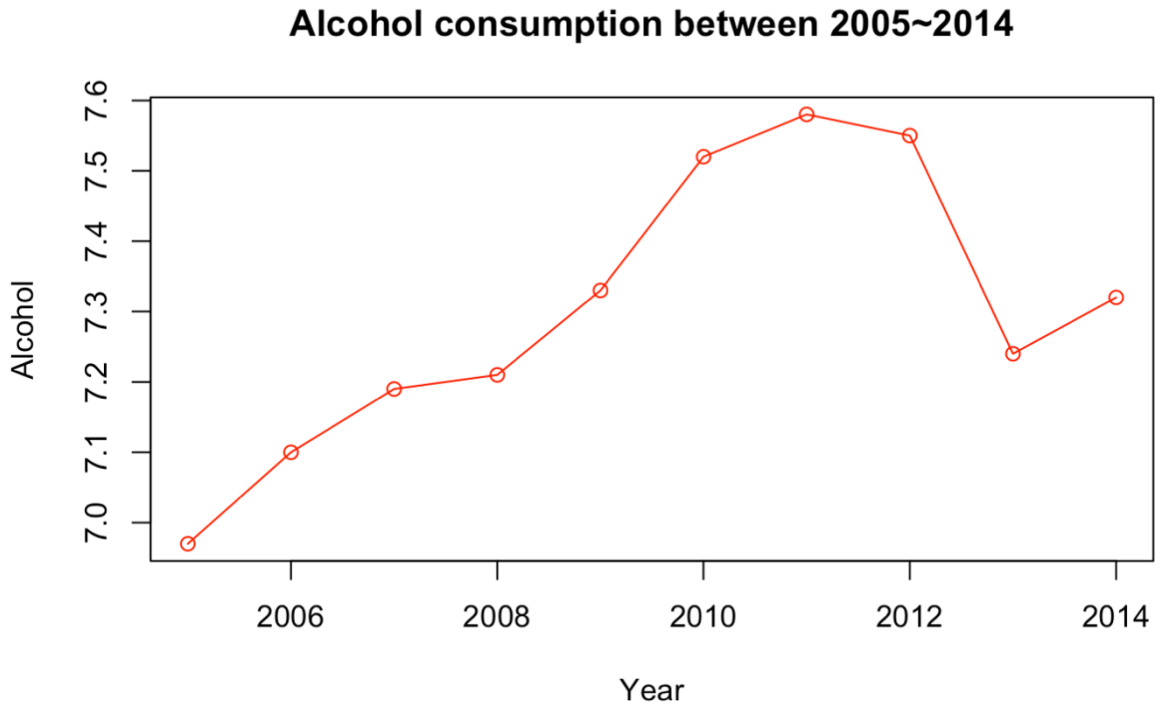
Methodology

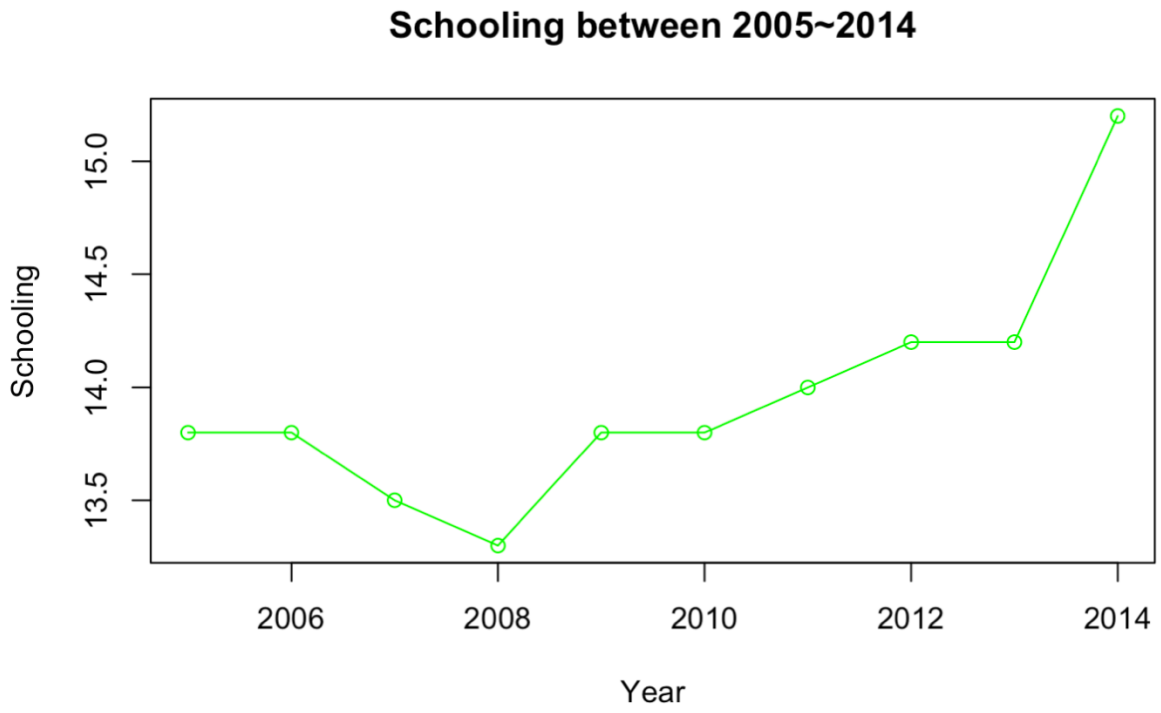
Data used in this study is obtained from open-source website, the original data was collected from World Health Organization and United Nations website. A sample of 10 years Life Expectancy from the year of 2005 to the year of 2014 in Brazil, the alcohol consumption, the number of years in school and total expenditure was selected from the data. Selected data sets can be found on table 1 of appendix at the end of this paper.

R studio is used in this study to draw graph and calculation: line charts display the trends of life expectancy changes in these 10 years, Spearman rank test is conducted to analysis the correlation between Life Expectancy and alcohol consumption, also the correlation between life expectancy and level of schooling. The assumption for Spearman rank test is that the data of the samples should be random or non-numeric observations, also each pair of samples represent two measurements taken on the same subject or individual.

Discussion of Results


Life Expectany between 2005~2014
Life Expectany
73.0
73.5
74.0
74.5
2006
2008
2010
2012
2014
Years


Alcohol consumption between 2005~2014
Alcohol
7.0
7.1
7.2
7.3
7.4
7.5
7.6
2006
2008
2010
2012
2014
Year



According to the line chart, the life expectancy between 2005 and 2014 was increasing slowly on the yearly basis. And from the second line chart, the alcohol consumption is significantly increased, but decreased in the year of 2013, then it back to increasing. For the Schooling, there is a small increasing in these 10 years, except from 2006 to 2008 they were decreasing.

```
> cor(LifeExpectancy,Alcohol, method ="spearman")
[1] 0.6969697
> cor.test(LifeExpectancy,Alcohol,method="spearman",exact = FALSE)

        Spearman's rank correlation rho

data:  LifeExpectancy and Alcohol
S = 50, p-value = 0.0251
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.6969697
```

For the Spearman rank test, assume one null hypothesis that Life Expectancy and Alcohol are independent. Alternative hypothesis is that Life Expectancy and Alcohol are related at 5% level of significance.

From the computational test analysis, there is a correlation coefficient of 0.6969697, which means variables sharing a positive relationship. With a small p-value of $0.0251 < \alpha = 0.05$, there is sufficient evidence to reject null hypothesis, comply with alternative hypothesis, therefore it can be concluded that there is direct positive relationship between Life Expectancy and Alcohol.

```
> cor(LifeExpectancy,Schooling, method ="spearman")
[1] 0.8029854
> cor.test(LifeExpectancy,Schooling,method="spearman",exact = FALSE)

        Spearman's rank correlation rho

data:  LifeExpectancy and Schooling
S = 32.507, p-value = 0.005157
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.8029854
```

For the Spearman rank test, assume null hypothesis that Life Expectancy and Schooling are independent. Alternative hypothesis is that Life Expectancy and Schooling are related at 5% level of significance.

From the computational test analysis, there is a correlation coefficient of 0.8029854, which means variables also share a positive relationship. With an extremely small p-value of $0.005157 < \alpha = 0.05$, there is sufficient evidence to reject null hypothesis, agree with alternative hypothesis, and can be concluded that there is direct positive relationship between Life Expectancy and Schooling.

Summary of Results

According to the result from R studio, Spearman rank test shows there is a positive relationship between life expectancy and alcohol consumption. In the range of this study data, the more litres of pure alcohol per capita consumed, the longer life expectancy it gets.
There is a positive relationship between life expectancy and Schooling. In the range of this study data, when there are more schooling years people performed, the longer life expectancy they can get.

Challenges/Future Plan

The most challenge part is there are only 10 sets of samples in this study, which is too small to conduct a very solid conclusion. There are other factors that could affect life expectancy are not included, due to the length limit of this study. The future plan is to search for more specific data from 2004 to 2015, for example taking large sample size, and looking for other factors aside of factors include in this study, that could affect life expectancy.

Reference:

1. Catharine Paddock, April 30, 2018, *These five habits will lengthen your lifespan*. https://www.medicalnewstoday.com/articles/321671

2. Elizabeth A. Dobis, Heather M. Stephens, Mark Skidmore, Stephan J. Goetz. 2020. *Explaining the spatial variation in American life expectancy*, Social Science & Medicine.

3. Producing Simple Graphs with R, Frank McCown, 2006 https://sites.harding.edu/fmccown/r/

4. Wayne W. Daniel, June 30, 2000*, Applied Nonparametric Statistics, second edition*, Duxbury Press

5. Siegel, S. and Castellan, N. J. Jr. 1988. *Nonpararmetric Statistics for the Behavioral Sciences,* 2nd ed

Appendix

Year	Life expectancy	Alcohol	Schooling
2014	74.8	7.32	15.2
2013	74.7	7.24	14.2
2012	74.5	7.55	14.2
2011	74.1	7.58	14
2010	73.8	7.52	13.8
2009	73.6	7.33	13.8
2008	73.4	7.21	13.3
2007	73.3	7.19	13.5
2006	73	7.1	13.8
2005	72.7	6.97	13.8

Table 1. Life expectancy, alcohol, schooling in 10 years (2005-2014) data in Brazil.

```r
#Brazil
Year=c(2005,2006,2007,2008,2009,2010,2011,2012,2013,2014)
LifeExpectancy=c(72.7, 73, 73.3, 73.4, 73.6, 73.8, 74.1,74.5, 74.7, 74.8)
Alcohol=c(6.97, 7.1, 7.19, 7.21,7.33, 7.52, 7.58,7.55,7.24, 7.32)

plot(Year, LifeExpectancy, type="o", col="blue", xlab="Years",
     ylab="Life Expectany",
     main="Life Expectany between 2005~2014")
plot(Year, Alcohol, type="o", col="red", xlab="Year", ylab="Alcohol",
     main="Alcohol consumption between 2005~2014")
plot(Year, Schooling, type="o", col="green", xlab="Year", ylab="Schooling",
     main="Schooling between 2005~2014")
```

R code. Draw 3 different line chart graphs.

```
> cor(LifeExpectancy,Alcohol, method ="spearman")
[1] 0.6969697
> cor.test(LifeExpectancy,Alcohol,method="spearman",exact = FALSE)

        Spearman's rank correlation rho

data:  LifeExpectancy and Alcohol
S = 50, p-value = 0.0251
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.6969697
```

R code. Spearman rank test to analysis the correlation between Life Expectancy and alcohol consumption.

```
> cor(LifeExpectancy,Schooling, method ="spearman")
[1] 0.8029854
> cor.test(LifeExpectancy,Schooling,method="spearman",exact = FALSE)

        Spearman's rank correlation rho

data:  LifeExpectancy and Schooling
S = 32.507, p-value = 0.005157
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.8029854
```

R code. Spearman rank test to analysis the correlation between Life Expectancy and Schooling.