Faculty of Computing and Information Technology (FOCS)

Academic Year 2020/2021

BAMS3043

Mathematical and Statistical Software

Programme : RMM3S1

Tutorial Group : Group 1

Date Submitted to Tutor : 19/9/2021

Team Members:

No	Student Name	Student ID
1.	Choy Mei Xin	19WMR01017
2.	Kenneth Koay Ting Foong	19WMR03488
3.	Tan Yin Yuen	19WMR05960

1. According to [1], changes in life expectancy are affected by changes in the educational structure. The people with the highest life expectancies are those who had more than high school education, followed by those with medium and low education attainment levels. Education helps individuals to develop health-related resources as described in fundamental social causes of health inequalities theory. People with higher education are willing to use their own money, knowledge prestige, power and beneficial social connections to gain access to life-saving innovations while a new health technology is developed.

   In addition, using Exploratory Data Analysis on all the Independent Variables, the correlation coefficient of $r = 0.727630$ using Pearson correlation indicates that there is a high degree of positive linear correlation between schooling and life expectancy as shown below

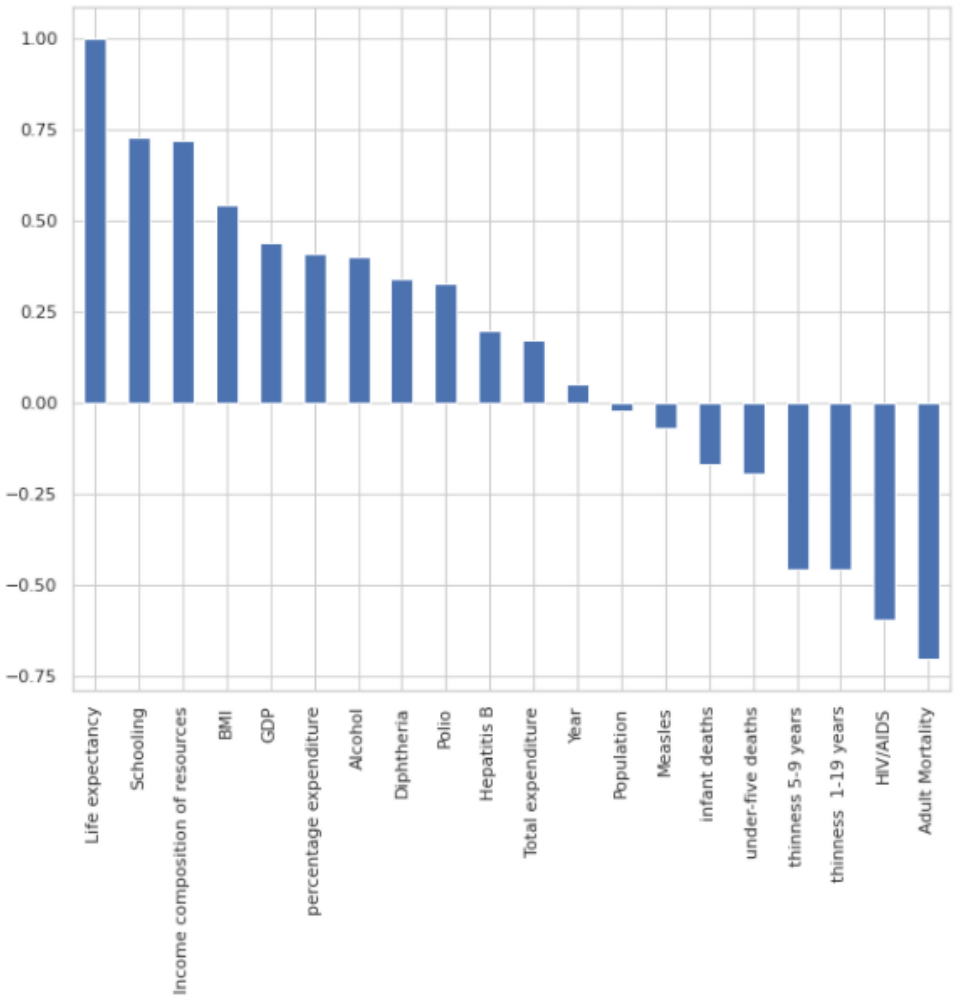

Figure 1: The Pearson correlation between the independent variables and life expectancy.

We let x_1 be the independent variable for Schooling and Life expectancy be the dependent variable y. After calculating using Python, we obtain the following simple linear regression model

$$y = 41.5503417453634 + 2.28978652x_1 \tag{1.1}$$

2. According to [2], life expectancy is affected by many factors. These factors can be categorised into socioeconomic status, the quality of the health system which includes the accessibility of health care, health behaviours, social factors, genetic factors and environmental factors. In [3], the experiment compared morbidity and mortality before and after widespread implementation of national vaccine recommendations for 13 vaccine-preventable diseases which are diphtheria, pertussis, tetanus, poliomyelitis, measles, mumps, rubella (including congenital rubella syndrome), invasive Haemophilus Influenzae Type B (Hib), acute hepatitis B, hepatitis A, varicella, Streptococcus Pneumoniae, and smallpox. It is concluded that the number of mortality cases of most vaccine-preventable diseases is at an all-time low, and has shown striking decreases in hospitalizations and deaths. In [4], the vaccine can prevent diseases beyond the specific infection they are designed to target and resulting in the reduction of secondary infections that complicate vaccine-Preventable Disease. For example, referring to HIV-negative women who are pregnant, their infants were protected against hospitalization with a vaccine efficacy of all-cause lower respiratory tract infections. Public health infrastructure and personnel are promoting these important messages and health education relating to preventable diseases such as malaria and HIV infection. We believe that people with higher education levels could understand the importance of the vaccine which led them to a higher life expectancy.

   We let x_2 be the independent variable for Hepatitis B, x_3 be the independent variable for Diphtheria, x_4 be the independent variable for Polio, x_5 be the independent variable for measles, and x_6 be the independent variable for HIV/AIDS, which are vaccine-preventable diseases among 0-4 years old, and Schooling as the Independent Variable x_1. After calculating using Python, our first multiple linear regression model as shown below

$$y = 44.87578424725493 + 1.88465451x_1 - 1.15627468(10)^{-2}x_2 + 3.01787360(10)^{-2}x_3 + 1.52802192(10)^{-2}x_4 + 1.02609345(10)^{-6}x_5 - 6.64448644(10)^{-1}x_6 \quad (1.2)$$

According to [6], health expenditure plays a role in life expectancy. In the year 2008, it is found that countries that are spending more on health had higher life expectancies. Besides, countries with a rapidly increasing health expenditure experience a steep increase in life expectancy too [6]. However, that is not always the case.

From a Clinical Perspective stated in [5], individuals could prolong their life expectancy by adherence to five low-risk lifestyle-related factors which are:

1. never smoke
2. maintain a healthy weight
3. engage in regular physical activity
4. maintain a healthy diet
5. adopt a moderate alcohol consumption

A healthy lifestyle can result in substantially reduced premature mortality and prolonged life expectancy. For example, although the United States was ranked first in the world of spending a total health expenditure per capita as a percentage of gross domestic product (17.7%), their healthcare system is focused on disease treatment rather than prevention. Their most commonly treated chronic diseases such as cardiovascular disease (CVD) are the costliest of all health problems but are preventable by maintaining a healthy lifestyle. Hence, unhealthy lifestyles are major risk factors for various chronic diseases and premature death. In [4], it reaffirms the health, economic and social benefits that vaccines programs bring to individuals and society. We think that education, vaccination, healthy lifestyle and the economy are factors that are affecting the life expectancy in this society.

Therefore, our second multiple linear regression model is built with the addition of BMI, Alcohol and Percentage Expenditure as our Independent Variables to our first multiple linear regression model.

We let x_7 be the independent variable for BMI, x_8 be the independent variable for Alcohol and x_9 be the independent variable for Percentage Expenditure. After calculating using Python, our second multiple linear regression model as shown below

$$\begin{aligned} y = {} & 45.9588556910422 + 1.55388746x_1 - 9.66648157(10)^{-3}x_2 \\ & + 3.14583993(10)^{-2}x_3 + 1.61839315(10)^{-2}x_4 + 1.60598438(10)^{-5}x_5 \\ & - 6.34128564(10)^{-1}x_6 + 6.50781541(10)^{-2}x_7 - 9.52277509(10)^{-2}x_8 \\ & + 6.42534528(10)^{-4}x_9 \end{aligned} \tag{1.3}$$

3. We will conduct Hypothesis Testing to determine which of the three model is the best model. We will first compare the second multiple linear regression model and the first multiple linear regression mode. We want to test the hypothesis is that BMI, Alcohol and Percentage Expenditure does not affect Life Expectancy at $\alpha = 0.05$.

   Hypothesis

$$\begin{aligned} & H_0 : \beta_7 = \beta_8 = \beta_9 = 0 \\ & H_1 : \text{Not all } \beta_i = 0 \text{ for } i = 7,8,9 \end{aligned} \tag{1.4}$$

   By using R programming, the p-value obtained is $< 2.2(10)^{-16}$. The value of $< 2.2(10)^{-16}$ in RStudio indicates that it is a value close to 0. Since p-value$= 2.2(10)^{-16}$ (near 0) < 0.05, we reject H_0 at $\alpha = 0.05$ and conclude that BMI, Alcohol and Percentage Expenditure cannot be removed together.

We will now compare the second multiple linear regression model and the simple linear regression model. We want to test the hypothesis is that Hepatitis B, Diphtheria, Polio, measles, HIV/AIDS, BMI, Alcohol and Percentage Expenditure does not affect Life Expectancy at $\alpha = 0.05$.

Hypothesis

$$H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$$
$$H_1 : \text{Not all } \beta_i = 0 \text{ for } i = 2,3,4,5,6,7,8,9 \tag{1.5}$$

By using R programming, the p-value obtained is $< 2.2(10)^{-16}$. The value of $< 2.2(10)^{-16}$ in RStudio indicates that it is a value close to 0. Since p-value $= 2.2(10)^{-16}$ (near 0) < 0.05, we reject H_0 at $\alpha = 0.05$ and conclude that Hepatitis B, Diphtheria, Polio, measles, HIV/AIDS, BMI, Alcohol and Percentage Expenditure cannot be removed together.

We will now compare the first multiple linear regression model and the simple linear regression model. We want to test the hypothesis is that Hepatitis B, Diphtheria, Polio, measles, HIV/AIDS does not affect Life Expectancy at $\alpha = 0.05$.

Hypothesis

$$H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$$
$$H_1 : \text{Not all } \beta_i = 0 \text{ for } i = 2,3,4,5,6 \tag{1.6}$$

By using R programming, the p-value obtained is $< 2.2(10)^{-16}$. The value of $< 2.2(10)^{-16}$ in RStudio indicates that it is a value close to 0. Since p-value $= 2.2(10)^{-16}$ (near 0) < 0.05, we reject H_0 at $\alpha = 0.05$ and conclude that Hepatitis B, Diphtheria, Polio, measles, HIV/AIDS, BMI, Alcohol and Percentage Expenditure cannot be removed together.

Therefore, the best model among the three models after doing hypothesis testing is the second multiple linear regression model shown below

$$\begin{aligned} y = {} & 45.9588556910422 + 1.55388746x_1 - 9.66648157(10)^{-3}x_2 \\ & + 3.14583993(10)^{-2}x_3 + 1.61839315(10)^{-2}x_4 + 1.60598438(10)^{-5}x_5 \\ & - 6.34128564(10)^{-1}x_6 + 6.50781541(10)^{-2}x_7 - 9.52277509(10)^{-2}x_8 \\ & + 6.42534528(10)^{-4}x_9 \end{aligned} \tag{1.7}$$

This is because Hepatitis B, Diphtheria, Polio, measles, HIV/AIDS, BMI, Alcohol and Percentage Expenditure cannot be removed when comparing between one model and another.

4. We will now construct a 95% Prediction Interval for Life Expectancy using a suitable set of reasonable x values. The suitable set of reasonable x values are the average values of each independent variable. Using Python, we obtain the following average for each of the independent variable of x_i

$$\begin{aligned} x_1(Schooling) &= 12.119891 \\ x_2(Hepatitis\ B) &= 79.217708 \\ x_3(Diphtheria) &= 84.155246 \\ x_4(Polio) &= 83.564585 \\ x_5(Measles) &= 2224.494239 \\ x_6(HIV\ /\ AIDS) &= 1.983869 \\ x_7(BMI) &= 38.128623 \\ x_8(Alcohol) &= 4.533196 \\ x_9(percentage\ expenditure) &= 698.973558 \end{aligned} \tag{1.8}$$

By using RStudio, the 95% Prediction Interval for Life Expectancy is as follow

$$60.89602 \leq y^* \leq 77.70859 \tag{1.9}$$

References

1. Luy, M, Zannella, M, Wegner-Siegmundt, C, Minagawa, Y, Lutz, W & Graziella Caselli, G 2019, 'The impact of increasing education levels on rising life expectancy: a decomposition analysis for Italy, Denmark, and the USA', Genus, vol 75, no. 11, pp. 1-21.

2. Australia Government Department of Health 2012, '*Tier 1—Life expectancy and wellbeing—1.19 Life expectancy at birth*', viewed 15 September 2021, <https://www1.health.gov.au/internet/publications/publishing.nsf/Content/oatsih-hpf-2012-toc~tier1~life-exp-wellb~119#:~:text=Life%20expectancy%20is%20affected%20by,nutrition%20and%20lack%20of%20exercise%3B>.

3. Roush, SW & Murphy 2007, 'Historical Comparisons of Morbidity and Mortality for Vaccine-Preventable Diseases in the United States', vol. 298, no. 18, pp. 2155-2153.

4. Rodrigues, C & Plotkin SA 2020, 'Impact of Vaccines; Health, Economic and Social Perspectives', *Frontiers in Microbiology*, vol. 11, p.1526.

5. Li, Y, Pan, A, Wang, DD, Liu, X, Dhana, K, Franco, OH, Kaptoge, S, Di Angelantonio, E, Stampfer, M, Willett, WC & Hu, FB 2018, 'Impact of healthy lifestyle factors on life expectancies in the US population', *Circulation*, vol. *138*, no.4, pp.345-355.

6. Rodolfo, G.Z., 2011, 'Society at a Glance 2011: OECD Social Indicators', OECD publishing, pp.86-87.

Link for Reference

[1] https://genus.springeropen.com/articles/10.1186/s41118-019-0055-0

[2] https://www1.health.gov.au/internet/publications/publishing.nsf/Content/oatsih-hpf-2012-toc~tier1~life-exp-wellb~119#:~:text=Life%20expectancy%20is%20affected%20by,nutrition%20and%20lack%20of%20exercise%3B

[3] https://jamanetwork.com/journals/jama/fullarticle/209448

[4] https://www.frontiersin.org/articles/10.3389/fmicb.2020.01526/full

[5] https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.117.032047

[6] https://www.oecd.org/berlin/47570194.pdf