MA2670 – Probability and Statistics II Assignment

Introduction:

There has been lots of debates and research being undertaken to investigate relationship between health and wealth across the world. According to one research an individual having the wealth is most likely to invest in health. This was based on a theoretical model Grossman (1972)[i]. Can the opposite to this statement be also true? i.e. The person with less wealth is unlikely to focus on health. This could mean for adult leading to low life expectancy and for a child growing in a household with low wealth may not be exposed to healthy eating[ii]. For the assignment, child obesity and life expectancy have been chosen for analysis based on the data being available.

According to Trust for London, childhood obesity is more common in London than England. In 2019/20, the obesity rate of year 6 children in London is higher than those in England. Barking and Dagenham had the highest proportion of childhood obesity out of all London boroughs according to Public Health England. There was a rise in most of the boroughs between 2009/10 and 2019/20, then there was a decline[iii].

As well as this, in the life expectancy in the different London boroughs, the life expectancy is higher for women than men. The borough with the lowest life expectancy is Barking and Dagenham for both men and women which shows how the low income is affecting them, whereas Westminster has the highest life expectancy for both men and women. However, there are some boroughs with longer healthy life expectancy for men than women[iv].

Methodology:

Descriptive Statistical Analysis

There is significant amount data available the Greater London Authority (GLA)website broken down into boroughs covering the whole of London. Each borough has further split into their respective wards. There is large number of data available on a range of health measures given in an excel table which helped to undertake further statistical analysis. The analysis on obesity has been done on primary children regardless of gender. As regards to life expectancy, the analysis is focused on male and female adults. The data from these attributes were analysed against median household income, which were also extracted from the GLA data.

For this research, the 5 boroughs in East London Boroughs have been chosen to investigate. As part of this suitable variables needed to be selected. To understand obesity related information regarding primary school children there are data available for the youngest and the oldest which would hopefully provide some meaningful information. In terms life expectancy, the GLA data have a fair amount of information both male and female adults.

The following variables have been chosen and used in R to have the graphs plotted and determine the correlation between them.

Dependent variables:

1. Percentage of obese children in reception
2. Percentage of obese children in year 6
3. Male life expectancy
4. Female life expectancy

Independent variable is Median Household Income Estimate (2012/13).

Inferential statistical analysis

With the help of R, the edited version of the given data can be used for a t-test to see if we can reject H_0.

For this hypothesis test, the data about the Median Household Income from the large data is used to be used. The point of this this is to see if the mean of the East London borough is different from the mean of London. The median household income estimate of London is going to be used as the mean. If H_0 is going to be rejected, the t_{stat} value must be less than the given value of the significance level.

The following data would be solved such as:

1. t_{stat}
2. df – degrees of freedom
3. p – value
4. Confidence Interval
5. Sample Estimates

The independent variable is Median Household Income Estimate (2012/13).

The effect of household income on the obesity rate of children in East London:

Dependent variables:

1. Percentage of children in reception year who are obese - 2011/12 to 2013/14 (see Figure 1)
2. Percentage of children in year 6 who are obese- 2011/12 to 2013/14 (see Figure 2)

The independent variable is Median Household Income Estimate (2012/13).

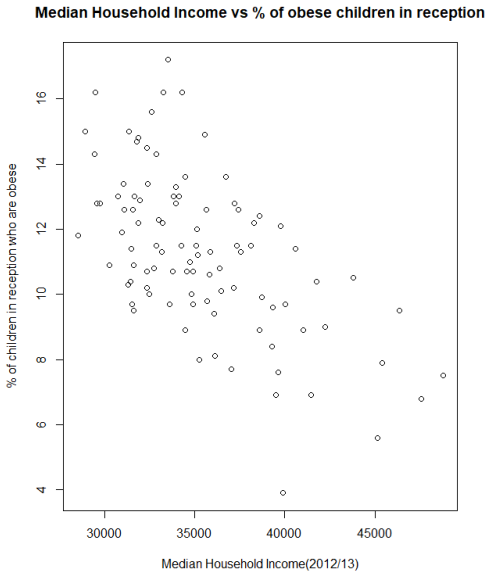

Figure 1: A graph that shows the relationship between the Median Household Income Estimate (2012/2013) and the percentage of children in reception who are obese

From the above scatter graph, it can be seen that there is a negative correlation between the median household income and the percentage of children who are obese in reception years. This seems to suggest that the median household income estimate increases, the percentage of obese children in reception supporting a statement that childen whose parents have lower income may not able to afford healthy food which are usually expensive. However, this may not apply to the wider audience. There can be some familiies who may not enough disposable income to spend on costly food items but they may be exposed to healthy living and the importance of choice of food. As for the children who parents are on low income, they can only afford the unhealthy food. It is also important to note that diet alone is not a contributory factor to rise in obesity as there can be other reasons for obesity such as inherent health or some genetic conditions.

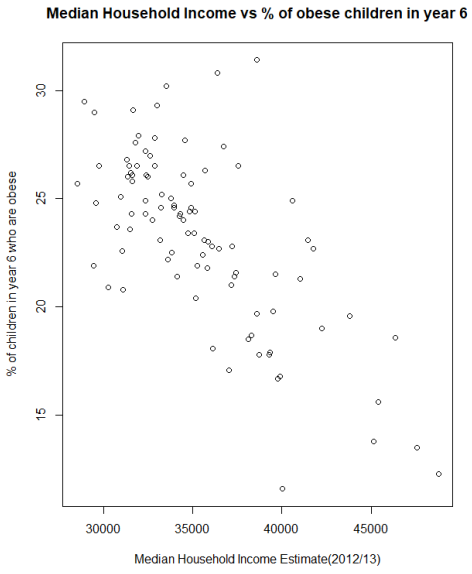

Figure 2: A graph that shows the relationship between the Median Household Income Estimate (2012/2013) and the percentage of children in reception who are obese

In Figure 2, there is also a negative correlation between median household income and the percentage of obese children in year 6. As the median household income increases, the percentage of obese children in year 6 also decreases. The correlation is stronger as indicated in Figure 2 compared to the analysis on reception children This is suggesting a stronger link between the household income and the obesity. It could be that the children develop their ealing habits during the primary schools hence showing stronger correlation. It can also be seen that the range of precentage obesity is higher than that of in children in reception.

The effect of household income on the life expectancy of adults in East London:

Dependent variables:

1. Male life expectancy -2009-13 (see Figure 3)
2. Female life expectancy -2009-13 (see Figure 4)

The independent variable is Median Household Income Estimate (2012/13).

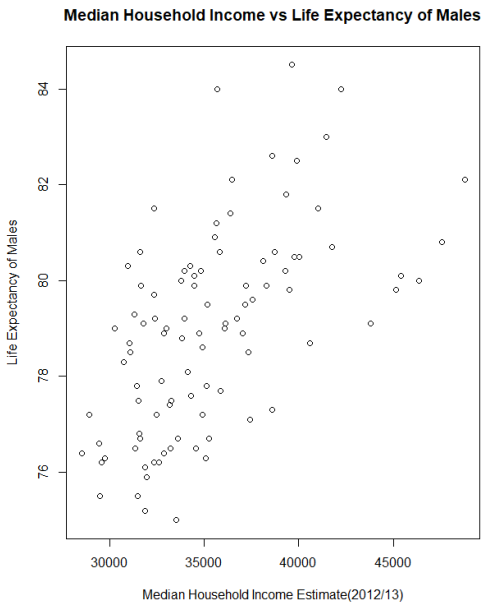

Figure 3: A graph that shows the relationship between the Median Household Income Estimate (2012/13) and the Life Expectancy of Males

From Figure 3, there is a positive correlation between the Median Household Income Estimate and the life expectancy of males. As the Median Household Income Estimate increases, the life expectancy of males increases. This indicates that the men who have high income are likely to live

longer as they have more disposable income to be able to look after their health. This means they spend their money wisely on food and supplements that helps them lead a healthy life. It can be inferred that the opposite will apply to men who are on low income may opt for cheaper and unhealthy food and consequently leading to health problems impacting on their life expectancy.

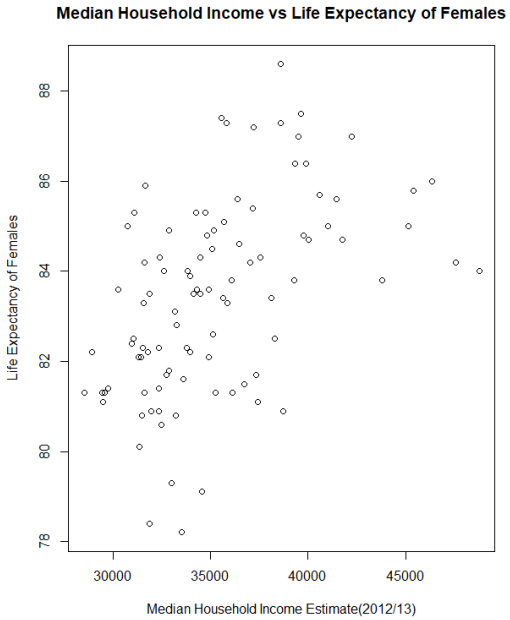

Figure 4: A graph that shows the relationship between the Median Household Income Estimate (2012/13) and the Life Expectancy of Females

In Figure 4, there is a positive correlation between the Median Household Income Estimate and life expectancy of females. As the Median Household Income Estimate increases, the life expectancy of females increases. This shows how females who have high income are more likely to live longer as they have more usable income to be able to look after their health. This also means they spend their money effectively on their essentials. In terms of longevity, it can be observed the females have a longer life expectancy compared to males. With these data it is difficult to draw any definitive conclusions to why there is a difference between men and women in this aspect.

Does the Mean of the Median Household Income Estimate (2012/13) of East London different from the Mean of the Median Household Income Estimate (2012/13) of London?:

```
summary(Borough_B_Data$`Median Household income estimate (2012/13)`)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
28500   32223   34515   35333   37440   48780
```

Figure 5: The summary statistics of Median Household Income Estimate (2012/13)

Let's assume that the data on the median household income is independent and continuous. There are no outliers. The dependent variables are normally approximated.

A one-tailed t-test is used.

H_0: The mean median household income estimate in London equals 51770. ($\mu = 51770$)

H_1: The mean median household income estimate in London is not equal to 51770. ($\mu \neq 51770$)

The following hypotheses will be tested at the 5% significance level.

```
> t.test(Borough_B_Data$`Median Household income estimate (2012/13)`, mu=51770)

        One Sample t-test

data:  Borough_B_Data$`Median Household income estimate (2012/13)`
t = -37.688, df = 95, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 51770
95 percent confidence interval:
 34467.53 36199.14
sample estimates:
mean of x
 35333.33
```

Figure 6: The results of the t-test obtained in R

The results of the one sample t-test on Figure 6 shows that $t_{stat} = -37.688$ and $p < 2.2 \cdot 10^{-16}$. As $p < 0.05$ we reject H_0 and conclude that there is evidence to suggest that the mean median household income estimate ($\mu = 35333$ as shown in Figure 5), the sample mean is less than the stated average median household income estimate ($\mu = 51770$). There is about 95% confidence that the mean median household income estimate is between 34467 and 36199, the interval that the median of the East London borough lies in. The hypothesis has been proved correct.

Conclusion:

Descriptive Statistical Analysis

Overall, the analysis indicates the median household income can impact on both the obesity rate on children and the life expectancy on adults. As can be on the scatter graphs there are outliers indicating exceptions to the general trends that were observed in both cases. From the findings, by some options could be explored to address these issues in the community. For example. How can we promote healthy living style at school level? The menu choice at school dinners to make sure the school are promoting healthy eating habits at primary school level.

As regards to educating adults, another approach could be via National Health Service (NHS) to promote healthy lifestyles. Accessibility to nutritionists can be made widely available. From basic dataset these are just a very crude analysis and not sufficient to draw any major conclusion. However, this gives a good starting point for further study should we wish to.

Inferential Analysis

After testing the median household income estimate of East London borough against the median household income estimate of London, it seems that there is a difference between one London borough compared to the whole of London. If we compare one borough to the city itself, there is more accuracy in the results.

---

[i] Tiainen, M. (2018). The relationship between wealth and health. [online] Available at: https://core.ac.uk/download/pdf/158607349.pdf.

[ii] Anon, (n.d.). The Grossman model – Healthcare Economist. [online] Available at: https://www.healthcare-economist.com/2019/12/04/the-grossman-model/#:~:text=The%20central%20proposition%20of%20the.

[iii] Trust for London. (n.d.). Child obesity. [online] Available at: https://www.trustforlondon.org.uk/data/child-obesity/.

[iv] Trust for London. (n.d.). Life expectancy by London borough. [online] Available at: https://www.trustforlondon.org.uk/data/life-expectancy-borough/.

Please refer to the R code in the appendix!