Effects of Education and Years on Obesity

Yuchang Bao, Yulin Huang, He Yang

B00782971, B00726567, B00758595

Group 6

Data Set 2

Dalhousie University

Date: December 10, 2020

Abstract

Obesity is a significant challenge faced by the medical circle. It is highly related to the incidence of a variety of diseases, such as sleep apnea and heart disease. Although more people are aware of the impact of obesity on physical health, the prevalence of obesity remains high in the United States. More than 60% of adults in the US are obese or overweight. The purpose is discussing the influence of education level and years on obesity rate of American adults. We hypothesized that obesity rate will decrease significantly as people's education level increases or as time goes by because of higher awareness of the issue. Through logistic regression, we found out the increase of education level correlates with the decrease of obesity rate, confirming our hypothesis; however, years were not significant to obesity rate, contradicting our hypothesis.

Introduction

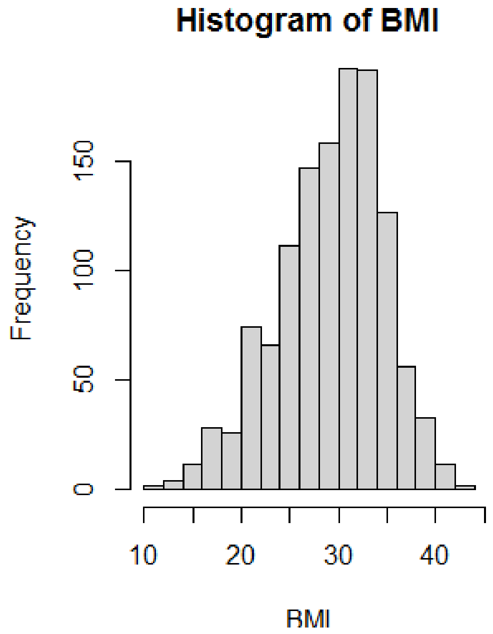
Obesity, defined as an abnormal physical condition or excessive fat accumulation (World Health Organization, n.d.) is one of the most difficult public health problems in our society (Mitchell et al., 2011). A body mass index (BMI) of more than 25 is considered overweight, while a BMI of over 30 is considered obese (World Health Organization, n.d.). A study of 195 different countries and regions in the New England Journal of Medicine (The GBD 2015 Obesity Collaborators, 2017) shows that about 603.7 million adults have obesity worldwide. According to the statistics of the U.S. National Health Interview Survey (2015), more than 70 million adults in the United States are obese. Nowadays, the obesity rate of adults in the United States ranks first in the world, and the obesity rates of adults and young people are as high as 42.4% (Galvin, 2020). The trend of obesity among American adults and adolescents has been increasing annually, especially in areas with low education levels (Ogden et al., 2017). Considerable studies have shown that obesity is an important dangerous factor for incidence rate and mortality rate of many diseases, such as hypertension, heart disease and diabetes (Xavier, 2009). Richard Carmona, an American surgeon, said: “Because of the increasing rates of obesity, unhealthy eating habits and physical inactivity, we may see that the first generation will be less healthy and have a shorter life expectancy than their parents” (Christopher & Caroline, n.d., page 8). It can be seen that the obesity problem is an urgent matter needed to be solved in the United States. We analyzed the Lionbridge website statistics of obesity rate in the United States from

2011 to 2016 using regression. Two factors, namely education level and time, have been examined. The hypothesis of this study is that the increase of education level and time will reduce the obesity rate significantly.

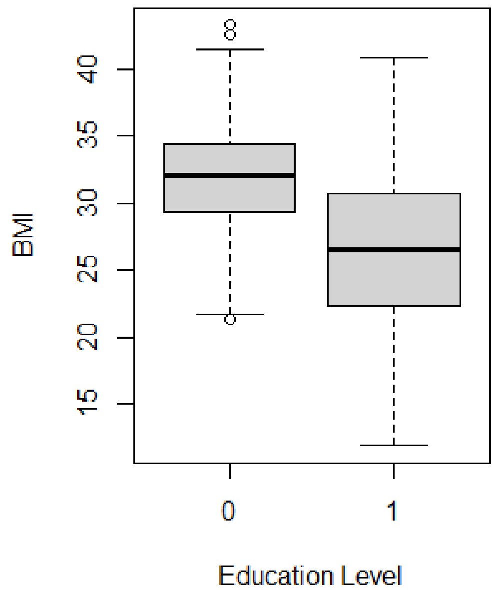
Data Description

The data set was collected on lionbridge website with 6 years of data, from 2011 to 2016. The data we used include 6 variables in total, namely Yearstart, LocationDesc, Question, Data_Value, Education, and Stratification Category 1. Yearstart denotes the year when data was collected and year is one factor examined to predict BMI. LocationDesc includes the states in the United States, and three other entries (National, Guma, Puerto Rico) which are removed since we are only interested in the US. We chose "percent of adults aged 18 years and older who have obesity" in the Question column as our research question. Data_Value is the body mass index (BMI) of each individual sampled. Education is the other factor examined to predict BMI, with four levels, i.e. less than high school, high school graduate, some college or technical school and college graduate. Finally, classification category 1 is the selection index used for data cleaning, among which we chose education as our factor of interest. For data visualization, we plotted a histogram for BMI (see *Figure 1(a)*) and two boxplots to examine the distribution of BMI across education levels (see *Figure 1(b)*) and across years (see *Figure 2*). From figure 1(a), we can see more than 50% of individuals in our sample are overweight or obese (i.e. BMI≥25). From figure 1(b), we could see there seems to be a significant difference in BMI scores between two education levels with higher education corresponding to lower BMI. From figure

2, we noticed a slight increase of BMI scores as time goes by. However, this needed a further test to confirm because of the large variances of data.



(a)



(b)

Figure 1. (a) Histogram of BMI (BMI≥25 indicates overweight or obese) in the US, across education levels and years. (b) Boxplot of BMI in the US in two education categories (1 represents a degree of "some college or technical school" or "college graduate"; 0 represents "less than high school" or "high school graduate").
Data from https://lionbridge.ai/datasets/10-open-datasets-for-linear-regression/. We chose dataset 2 here (National, Guma, Puerto Rico removed).

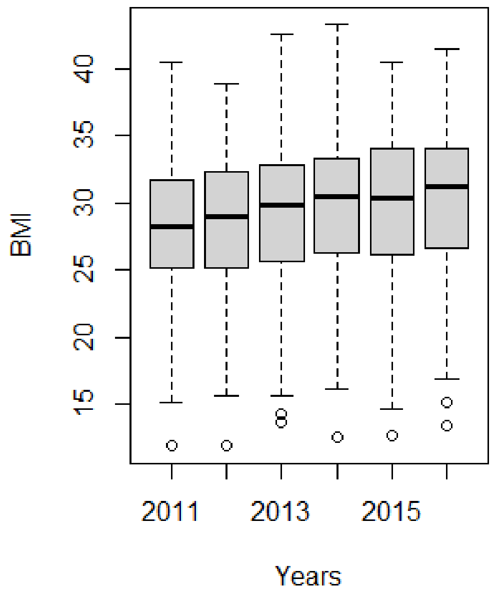


Figure 2. Boxplot of BMI in the US in each of the 6 years from 2011 to 2016 repectively.
Data from https://lionbridge.ai/datasets/10-open-datasets-for-linear-regression/. We chose dataset 2 here (National, Guma, Puerto Rico removed).

Methods

We selected R as the statistical analysis tool for this study. Relevant data were selected from the original data set as detailed in the above "Data Description" section. The first thing we did was transforming the non-numerical education categories (less than high school etc.) into some numerical levels (0 and 1). We classified education levels into two numerical categories: 1 denotes a degree of "some college or technical school" or "college graduate" while 0 denotes "less than high school" or "high school graduate". We also classified our BMI value into two categories: 1 indicates overweight or obese (BMI≥25) while 0 indicates normal weight underweight (BMI<25). Year data was kept the same. We used Logistic Regression/Generalized Linear Model(GLM) to analyze the new binary data (model:BMI~year+education), after using Shapiro-Wilk test to check normality on the original dependent variable

BMI(before binary transformation) and Variance Inflation Factor (VIF) to check the multicollinearity in our model. We chose alpha=0.01 as the significance level before the analysis.

Results

From the Shapiro-Wilk test, we found out our dependent variable is approximate to be normally distributed (see *Figure 3*, p-value = 4.897e-10), which met the normality assumption. Moreover, Variance Inflation Factor (VIF) was very close to 1 (see *Figure 4*), suggesting the two independent variables (years and education) are independent, which met another assumption. From the Logistics regression analysis, education (p-value <2e-16) was significant to the BMI scores (see *Figure 5*). However, years (p-value = 0.0213>0.01=alpha) was not significant.

```
        Shapiro-Wilk normality test

data:  finaldata$Data_Value
W = 0.98494, p-value = 4.897e-10
```

Figure 3. Result from the Shapiro-Wilk normality test.

```
finaldata$YearStart finaldata$Education
           1.001742            1.001742
```

Figure 4. Result from the Variance Inflation Factor (VIF) calculations for multicollinearity check.

```
Call:
glm(formula = (finaldata$Data_Value) ~ finaldata$YearStart +
    finaldata$Education, family = binomial())

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6515   0.2457   0.2865   0.9160   1.0796

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)         -206.71884   91.14580  -2.268   0.0233 *
finaldata$YearStart    0.10427    0.04527   2.303   0.0213 *
finaldata$Education   -2.72934    0.22447 -12.159   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1270.9  on 1239  degrees of freedom
Residual deviance: 1020.6  on 1237  degrees of freedom
AIC: 1026.6

Number of Fisher Scoring iterations: 6

Analysis of Deviance Table

Model: binomial, link: logit

Response: (finaldata$Data_Value)

Terms added sequentially (first to last)


                    Df Deviance Resid. Df Resid. Dev
NULL                                 1239     1270.9
finaldata$YearStart  1    4.402      1238     1266.5
finaldata$Education  1  245.844      1237     1020.6
        (Intercept) finaldata$YearStart finaldata$Education
       -206.7188357           0.1042678          -2.7293382
```

Figure 5. Result from the Generalized Linear Model (Logistics Regression) fitting: ANOVA output.

Conclusion

From the results, different education levels have significantly different BMI scores (see *Figure 5*). Combining with *Figure 1(b)*, we can see that higher education level corresponds to lower BMI (or lower obesity rate). On the other hand, years were not significant to obesity. With this in mind, the US government may consider increase the education level nationwide by various measures, such as allocating more education budget, reducing tuition fees and advertising the various benefits of education, not only because of the advantage of reducing the likelihood to obesity but many other personal benefits as well. However, we cannot overly conclude that education difference caused changes in people’s obesity in the US because correlation is not equivalent to causation. Many more factors, such as genetics and eating habits, interacting with education, contribute to BMI score as well. Further analysis could examine these factors in order to make it more complete.

References

Christopher, M., Marie, N., Ali, M. (2014). The vast majority of American adults are overweight or obese, and weight is a growing problem among US children. *Institute for Health Metrics and Evaluation.* http://www.healthdata.org/news-release/vast-majority-american-adults-are-overweight-or-obese-and-weight-growing-problem-among

Galvin, G. (2020). The U.S. Obesity Rate Now Tops 40%. *U.S. News*. https://www.usnews.com/news/healthiest-communities/articles/2020-02-27/us-obesity-rate-passes-40-percent

Mitchell, N., Victoria, C., Holly, W., James, H. (2011). Obesity: Overview of an Epidemic. *ScienceDirect*, 34(4), 717-732. https://www.sciencedirect.com/science/article/abs/pii/S0193953X11000827?via%3Dihub

National Health Interview Survey. (2015). Frequency distribution (inthousands) of body mass index among adults aged 18 and over. *The U.S. Department of health and humanservices.*https://ftp.cdc.gov/pub/Health_Statistics/NCHS/NHIS/SHS/2015_SHS_Table_A-15.pdf

Ogden, C., Fakhouri, T., Margaret, C., Craig, H., Cheryl, F., Xianfen, L., David, F. (2017). Prevalence of Obesity among Adults, by Household Income and Education — United States, 2011–2014. *Centers for Disease Control and Prevention*. https://www.cdc.gov/mmwr/volumes/66/wr/mm6650a1.htm

The GBD 2015 Obesity Collaborators. (2017). Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *The New England Journal of Medicine*, 377(1), 13-27.https://www.nejm.org/doi/10.1056/NEJMoa1614362

World Health Organization. (n.d.). Obesity.

https://www.who.int/health-topics/obesity#tab=tab_1

Appendix

See the attached R markdown files (NEW statproject(1).Rmd) for all data and R codes.