The correlation between childhood obesity ratios and six different governments' budgets

November 2020

Introduction

Childhood obesity is related to many environmental factors. Since 2008, 152 local governments in England allocated funds in six different fields (air quality, public places, health professionals, awareness-raising in schools, awareness-raising through media, and consulting services) to try to mitigate this situation. This paper investigates whether governments' investment has worked, and which kind of investment is the most effective.

Data

The data is collected from LondonDataStore. Ten variables are chosen for analysis in this paper: the number of people who developed childhood obesity in 2008 and 2018, the total population in 2008 and 2018, governments' budget in 6 fields every year.

Methodology

Hypothesis Testing

This paper chooses hypothesis testing to find out whether governments' investment work worked. The hypothesis testing implementation is mainly based on null hypothesis significance testing mode, where the null hypothesis and alternative hypothesis are established. The test statistic value is selected and calculated, and the statistical judgment on whether the null hypothesis is rejected is made(Zacks, 1981).

Multiple linear regression

Regression analysis is used to find out which kind of investment is the most effective one in this paper. Multiple linear regression model is one of the common ones, which can be described as follows.

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \delta$

Where y is the response variable and x_1,x_2 are the independent variables. β is the intercept and β_1, β_2 are the coefficients. δ is error term (Rath et al., 2020).

Results

Result of hypothesis testing

The childhood obesity ratios in 2008 and 2018 are calculated based on existing data. Based on summary statistics (Table 1), the mean of childhood obesity ratios increased, which means there might be a higher level of childhood obesity rate in 2018.

However, does this increase be significant enough to indicate that the overall level of childhood obesity has increased? Are they just occasional fluctuations? The mean comparison test is used to investigate.

Table 1: Summary statistics of childhood obesity ratios

	Childhood obesity in 2008	**Childhood obesity in 2018**
count	152	152
mean	0.309725	0.344673
std	0.137058	0.138712

The hypotheses are set as follows.

H_0: The childhood obesity ratio in 2008 is the same as that in 2018.

H_1: The childhood obesity ratio in 2008 is not the same as that in 2018.

The significance level is set to 0.05, and the p-value is calculated as 0.028, which is less than the significance level, indicating that H_0 should be rejected. Therefore, the childhood obesity ratio in 2008 is different from that in 2018, which means that the increase in the ratio is not random fluctuation but significant.

Result of MLR

The standard deviation of the child obesity ratios increased, indicating that there may be some places where the ratio has changed considerably, while others have not. Is it because of the differences in governments' investment? MLR model is chosen to investigate. The six kinds of government funding are selected as independent variables, and the change in childhood obesity ratios from 2008 to 2018 is selected as the dependent variable.

Explore the data

The six independent variables' frequency distribution graphs are drawn and find that they do not conform to the normal distribution. Therefore, they are then transformed by taking the square root. The new frequency distribution diagrams indicate that the transformed variables conform to the normal distribution better than their original counterparts (Figure 1).

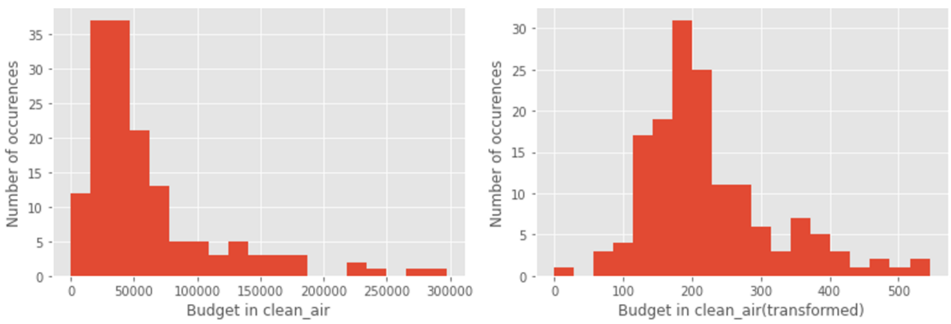

Figure 1: An example of frequency distributions before and after transform

Table 2: Relevant coefficients of oringinal model

R-squared:	0.061	**F-statistic:**	1.575
Adj.R-squared:	0.022	**Durbin-Watson:**	2.125
		coef	**p**
$\sqrt{x_1}$	clean_environ(transformed)	1.37E-06	0.964
$\sqrt{x_2}$	health_training(transformed)	-2.01E-05	0.689
$\sqrt{x_3}$	sub_counselling(transformed)	3.23E-05	0.612
$\sqrt{x_4}$	clean_air(transformed)	-4.68E-05	0.173
$\sqrt{x_5}$	school_awareness(transformed)	-5.72E-05	0.094
$\sqrt{x_6}$	media_awareness(transformed)	8.93E-05	0.051
	Intercept	0.0421	0

Table 3: Relevant coefficients of refined model

R-squared:	0.045	**F-statistic:**	3.505	**Durbin-Watson:**	2.118
Adj. R-squared:	0.032	**Prob (F-statistic):**	0.0325	**No. Observations:**	152
		coef	**std err**	**P**	**VIF**
	Intercept	0.04	0.006	0	
$\sqrt{x_5}$	school_awareness (transformed)	-7.49E-05	2.87E-05	0.01	10.0
$\sqrt{x_6}$	media_awareness (transformed)	7.43E-05	3.70E-05	0.047	10.0

Multicollinearity analysis

With one dependent variable and six new independent variables, an MLR model is established:

$$y = 0.000001367\sqrt{x_1} - 0.0000201\sqrt{x_2} + 0.00003227\sqrt{x_3} - 0.00004677\sqrt{x_4}$$
$$-0.00005716\sqrt{x_5} + 0.00008929\sqrt{x_6} + 0.0421$$

Relevant coefficients are shown in Table 2.

The variation inflation factors (VIF) method is chosen to examine the multicollinearity. The threshold is set to 10, while the maximum VIF is 21.520. Therefore, variables selection is required. Deleting variables in descending order of p-value until the maximum VIF is less than 10, $\sqrt{x_5}$ and $\sqrt{x_6}$ remain, and now the maximum VIF is 9.999.

Fit a model

The refined model is established:

$$y = -0.0000749\sqrt{x_5} + 0.0000743\sqrt{x_6} + 0.04$$

Relevant coefficients are shown in Table 3.

Diagnose the model

Durbin-Watson (DW) test has always been the standard diagnostic test for regression analysis (Grose and King, 1991). The DW value between 2 and 4 can be considered as the residual comfort to the normal distribution (King, 1983). While the DW value is 2.118, so the residuals conform to the normal distribution.

Compared to Table 3 with Table 2, the adjusted R-squared is increased from 0.022 to 0.032, showing that the new model can better explain childhood obesity changes. However, the value 0.032 is still meagre, meaning that these independent variables can only explain 3.2% of the change in the dependent variable. Therefore, the model is not fit well.

Interpret the model

According to Table 3, an increase in the funds spent on awareness-raising in schools by 1£/yr leads to a decrease in the predicted change of the ratio by 0.0000749%. While an increase in the funds spent on awareness-raising through media by 1 £/yr leads to an increment in the predicted change by 0.0000743%. In addition, when the government's funding for these two fields is 0, the predicted change of the ratio is 0.04%.

Discussion

Hypothesis testing indicates that the government's investments to improve childhood obesity have not worked. However, it is possible that it did work, but there may be other factors that caused a more considerable effect. The coupling effect of all these factors is the rise in childhood obesity ratio. In addition, the low adjusted R-squared means that these variables can only explain a tiny part of the change in ratios. In order to improve the model, more relevant variables should be selected for the regression analysis.

Conclusion

Based on hypothesis testing and MLR model, this paper analyzes the changes in childhood obesity ratios between 2008 and 2018, and the correlation between the changes and different governments' budget. Hypothesis testing indicates that the childhood obesity ratio has increased significantly. In addition, regression analysis indicates that the investment of awareness-raising in schools is most significant among the six. However, it is also found that governments' investment cannot explain the changes in ratios very well. Therefore, it is ungranted to assert that governments' investment has had a counterproductive effect on tackling the issue of childhood obesity. There may be other reasons that have contributed to the increase in childhood obesity, such as the age structure or the mortality rate.

References

S. D. Grose and M. L. King. The locally unbiased two-sided durbin—watson test. *Economics Letters*, 35(4):401 – 407, 1991. ISSN 0165-1765. doi: https://doi.org/10.1016/0165-1765(91)90010-I.

M. L. King. The durbin-watson test for serial correlation: Bounds for regressions using monthly data. *Journal of Econometrics*, 21(3):357 – 366, 1983. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(83)90050-7.

S. Rath, A. Tripathy, and A. R. Tripathy. Prediction of new active cases of coronavirus disease (covid-19) pandemic using multiple linear regression model. *Diabetes Metabolic Syndrome: Clinical Research Reviews*, 14(5):1467 – 1474, 2020. ISSN 1871-4021. doi: https://doi.org/10.1016/j.dsx.2020.07.045.

S. Zacks. Testing statistical hypotheses. *Parametric Statal Inference*, pages 113–175, 1981.

Appendix

Word count: 999
Repository with Data and Scripts: https://github.com/akiakutaji/QM/tree/main/Assessment%201