Research on the Influencing Factors that May Lead to Obesity

Nike Qiu*

College of Business and Public Management, Kean University, Wenzhou, 325060, China

*Corresponding author: qiun@kean.edu

Abstract. The global obesity problem is getting more and more serious, which not only affects the quality of life but also induces a series of complications. In this study, multiple linear regression method was used to study the influencing factors of obesity, and their answers were linked to the literature review to draw conclusions. And their responses were linked to literature Review to draw conclusions. Calorie consumption monitoring, physical activity frequency, and vehicle use also have significant positive effects on obesity levels. Frequent consumption of vegetables, main meals, smoking times, and alcohol consumption will also have a significant negative impact on the obesity level. Due to the limitations of this data set and other reasons, the model test results will be inaccurate. In the future, people from other regions may be included, making the research more scientific. The research in this paper has a positive effect on the treatment of obesity, which will help people to find it earlier, treat it as soon as possible, and improve the survival rate and quality of life of patients.

Keywords: Obesity; multiple factors; multiple linear regression.

1. Introduction

Obesity has grown throughout the world's population and is now a substantial contributor to health issues. According to the report, more than 22 million children under the age of five are obese worldwide, and a tenth of them are overweight [1]. Despite being significantly lower than in other nations like the US and the UK, China's obesity rate is rising dramatically yearly [2]. In China, the problem of childhood obesity has become increasingly serious, and 90% of children will develop adult obesity when they grow up [2]. Many diseases are chronic, progressive processes that begin in adolescence. Controlling the occurrence of childhood obesity can reduce a lot of unnecessary medical expenses in the future, which is beneficial to families and society. Future productivity may suffer as a result of childhood obesity; thus, we must take action to address this issue. The purpose of this study is to examine the various potential factors that may contribute to obesity, to help people assess their own risk of developing obesity, and to take meaningful preventive measures based on their risk level.

Obesity is the result of many factors. Some literature has shown that the occurrence of obesity is related to many factors, such as excessive intake of high-calorie food [1], low economic level [3], psychosocial issues [4], anxiety symptoms [5], longevity [6] and so on. However, there are few studies on obesity-related factors in this literature, lacking systematic and complete statistics. Therefore, this paper aims at some factors to study whether they have an impact on obesity and select an appropriate model to study the degree of correlation between these factors and obesity.

According to experts, several adult cardiovascular risk factors are already linked to childhood and adolescent obesity, and at the current rate of obesity increase, the number of cardiovascular disease events will rise by almost 14% in 2035, with an increase of 7% for men and 10% for women predicted [7]. This is because research on young people who pass away from non-cardiac causes has demonstrated that obesity is one of the main predictors of heart pathology. This study lends credence to the hypothesis that cumulative cardiovascular effects of childhood obesity on adult cardiovascular outcomes may exist.

To establish whether a child's body mass is appropriate, BMI charts are frequently employed. Weight divided by height squared (kg/m2) is how BMI is calculated. A chart like this can be used to compare a child's BMI to the typical values for other kids their age. A bmi greater than 95 percentile indicates obesity and a bmi between 85 and 95 percentile indicates obesity in children [8]. Continually

high global quality of life can be judged in terms of physical ability, appearance, and social functioning, according to child completion [9].

Obesity is not permanent, and there are ways to deal with it. The long-term effects of childhood obesity can be reduced with early obesity therapies. Starting at age 6, school health inspections should include annual measurements and weight checks of students. The standard protocol for diet restriction suggests a fat intake of dropping to 25% to 35% for kids 4 to 18 [10]. Strong data backs up the advantages of cutting back on salt, sugary drinks, and trans fats. Strong data backs up the advantages of cutting back on salt, sugary drinks, and trans fatty acids in the diet [11]. According to recent research, adolescents who undergo bariatric surgery experience significant weight loss and a significant drop in problems [1]. When treating children with severe obesity, bariatric surgery performed during adolescence may be more successful than waiting until adulthood.

This paper will undertake a thorough examination of the elements that contribute to obesity and develop effective solutions to combat the problem. This will help us better appreciate the severity of the harm caused by obesity and lay the groundwork for paying more attention to our own health.

2. Method

2.1. Data Sources

A machine learning database at the university of California, Owen provided data from the study, which analyzed the data to estimate obesity levels based on dietary and health conditions in Colombia, Peru, and Mexico. 2020 saw 2111 records and 16 features added to this data.

2.2. Variable Selection

The dataset consists of 2111 records and 16 features (Figure 1).

Fig. 1 Variable-related schematic

The data are identified by the class variable Obesity Level, which permits classification into seven categories.

Table 1. Logogram and numbers of the 16 factors

Element	Logogram	variable assignment and meeting
Gender	X_1	Female, Male
Age	X_2	14-61 years old
Height	X_3	1.45m - 1.98 m
Weight	X_4	39 kg -173kg
FHO	X_5	Yes, No
FAVC	X_6	Yes, No
FCVC	X_7	1<FCVC<3
NCP	X_8	1<NCP<4
CAEC	X_9	Always, Frequently, Sometimes, No
SMOKE	X_{10}	Yes, No
CH2O	X_{11}	1<CH2O <3
SCC	X_{12}	Yes, No
FAF	X_{13}	0<FAF<2
TUE	X_{14}	1<TUE<2
CALC	X_{15}	Always, Frequently, Sometimes, No
MTRANS	X_{16}	Automobile, Bike, Motorbike, public transport, Walk
NObeyesdad	Y	Insufficient, Normal, Level I, Level II, Type I, Type II, Type III

Table 1 shows the allocation of 16 factors and variables and the linguistic signs and figures of the meeting. In order to facilitate writing, the identification of factors is mentioned above, as shown in Table 2. The model sets women and men for gender. The sample size is 2111 people, 972 of whom are obese. Keep the original values of age, height, and weight variables, and don't classify or evaluate the participants in any way. The answers to FHO, FAVC, SMOKE, and SCC are yes or no. The range of FCVC is 1-3, NCP is 1-4, CH2O is 1-3, FAF is 0-2, and TUE is 1-2. In order to classify whether the sample people eat other foods between meals, they are always, frequently, sometimes, and No. The frequency of drinking is divided into four grades. In this paper, the vehicles used are divided into five categories, 1 for cars, 2 for bicycles, 3 for motorcycles, 4 for public transportation, and 5 for walking. The dependent variable obesity level is set to Y1, and the sample results are assurance weight, normal, level I, level II, type I, type II, and type III.

2.3. Research Protocol

This paper uses the multiple linear regression model. This formula explains the relationship between the explanatory independent variables (X) and dependent variables (Y). The relationship between the effects of X on Y, or the relationship between the 16 factors on obesity, is then analyzed in this article using SPSSAU.

Deforming multiple linear regression model, it can be seen that:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon \quad (1)$$

Where, for n = i observations: y_i= Obesity Level; x_i= 16 variable; β_p= coefficients.

2.4. Model testing

The model passed the F test (F=52.543, p=0.000<0.05) when carrying out the F test on the model, that is to say, at least one of the independent variables will have an influence on NObeyesdad, which means that the model construction is meaningful.

Table 2. ANOVA Table

	Sum of squares	df	Mean Square	F	p Value
Regression	2303.321	16	143.958	52.543	0.000
Residual value	5737.163	2094	2.74		
Sum	8040.484	2110			

3. Results and Discussion

Figure 2 shows the number of men and women in different age groups and the number of obese people. As can be seen in Figure 1, the obese population is concentrated in the 21–30 age group, with 649 people in the sample, while the 11–20 and 31–40 age groups are the second and third most affected, with 176 and 127 people, respectively.

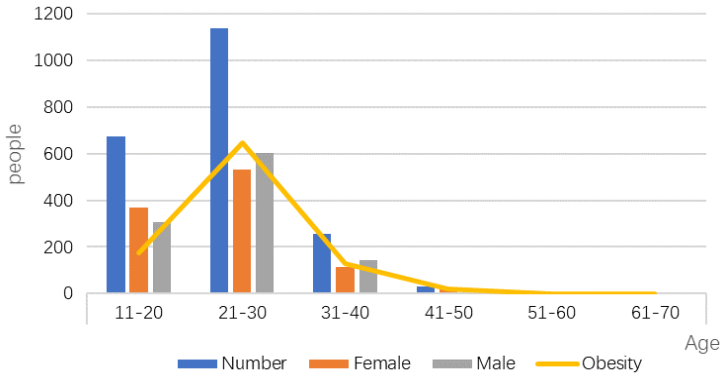

Fig. 2 Gender distribution by age group

Figure 3 shows a range of factors contributing to obesity as a proportion of obese patients. In total there were 2111 samples from which the proportion of samples suffering from obesity was determined and categorized according to the different influencing factors in order to determine the probability of each risk contributing to obesity.

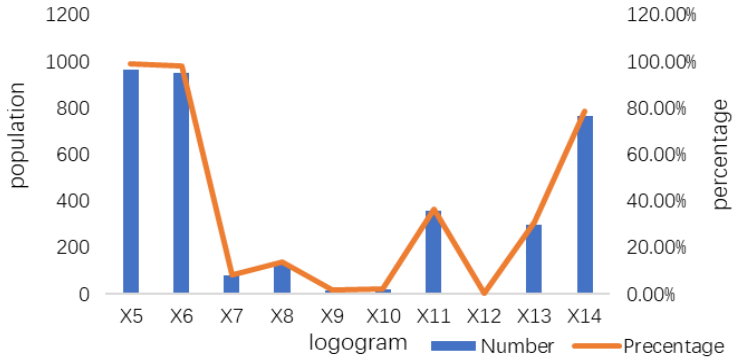

Fig. 3 Factors in the proportion of people with obesity

As can be seen from Figure 3 above, family history, high calorie foods, and Physical activity frequency accounted for the highest proportion, 98%, 99%, and 93% respectively. Calorie consumption monitoring, on the other hand, accounted for the lowest proportion, less than 1 per cent. Secondly, eating between main meals and smoking did not have a particularly large impact on obesity, accounting for only 1.65 per cent and 2.26 per cent. Thus, the hidden risk of obesity comes mainly from family history, high-calorie food and physical activity.

Table 3. Summary of basic information on decision tree classification

Name	Option	Frequency	Percentage
NObeyesdad	Insufficient	272	12.88%
	Normal	287	13.60%
	Type I	351	16.63%
	Type II	297	14.07%
	Type III	324	15.35%
	Level I	290	13.74%
	Level II	290	13.74%
Aggregation	Sum	2111	100.00%
	Valid	2111	100.00%
	Missing	0	0.00%
	Sum	2111	100.00%

Table 3 The items X_1-X_{16} are used as the independent variables, while the items obesity level as the dependent variable for decision tree modelling, as can be seen from the table above, a total of 2111 samples were involved in the analysis.

A decision tree is a commonly used classification algorithm. A decision tree consists of one root node, several internal nodes, and several leaf nodes.

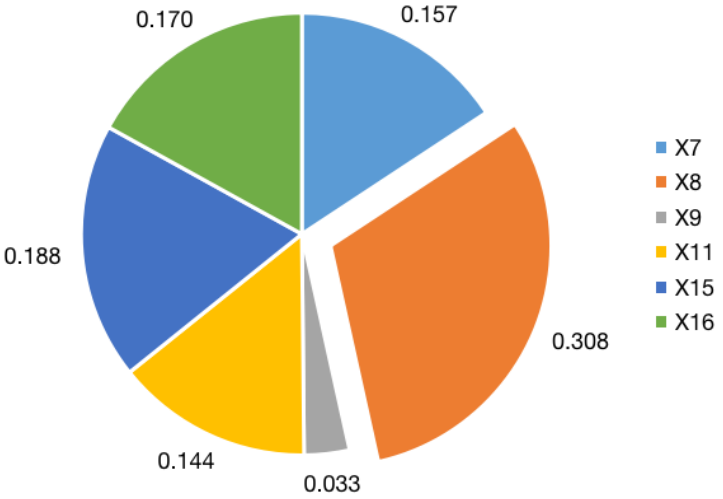

Fig. 4 feature weight graph (%)

Figure 4 which has a cumulative value of 1, demonstrates the significance of each heading in the model. It is evident from the preceding table that, NCP has a weight of 30.78%, making it the feature with the highest weight and one that is crucial to the creation of the model. The weight of CALC is 18.77%, that of MTRANS is 16.98%, and that of FCVC is 15.75%. Together, the weight of the four characteristics accounted for 82.27 %. Since the weight of other variables is 0, it will not be shown in the chart.

Table 4. Linear regression analysis results (n=2111)

	B	SE	Beta	T Value	p value	VIF	tolerability
Constants	3.816	1.16	-	3.289	0.001**	-	-
Gender	0.094	0.1	0.024	0.939	0.348	1.91	0.523
Age	0.051	0.008	0.165	6.362	0.000**	1.968	0.508
Height	-2.423	0.626	-0.116	-3.872	0.000**	2.625	0.381
Weight	0.022	0.002	0.295	10.68	0.000**	2.234	0.448
FHO	0.837	0.112	0.166	7.469	0.000**	1.443	0.693
FAVC	0.588	0.121	0.097	4.843	0.000**	1.166	0.858
FCVC	-0.193	0.075	-0.053	-2.579	0.010**	1.228	0.814
NCP	-0.199	0.049	-0.079	-4.073	0.000**	1.108	0.902
CAEC	0.854	0.084	0.205	10.206	0.000**	1.185	0.844
SMOKE	-0.602	0.257	-0.044	-2.344	0.019*	1.039	0.962
CH2O	0.092	0.062	0.029	1.481	0.139	1.124	0.89
SCC	0.514	0.182	0.055	2.825	0.005**	1.105	0.905
FAF	-0.106	0.047	-0.046	-2.247	0.025*	1.237	0.809
TUE	0.045	0.063	0.014	0.718	0.473	1.137	0.88
CALC	-0.349	0.074	-0.092	-4.721	0.000**	1.119	0.894
MTRANS	0.09	0.038	0.058	2.386	0.017*	1.74	0.575
F	F (16,2094) =52.543, p=0.000						
D-W value	1.78						
Dependent variable：NObeyesdad							
* p<0.05 ** p<0.01							

As can be seen from Table 4, a linear regression analysis was carried out with obesity level was used as the dependent variable as it can be seen in the above table that, the model equation is:

$$Y = 3.816 + 0.094X_1 + 0.051X_2 - 2.423X_3 + \ldots + 0.090X_{16} \quad (2)$$

Table 4 show that the regression coefficient value of family history of overweight is 0.837, and the regression coefficient value of food consumption between two meals is 0.854, which shows a significant positive correlation with obesity level. To sum up, Age, Weight, FOH, FAVC, CAEC, SCC, FAF, and MTRANS will have a significant positive effect on obesity level. Height, FCVC, NCP, smoking, and CALC will have a significant negative effect on obesity level. However, Gender, CH2O, and TUE do not affect obesity level.

4. Conclusion

In this study, various data were selected to focus on the influential factors that may be associated with developing obesity. It was concluded that Age, Weight, FOH, FAVC, CAEC, SCC, FAF, MTRANS will have a significant positive effect on obesity level. As well as Height, FCVC, NCP, SMOKE, CALC will have a significant negative effect on obesity level. However, Gender, CH2O, and TUE do not affect obesity level.

Due to the small sample size of only 2111 samples, 77% of which were generated using the Weka tool combined with the SMOTE filter, and 23% of which were collected directly from users via the web platform, there may be some error in the model. And the sample did not cover all other factors, such as age and race, that could have contributed to the differences, which could also affect the accuracy of the results. However, this study still has great value. First, the research method adopted in this study is innovative. On the one hand, this study visually analyzed the difference in the proportion of each factor in the lung cancer population and the non-lung cancer population by using the bar chart. This allows the experiment to be visualized, and the results are clearer and more intuitive. On the other hand, this study uses multiple linear regression and decision tree methods to make the experiment more comprehensive. Secondly, the study in this paper has a positive effect on the treatment of obesity, which can effectively prevent obesity.

In addition to the known obesity-related factors such as CAEC, SCC, and MTRANS, there may be more obesity-related factors worthy of attention and consideration, such as endocrine factors, psychological and neurological factors, sleep, and so on. Whether these factors are related to the development of obesity requires further medical research, which means that these findings will point the way to further related research in the future. Once a new pathogenic factor is discovered, it will help us to detect obesity earlier and treat it as early as possible, improving the survival rate and quality of life of patients.

References

[1] Raj M, Kumar R K. Obesity in children & adolescents. The Indian journal of medical research, 2010, 132(5): 598.

[2] Chen C M. Overview of obesity in Mainland China. Obesity reviews, 2008, 9: 14-21.

[3] Murasko J E. Socioeconomic status, height, and obesity in children. Economics & Human Biology, 2009, 7(3): 376-386.

[4] Canoy D, Bundred P. Obesity in children. BMJ clinical evidence, 2011.

[5] Değirmenci T, et al. Psychological symptoms in obesity and related factors. Nöro Psikiyatri Arşivi, 2015, 52(1): 42.

[6] Kopelman P G. Obesity as a medical problem. Nature, 2000, 404: 635-643.

[7] Bastien M, et al. Overview of epidemiology and contribution of obesity to cardiovascular disease. Progress in cardiovascular diseases, 2014, 56(4): 369-381.

[8] Kuźbicka K, Racho D. Bad eating habits as the main cause of obesity among children. Pediatr Endocrinol Diabetes Metab, 2013, 19(3): 106-110.

[9] Griffiths L J, et al. Self‐esteem and quality of life in obese children and adolescents: A systematic review. International Journal of Pediatric Obesity, 2010, 5(4): 282-304.

[10] Kawulich B B. Participant observation as a data collection method. Qualitative social research, 2005.

[11] Bradwisch S A, et al. Obesity in children and adolescents: An overview. Nursing, 2022, 50(11): 60-66.