Analysis of Influencing Factors of Heart Disease

Shuyi Li *

School of mathematic and information science, Nanjing Normal University of special education, Nanjing, 210000, China

* Corresponding author: 631402090104@mails.cqjtu.edu.cn

Abstract. In recent years, heart disease has gradually become one of the important factors endangering human health. There are many factors that can trigger heart disease in everyday life. It is of important practical value to explore the specific influencing factors of heart disease. In this paper, the linear regression model is selected to analyze the influencing factors of heart disease; the function in SPSS is used to determine the order. 463 groups of patients that have heart disease are examined on five different dimensions. SPSS was used to test the degree to which these five variables fit the dependent variable. This paper will explore the influencing factors of the onset of heart disease through these five aspects. To find out which attribute is most likely to cause heart disease. According to the data analysis results among the five aspects, family history is the main influencing factor, followed by the measure of cholesterol and the age of study subject. Moreover, systolic blood pressure and a measure of obesity have no significant effect on the heart disease. The research result accords with the current reality.

Keywords: Linear regression, F tes, influencing factors, heart disease.

1. Introduction

Heart disease has already been an eye-catching problem. In recent years, new process is being made in the study of the factors that influence heart disease. Zhen, et al. used the regression model to analyze the time change trend of CHD incidence in perinatal infants in Henan Province from 2011 to 2020 [1]. The results show that the incidence of CHD in urban infants is significantly higher than that in rural areas from 2011 to 2020 in Henan Province. The incidence of CHD in male infants is higher than that in female infants significantly, and the incidence of CHD in older women is significantly higher than that in younger women. The overall incidence is increasing year by year.

Zhang, et al. tried to analysis the dominant factors of heart disease. Children with congenital heart disease born in Lishui Central Hospital between 2018 and 2020 are selected as the study subjects [2]. This paper use univariate and multivariate logistic regression models to analyze parental age, infection in early pregnancy, gestational diabetes, etc. The results show that congenital heart disease is associate with advanced maternal age, gestational diabetes, early pregnancy infection, parental smoking/alcohol consumption, vaginal bleeding, neonatal low body weight, premature delivery, and negative sex life events. Multigrain intake in early pregnancy, folic acid intake in the first year of pregnancy and natural conception are associate with a lower risk of congenital heart disease.

Des illustrated that the number of major vessels and the type of chest pain and the ST slope are three main factors that caused heart disease [3]. The amount of cholesterol is not a dominating reason that cause heart disease. The disease is not related to gender. The author Kamil Pytlak investigated many Americans with heart disease [4]. These Americans surveyed include African Americans, whites, and native Americans. According to the survey results, half of the study object (47%) have at least one negative lifestyle habit that can lead to heart disease. These bad living habits above include: excessive cholesterol intake, excessive fat intake (obesity and high BIM), and smoking habit. In addition, people who don't exercise regularly, and people who like to drinking too much alcohol are more likely to develop heart disease.

Mente, et al. used the Join point regression model to analyze the time change trend of CHD incidence in perinatal infants in Henan Province from 2011 to 2020 [5]. The results showed that in urban infants, the incidence of CHD was conspicuously higher than that in rural areas from 2011 to 2020 in Henan Province. The incidence of CHD in male infants was prominently higher than that in

female infants. Besides, the incidence of CHD in older women was conspicuously higher than that in younger women. The overall incidence is increasing year by year. Greenfield S, et al. investigated the congenital heart disease in newborns from 2018 to 2020 and its influencing factors. Children who suffered congenital heart disease born in Lishui Central Hospital between 2018 and 2020 were selected as the study subjects [6]. Univariate and multivariate logistic regression models were used to analyze parental age, infection in early pregnancy, parental smoking/drinking, negative sex life events, vaginal bleeding, premature birth, birth weight, gestational diabetes, etc. The results showed that congenital heart disease was associated with advanced maternal age, gestational diabetes, early pregnancy infection, parental smoking/alcohol consumption, vaginal bleeding, neonatal low body weight, premature delivery, and negative sex life events. Multigrain intake in early pregnancy, folic acid intake in the first year of pregnancy and natural conception were associated with a lower risk of congenital heart disease.

In analysis of birth defects and influencing factors of congenital heart disease in perinatal infants in Lanzhou from 2012 to 2018, the authors collected data of perinatal infants diagnosed with congenital heart disease from 28 weeks gestation to 7 days postpartum during hospitalized delivery in Lanzhou from 2012 to 2018 for univariate and multivariate logistic regression analysis [7]. The results showed that the incidence of congenital heart disease has increase tendency. Multivariate Logistic regression analysis showed that maternal first pregnancy and higher family income were protective factors for congenital heart disease, and gestational age less than 37 weeks was a risk factor for congenital heart disease. In analysis of influencing factors leading to congenital heart disease in fetus. The authors selected 3970 pregnant women who underwent prenatal examination in Shapingba Hospital of Integrated Traditional Chinese and Western Medicine in Chongqing as study subjects [8]. SPSSAU multifactor analysis showed that smoking or drinking during pregnancy and exposure to harmful environment during pregnancy were independent factors leading to congenital heart disease in their fetuses.

In a research report on the Jilin region. Seventy pairs of CHD children and their mothers in Jilin Province from 2019 to 2020 were selected as the experimental group, and 140 pairs of healthy children and their mothers of the same sex, age and region were selected as the control group for 1:2 case matching [9]. Multivariate Logistic regression analysis was used to analyze the dominating factors of CHD in children. The results show that: There was no significant difference in gender and age distribution between experimental group and control group ($P>0.05$), the results of multivariate analysis showed that preterm birth (OR= 6.224), alcohol consumption within 3 months prior to pregnancy (OR=2.476), and BMI≥25(OR=8.372) before the pregnancy were the risk factors for CHD in children ($P<0.05$), multivitamin supplementation during pregnancy (OR=0.342) was a protective factor for CHD in children ($P<0.05$).

In epidemiological investigation of congenital heart disease in preschool children in Nanjing area [10]. Multivariate Logistic regression analysis showed that CHD family history (OR=4.941,95% CI:2.344-9.92), maternal history of viral infection in early pregnancy (OR=4.148,95% CI:2.232-8.310) and drug use in early pregnancy (OR=3.007,95% CI:2.473~4.304) are independent risk factors for CHD ($P<0.05$).

2. Method

The data used in this paper are collected from patient data from the local hospitals by the author, and send the patient questionnaires to acquiring the family history of heart disease. There are five variables used in this paper. Denote x1 represents systolic blood pressure, x2 represents a measure of cholesterol. x3 represents family history (integer valued 0=no disease and 1=disease.). x4 represents a measure of obesity. x5 represents age of study subject. And y=heart disease (It is integer valued 0 = no disease and 1 = disease).

The purpose of this paper is to study x_1, x_2, x_3, x_4, x_5 five argument variables which will have an impact on the response variable. Thus, the linear regression model was chosen to analyze the data.

Linear regression is a model that can analysis the relationship between one or more independent variables and response variable. In addition to the linear regression equation, there are also B values, VIF test result, F test result and R^2values in the analysis results of SPSS.

In the process of fitting synthesis parameter, the R^2 indicates how much of the change in the response variable can be explained by the independent variable. The B value indicates how much x affects y. In the analysis results, only when p<0.05 of F test, the analysis results have reference significance. If p>0.05, it means that the error is obviously large, and the analysis results are of no reference value at this time. VIF values are used to test the collinearity problem of the model. If the model has collinearity problems, the ridge regression model should be replaced for analysis.

In the analysis results, if the regression coefficient is greater than 5%, it indicates that the results have reference value. Since linear regression is calculated directly, its error can be determined. This paper adopts the linear regression to analyze the correlation degree of x_1, x_2, x_3, $x_4$and $x_5$to y. After obtaining the agreement of the patients, the author investigated the data of various indicators of 463 local patients with heart disease. There are 2315 pieces of data in total. Firstly, Univariate linear regression was used to calculate the fitting degree of each attribute to the response variable. Then using multiple variable linear regression to analysis, the entire five attributes. The linear regression conclusion was obtained through SPSS analysis, and the regression coefficients were compared to determine which variable (x_1, x_2, x_3, x_4, x_5) had a greater correlation with the response variable.

3. Results and Discussion

With x1, x2, x3, x4and x5 as independent variables and y=heart disease (It is integer valued 0 = no disease and 1 = disease) as dependent variables. We use SPSSAU to conduct multiple linear regression analysis. Table 1 shows the results of regression analysis.

Table 1. Analysis results with *n*=462

	Unstandardized Coefficients		standardized Coefficients	*t*	*p*	VIF
	B	standard error	β			
constant	-0.278	0.166	-	-1.677	0.094	-
x1	0.001	0.001	0.061	1.317	0.188	1.202
x2	0.037	0.011	0.161	3.506	0.001**	1.199
x3	0.177	0.042	0.183	4.224	0.000**	1.071
x4	-0.008	0.005	-0.069	-1.506	0.133	1.197
x5	0.009	0.002	0.276	5.657	0.000**	1.352
R^2	0.200					
Adjust R^2	0.191					
F	F (5, 456)=22.757, p=0.000					
D-W value	1.981					
Dependent variable: heart disease						
* p<0.05; ** p<0.01						

The linear regression equation can be obtained from the table above:

$$y = -0.278 + 0.001x_1 + 0.037x_2 + 0.177x_3 - 0.008x_4 + 0.009x_5 \quad (1)$$

The R^2 value of the model is 0.200, which means that the independent variables x_1, x_2, x_3, x_4, and x_5 can explain the 20.0% change of response variable. When conducting the F test on the model, the result shows that F=22.757, p=0.000<0.05. It indicates that the model has passed the F-test. which also indicates at least one of x_1, x_2, x_3, x_4, x_5 has effect on heart disease.

In addition, The VIF value in the analysis results is less than 5, which indicates that there is no collinearity problem in this model. Moreover, the D-W value is used to determine whether there is an interaction between the sample data of the model. In this paper, the D-W value stabilizes around

number 2. This indicates that there is no interaction between the independent variables and the model has no autocorrelation.

According to the results, the influence degree of each independent variable is analyzed separately. The value of regression coefficient x1 was 0.001 (t=1.317, p=0.188> 0.05), indicating that x1 has no effect on heart disease. The value of regression coefficient x2 was 0.037(t=3.506, p=0.001< 0.01), indicating that x2 has a prominent positive effect on heart disease. By analyzing the regression coefficients of other independent variables. It is clear that x3 and x5 also has a very significant effect on response variable, and x4 has no effect on response variable. Therefore, the influencing factors represented by x2, x3, x5 are the main causes of heart disease.

The conclusion of this linear regression analysis shows that the independent variables x2, x3 and x5 has prominent positive effects on heart disease. But x1 and x4 did not have an effective on heart disease.

Table 2. Results of linear regression analysis

	Regression coefficient	95% CI	VIF
constant	-0.278 (-1.677)	-0.604 ~ 0.047	-
x1	0.001 (1.317)	-0.001 ~ 0.003	1.202
x2	0.037** (3.506)	0.016 ~ 0.058	1.199
x3	0.177** (4.224)	0.095 ~ 0.259	1.071
x4	-0.008 (-1.506)	-0.018 ~ 0.002	1.197
x5	0.009** (5.657)	0.006 ~ 0.012	1.352
Sample size	462		
R^2	0.200		
Adjust R^2	0.191		
F value	F (5,456)=22.757,p=0.000		
dependent variable: heart disease			
* p<0.05; ** p<0.01 (The t value is in bracket)			
Value of D-W: 1.981			

According to the analysis results in Table 3, it can be seen that the R^2 value of this model is 0.200, which indicates the independent variable x1, x2, x3, x4and x5 have 20% fit degree to the model, and it can explain 20.0% of the changes in heart disease.

Table 3. Model profile

R	R^2	Adjust R^2	Root-mean-square error	AIC	BIC	DW
0.447	0.200	0.191	0.426	533.866	558.680	1.981

As table 4 shows, F test is used to test whether the results of linear regression analysis have reference value. To be precise, if the analysis result shows p<0.05, it means that the model passes the F test. Furthermore, it shows that at least one independent variable can affect the dependent variable. This shows that the analysis results are accurate and valuable for reference. Moreover, if the analysis result shows p>0.05, it means that the model does not pass the F test. It indicates that all independent variables have no influence on dependent variables, and the analysis result is meaningless at this time. From the result table below, it can be clearly seen that the model used in this paper passes the f test. Thus, the analysis results in Table 4 show that the data selected in this paper are accurate and reasonable, the results have reference value.

Table 4. Variance analysis

	Sum of squares	*df*	Mean square	*p*	*F*
regression	20.886	5	4.177	0.000	22.757
Residual error	83.703	456	0.184		
summation	104.589	461			

4. Conclusion

In conclusion, the model in Section 3 passes the F test ($p<0.05$), and there are three remarkable results; which indicates that a successful model is built. As shown in the figure: family history is the main influencing factor, followed by the measure of cholesterol and the age of study subject. The systolic blood pressure and a measure of obesity have no significant effect on the heart disease. However, this model also has some shortcoming. 463 samples may not be representative of all the people with heart disease. Overall model could be improved with more data.

References

[1] Zhen Du, et al. Journal of Xinxiang Medical University: The incidence and influencing factors of perinatal congenital heart disease in Henan Province, 2023, 40 (1): 45 - 49.

[2] Barassi M R. Micro econometrics; Methods and Applications by A. Colin Cameron; Pravin K. Trivedi. 2005.

[3] Houcui Zhang, et al. Modern psychology and Educational Statistics. 3rd edition. Beijing Normal University Press, 2009.

[4] William H G. Econometric analysis. Beijing: China Social Sciences Press, 1998.

[5] Mente A, et al. A systematic review of the evidence supporting a causal link between dietary factors and coronary heart disease. Archives of Internal Medicine, 2009, 169 (7): 659 - 669.

[6] Greenfield S, et al. Factors influencing clinicians' decisions to prescribe medication to prevent coronary heart disease. Journal of Clinical Pharmacy & Therapeutics, 2010, 30 (1): 77 - 84.

[7] Borecki I B, et al. Genetic factors influencing apolipoprotein AI and AII levels in kindred with premature coronary heart disease. Genetic Epidemiology, 1988, 5(6): 393 - 406.

[8] Pryor W W, et al. Factors influencing pulmonary compliance in heart disease. The American Journal of Medicine, 1955, 19 (1): 149 – 150.

[9] Park K O, Seo J Y. Gender Differences in Factors Influencing the Framingham Risk Score-coronary heart disease by BMI. Journal of Korean Academy of Community Health Nursing, 2014, 25 (4): 248.

[10] Moon J R, et al. Influencing factors on depression of adolescent with congenial heart disease. Heart and Lung, The Journal of Acute and Critical Care, 2009, 38: 419 - 426.