The Analysis of Stroke Incidence Based on the Multivariate Successive Regression Analysis

Gaofeng Hui[1, a]

[1]Xi'an Railway Vocational and Technical College, Xi'an, Shaanxi, 710014

Keywords: Stroke, MATLAB Programming, Residuals, Multivariate Successive Regression Analysis

Abstract. Stroke is a common serious threat to human health and life. This article uses multivariate successive regression analysis methods discussed the relationship between incidence and average pressure, pressure difference, average temperature, temperature difference and the average humidity. Through statistical data we obtained the number of cases each year for each month and the incidence of personnel, as well as the number of cases with different gender, age and occupation to obtain a four-year comprehensive monthly temperature data. Finally, summary the results of the etiology of stroke and proposed early warning and intervention programs for high-risk groups.

Introduction

Stroke (commonly known as stroke) is a serious disease that currently threatens human life and its occurrence is a long process, once the disease is difficult to reverse. Induce the disease has been linked to environmental factors, including the presence of a close relationship between air temperature and humidity. Environmental factors on the incidence of stroke were analyzed, its purpose is to conduct a risk assessment of the disease, stroke risk groups to take timely interventions, but also to the health of people not yet sick, or sub-health people have to know their level of risk of stroke to protect themselves. At the same time, through the establishment of a data model, master morbidity of law, for health authorities and medical institutions, medical reasonable allocation of power, improve diagnosis and treatment environment, the configuration of beds and medical drugs and so it has practical significance.

Data (see Appendix-C1) from some city hospitals Chinese stroke patients' information from January 2007 to December 2010 and the corresponding period of local daily meteorological data (Appendix-C2). According to data provided by the subject, carried out on the following issues:

1. Based on the patient's information, statistical description of the incidence of people.
2. Mathematical model of relationship incidence of stroke and air temperature, barometric pressure, relative humidity between.
3. Read and collect documents important features and key indicators related to stroke risk populations, combined with 1 and 2 in the conclusions of the high-risk groups to make recommendations early warning and intervention programs.

The Description of Symbols

Symbolic	Significance
x_1	The average monthly air pressure
x_2	Monthly air pressure difference
x_3	The average monthly temperature
x_4	Monthly temperature difference
x_5	The average monthly humidity
y	Monthly incidence

The Analysis of Problems and Solutions

Data processing and statistics. In the process of statistical data processing, the use of EXCEL and VB program for statistical data. Due to the large amount of data errors, so we put all the wrong data in accordance with the ratio of the total of valid information is allocated validity of information, reached the following types of statistics.

1, 2007 - 2010, the monthly number of cases and the incidence of personnel statistics are as follows:

Table 1

	2007		2008		2009		2010			
Month	Num	Incidence	Num	Incidence	Num	Incidence	Num	Incidence	Total	Mean
Jan	934	0.0705	1825	0.0958	869	0.0803	1745	0.0876	5373	1343
Feb	732	0.0552	1955	0.1026	847	0.0783	1798	0.0903	5332	1333
Mar	1018	0.0768	1917	0.1006	789	0.0729	1725	0.0866	5449	1362
Apr	1069	0.0807	1763	0.0926	860	0.0795	1700	0.0854	5392	1348
May	1090	0.0822	1771	0.093	871	0.0805	1887	0.0948	5619	1405
Jun	1030	0.0777	1499	0.0787	792	0.0732	1614	0.081	4935	1234
Jul	1013	0.0764	1496	0.0785	962	0.0889	1761	0.0884	5232	1308
Aug	1196	0.0902	1372	0.072	1783	0.1648	1683	0.0845	6034	1509
Sep	1221	0.0921	1280	0.0672	827	0.0764	1632	0.082	4960	1240
Oct	1374	0.1037	1470	0.0772	757	0.07	1722	0.0865	5323	1331
Nov	1207	0.0911	1380	0.0724	664	0.0614	1568	0.0787	4819	1205
Dec	1369	0.1033	1321	0.0693	797	0.0737	1079	0.0542	4566	1142
Total	13253		19049		10818		19914		63034	

2, Different gender (Table 2), different ages (Table 3), different occupations (Table 4) the incidence of statistics as follows:

① Annual incidence of different sexes

Table 2

Gender	2007	2008	2009	2010	Total	Proportion
Male	6571	10653	5143	11025	33392	53.93%
Female	5351	8857	4712	9611	28531	46.07%

② Annual incidence of different ages

Table 3

Age	2007	2008	2009	2010	Total	Proportion
0-44	377	563	660	600	2200	3.55%
45-59	1790	2806	2044	2987	9627	15.55%
60-74	4744	7830	3733	8032	24339	39.31%
75-89	4765	7903	3278	8573	24519	39.60%
⩾90	246	408	140	444	1238	2.00%

③ Annual incidence of different occupational

Table 4

Occupation	2007	2008	2009	2010	Total	Proportion
Farmer	8246	13338	3477	14840	39901	64.44%
Worker	905	1454	5039	1762	9160	14.79%
Retirees	2169	3381	0	2884	8434	13.62%
Teacher	44	82	213	61	400	0.65%
Fisher	19	47	28	6	100	0.16%
Medical staff	25	23	122	16	186	0.30%
Staff	56	101	976	386	1519	2.45%
Retirees	458	1084	0	681	2223	3.59%

3, Based on the statistical results, we got the crowd to stroke the following information:

(1) For different sex, the male is larger than the incidence of women;

(2) for different age groups, the 75-89 year-olds most likely the disease, followed by 60-74

year-olds;

(3) for different occupations, the highest incidence of farmers, workers, ranking second in the incidence of retirees is higher.

In summary, the results of the incidence of people, mostly elderly, and to focus on the living standards of low farmers, workers and other groups.

Analyze the relationship between the incidence of stroke and air temperature, barometric pressure, and relative humidity.

Through access to information we selected five main indicators, namely the average pressure, pressure difference, the average temperature, temperature, average humidity indicators as the five factors, through the establishment of multiple linear regression model to analyze the average pressure, pressure difference, the average temperature the relationship between temperature, average humidity and disease incidence.

1, Firstly to analyze the relationship between roughly incidence of various environmental factors, we are in the first four years of the first month of the mean pressure summed and then divided by four, the value obtained as the average pressure, the first data that is It represents the average pressure of the first month, whereby the method to calculate the pressure difference per month for four years, data on average temperature, temperature, average humidity is shown in Table 5.

Table 5

Month	Average Pressure	Pressure difference	Average temperature	Temperature difference	Average humidity	Number of cases	Disease incidence
Jan	1027.22	5.37	3.76	6.76	67.83	1343	0.0852
Feb	1022.15	5.88	6.74	7.40	70.71	1333	0.0846
Mar	1019.23	6.38	10.35	8.15	67.25	1362	0.0864
Apr	1016.18	5.73	14.86	8.45	65.47	1348	0.0855
May	1009.71	4.52	21.59	9.39	64.42	1405	0.0891
Jun	1005.70	3.52	24.48	6.70	77.16	1234	0.0783
Jul	1003.92	3.45	29.15	7.26	73.84	1308	0.0830
Aug	1006.03	3.48	28.89	6.94	74.89	1509	0.0957
Sep	1011.34	3.41	24.78	6.50	78.18	1240	0.0787
Oct	1018.21	3.83	19.44	7.58	73.17	1331	0.0844
Nov	1023.17	4.49	12.17	8.02	70.98	1205	0.0765
Dec	1023.33	5.53	6.81	7.64	66.86	1142	0.0724

2, According to the distribution of data, approximately (quadratic fit), the initial establishment y regression model with the following:

$$y = b_1 + b_2 x_1 + b_3 x_2 + b_4 x_3 + b_5 x_4 + b_6 x_5 + b_7 x_2^2 + b_8 x_4^2 + b_9 x_5^2$$

Successive multivariate regression analysis using MATLAB command program (see appendix) to solve the regression coefficient estimates and confidence intervals for the above model (confidence level check), the results of the test statistic in the table below:

Table 6

Parameter	Parameter estimates	Confidence Intervals
b_1	-3.0459	[-7.8206 1.7288]
b_2	1.3783	[-0.8687 3.6252]
b_3	4.0026	[-4.2724 12.2775]
b_4	3.7463	[-0.4289 7.9216]
b_5	-4.0407	[-7.6263 -0.4550]
b_6	3.4056	[-0.7001 7.5114]
b_7	-1.6692	[-7.3322 3.9939]
b_8	3.5626	[-0.0061 7.1312]
b_9	-3.5182	[-6.9482 -0.0881]
$R^2 = 0.8835$ F=2.8426 $p = 0.0300$		

The table shows, $R^2 = 0.8835$ the dependent variable y (incidence) of 88.35% determined by the model, F value exceeds the critical value of the F-test, $p = 0.0300$, less than $\alpha = 0.05$, so regression model was established.

Because the value of R^2 is not very ideal, and therefore we consider the possibility of a few of the five factors in the incidence of interactions on impact, so we improve the above model, the establishment of a new model is as follows:

$$y = b_1 + b_2 x_1 + b_3 x_2 + b_4 x_3 + b_5 x_4 + b_6 x_5 + b_7 x_1 x_3 + b_8 x_3 x_5 + b_9 x_2^2 + b_{10} x_4^2 + b_{11} x_5^2$$

We use MATLAB successive multivariate regression analysis program (see appendix) to solve the above model of regression coefficients give estimates and confidence intervals (confidence level $\alpha = 0.05$), the results of the test statistic R^2, F, p follows:

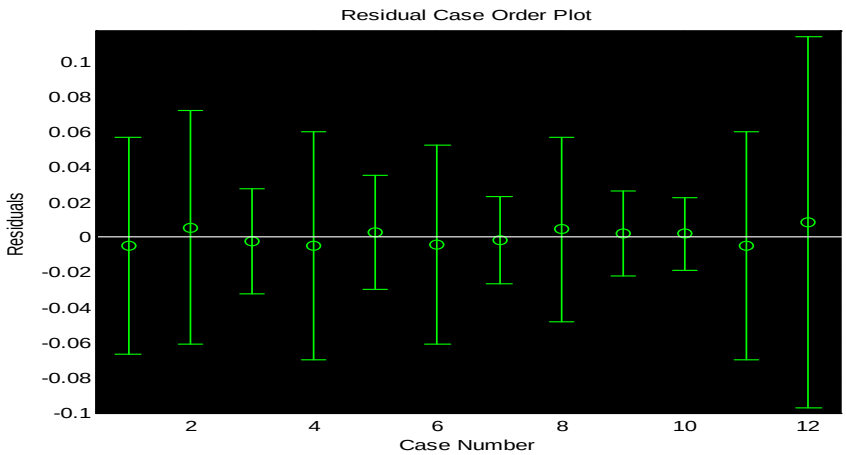
Table 7

Parameter	Parameter estimates	Confidence Intervals
b_1	1.3871	[-1.9496 4.7238]
b_2	-3.1452	[-6.1968 -0.0936]
b_3	5.2221	[1.4924 8.9517]
b_4	-2.469	[-7.2337 2.2957]
b_5	-7.9356	[-10.8244 -5.0469]
b_6	5.9298	[3.5481 8.3115]
b_7	5.13	[1.7672 8.4928]
b_8	5.0461	[-0.3312 10.4233]
b_9	-3.1303	[-5.8704 -0.3902]
b_{10}	7.3685	[3.9969 10.7401]
b_{11}	-9.5329	[-14.2548 -4.8109]
R^2=0.9997 F=333.5802 p <0.0002		

Analysis: The table shows $R^2 = 0.9997$, the dependent variable y (incidence) of 99.97% determined by the model, F value far exceeds the threshold value F-test, $p < 0.0002$, far less than α, so this model is better.

This model residual analysis test results, obtained:

The results of the second regression residuals analysis chart:



Residual analysis chart displays the selected data have been zero and no outliers, the model results are reasonable.

Based on the results of these two models are compared, there is a significant increase, and residual analysis results show that the improved model is more in line with reality than the previous model and

can more accurately reflect the different average pressure, pressure difference, the average temperature, temperature, humidity and the relationship between the average incidence.

Popular to say that, Interactions average temperature, the average temperature and average humidity between the incidence of stroke was not significant; the interaction of the average pressure, pressure difference, temperature, average humidity, barometric pressure and average temperature average between the pressure difference of amplification, amplification of the temperature difference, the amplification of the average humidity affect the incidence of stroke is significant. With the function expressed as:

$$y = 1.3871 - 3.1452x_1 + 5.2221x_2 - 2.4690x_3 - 7.9356x_4 + 5.9298x_5 + 5.13x_1x_3 + 5.0461x_3x_5 - 3.1303x_2^2 + 7.3685x_4^2 - 9.2329x_5^2$$

In accordance with the results of (a), (b), we are proposed early warning and intervention advice.

1, for the three outcomes (a) obtained: male incidence rate (53.93 percent) than women (46.07%) large; people (39.60%) 75-89 years of age most vulnerable to disease, followed by 60-74 year-olds (39.31%); the incidence of farmers (64.44%) Highest workers ranked second (14.79%), the incidence of retirees (13.62%), followed by;

For analysis, so-called high-risk groups are a high incidence of people. We can see that:

The incidence for the impact of gender is an objective reality, we can not intervene.

For the second result, we can take some preventive measures to reduce or even avoid the occurrence of stroke. For over 60 years, especially the elderly and should pay more attention and the need for prevention through the following programs:

(1) Periodic medical examinations, to understand whether the abnormal cardiovascular function was found to be active after the exception therapy;

(2) Maintain a good mood thinking, adherence to the law of the biological clock; reasonable diet, the elderly should eat high-fat, high-cholesterol as a diet should be rich in vitamins and vegetarian foods. This prevents excessive blood lipid, prevent arteriosclerosis, reduce the possibility of thrombosis;

(3) emphasis on threatened, if there is blood pressure, high blood lipids and other phenomenon, should go to regular hospital for examination.

For the third result, the peasants, the workers, the higher the incidence of retirees, so for these high-risk groups, should:

(1) Develop a healthy, scientific habits;

(2) Note that the abnormal changes in the body, regular medical examinations.

2. For the results of (b):

$$y = 1.3871 - 3.1452x_1 + 5.2221x_2 - 2.4690x_3 - 7.9356x_4 + 5.9298x_5 + 5.13x_1x_3 + 5.0461x_3x_5 - 3.1303x_2^2 + 7.3685x_4^2 - 9.2329x_5^2$$

Description Average pressure, pressure difference, temperature, average humidity, barometric pressure and average temperature average interaction between the pressure difference of amplification, amplification of the temperature difference, the amplification of the average humidity affect the incidence of stroke was significant .

Therefore, we propose the following recommendations:

(1) In every winter, high pressure, low humidity, should keep warm, alert, reduce adverse stimuli external environment and to improve their ability to adapt.

(2), in the annual seasonal temperature, pressure difference is large, it should pay attention to changes in weather conditions, to take timely measures to reduce the possibility of stroke induced by environmental factors.

3, promotion

Sales of the goods, the salary problem of company personnel, chemical catalytic reaction

References

[1] H.Y.Xiao, Practical mathematical modeling and software application,Northwestern Polytechnical University Press, 11(2008).

[2] Q.Y.Jiang, Mathematical model t, Higher Education Publisher Press. 8(2003).

[3] Information on http://www.jhhospital.com/wsfw/ShowArticle.asp?ArticleID=657

[4] Information on http://www.365heart.com/show/79521.shtml

[5] Information on http://zx.wuxi.gov.cn/web101/xwzx/dtxx/tagz/taxb/816008.shtml

[6] Information on http://baike.baidu.com/view/349682.htm