IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 15, Issue 6 Ser. VI (Nov – Dec 2019), PP 01-08
www.iosrjournals.org

Analysis and Prediction of Influencing Factors of Pediatric Pneumonia Based on Association Rules

Chao Wei, Xiaoyin Li, Zhaoqian Liu, Xiang Zhang
(School of Mathematics and Statistics, Anyang Normal University, Anyang, China,455000)

Abstract: *In this paper, association rules, logistic regression model and ARMA model are established for reflecting the internal factors of influencing the Pediatric Pneumonia. It is concluded that age, delivery, feeding style, current address and normal physical condition have significant influence on Pediatric Pneumonia. The prediction results showed that the change trend of the predicted value is basically consistent with that of the actual number.*

Keywords: *Pediatric Pneumonia, association rules, logistic regression, ARMA model.*

Date of Submission: 27-11-2019 Date of Acceptance: 12-12-2019

I. Introduction

Recent years, people's pace of life is getting faster and faster. Life pressure is increasing. Raising children has become a family focus, so the health of children can't be ignored. According to global data, more than 2 million children die of pneumonia each year, and 99 percent of children in developing countries are given great national attention because of the serious threat to their lives posed by the constant occurrence and outbreak of new and sudden infections or infectious diseases. Therefore, the analysis of the internal factors of the incidence of Pediatric Pneumonia can provide a scientific basis for the government and the center of disease control to formulate and improve the prevention strategies and measures of Pediatric Pneumonia.

II. The innovation of this paper

Although there have been some research results on the influencing factors of pediatric pneumonia, most of them only discussed from unilateral (internal or external causes), with relatively few contents. This paper studies the influence of environmental factors and internal factors such as childbirth on pediatric pneumonia from both external and internal factors, calculates 10 predicted values from September 2017 to June 2018 through ARMA model, and compares them with the original data. The results show that the predicted values are basically consistent with the original data.

The existing research results only used some simple correlation analysis methods to study the influencing factors of pediatric pneumonia. In this paper, multiple linear regression, linear proportional transformation and range method are used to analyze the influence of environmental factors on the incidence of pediatric pneumonia, and the weight of each factor is given. Then, Apriori association rule algorithm is adopted to conduct data mining of association rules for pediatric pneumonia case information. Then, logistic regression model is constructed to conduct in-depth analysis on the relative influence degree of internal causes. Finally, time series ARMA model is used to predict the number of children with pneumonia.

Analysis of Intrinsic Factors Based on Association Rules

The data used in this paper are from the real pediatric inpatient case of a hospital in Anyang City, Henan Province, 2017-2019, including the basic information of patients, the condition of hospitalization and so on. A total of 391 cases covering nearly 30 indicators, including:

5 indicators of patient's personal information (name, gender, age, place of birth, current address, etc.); 1 hospital admission indicator;

2 medical history indicators (pre-illness warning, accompanying illness);

5 clinical performance indicators (mental response, appetite, sleep, bowel, weight);

3 personal history indicators (birth, feeding, development (language, motor development, pronunciation,));

4 patient body-like indicators (body temperature, pulse rate, breathing rate, weight);

5 past historical indicators (pinin, history of infection, history of allergies, history of surgery, vaccination)

There are also some family medical history indicators and notes and admission times.

Case data is a true record of patients' information, including personal privacy information such as the patients' name. Therefore, to protect the rights and privacy of pediatric patients, Some information can not be disclosed to researchers. However, in order to maintain the integrity of the data, We use letter "N" to cover the patient's

real name. According to the actual meaning of the variable and the requirements of the association rules, the data is processed as follows:

Some cases don't have indicator information related to the subject of this research, such as gender, parturition condition, feeding patterns, etc., so such cases are deleted. Extract the required fields based on the need of the research. The focus of this article is on the fields related to the factors affecting childhood pneumonia. Therefore, the basic information fields of the patients are extracted such as "gender", "age", "birth place", "current address", "admission" and so on.

In order to explain the analysis results reasonably, it is necessary to observe the data distribution, some variable distribution description is shown in Table 1.

Table 1 Distributions Description of Some Indicator

Item	Category	Number	Proportion	Item	Category	Number	Proportion
Gender	Male	233	50.59	Mental Response	Bad	255	65.22
	Female	158	40.41		Well	136	34.79
	Total	391	100.00		Total	391	100.00
Hospitalization Condition	Severe	42	10.74	Appetite	Bad	283	72.38
	General	349	89.26		Well	108	27.62
	Total	391	100.00		Total	391	100.00
Sleep Quality	Well	65	12.62	Is weight change significant?	No	385	98.47
	Bad	326	83.78		Yes	6	1.53
	Total	391	100.00		Total	391	100.00
Defecation	Abnormal	43	11.00	Parturition Condition	Cesarean Section	233	59.59
	Normal	348	89.00		Natural Labor	158	40.41
	Total	391	100.00		Total	391	100.00
Feeding Regimens	Mixed Feeding	108	27.62	Development State	Delayed	3	0.77
	Breast Milk	261	66.75		Normal	388	99.23
	Bottle Feeding	22	5.63		Total	391	100.00
	Total	391	100.00				
Constitution	Well	41	10.49	Infection History	No	391	100.00
	Bab	350	89.51		Yes	0	0
	Total	391	100.00		Total	391	100.00
Allergic History	Yes	13	3.32	Surgery History	Yes	9	2.30
	No	378	96.68		No	382	92.70
	Total	391	100.00		Total	391	100.00
Have ever been vaccinated?	No	29	7.42	Inherited Diseases History	No	383	97.95
	Yes	362	92.58		yes	8	2.05
	Total	391	100.00		Total	391	100.00

In order to apply the association rules, the relevant variables in the case are discretized and grouped based on the requirements of the association rules and the connotation of the case indicators (Table 2).

Table 2 Discretization of Variables

indicator	groups	values
Gender	2	0=Female，1=Male
Age	4	1={1}， 2={2,3}， 3={4,5,6}，4=[7,14] Year
Current Address	9	Beiguan District, Anyang City - HAS1 Longan District, Anyang City - HAS2 Wenfeng District, Anyang City - HAS3 Yindu District, Anyang City - HAS4 Anyang City Municipal District - HAS5 Anyang County, Anyang City - HAX1 Linzhou City, Anyang City - HAX2 Anyang City, Huang County - HAX3 Tangyin County, Anyang City - HAX4
Hospitalization Condition	2	0=severe，1=general
Mental Response	2	Bad，Well
Appetite	2	Bad，Well
Sleep Quality	2	Bad，Well
Defecation	2	Normal，Abnormal
Weight	2	Yes，No
Parturition	2	Cesarean Section，Natural Labor

Condition		
Feeding Regimens	3	Mixed Feeding，Breast Milk，Bottle Feeding
Development State	2	Normal，Abnormal
Constitution	2	Bad，Well

Load "arules" package in R, use the function "apriori ()" in the package to apply association rules.
1. Generate association rules. The preprocessed data is passed to the R via the "read. csv ()" function, and the function "apriori ()"is used to generate the association rules.
2. View the rules. Use "inspect ()" to view the returned association rules, and then use the "summary ()" function to view the association rule summary.
With a certain degree of support and confidence, the lift level bigger than 1 indicates that the previous item is positively correlated with the latter item, and the lift level equal to 1 indicates that the previous item is independent from the latter, and the lift less than 1 indicates that the previous item is negatively correlated with the latter item.
In the analysis of the correlation between admission and age based on association rules, it is found that the younger the children are, the more serious the hospitalization condition is (Table 4).

Table 3 Age Group and hospitalization condition

The previous term		The Latter term	Support	Confident	Lift
{Age Group =4}	=>	{hospitalization condition =1}	0.11253197	0.9166667	1.0269819
{Age Group =3}	=>	{hospitalization condition =1}	0.23273657	0.9680851	1.0845882
{Age Group =2}	=>	{hospitalization condition =1}	0.30946292	0.9837398	1.1021269
{Age Group =1}	=>	{hospitalization condition =1}	0.23785166	0.7380952	0.8269205

Data of hospitalization condition and age find that: the children suffering from pneumonia between February 2017 and February 2019 is mainly in 1-4 years old, accounting for 76.98 percent of all, of which 1-year-olds can be affected most easily, accounting for 32.23 percent. That indicates the younger the children are, the weaker the immunity is, so that they are more likely to develop the childhood pneumonia. Intuitive conclusions can be easily obtained from the echidna chart.

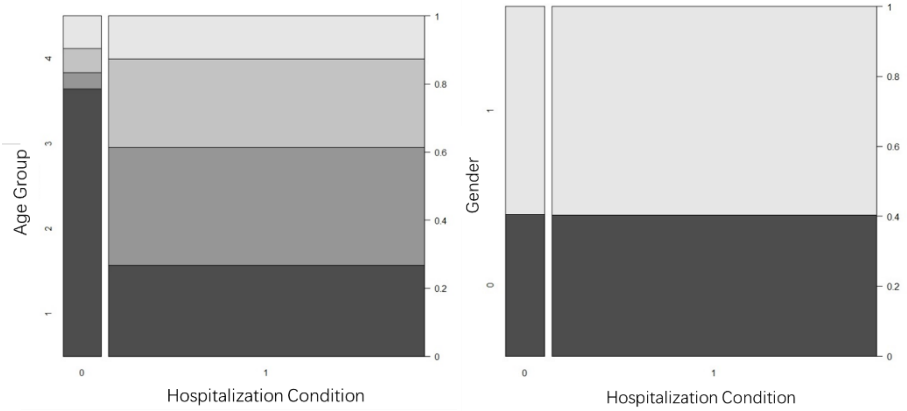


Figure 1 Echino chart of hospitalization condition, age groups and gender

Similar conclusion can be drawn in the echidna chart of gender and age. It can also be seen that: from the age group and the change of the ratio of men and women, with the increase of age, boys' physical immunity increased, the proportion of men and women suffering from childhood pneumonia gradually decreased.

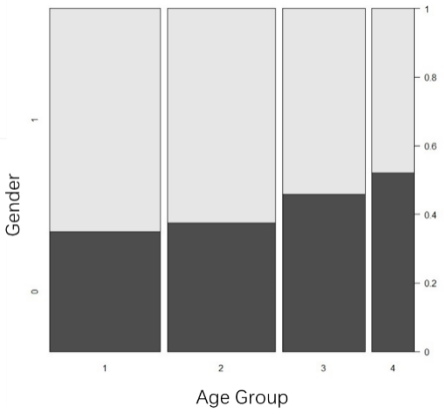


Figure 2 Echino chart of Gender and Age

WithTable 4 and Figure 2 Echino chart of Gender and Age, it can be assumed that girls are more likely to develop the disease than boys and that girls are more likely in a bad condition.

Table 4 Hospitalization Condition and Gender

The previous term		The Latter term	Support	Confident	Lift
{Gender=0}	=>	{Hospitalization Condition =1}	0.36061381	0.8924051	0.9998005
{Gender=1}	=>	{Hospitalization Condition j=1}	0.53196931	0.8927039	1.0001353

In the association rules of hospitalization condition and mental response, it was found that the mental response of children was obviously related to the hospitalization condition.

Table 5 1.4.3 Hospitalization Condition and Mental Response

The previous term		The Latter term	Support	Confident	Lift
{Hospitalization Condition =0}	=>	{Mental Response = bad}	0.08439898	0.7857143	1.2047619
{Hospitalization Condition =1}	=>	{Mental Response = bad}	0.56777494	0.6361032	0.9753582

In the analysis between hospitalization condition and current address, it was found that most patients with pediatric pneumonia are in Anyang city central district and surrounding areas, and the reasons are related to urban pollution which meets people's perception.

Table 6 Hospitalization Condition and Current Address

The previous term		The Latter term	Support	Confident	Lift
{Current Address =HAX4}	=>	{Hospitalization Condition =1}	0.02301790	1.0000000	1.1203438
{Current Address =HAS2}	=>	{Hospitalization Condition =1}	0.02301790	1.0000000	1.1203438
{Current Address =HAS4}	=>	{Hospitalization Condition =1}	0.04092072	0.9411765	1.0544413
{Current Address =HAX3}	=>	{Hospitalization Condition =1}	0.06393862	0.7352941	0.8237822
{Current Address =HAS1}	=>	{Hospitalization Condition =1}	0.08184143	0.9142857	1.0243144
{Current Address =HAS5}	=>	{Hospitalization Condition =1}	0.10230179	0.9302326	1.0421803
{Current Address =HAX2}	=>	{Hospitalization Condition =1}	0.08439898	0.7173913	0.8037249
{Current Address =HAX1}	=>	{Hospitalization Condition =1}	0.12020460	0.9038462	1.0126185
{Current Address =HAS3}	=>	{Hospitalization Condition =1}	0.35294118	0.9452055	1.0589551

It can be found that parturitioncondition is closely related to hospitalization condition, natural-labor born children tend to be healthier than the children born through cesarean section.

Table 7 Hospitalization Condition and Parturition Condition

The previous term		The Latter term	Support	Confident	Lift
{Parturition Condition =natural-labor }	=>	{Parturition Condition =1}	0.36317136	0.8987342	1.2068913
{Parturition Condition =cesarean section }	=>	{Parturition Condition =1}	0.52941176	0.8884120	0.8953269

In this association rule, two lift values are very close to 1, so that it can be considered that there is no significant correlation between hospitalization condition and weight change.

Table 8 Hospitalization Condition and Weight Change

The previous term		The Latter term	Support	Confident	Lift
{Hospitalization Condition =0}	=>	{Weight Change = No}	0.1048593	0.9761905	0.9914038
{Hospitalization Condition =1}	=>	{Weight Change = No}	0.8797954	0.9856734	1.0010345

In this association rule, two lift values are very close to 1, so that consumed that there is no significant correlation between hospitalization condition and appetite.

T**able 9** hospitalization condition and appetite

The previous term		The Latter term	Support	Confident	Lift
{Appetite =well}	=>	{ Hospitalization Condition =1}	0.24808184	0.8981481	1.0062347
{Appetite =bad}	=>	{ Hospitalization Condition =1}	0.64450128	0.8904594	0.9976207

In the association rule between hospitalization condition and feeding pattern, it is found that the bottle-fed children are more likely to develop pneumonia. The bottle-fed children tend to have lower immunity compared to breast-milk-fed children.

The previous term		The Latter term	Support	Confident	Lift
{feeding pattern = bottle-fed }	=>	{hospitalization condition =1}	0.03836317	0.6818182	0.7638708
{feeding pattern =mixed}	=>	{hospitalization condition =1}	0.26342711	0.9537037	1.0684761
{feeding pattern =breast-milk }	=>	{hospitalization condition =1}	0.59079284	0.8850575	0.9915687

III. Analysis of Intrinsic Factors Based on Logistic Regression

In this paper, there are two types of hospitalization condition for childhood pneumonia: severe and general.In the Logistic model, when the hospitalization condition is severe, Y is 0, and when the admission is general, Y is 1.According to the analysis of the association rules for intrinsic causes in Part 4, four variables, age group, parturition condition, feeding pattern, constitution are taken into account, and the symbols and their names are shown in Table 17:

Names	Symbols
Age group（1）	X_{11}
Age group（2）	X_{12}
Age group（3）	X_{13}
parturition condition（1= natural-labor）	X_2
feeding pattern（1= breast-milk）	X_{31}
feeding pattern（2=Mixed）	X_{32}
constitution（1=well）	X_4

Establish regression model

$$y=\frac{\exp(\beta_0+\beta_{11}X_{11}+\beta_{12}X_{12}+\beta_{13}X_{13}+\beta_2X_2+\beta_{31}X_{31}+\beta_{32}X_{32}+\beta_4X_4)}{1+\exp(\beta_0+\beta_{11}X_{11}+\beta_{12}X_{12}+\beta_{13}X_{13}+\beta_2X_2+\beta_{31}X_{31}+\beta_{32}X_{32}+\beta_4X_4)}$$

The Logistic model can be calculated through SPSS 24, and the confidence level is set as 95%. 7 independent variables are substituted into the model, and the analysis results are shown in.

	coeficients	Standard error	Wals	df	Sig.	Exp (B)
Age group（1）	1.485	.576	6.654	1	.010	.227
Age group（2）	1.651	.894	3.413	1	.005	5.214
Age group（3）	.820	.798	1.056	1	.004	2.270
parturition condition（1= natural-labor）	3.138	.377	.133	1	.000	1.147
feeding pattern（1= breast-milk）	2.512	.611	6.135	1	.013	4.538
feeding pattern（2=Mixed）	2.651	.750	12.498	1	.000	14.169
constitution（1=well）	1.888	1.077	3.076	1	.029	6.609
Age group（1）	.713	.791	.812	1	.367	2.039

Therefore, the model is

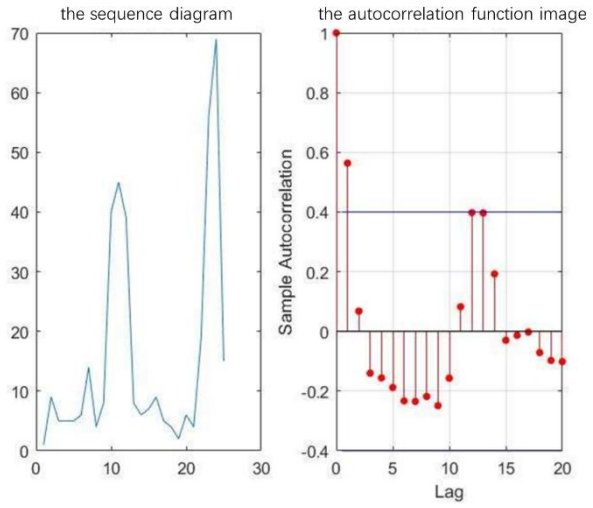
$$y=\frac{\exp(0.713+1.485X_{11}+1.651X_{12}+0.820X_{13}+0.138X_{2}+1.512X_{31}+2.651X_{32}+1.888X_{4})}{1+\exp(0.713+1.485X_{11}+1.651X_{12}+0.820X_{13}+0.138X_{2}+1.512X_{31}+2.651X_{32}+1.888X_{4})}$$

Parturition condition and feeding pattern have larger coefficients. As a result, parturition condition and feeding pattern on the influence degree of the pneumonia are the largest, second is usually constitution and age.

IV. Prediction of Childhood Pneumonia Based on Time Series Model

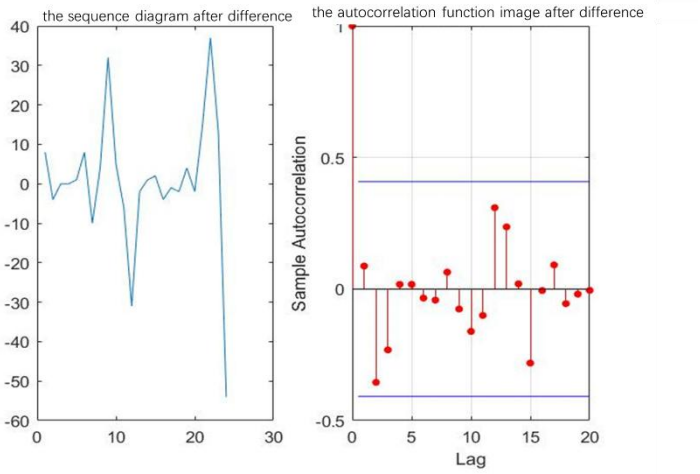
Due to the increasing influence of pneumonia on children's health, parents pay more and more attention to the probability of pediatric pneumonia in the future. In addition, one of the important task of the relevant departments of the hospital is to predict the probability of having Pneumonia and provide reference data for people to prevent Pneumonia. The severity of infantile pneumonia can be measured by the number of childhood pneumonia patients. Therefore, an accurate prediction of the number of childhood pneumonia patients in advance is of great significance for parents to prevent diseases in the high incidence period of pneumonia.Time series analysis is a theory and method to establish mathematical model by curve fitting and parameter estimation based on time series data obtained from systematic observation. The full name of ARMA model is auto regression moving average model, which is currently the most commonly used model for fitting stationary series. ARMA model is often used to fit time series and predict its future value. At present, infantile pneumonia is a common disease in infantile period. Therefore, it is of great practical significance to predict the number of children with pneumonia in advance.

The premise of establishing ARMA model is a stable time series. Therefore, before establishing the model, it is necessary to preprocess the sequence and especially to conduct stationarity test.



The timing chart of the original data showed that the number of children with pneumonia fluctuated greatly with months and the period was unstable. Moreover, it can also be seen from the autocorrelation function image (as shown in FIG. 7) that the fluctuation range of autocorrelation coefficient is also large, so that the sequence is determined to be an unstable time series. Therefore, it is necessary to perform automatic difference operation and difference sequence stationarity test on the original data. When the original data is stable, stop the difference and break out of the loop. The obtained sequence is a stationary time series.From the sequence

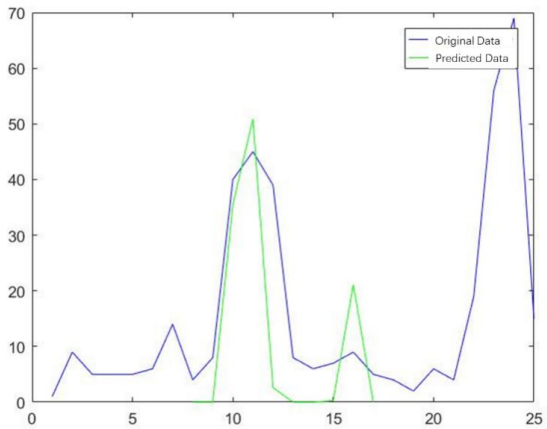
diagram after difference, it can be seen that the sequence after difference is stable and is a sequence without white errors, so the model can be solved. Before the establishment of the model, the order of the model needs to be identified, and the final order can be determined with the help of the correlation graph of time series, namely the autocorrelation function image after difference.



According to the autocorrelation coefficient displayed in the autocorrelation function image after difference, the fitting model can be carried out.
After identifying the form of the model, specify the structure of the model, namely the ARIMA (2,1,2) model, and estimate its parameters.

Parameter	Value	Error	Statistic
Constant	-0.532067	0.645733	-0.823973
AR{1}	0.675762	0.384493	1.75754
AR{2}	-0.982124	0.218172	-4.50161
MA{1}	-1.20764	0.327756	-3.68457
MA{2}	0.20764	0.396778	0.523317
Variance	204.986	103.633	1.978

The white noise and significance test of the residual sequence showed that stat =16.2660, p-Value =0.7000, c-Value = 31.4104, and P value was far greater than 0.05. Therefore, it could be considered that the residual sequence was classified as white noise sequence, with significant parameters. And the model information is adequately extracted , and the model is simplified enough.According to the above analysis, ARIMA (2,1,2) model has a good effect. Therefore, by reducing the difference, we can calculate 10 predicted values from September 2017 to June 2018 and compare them with the original data.



As can be seen from figure 9 (the x-coordinate represents the number of children with pneumonia from

February 2017 to February 2019, and the y-coordinate represents the number of children with pneumonia), the change trend of the predicted value is basically consistent with the trend of the actual number of children with pneumonia (see appendix 5 for details). From the figure, we can observe that winter is a high incidence of pediatric pneumonia. Parents should strengthen prevention and pay attention to keep warm in winter to reduce the incidence of pediatric pneumonia.

References

[1]. Seyed Mohssen Ghafari,Christos Tjortjis. A survey on association rules mining using heuristics[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,2019,9(4).
[2]. Hanzhu Xing, Effects of haze air exposure on bronchial pneumonia control in children[J]. Primary Medical Forum, 2015,19(33):4630-4631.
[3]. Sijing Yu, Jun Zheng, Using COX regression analysis to explore the influencing factors of pediatric pneumonia[J]. Practical Preventive Medicine, 2001(04):306-307.
[4]. Yan Liu, Ying Wang, Impact of climate factors on pediatric pneumonia[J].Heilongjiang Medicine,2000(03):78-79.
[5]. Guoguang Zhao.Investigation on the clinical characteristics and prognostic factors of children with streptococcal meningitisin different age groups[J].Armed Police Medicine ,2018,29(03):229-232.