Statistical Analysis on the Book Borrowing Quantity of University Library—Taking Qilu University of Technology as an Example

Yulin Huang[1], Zhongshang Liu[1], Yongxue Liu[1] and Wenrui Qu[1,*]

[1]*School of Mathematics and Statistics, Qilu University of Technology (Shandong Academy of Sciences), Jinan, Shandong 250353, China*
**Corresponding author. Email: qwr@qlu.edu.cn*

ABSTRACT
According to the library book borrowing amount data from June 2013 to December 2018 at Qilu University of Technology (Shandong Academy of Sciences), this work studied the statistical laws of data series of borrowing quantity. Based on the combination of seasonal effects, long-term trends, and random factors, a deterministic seasonal model and an ARIMA seasonal model were established for the loan quantity series. The modeling and simulation obtained well-fitting and forecasting effects, and the borrowing quantity showed periodic fluctuation and rising slowly. Furthermore, a possion logarithmic linear model was established to obtain a significant difference between male and female students in different grades. The results disclosed that the volume of books borrowed is lower in freshman, higher in sophomore and junior, and lower in the senior year. Besides, female students generally borrow more books than male students. The research results are expected to offer a useful reference for researchers majoring in book management.
***Keywords**: Library Borrowing Loan Amount; Deterministic seasonal model; ARIMA model; Possion log-linear model*

1. INTRODUCTION

The data selected in this work is the paper book borrowing data of Qilu University of Technology from June 2013 to December 2018. Simple statistical analysis was completed by writing SQL statements in Access, and the total borrowing amount (unit: time) from 2014 to 2018 was 136255, 120304, 134138, 143127, 153824, respectively. The total book borrowing amount increased by about ten thousand times in the next four years. The top four books borrowed in each year are literature (I), automation and computer technology (TP), English language (H), history and geography (K). The top 10 books borrowed in 2018 are shown in Table 1.1. Among them, the borrowing quantity of literature books (66397) is the highest, which shows that literature, history and geography books are the books that students enjoy after class. Automation and computer technology are not only professional courses for science students, but also skills for teachers and students of engineering and economic management. Therefore, TP books are also popular with teachers and students. Before sophomore year, college students should take CET-4 and CET-6. At the same time, English is the main tool for us to communicate with all over the world, and its importance is self-evident. Also, English reading is widely concerned by teachers and students.

Table 1. Top 10 books borrowed in 2018

Ranking	Class number	Class name	Borrowing quantity (times)
1	I	Literature	66397
2	TP	Automation and computer technology	11993
3	H	English language and text	11328
4	K	History and geography	9874
5	O	Mathematical science and chemistry	9716
6	F	Economic	6088
7	B	Philosophical religion	5932
8	J	Arts (popularity)	5544
9	TS	Light industry, handicraft industry and living service industry	3693
10	D	Political law	3409

The borrowing of paper books in university library are greatly affected by holidays and school curriculum arrangement. There are great differences between different months of the year and highly similar periodicity between each year. The borrowing quantity of university books has been widely concerned by many scholars. Wang Jiasheng, et al. pointed out that the borrowing quantity of university books has periodicity when they studied the borrowing quantity of university library [1]. Wu Hongyan, Xu Zhirong, et al. established the ARIMA season model for the borrowing quantity of university library [2,3]. Chen Yuehua put forward a prediction model of book borrowing flow based on particle swarm optimization RBF neural network [4]. Tan Wenhua and Sun Bao used GM model in grey theory to model and predict the trend of book borrowing in colleges and universities. Many scholars also did a lot of detailed research, which is not listed here.

Taking the monthly data of book borrowing quantity in Qilu University of Technology in recent five years as the research sequence, this work studied its periodic variation law, established ARIMA seasonal model and gave the forecast value in the future. On the other hand, Poisson logarithm linear model is used to study whether there is significant difference between male and female students in different grades. Based on the research results, constructive suggestions are provided for the purchase and borrowing management of paper books in university library combined with the actual situation.

2. INTRODUCTION TO SEASONAL MODEL

2.1. Seasonal index

When a sequence presents a regular periodic variation law, it is called seasonal effect. To investigate its seasonal variation law, the seasonal index generally needs to be calculated. The so-called seasonal index is the relative number between the moving average value of each period and the annual average value in the cycle. In order to get more accurate seasonal index, we use multiple moving average method to eliminate the influence of random factors and seasonal factors. The specific calculation steps are as follows. It is supposed that the period of the original sequence $\{x_t\}$ is m.

Step 1. In order to eliminate the influence of random factors on the current sequence value, the short-term composite moving average is used to estimate the current sequence value for the original sequence $\{x_t\}$.

$$\hat{x}_j = M_{P\times Q} \tag{1}$$

Where $M_{P\times Q}$ is the moving average of the first Q period and the second P period for $\{x_t\}$.

Step 2. In order to eliminate the influence of seasonal factors on the current sequence value, the periodic composite moving average is used to estimate the average value of the current sequence value for the sequence $\{\hat{x}_t\}$ obtained in the first step.

$$\hat{\bar{x}}_j = M_{2\times m} \tag{2}$$

Step 3. The following is the seasonal index of the current sequence value.

$$S_j = \frac{\hat{x}_j}{\hat{\bar{x}}_j} \tag{3}$$

The seasonal index applies to the following deterministic seasonal model.

$$x_j = \bar{x}_j S_j + I_j \tag{4}$$

Where, $\bar{x}_j$ is the average value of period j and I_j is a random factor.

2.2. ARIMA seasonal model

When the sequence contains long-term trend and seasonal effect, and the two effects are relatively independent, it can be assumed that they meet the additive relationship. By simple trend difference and seasonal difference, the sequence can be transformed into stable. Then, ARMA model is established for the stable sequence, which is the construction principle of ARIMA model. Its structure is as follows:

$$(1-B^D)(1-B)^d x_t = \frac{\Theta(B)}{\Phi(B)}\varepsilon_t \tag{5}$$

Among them, B is the delay operator, $(1-B)^d$ is the d-order difference. $(1-B^D)$ is the seasonal difference with step D, and ε_t is the white noise with zero mean. $\Phi(B) = 1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p$ is the autoregressive coefficient polynomial, and $\Theta(B) = 1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q$ is the moving average coefficient polynomial.

3. THE CONSTRUCTION OF BORROWING QUANTITY SEASONAL MODEL

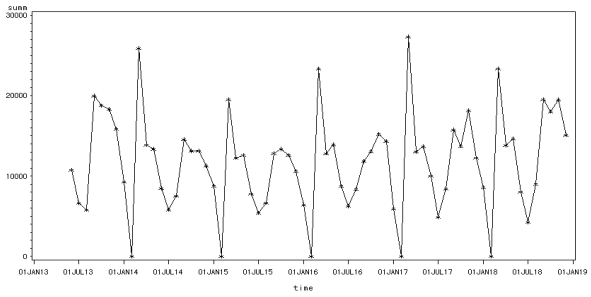

Figure 1. Borrowing quantity sequence chart

The borrowing quantity of Qilu University of Technology library was selected as the sample data from June 2013 to December 2018 to construct a time sequence, and draw its sequence diagram (Fig. 1). It is not difficult to find that the sequence fluctuates greatly and has significant seasonal effect. The seasonal model can be used to study the variation law of the sequence.

3.1. Deterministic seasonal model

The seasonal index can be obtained by using the deterministic factor decomposition method. The X-11 seasonal adjustment model in the statistical software SAS is the most popular standard method used by global statistical institutions and commercial institutions in factor decomposition. In this work, the X-11 seasonal adjustment model is used for factor decomposition to obtain the seasonal index of each period sequence (see Table 2 and Fig. 2).

Table 2. Seasonal index table

Year	JAN	FEB	MAR	APR	MAY	JUN
2013	.	.	.	.	.	75.643
2014	66.106	0.031	207.953	115.921	119.931	76.289
2015	66.771	0.031	208.290	115.541	119.634	76.221
2016	67.901	0.031	207.991	115.243	119.378	76.129
2017	67.771	0.031	209.421	115.020	120.061	75.328
2018	67.710	0.031	210.659	114.713	120.242	75.237
Avg	67.252	0.031	208.863	115.287	119.849	75.808

Year	JUL	AUG	SEP	OCT	NOV	DEC
2013	51.150	67.984	131.223	126.563	126.941	110.423
2014	50.704	68.353	130.371	124.966	128.761	110.483
2015	50.375	68.829	129.859	123.317	131.049	109.880
2016	49.663	69.116	129.855	122.197	133.613	108.362
2017	49.214	69.277	129.587	120.976	135.216	107.678
2018	48.685	69.365	129.126	120.034	136.474	107.788
Avg	49.965	68.821	130.004	123.009	132.009	109.102

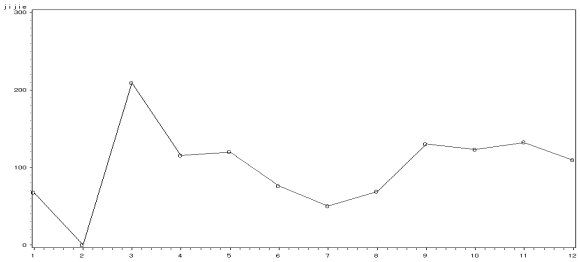

Figure 2. Seasonal index chart

It can be seen from the seasonal index of each month that the borrowing quantity has a significant seasonal effect. The borrowing quantity in March is about twice the annual average, and it is the highest peak of the whole year. The seasonal index in September, October and November is higher than that in other months. Due to March and September are the beginning of the new semester, teachers and students borrow various books frequently, resulting in the high seasonal index. However, the seasonal index in February is close to 0. Due to January, February, June, July and August are in winter and summer vacation, the borrowing activities are relatively reduced, and the seasonal index is relatively low. Especially, the whole school is in the winter vacation period in February, and the borrowing work is completely stopped. These phenomena are closely related to the school teaching law.

The deterministic factor decomposition method can not only get the seasonal index $\{S_t\}$, but also get the trend $\{T_t\}$ and the random fluctuation $\{I_t\}$ of the sequence, and thus the deterministic seasonal model of the original sequence is

$$\hat{x}_j = T_j \times S_j \tag{6}$$

The fitting effect diagram of model (6) (see Fig. 3) shows that the fitting value is highly consistent with the actual value. It shows that the model can reflect the variation law of borrowing quantity month by month, and once again shows that the borrowing quantity of university book has a significant seasonal effect.

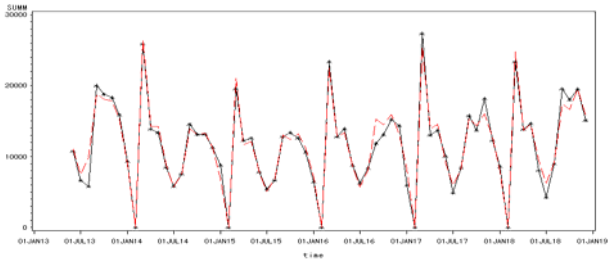

Figure 3. Fitting effect diagram of the deterministic seasonal model (6)

*The black solid line is the actual value and the red dashed line is the model (6) fitting value.

3.2. ARIMA seasonal model

It is noted that the borrowing quantity in February 2014, 2015, 2018 was 0. However, there were 5398 times of borrowing on February 26-28, 2017, and there were 1909 times of borrowing on February 29, 2016. The borrowing in February of these two years occurred at the end of the month and was caused by the opening of school. It is advisable to add them to the borrowing quantity in March.

Although the above-mentioned deterministic seasonal model can well reflect the actual value, it can't get the future forecast value of borrowing quantity. Considering that the sequence is the result of long-term trend, seasonal effect and random factors, ARIMA seasonal model should be established.

The first-order difference and the seasonal difference with step size of 11 are applied to the sequence, and the sequence diagram of the difference sequence is basically stable. The ADF unit root test p value is less than 0.05, which can be determined that the difference sequence is stable. In the white noise test of the difference sequence, the corresponding p-values of df = 6, 12 are 0.0024 and 0.0059, respectively, which are less than 0.05, indicating

that there is a strong correlation between different sequences. Therefore, it is necessary to build a model to extract their correlation. The auto-correlation coefficient and partial correlation coefficient of the different sequences are calculated (see Fig. 4 and Fig. 5). The correlation coefficient chart shows that they are significantly non-zero in the first order. After the second order, the autocorrelation coefficient almost falls within 2 times of the standard deviation, showing the nature of the first order truncation. Also, the partial correlation coefficient falls within 2 times of the standard deviation after the second order. However, the distance from 0 is slightly larger and the speed towards 0 is relatively slow, showing the nature of tailing. According to the above characteristics of the correlation coefficient, the relative optimal model selected from multiple alternative models (see table 3) is ARIMA (0, (1,11), 1) combined with the parameter test of the model, the white noise test of the residual and the AIC statistics:

$$(1-B^{11})(1-B)x_t = (1-0.78705B)\varepsilon_t, \quad (7)$$

$$\sigma = 2333.736$$

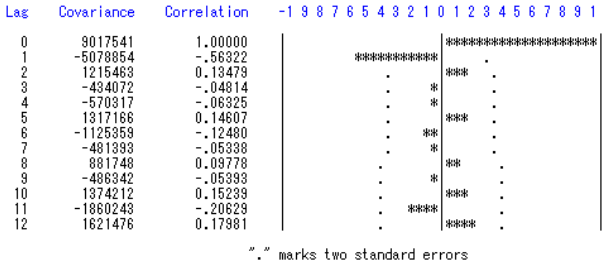

Figure 4. Auto-correlation coefficient diagram

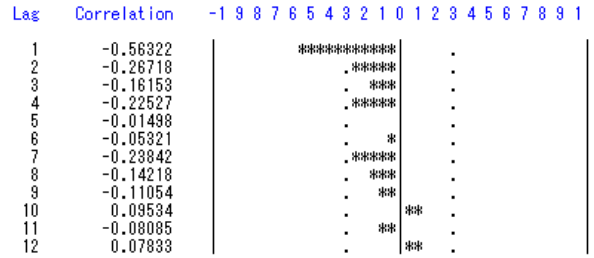

Figure 5. Partial correlation coefficient

Table 3. Comparison of models

Models	AIC	Parameter testing	Residual white noise test
1 p=0, q=1	918.4063	θ_1, p<0.0001	df=5, p=0.7987
			df=11, p=0.8452
2 p=1, q=0	925.5907	ϕ_1, p<0.0001	df=5, p=0.2419
			df=11, p=0.4394
3 p=1, q=1	920.1825	ϕ_1, p=0.6216	df=4, p=0.7075
		θ_1, p<0.0001	df=10, p=0.8123
4 p=0, q=2	920.2402	θ_1, p<0.0001	df=4, p=0.7004
		θ_2, p<0.7166	df=10, p=0.8064

3.3. Prediction of ARIMA model

From Table 3, the parameter test and white noise test of model 1 and model 2 have passed, and they are all effective models. The AIC of model 1 is smaller than that of model 2, which shows that model 1 can interpret data better. However, does model 1 have a good prediction effect? Therefore, we use the data from June 2013 to June 2018 for modeling and take the data from July 2018 to December 2018 as the test set to test the prediction effect of the model. First of all, the data from June 2013 to June 2018 are used for modeling. The prediction value of 2018.7 is given by the model, and the actual value of 2018.7 is supplemented before modeling. Then, the prediction value of 2018.8 is given by the model. In this way, all the predicted values for the six months from July 2018 to December 2018 (the data in brackets are the upper and lower limits of confidence) are obtained. Compared with the actual value, the relative error and average relative error are calculated (see Table 4).

Table 4. Model prediction

2018	July	August	September
Actual value	4213	8976	19554
Predicted value (confidence interval)	4764 (171, 9356)	8203 (3660, 12747)	15675 (11177, 20173)
Relative error	0.131	0.086	0.198
2018	**October**	**November**	**December**
Actual value	18016	19493	15101
Predicted value (confidence interval)	14400 (9813, 18987)	19599 (14946, 24253)	13697 (9092, 18301)
Relative error	0.201	0.005	0.093

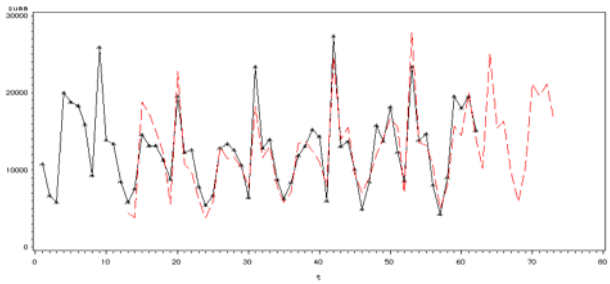

Figure 6. Model (7) predictive effectiveness diagram

*The black solid line is the actual value, and the red dashed line is the model (7) fitting value and the predicted value.

4. POISSON LOGARITHMIC LINEAR MODEL

Whether there is a significant difference in the amount of borrowing by boys and girls in different grades will be investigated. Accordingly, more than 27000 students were selected from the whole school. Considering the equilibrium of indicators at all levels, 200 boys and girls

were randomly selected from 2015, 2016, 2017 and 2018, with a total of 1600 people. By observing their borrowing quantity in 2018, the borrowing quantity of a student changes from 0 to 140, with an average of 4.6575 times/person. Among them, the proportion of borrowing quantity 0 is 44.81% (i.e., about 44.81% of students did not borrow one book in 2018). The proportion of borrowing quantity 1-14 is 47.63%, and the proportion of more than 15 is 7.56%. After preliminary arrangement, the following frequency statistical Table 5 is obtained.

Table 5. Borrowing quantity statistics for 1600 students in 2018

Total loans	Grade 2015	Grade 2016	Grade 2017	Grade 2018	Subtotal
Boys	550	886	833	672	2941
Girls	721	1263	1487	1040	4511
Subtotal	1271	2149	2320	1712	7452

Throughout the four years of university, the borrowing behavior of freshmen is the least. Then, the number of sophomores and juniors increases year by year, and the number of senior declines. This regularity is closely related to the four-year learning activities of college students. Freshmen mainly learn basic courses, and the knowledge is relatively narrow and the ability of self-study is insufficient, resulting in low borrowing quantity. With the continuous learning of major courses in sophomores and juniors, students' knowledge range has been further broadened, and the desire for knowledge and the ability of self-learning have also been enhanced. Therefore, the library has become the main place for students to acquire knowledge, and the borrowing quantity has gradually increased. In the senior year, most of the students are busy with postgraduate entrance examinations, civil servant examination or off campus internship, etc. they do some preparatory work for the society in the future, and the book borrowing quantity has declined. Looking at the borrowing situation of girls and boys, no matter which period, the borrowing quantity of girls is higher than that of boys, which is 1570 / 2941 = 53.4% higher overall. Especially in the junior year, the borrowing quantity of girls is as high as 78.5% compared with that of boys, indicating that girls are more enthusiastic about learning than boys. This is a common phenomenon in various domestic universities at present.

In order to discuss the influence of girls and boys in different grades on book borrowing quantity, Poisson logarithm linear model of book borrowing quantity should be established. In 2018, the students' book borrowing quantity was a uniform Poisson distribution. Due to the influence of different grades and genders, the mean value λ of the Poisson distribution will change with each other. The mean value λ of the positive real number will be logarithmically transformed into the whole real number field, and then a linear regression will be established. Therefore, the structure of the Poisson logarithmic linear model is as follows:

$$\ln \lambda = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \tag{8}$$

Where λ is the mean value of Poisson distribution, and $x_1, \cdots, x_p$ are the independent variables. β_0 is the intercept term of the model, and $\beta_1, \cdots, \beta_p$ are the coefficient of the independent variables. Due to two levels of gender and four levels of grade, the dumb variables are defined as follows:

$$x_1 = \begin{cases} 1, & Girls \\ 0, & Otherwise \end{cases}, \quad x_2 = \begin{cases} 1, & Grade\ 2016 \\ 0, & Otherwise \end{cases},$$

$$x_3 = \begin{cases} 1, & Grade\ 2017 \\ 0, & Otherwise \end{cases}, \quad x_4 = \begin{cases} 1, & Grade\ 2018 \\ 0, & Otherwise \end{cases}$$

The parameter estimation of the model (8) is obtained by using the GLM ( ) function of R software.

Table 6. Parameter estimates

	Estimate	Standard error	z value	Probability (>|z|)
(Intercept)	0.91951	0.03151	29.186	< 2e-16 ***
Female	0.42777	0.02370	18.049	< 2e-16 ***
Grade 2016	0.52520	0.03538	14.842	< 2e-16 ***
Grade 2017	0.60176	0.03490	17.244	< 2e-16 ***
Grade 2018	0.29786	0.03703	8.045	8.64e-16 ***

The significance test p value of parameters is less than 10^{-16}, indicating that all coefficients are highly significant and non-zero (i.e., the borrowing quantity of girls in the same grade is significantly higher than that of boys). In the category of girls, the borrowing quantity of different grades is significantly different. Similarly, the borrowing quantity of different grades is also significantly different in the category of boys. From the above results, Poisson log linear model is obtained:

$$\ln \lambda = 0.91951 + 0.42777 x_1 + 0.5252 x_2 + 0.60176 x_3 + 0.29786 x_4 \tag{9}$$

The deviation of the zero model is 16771, with 1599 degrees of freedom. The residual deviation of model (9) is 16068, with 1595 degrees of freedom. The AIC is 19167, and the dispersion parameter is 1, indicating that no dispersion has occurred. The likelihood ratio test (see table 7) is used to compare whether the difference between the model (9) and the zero model is significant. The following output results show that the model (9) is significantly better than the zero model. Once again, it shows that the influence of grade and gender on borrowing quantity is very significant.

Table 7. Likelihood ratio test

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	1599	16771			
2	1595	16068	4	703.56	< 2.2e-16 ***

5. CONCLUSION AND SUGGESTION

This work established the deterministic seasonal model and ARIMA stochastic seasonal model of library borrowing quantity at Qilu University of Technology. At the same time, the Poisson logarithmic linear model of borrowing quantity was established. It can be concluded that the book borrowing quantity of university has obvious periodicity, which will fluctuate and rise slowly in the next year. There are significant differences in the amount borrowed by boys and girls at different grades. Based on these research results, some suggestions are given combined with the characteristics of university library.

First, the purchase of books should be increased. The borrowing situation of teachers and students reflects the readers' learning tendency and demand from the side. Students have a huge demand for literature books and great potential for computer and English books. The library should increase the purchase of these books, and purchase the books published recently, issued by the core publishing house and compiled by famous scholars to meet the readers' reading needs.

Second, it should reasonably arrange the management of the library. The book borrowing quantity shows the periodic variation law. For example, it is the peak period for students to borrow and return books in the opening season (March and September), and the borrowing machines and staff can be increased appropriately. At the end of the semester, some borrowing machines can be shut down properly to reduce borrowing costs and improve work efficiency.

Third, it is necessary to carry out regular library publicity activities and launch personalized service methods. Freshmen should be introduced to library collection resources and borrowing methods. The library should regularly release the purchased paper books and digital resources, push new books and popular books in real-time, and improve the mobile library client with the help of WeChat, etc. It is convenient for students to know the library book resource information in time and increase the students' borrowing quantity.

Fourth, colleges and universities should pay attention to the cultivation of students' reading literacy. It is necessary to carry out various social practice, scientific and technological competition activities, enrich students' daily life, and activate the atmosphere of academic research. Boys who are addicted to online games should be given attention and education, and relevant measures should be formulated to encourage them to join the team that loves learning and reading.

ACKNOWLEDGMENT

1) National college student innovation and entrepreneurship training program project: Research on the reading tendency of university students based on data mining(S202010431018);
2) Key teaching and research projects at school level of Qilu University of Technology (Shandong Academy of Sciences): Constructing mathematics linking courses between middle schools and universities, deepening credit system and promoting the development of new engineering (2019ZD28).

REFERENCES

[1] Wang Jiasheng, Mou Xiaoguang, “Statistical Analysis on the Distribution of Borrowing Quantity in University Library Based on Time Series, ” Journal of Agricultural Library and Information Science, vol.23 (4), pp. 72-75, 2011.

[2] Wu Hongyan, “Seasonal Prediction Model of Book Borrowing Quantity Behavior, ” Library and Information Work, vol.11, pp.98-101, 2007.

[3] Xu Zhirong, Chen Qian, Guo Liuxiao, “Research on the Borrowing Quantity of University Library Based on Time Series Analysis—Taking Jiangnan University Library as an Example, ” Journal of Agricultural Library and Information Science, vol.30 (10), pp.107-110, 2018.

[4] Chen Yuehua, “Application of Neural Network in Prediction of Book Borrowing Quantity in University Library, ” Modern Electronic Technology, vol.40(19), pp. 115-118, 2017.

[5] Tan Wenhua, “Grey Prediction of Book Circulation in University Library, ” Modern Information, vol.10, pp.119-120, 2003.

[6] Sun Bao, Wang Zhili, Liu Lin, “Research on the Prediction Model of Borrowing Quantity in University Library Based on GM, ” Modern Information, vol.4, pp.186-188, 2008.

[7] Zhang Qunche, Zhu Jiafu, “Nonlinear Characteristics of Book Circulation Statistical Time Series , ” Journal of Southwest Normal University (NATURAL SCIENCE EDITION), vol.2, pp.149-152, 2009.

[8] Wang Jing, Li Pishi, “ Chaos Prediction of Book Borrowing Flow in University Library Based on Lyapunov Index , ” Modern information, vol.9, pp.7-10, 2009.

[9] Zhao Shaojun, Yang Weiping, “Computer Statistical Analysis on Book Borrowing, ” University Library Work, vol.1, pp.70-73, pp.89, 2002.

[10] Song Ailin, Zhu Yunxia, Yuan Siben, “Acquisition Improvement of University Library Based on Multidimensional Data Analysis, ” Library Forum, vol.5 (39), pp.95-102, 2019.

[11] Tian Mei, “Prediction Research on Book Borrowing Quantity Based on Chaotic Time Series Model, ” Library Theory and Practice, vol.7, pp.1-3, pp.26, 2013.

[12] Gong Xingang, Zhang Ya, Shen Lijuan, “Analysis and Prediction of Book Borrowing Historical Data, ” Library and Information Work, vol.59 (Increase 1), pp.161-165, pp.87, 2015.

[13] Wang Juan, “Analysis of Seasonal Patterns and Influencing Factors of Book Circulation in University Library, ” Journal of Library Science, vol.6, pp.41-43, 2005.

[14] Cai Zhicheng, Zhang Genbin, “Reader Distribution Density and Regression Analysis of Library Circulation System, ” Information Journal, vol.3, pp.114-115, 2004.

[15] Wang Yan, “Applied Time Series Analysis, ” Beijing: China Renmin University Press, 2015.

[16] Wu Xizhi, “Application Regression and Classification Based on R, ” Beijing: China Renmin University Press, 2016.

[17] He Xiaoqun, “Multivariate Statistical Analysis, ” Beijing: China Renmin University Press, 2015.

[18] Wang Shuang, Ma Jingyi, “ Research on Innovation Output Capacity of Enterprises Based on Poisson Logarithm Linear Model,” Statistics and Decision, vol.19, pp.59-6, 2014.