ISSN: 1813-4890

Research on Consumer Preference based on Regression and ARIMA Model

Xilai Chen[1], Ling Jiang[2,*], Yimeng Ding[1], Yahui Li[1]

[1]School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, Anhui, China

[2]School of Economics, Anhui University of Finance and Economics, Bengbu, Anhui, China

*tiaowudejiangling@163.com

Abstract

With the structure of traditional consumerism and the rise of the "new retail" industry, consumers' purchasing needs and purchasing motives have undergone drastic changes, and the sales idea centered on "cost-effectiveness" is no longer suitable for new market conditions. Faced with this situation, merchants are bound to establish a new predictive model to describe consumers' purchasing hearts and determine the direction of future transformation and structural optimization. This paper analyzes the statistical data of the merchants and uses Excel, SPSS and other software to carry out mathematical modeling, to explore the factors that affect sales and predict future sales. By sorting and filtering the statistical data, this paper extracts the 50 kinds of SKCs with the highest sales volume, and analyzes the influencing factors of their sales in certain festivals. After removing the influence of time series autoregression, we determined that there is a variable influence on the festival itself through variance analysis, and then used SPSS to carry out the correlation test and regression analysis of the variables to study the economic significance of the coefficients. Finally, it is found that merchants will actively sell products with high inventory, and consumers tend to buy products with a price close to the actual price.

Keywords

Regression Models; ARIMA Models; ANOVA; MAPE.

1. Introduction

With the upgrading and development of the overall consumer market, the characteristics of stratification, personalization, niche, facilitation and communityization of consumer demand will become more prominent. How to better meet new consumer demands has become the main driving factor for retail reform. Facing the current new consumer demand, the transformation of new retail will definitely pay more attention to consumer experience. Redefining the new value of retail stores and creating new stores with more experiential and social attributes is an important direction of change. Facing the continuous development of my country's consumer market, the consumption pattern in the market has gradually changed from "things-oriented" to "customer-oriented". Only accurate and proactive management methods can help companies to actively communicate with customers, preemptively occupy positions to intercept competitors, accurately provide products that meet customer needs, provide personalized promotions and preferential plans that meet customer value, improve customer experience, and improve business efficiency. Driven by this model, new retail enterprises are moving towards a multi-variety and small-batch production model. There are thousands of products entering the market. How to choose the best combination of products among the many products is very important for merchants. It is a great problem, and at the same time, the mass

ISSN: 1813-4890

production of commodities has also brought great problems to the inventory management of the retail industry.

To solve the problems of the retail industry and promote the development of the retail industry, we will provide accurate demand forecasts in different aspects with the historical sales data of complex levels and various categories. In order to better allocate the types and inventory of commodities, save costs, meet consumer needs, and pursue maximum benefits.

2. Literature Review

In recent years, business model theory has been developing, and its content and research perspectives have become more and more rich, but its definition has always been controversial. Based on the summarization and analysis of early domestic and foreign literature, there are three main perspectives of business models: From the perspective of operation, American management scientist Hammer believes that a business model is a unique operating model formed in the process of production and operation of an enterprise, that is, an enterprise A series of activities to achieve business goals [1]; from the perspective of value, Zhixin Qian believes that enterprises should use their own business models to bring out the value of the enterprise, realize the accumulation and transmission of value, and finally achieve the goal of maximizing enterprise value [2]. From the perspective of the overall system, Min Luo et al believe that based on the Internet background, the business model has evolved into a description of the overall system of community, platform, resource integration, and product design [3].

Scholars have done some research on the business model of new retail enterprises. FukunYang and others took campus new retail as the research object, analyzed the innovation of campus new retail business model under the Internet background, and deepened the concept and model of new retail [4]. Chaoran Zhang selected three squirrels as representatives of new retail to study the impact of digital empowerment on business model innovation of new retail companies from the nine elements of the business model, emphasizing the important role of digitalization in promoting [5]. From the perspective of the entire market, Xu Hong and Li Li studied how traditional retail enterprises should realize business model innovation in the context of new retail, aiming to provide an innovative path for retail enterprise business model innovation [6].

3. Model Building

3.1. Basic Assumptions of the Model

The flow of people in the shopping mall is only affected by the time period and is not affected by irreversible factors such as changes in the surrounding environment or policy changes. Inflation and income growth are synchronized.

3.2. Preparation of the Model

3.2.1. The Analysis of the Model

Analyze the impact of various related factors on the sales volume of target skc (stock keeping color) during the four holidays of National Day, Double Eleven, Double Twelve and New Year's Day in 2018. We analyzed the relevant data given and extracted some representative economic variables. Then observe a large number of skc time series data to study whether it has the periodicity of time series. After removing cyclical effects, we performed an ANOVA on skc and determined that festivals have an effect on sales. After that, the selected variables are tested for correlation, and then regression analysis is performed to obtain the results and explain them in conjunction with economics.

ISSN: 1813-4890

3.2.2. Preparation of the Model

First, data preprocessing is performed to screen out the top 50 sub-categories with cumulative sales from July 1, 2018 to October 1, 2018. We use Excel for coarse screening, and the screening results are shown in the following table.

Table 1. Top 50 SKCs by Sales

796573650638	390572118052	902573870176	696572115379
396573870660	696572117667	490572114587	102573321111
690572114917	402573650792	902572118613	396573870484
996573870572	502573650759	896572118954	296572114433
996572333025	608572774763	496572117700	596573650847
802573650495	296572551749	596572118723	702573870715
496572333586	102572118283	902573320275	702572225797
302572775577	690572118833	433572114917	696573650781
408572660891	496573321265	896573870660	396572118745
499572333179	596572118162	602572551672	802573650803
196573870759	802573870715	391572223674	702573870858
100572118316	496572118514	890572114477	790572118833
102573650748	702573650594		

Note: All data in this article comes from the attachment of the 2020 Mathorcup College Mathematical Modeling Challenge.

We randomly select three target subclasses and observe their time series images:

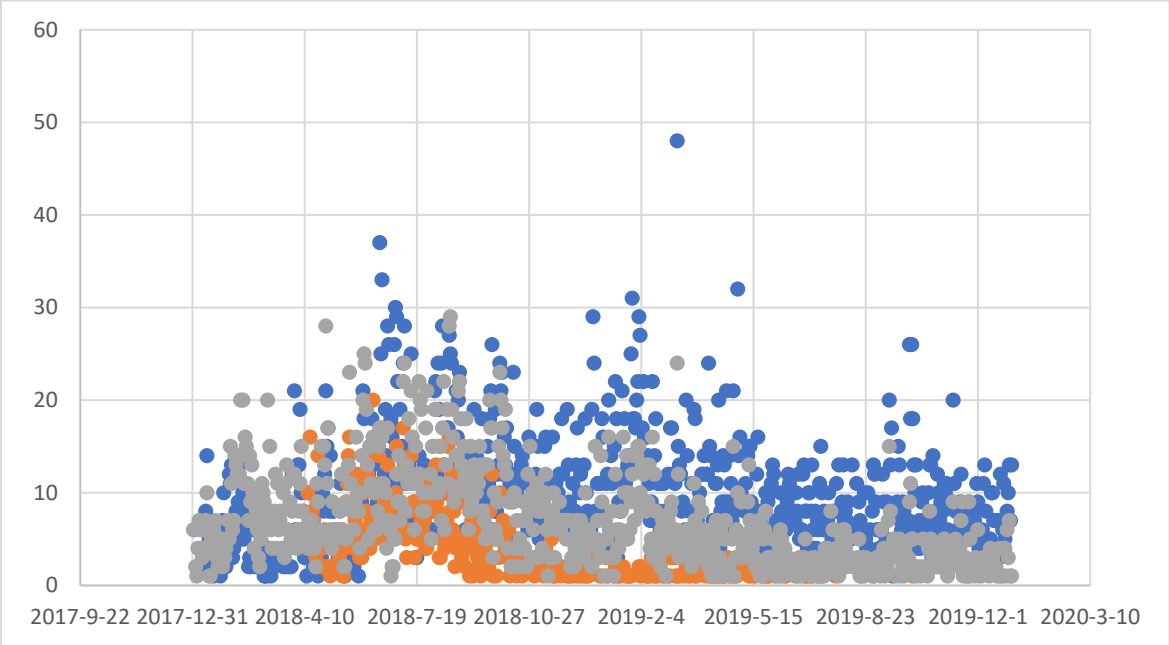

Figure 1. Time series image of random subclasses

Note: Blue is 796573650638, Orange is 196573870759, Gray is 690572114917

Obviously, we can find that its trend has no cyclical characteristics, coupled with our assumption that the inflation rate and income growth rate are synchronized, we can think that it does not need to consider its autoregressive phenomenon in the time series. Next, we perform

ISSN: 1813-4890

variance analysis on skc during festivals and non-holidays, and the results are shown in the following table.

Table 2. ANOVA Results Table

Skc number	P-value
996572333025	0.000356601
996573870572	0.00000000012
796573650638	0.00000001

If the P value is less than 0.05, it means that the impact is significant. Obviously, whether the holiday has a huge impact on the sales volume, the preliminary preparation has been completed.

4. Model Solution

4.1. Numeral Calculations

Combining the requirements of the question and the actual data provided in the table, we have sorted out three groups of indicators that may affect sales, namely discount, storage and holiday length. Next, we use SPSS to test the correlation between each indicator and its corresponding skc sales, and randomly select ten results as shown in the following table.

Table3. Correlation test table p-values

Skc number	1	2	3	4	5	6	7	8	9	10
Discount	0.122	-0.4	0.188	-0.57	0.539	-0.34	0.513	0.432	0.478	0.239
Warehousing	0.844	0.797	0.712	0.417	0.124	0.612	0.274	0.768	0.679	0.804
Length of Vacation	0.750	0.743	0.630	-0.99	0.411	0.475	-0.41	0.685	0.647	0.658

Note: The above 10 skc numbers are 499572333179, 102573650748, 490572114587, 433572114917, 890572114477, 696572117667, 8025736510803, 396573870484, 49657114917, 696572117667, 8025736510803, 396573870484, 64962555

We select some of the parameters that pass the Pearson test for regression, and the results are shown in the following table:

Table 4. Regression Results

Skcnumber	1	2	3	4	5
Discount				-28.199/0.09	18.812/0.212
Warehousing	0.096/0.026	0.022/0.129	0.093/0/165	0.034/0.289	
Length of Vacation	1.035/0.245	0.308/0.461	-0.473/0/66		0.106/0.651
Intercept	-36.6/0.028	3.103/0.321	15.92/0.661	24.001/0/215	-10.356/0.40
Skcnumber	6	7	8	9	10
Discount		21.039/0.153	9.036/0.691	28.011/0.76	
Warehousing	0.075/0.165	-0.011/0.582	0.063/0.145	0.007/0.637	0.083/0.044
Length of Vacation	0.132/0.743	-0.454/0.292	0.431/0.490	0.469/0.421	0.052/0.898
Intercept	14.937/0.23	-10.07/0.395	12.549/0.55	27.760/0.68	16.993/0.07

Note: The table content is coefficient/significance (sig.)

ISSN: 1813-4890

According to the correlation test and regression results of the above ten SKCs, it can be found that the three parameters of discounted storage and vacation length have great differences in the influence of different SKCs. These differences are reflected in real life as the demand differences caused by the essential differences of products For example, necessities are relatively less sensitive to discounts, while luxury products are relatively more sensitive to discounts. The tourism and entertainment industries are relatively sensitive to the length of holidays.

Based on the research results and the actual situation as well as the consideration of the data structure, we decided to use the price factor to distinguish the various skc indicators, and then introduced the price as a parameter into the model. The correlation test results are as follows:

Table 5. Correlation tests for overall skc

Parameter	Pearson correlation coefficient
In Stock	0.601
Holiday Time	0.207
Tag Price	-0.346
Actual Price	-0.324

After comparison, it is found that except for the holiday time, the rest of the parameters have passed the Pearson test, so we select the remaining three parameters to perform regression on them, and the regression analysis results are as follows:

Table 6. Regression Results

Parameter	Coefficient	Significance
Intercept b	3.496	0.000
Warehousing i	0.018	0.000
Actual price P_1	0.011	0.469
Label price P_0	-0.026	0.048

4.2. Result Analysis

Combining the analysis and results in 4.1, we can get the regression equation about skc as:

$$y = 0.018i + 0.011p_1 - 0.026p_0$$

This conclusion is in line with the principles of economics. People's purchasing desire is negatively related to the price, which is in line with the consumption psychology of rational economic people. The positive correlation between the sales price after the same discount indicates that the profit of the merchant itself is low, which is lower than people's expectation of cost. The positive correlation between warehousing and sales also meets the needs of merchants to reduce inventory costs, and also meets the assumptions of rational economic people.

5. Policy Suggestion

Based on the regression results of parameter derivation, we give the following suggestions: try to increase the inventory to ensure that there is enough inventory at any time; the listed price and the actual selling price will jointly affect the value judgment of consumers. According to the data given by your company, we found that the consumption The purchase desire of consumers is negatively correlated with the listed price, that is, the listed price guides consumers'

horizontal comparison, so it is recommended to lower the listed price, which is lower than that of similar products in other stores; while the actual selling price is positively related to consumers' purchase, that is, the actual selling price guides Consumers' value judgments, in other words, consumers believe that merchants with high discounts earn higher profits than merchants with low discounts, so it is recommended to increase the actual price, reduce discounts, and strengthen service quality control and other factors to further create "quality control". Excellent quality, good service, perfect after-sale service, and low profit", thus consolidating consumers' value judgments, turning the phenomenon of low discounts into the concept of good quality control, making consumers think that their purchases are worth the money, and ultimately improving sales.

In addition, the company can add some indicators into the new model, such as the flow of people at the sales location or the policy tendency of the local government to quantify the long-term impact on the time period, or introduce the income growth rate and inflation rate to adjust the real income value of the product , you can also introduce product cost to observe the difference between consumers' estimated cost and actual product cost, adjust pricing strategy, and focus on selling products with high gross profit margins; you can even count the online discussion heat and entry search rate of small categories of goods, etc. Determine the preferences of target customer groups, etc.

Acknowledgments

This work is supported by the project of School of Statistics and Applied Mathematics, "*Research on the Collaborative Evolution of Data-Driven Industrial Structure and Chinese Residents' Consumption Capacity*" (Grant No: XSKY2137ZD), and the project of Anhui Provincial Federation of Social Sciences Project "*Data-Driven* Research *on the Improvement Path of Residents' Consumption Level in Anhui Province*" (Grant No: 2020CX098).

References

[1] Hammer M, Champy J. Reengineering the Corporation: A Manifeste for Business Revolution[J]. Elsevier, 1993,36(05):90-91.

[2] Qian Zhixin. An Analysis of Innovative Business Models [J]. Modern Management Science, 2007(08):3-4+105.

[3] Luo Min, Li Liangyu. Business Model Innovation in the Internet Age: Value Creation Perspective [J]. China Industrial Economy, 2015(01): 95-107.

[4] Yang Fukun, Liu Junjun, Wei Jin. Research on campus "new retail" business model innovation under the background of "Internet +" [J]. Business Economics, 2021(12):57-59.

[5] Zhang Chaoran. Research on digital empowerment of new retail business model optimization--taking three squirrels as an example [J]. Modernization of shopping malls, 2021(21):1-3.

[6] Hong Xu, Li Li. Research on the innovation path of retail enterprise business model under the background of new retail [J]. China Collective Economy, 2021(26):75-76.

[7] Fan Xiang, Li Jianling, Zhang Jingjing. Analysis of business model innovation of new retail enterprises under the background of digital economy: Taking Hema Xiansheng as an example [J]. Modern Business, 2022(09):30-32.

[8] Chen Lingjie. Risk analysis in the operation and management of fresh food e-commerce under the new retail model: Taking Hema Fresh and Super Species as examples [J]. Logistics Engineering and Management, 2021, 43(12): 66-68.