Team Number :	apmcm24205869
Problem Chosen :	C

2024 APMCM summary sheet

In response to the first question, we analyzed the enterprise market and consumer groups of China's pet industry in recent years, and visualized the data through a line chart. Using the scale of the pet industry, investment and the number of pet cats and dogs from 2019 to 2023, the ARIMA and Holt models are constructed to predict the trend in the next three years. The results show that although the investment has decreased slightly, the scale of the industry has increased, and the overall situation is improving; The number of pet cats has increased, and the number of pet dogs has decreased.

In response to question 2, we analyzed the number, food and market size of pets in the United States, Germany, France and other countries, and found that the number of pet cats and dogs has increased linearly and accounted for overseas market players. Using the Holter linear trend method to predict the development trend of pet food in the next three years, the results show that except for the Americas, the global pet food and indicators are on the rise, and it is inferred that the global pet industry will develop well in the next few years.

In response to the third question, we analyzed the current situation of pet food production and export in China, and found that the situation is good and the results are remarkable. Using the Holter linear trend model to predict the trend in the next few years, we are optimistic that China's pet industry will continue to grow, with production, exports and sales all rising.

In response to the fourth question, we analyze the factors that affect the export value of pet food in China, including domestic and foreign economic policies and trade volume. Through principal component analysis and multiple regression analysis, a quantitative analysis model was established, and it was found that the U.S. economic policy had a great impact on China's pet food exports. Therefore, it is recommended to formulate economic policies that are conducive to exports to the United States to promote the development of China's pet food industry.

Key words: Pet industry; Market size; ARIMA model; Holt model; Principal Component Analysis; Multiple regression analysis

Contents

I. Introduction

1.1 Background

In the past few decades, with the rapid development of the world economy and the continuous improvement of per capita income, people's consumption concepts have undergone a fundamental change, with the strengthening of personalized and emotional consumption concepts, and consumers are more willing to pay for goods that provide high emotional value. Under this background, the pet industry, a new emerging industry, has risen rapidly around the world with an unstoppable momentum, presenting a thriving development trend.

In 1992, the China Small Animal Protection Association was established, marking the awakening of China's awareness of protecting and caring for pets. Subsequently, international well-known pet brands entered China, bringing advanced management experience to the just-starting Chinese pet industry. As time went by, the Chinese pet industry continued to innovate and develop, driving the development of industry subdivisions such as "pet clinics" and "pet care". According to the 2024 China Pet Industry White Paper, the Chinese pet industry is expected to surpass the 300 billion yuan mark in 2024, coupled with the compatibility of the pet's companionship attribute and the current trend of single economy, the Chinese pet industry market will continue to maintain a stable expansion trend, with a bright future ahead.

1.2 Problem Restatement

After fully understanding the background information of the problem, we need to establish a model and solve the following problems based on the potential constraints and attached information:

Problem 1

Question 1 requires the use of Attachment 1 and additional data sets to analyze the development of China's pet industry in the past five years by pet type, and analyze the

factors affecting the development of the pet industry. Finally, it is necessary to build a mathematical model to predict the development of the industry in the next three years.

Problem 2

Question 2 turns the focus from domestic to global, asking for an analysis of the development of the global pet industry by pet type using Attachment 2 and additional data. A mathematical model is also needed to predict global pet food demand over the next three years.

Problem 3

Question 3 on the basis of excluding the impact of economic policy changes, it is necessary to use the relevant data in Attachment 3 to establish a suitable model to analyze the development trend of China's pet food industry and market demand.

Problem 4

Question 4 takes the economic policy variables excluded in question 3 as separate objects of investigation, and requires us to establish a suitable mathematical model, combine the attachments and additional data sets collected, and comprehensively consider the calculation results of the above questions to quantitatively analyze the changes brought about by this influence. In addition, it is also required to formulate a sustainable development strategy for the pet food industry in China based on the calculation results

II. Assumptions and Justifications

1. Assuming the data source is authentic

2. It is assumed that there are no major world events in the next three years

3. It is assumed that the collected data sample is representative

4. It is assumed that the cost of raising a pet will remain constant for the next few years

5. It is assumed that the characteristics of pet owners are directly related to the pet market

Ⅲ. Modeling and Solving

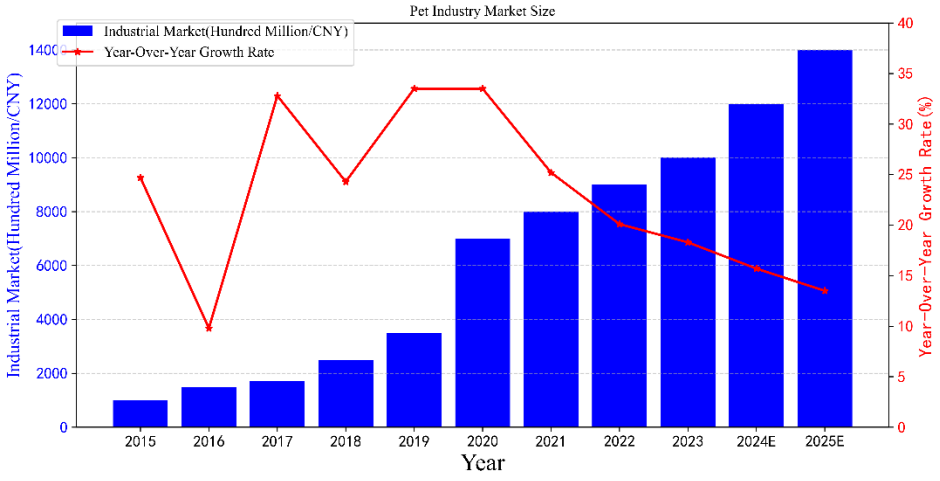
3.1 Model for problem 1

3.1.1 Problem analysis and solution ideas

The first question requires us to analyze the development of China's pet industry in the past five years by pet type. After observing and analyzing the existing data sets, we find the following two characteristics in the domestic market: 1. Domestic big cities, mainly in the first and second tier cities, are the main target market of the pet industry, accounting for the majority of PET owners; 2. Pets are mainly raised by cats and dogs. In order to facilitate our research, we choose cats and dogs as the target research objects. Finally, the number of cats and dogs, pet market size and market investment are taken as independent variables to predict the pet industry situation in the next three years.

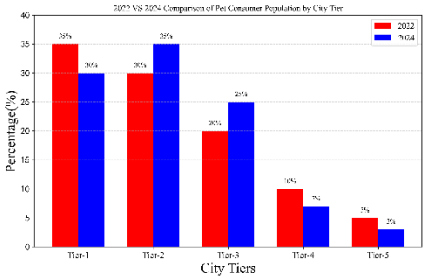
3.1.2 Background analysis

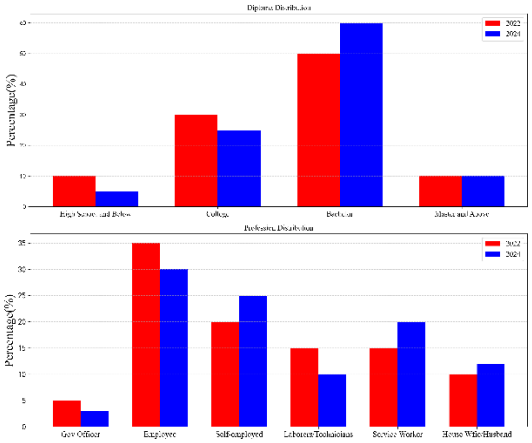
3.1.2.1 Development of China's pet industry in the past five years

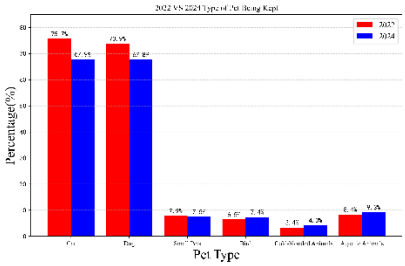


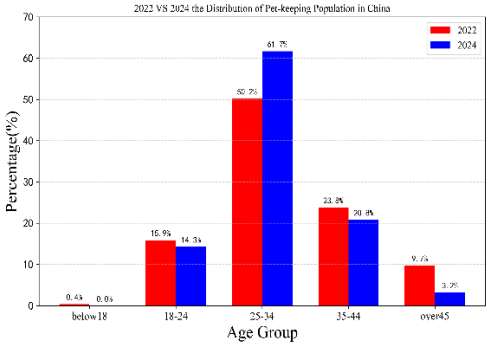
Through the analysis of the data of China's pet industry in the past five years, we draw the following conclusions about the development situation: The market size of the pet industry has been growing year by year since 2015, it is expected to maintain a continuous growth trend in the future.

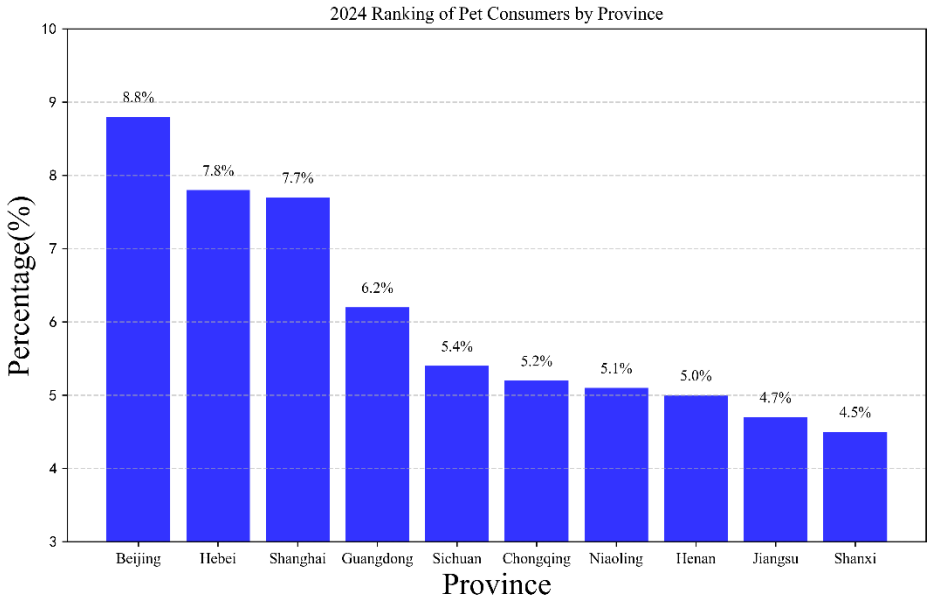
3.1.2.2 Analysis of influence factors











In view of the many factors affecting the development of the pet industry, we mainly analyze from the perspective of consumer groups, the following conclusions are drawn:

1. Consumer groups in China's pet industry are mainly distributed in China's first and second tier cities such as Beijing, Shanghai and Guangzhou, and most of them have college and Bachelor degrees and employees. As urban middle class, most of them have the economic conditions to raise pets, which has a great impact on the industry.
2. Through comparative analysis, it can be seen that the type of pet also affects the market development. Consumers' pet preference is for cats and dogs, because the supporting maintenance of cats and dogs has formed a certain scale, and their appearance, especially the accompanying attribute, is more in line with the spiritual needs of consumers.
3. The age of consumers is also a factor that cannot be ignored. The pet owners are mainly middle-aged and young people, especially the changes in social concepts in recent years, and they are more willing to accept the emotional feedback and support brought by pet raising.

3.1.3 Model Preparation Ⅰ——ARIMA

ARIMA model is a widely used time series analysis technique to predict future values based on historical data patterns. The mathematical formula of ARIMA (p, d, q) model is expressed as follows:

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_2 B^p)(1 - B)^d X_t = \left(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q\right) Z_t$$

Where, x_t is the time series data, B is the lag operator, ϕ and θ is the autoregressive and moving average parameters, P is the autoregressive order, D is the difference degree, and Q is the moving average order.

In order to construct an ARIMA model, it is necessary to select the appropriate parameters P, Q, and D.

ADF

Difference is designed to convert non-stationary time series to stationary time series for better model fitting. ADF test is performed after each difference. If the P-value is less than 0.05, the test is successful. Otherwise, the difference continues until the p-value is less than 0.05.

Determination of order

In ARIMA (Autoregressive Integrated Moving Average) model, AIC (Akaike Information Criterion) can be used to select the best ARIMA model.

$$AIC = -2\log(L) + 2k$$

Where, represents the maximum likelihood estimate of a given model, and represents the number of parameters in the model.

In the ARIMA model, AIC can be used to compare different ARIMA models to help select the best model. Typically, we build multiple ARIMA models by trying different AR, MA, and difference orders. Then, according to the maximum likelihood estimate and the number of parameters for each model, the corresponding AIC value is calculated.

In the AIC test, we choose the ARIMA model with the minimum AIC value as the best model. A smaller AIC value indicates that the model finds a good balance between fitting ability and model complexity when fitting data.

It should be noted that the smaller the AIC value, the better. AIC is only a relative evaluation criterion and can only be used as a reference in comparison of models on the same problem and the same data set. Therefore, when conducting AIC test, comprehensive consideration should be made according to specific problems and characteristics of data sets, and the model with smaller AIC value should be selected as the best ARIMA model.

According to the AIC , we can obtain the values of p and q through the algorithm to minimize the AIC value, and the ARIMA effect is optimal.

3.1.4 Model solving and analysis

From the above analysis, we can see that ARIMA is suitable for analyzing and predicting time-dependent data. By combining the three main parts of autoregression (AR), moving average (MA) and difference (I), the model can adapt to the characteristics of various time series data, especially non-stationary series. According to the data collected from literature review and the visualized images drawn accordingly, it can be seen that the development scale and investment of China's pet industry in recent years have significant time-dependent and non-stationary characteristics. It can be seen that the ARIMA model coincides with our data, so the ARIMA model is used

to predict the industrial scale and investment of China's pet industry, and according to this, the overall development of China's pet industry in the next three years is predicted.

3.1.5.1 Industry scale forecasting model

(1) ADF test results

```
Test Statistic: 0.265391
p-value: 0.975700
```

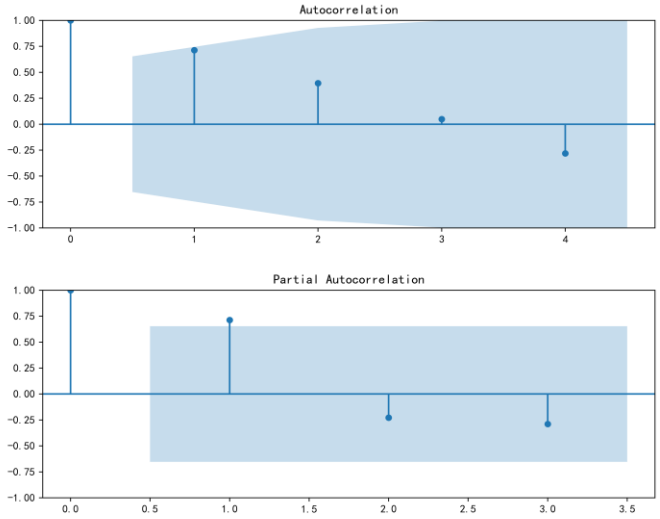
From the above results, it can be seen that the model needs to be differentiated.

(2) AIC test results

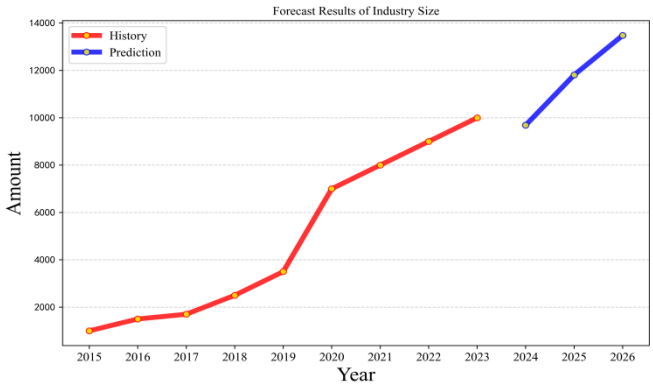
```
min aci: (1, 1, 1)
```

From the above results, the values of p,q and d can be calculated as (1,1,1).

(3) Calculation results of ACF and PACF



(4) The prediction results of the model



3.1.5.2 Investment forecasting model

(1) ADF test result

```
Test Statistic: 0.265391
p-value: 0.975700
```

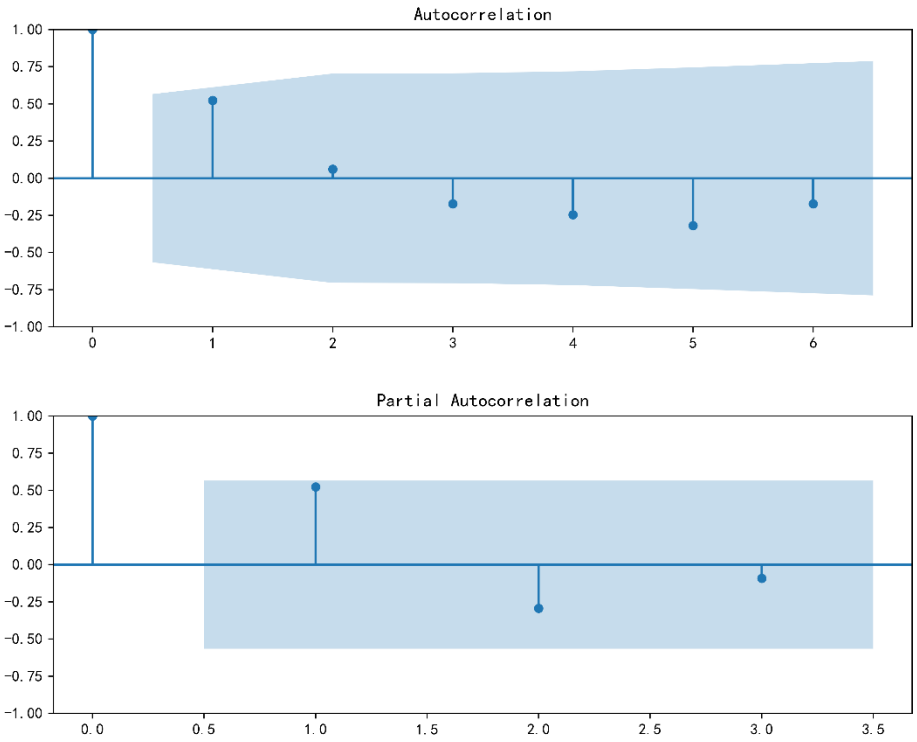
From the above results, it can be seen that the model needs to be differentiated.

(1) AIC test results

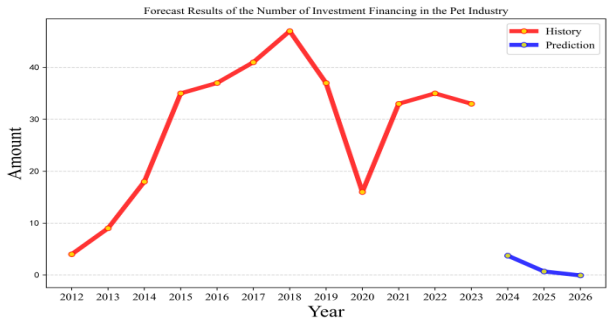
```
min aci: (1, 1, 1)
```

From the above results, the values of p,q and d can be calculated as (1,1,1).

(3) Calculation results of ACF and PACF



(2) The prediction results of the model



3.1.5.3 Result analysis and interpretation

According to our expectations and subjective cognition, the industrial scale and investment of China's pet industry will gradually increase in the next three years. It can be seen from the operation results of the model that the scale of the pet market is expanding, which is in line with our expectations and cognition, but the market financing situation is showing a downward trend. Through consulting relevant literature, we know that during the period of the novel coronavirus epidemic, the market is unstable, the financial market is unstable, and many investors are faced with asset

shrinkage or financial crisis. Therefore, the prediction result of the model of the financial investment market is not particularly ideal, and the financial market has a huge shock in the early stage of the epidemic, so the trend is not stable, which will bring errors. If the sample after the outbreak of the epidemic is taken or the unexpected decline caused by the epidemic, financial crisis and other volatility factors is removed, the error of time series estimation will be reduced.

3.1.6 Model Preparation II——Holt's linear trend method

Holt's linear trend method is a time series prediction method that applies to data with a linear trend. This method was proposed by Charles C. Holt in 1957 to extend the simple exponential smoothing method to handle time series data with trends.

(1) Horizontal component: Used to describe the base level or mean of a time series.

(2) Trend component: represents the linear change trend of data over time.

The method estimates the horizontal component and the trend component by analyzing the historical data, and predicts the future value based on these estimates. For example, if we have a set of time series data on sales volume, through the Hotelling linear trend method, we can determine the approximate average level of sales volume and the trend of sales increase or decrease. This information is then used to predict what range sales are likely to be in over the next few periods. Its advantages include relative simplicity and ease of use, and its ability to handle data with linear trends well. However, it also has some limitations, such as data with non-linear trends or complex fluctuations may not be suitable.

Components：

Prediction equation: A prediction used to generate future values.

Level equation: Used to smooth time series data, reflecting its underlying level.

Trend equation: Used to smooth the trend part of a time series.

Update formula for horizontal estimates:

$$L_t = \alpha y_t + (1-\alpha)\left(L_{t-1} + T_{t-1}\right)$$

Updated formula for trend estimates:

$$T_t = \beta\left(L_t - L_{t-1}\right) + (1-\beta)T_{t-1}$$

The horizontal and trend estimates are added together to get the predicted value：

$$\hat{y}_{t+h} = L_t + hT_t$$

3.1.7 Model solving and analysis

From the previous background analysis and related data, we can see that the main body of China's pet industry is pet cats and pet dogs, which occupy the majority of the pet market share and are representative. Therefore, the development trend of pet cats and pet dogs can be further analyzed. Based on background analysis and relevant data, it can be seen that the number trend of pet cats and pet dogs in China is in line with the linear trend, so the Hotelling's linear trend model is adopted for analysis.

(1) Let y_t be the number of pet dogs/cats in a given year,

Update formula for horizontal estimates:

$$L_t = \alpha y_t + (1-\alpha)\left(L_{t-1} + T_{t-1}\right)$$

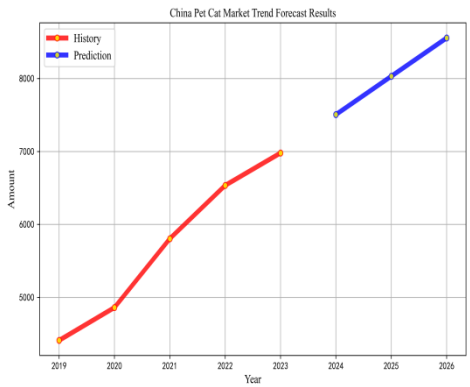
Updated formula for trend estimates:

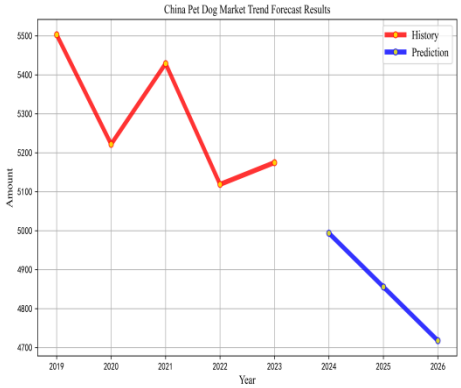
$$T_t = \beta\left(L_t - L_{t-1}\right) + (1-\beta)T_{t-1}$$

The horizontal and trend estimates are added together to get the predicted value：

$$\hat{y}_{t+h} = L_t + hT_t$$

(2) Model result





(3) Result analysis and interpretation

As can be seen from the figure, the number of pet cats is on the rise in the next

three years, which is consistent with the yearly increase in the number of pet cats in the five years from 2019 to 2023. And because the growth rate in each of the past five years has not changed much from the previous year (generally showing linear growth), we have reason to believe that the Holt linear model forecast is reliable.

The number of pet dogs is expected to decline over the next three years. Although there is a brief small rebound in the two stages from 2020 to 2021 and from 2022 to 2023, in the five years from 2019 to 2023, the number has an obvious downward trend on the whole, so it can be approximately fitted with a linear decline. That is, the predicted trends we get through modeling are generally consistent with the original data. Therefore, we can conclude that the downward trend obtained by Hotelling's linear trend method is valid.

3.1.8 Summary of Question 1

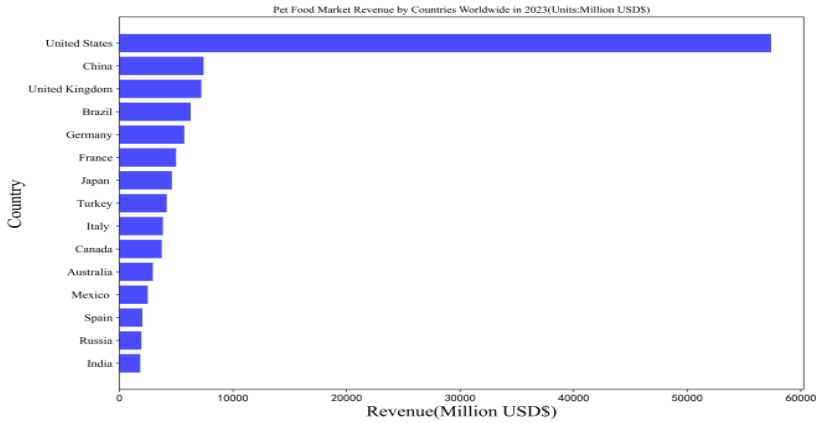
In question 1, we conduct data analysis and visualization on the development of China's pet industry market in the past five years, and draw four conclusions describing the development of the industry. By using ARIMA model and Hotelling linear trend method, we conclude that the market size of China's pet industry will expand in the next three years, and the financing situation will not be stable. And the conclusion that pet cat numbers are up and dog numbers are down

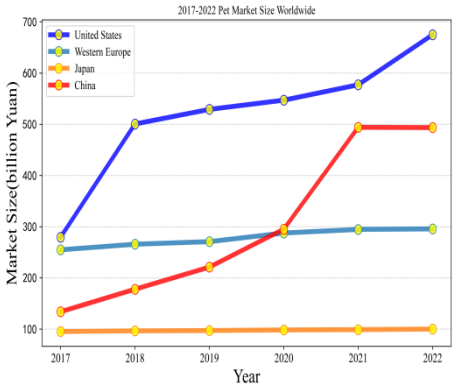
3.2 Modeling and Solution of Problem 2

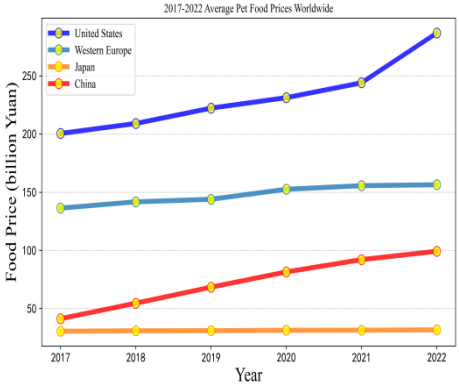
3.2.1 Problem analysis and solution ideas

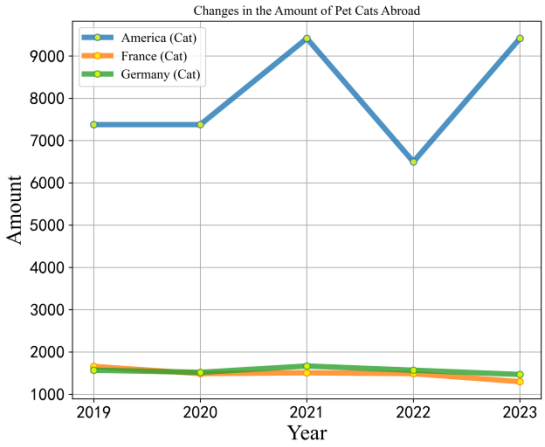
On the basis of the first question, the second question extends the research scope of the question from domestic to foreign countries. Since the international market has the characteristics of large differences and wide range among countries and regions, in order to solve this problem, we first visualize the existing data to observe the data characteristics, then find and demonstrate the relationship between food and market size of countries in the world, and finally, similar to question 1, based on the Holt linear trend method to predict the global food demand of countries in the world.

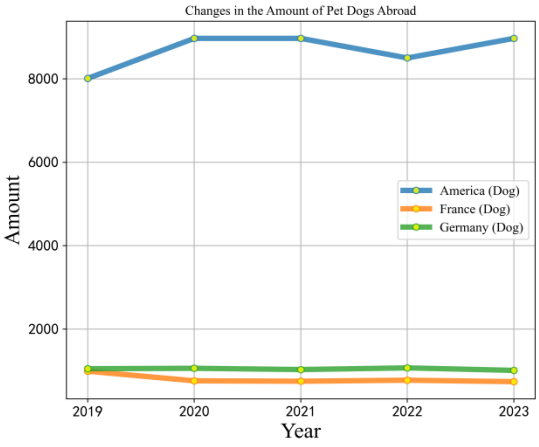
3.2.2 Background analysis











Through the analysis of the above chart, the following conclusions are drawn about the development of the global pet industry:

1. The United States occupies an absolute leading position in the pet food market, and its market income far exceeds that of other countries; China is a close second; And other developed countries such as Europe also occupy a considerable market share and revenue.

2. In recent years, the market size of the pet industry in the United States and China has generally shown an upward trend, while the growth of Western Europe and Japan is slow, while the price of pet food in China and the United States has risen rapidly, while the price of pet food in Western Europe, especially Japan, is relatively stable

3.The US pet industry has large size, growth potential, and volatility, especially for cats, while France and Germany have more stable, smaller - sized markets.

3.2.3 Model solving and analysis

According to 3.1.6:

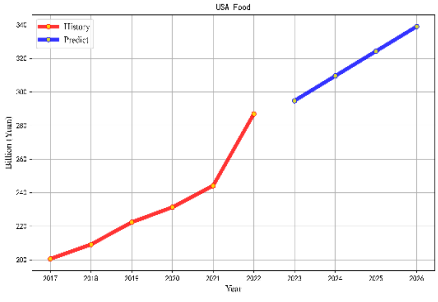
The horizontal and trend estimates are added together to get the predicted value：

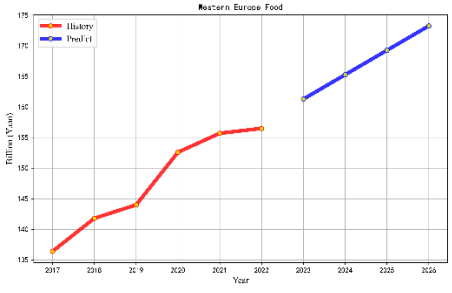
$$\hat{y}_{t+h} = L_t + hT_t$$

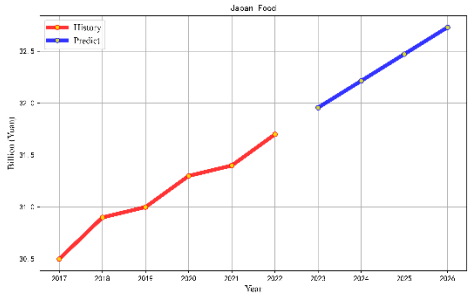
Let y_t be the Sales of pet food in a country in different years

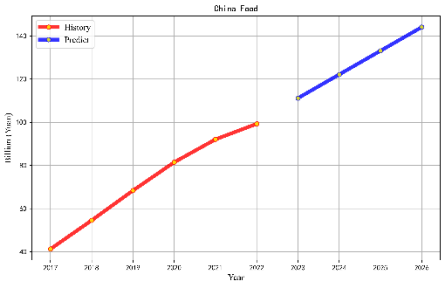
3.2.3.1 Model result

According to the data of background analysis, it can be seen that the overseas pet market is on the rise as a whole, and the number of pet cats and pet dogs has increased year by year in recent years, with an approximate linear trend increase. Therefore, the Holt linear trend model is adopted to predict the development of the overseas pet market. Therefore, the number of overseas pet cats and dogs and the market share of pet food are used as indicators to measure the overseas pet market, which is taken as the research object.









3.2.3.2 Result analysis and interpretation

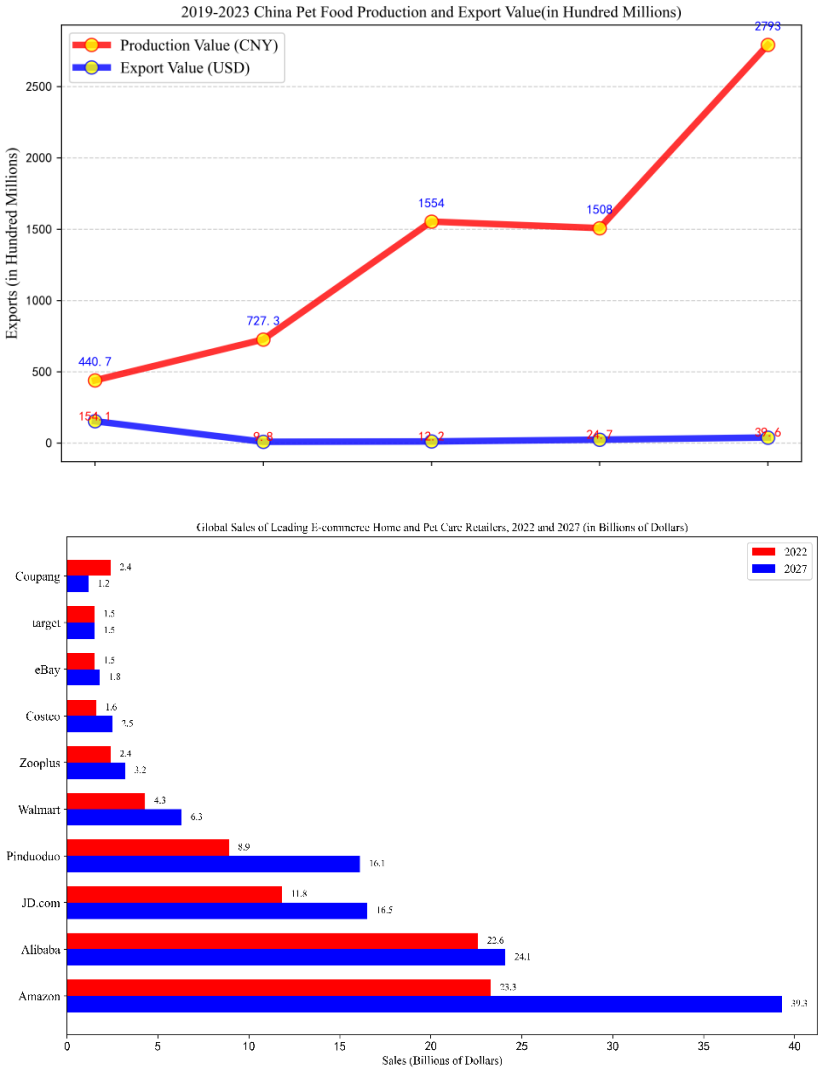
According to the relevant data and literature we have collected, as well as the analysis of the prediction results of the established model, the number of pet cats and dogs worldwide will increase steadily in the next few years, which is one of the representative indicators to evaluate the overall pet industry. At the same time, we estimate that the world's demand for pet food will increase year by year in the next few years through the Holter linear trend method. There is a strong positive correlation between the market size of pet food and pets, so we can infer that the global pet market as a whole is on an upward trend in the next few years, and we are optimistic about the prospects of the industry.

3.3 Modeling and Solution of Problem 3

3.3.1 Problem analysis and solution ideas

Question 3 requires us to analyze the development trend of the domestic pet food industry and market demand without considering the economic policy changes. The method of thinking and reconciliation is similar to the previous two questions. Therefore, we also adopt the Holt linear trend model for analysis, and predict the future trend on this basis.

3.3.2 Background analysis



From the chart above on the production and export value of pet food in China from 2019-2023, we can draw the current development status of the pet food industry in China:

1.In the past five years, China's pet food industry has shown an upward trend, with a large increase in output. This reflects the strong market demand of China's pet industry, and also reflects the rapid expansion trend of the food industry.

2.Exports showed a steady and positive trend. Although exports declined in 2019-2020 due to the impact of the global COVID-19 epidemic, it gradually recovered and maintained stable growth. This fully confirms the development potential and strong resilience of China's pet food market industry.

3. Of the top three export e-commerce companies, two are from China. This situation

strongly confirms that China occupies a dominant position in the export field related to the pet industry from the side.

3.3.3 Model solving and analysis

According to the relevant data and literature analysis collected by us, the output and export volume of pet food in China showed an approximate linear trend in recent years, so we adopted the Holt linear trend model to analyze it.

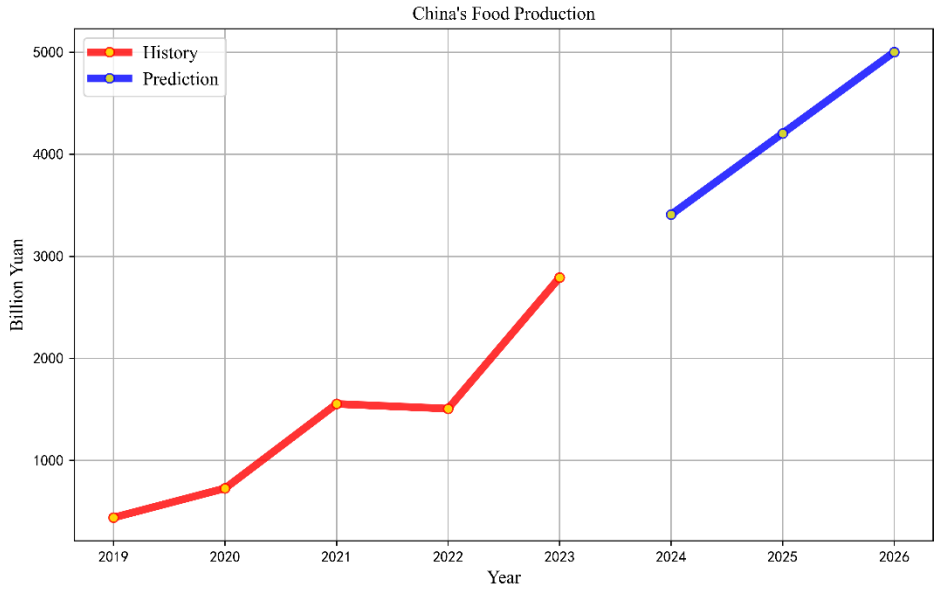
According to 3.1.6:

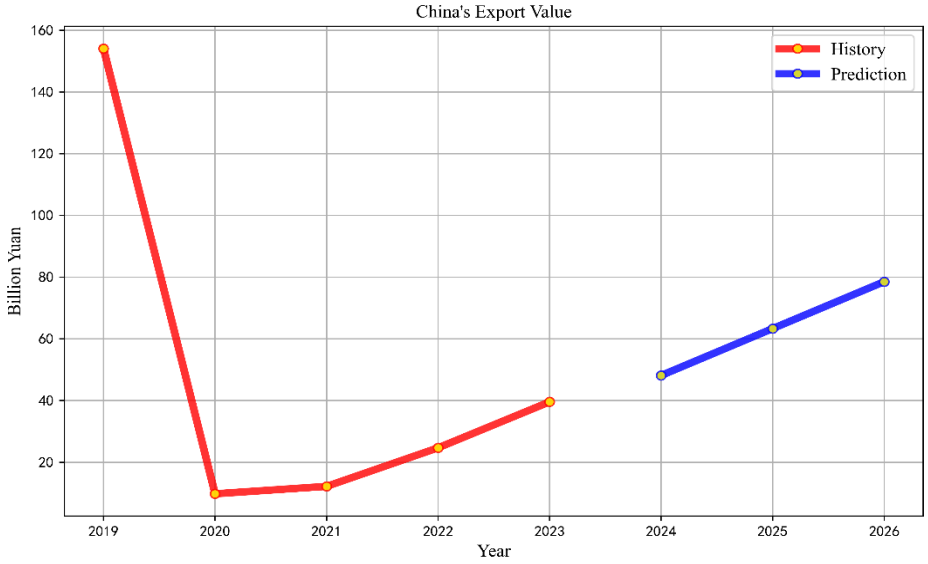
The horizontal and trend estimates are added together to get the predicted value：

$$\hat{y}_{t+h} = L_t + hT_t$$

Let y_t be the Production of pet food in China in different years or Total exports of pet food from China in different years

3.3.3.1 Model Result





3.3.3.2 Result analysis and interpretation

Through the analysis of the model operation results, it can be found that China's pet food sales and exports are on the rise. It is worth noting that although in 2020, affected by the epidemic, shipping outage and other factors led to a relatively large decline in exports, from the data of the following years, the overall trend is still roughly linear, which is in line with expectations.

3.4. Modeling and Solution of Problem 4

3.4.1 Problem analysis and solution ideas

Question 4 takes the economic change factors excluded from question 3 as the research object. Considering the different economic policies of different countries and the characteristics of dynamic changes, and the different impact of different indicators on the final result, we intend to establish an analysis model through principal component analysis and multiple regression analysis to quantitatively evaluate the impact of different factors on China's pet food exports.

3.4.2 Model preparation

(1) PCA

PCA is a dimensionality reduction algorithm that converts multiple indicators

into a small number of principal components, which are linear combinations of the original variables and are unrelated to each other, reflecting most of the information of the original data. In general, when the research problem involves multiple variables and there is a strong correlation between the variables, we can consider using the principal component analysis method to simplify the data.

In general, PCA consists of the following steps:

1.Arrange the raw data in rows to form a matrix X.

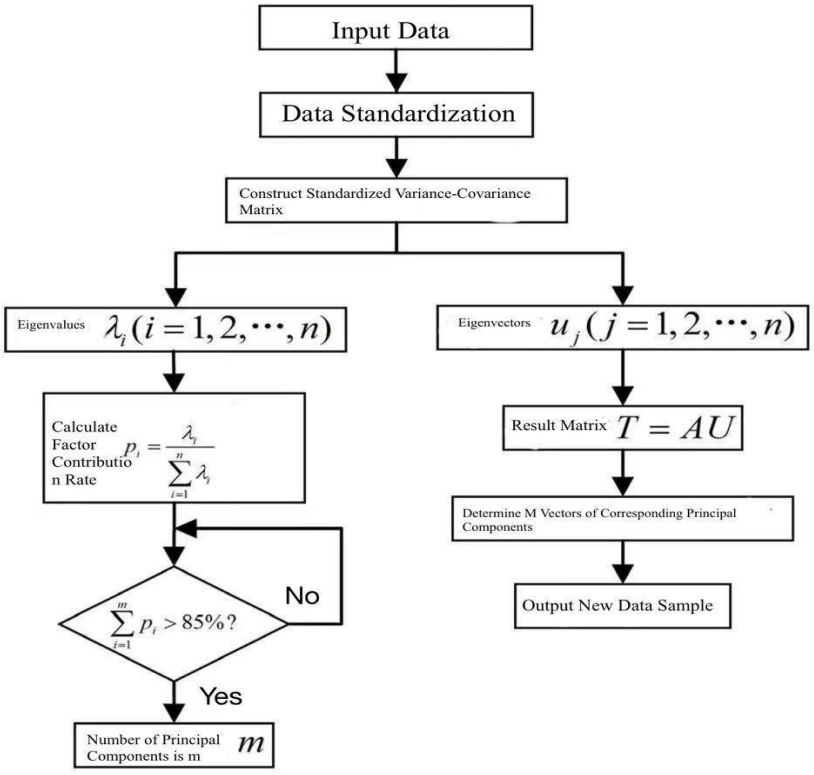
2.The matrix X is normalized so that its mean is zero.

3. Calculate the covariance matrix C of matrix X.

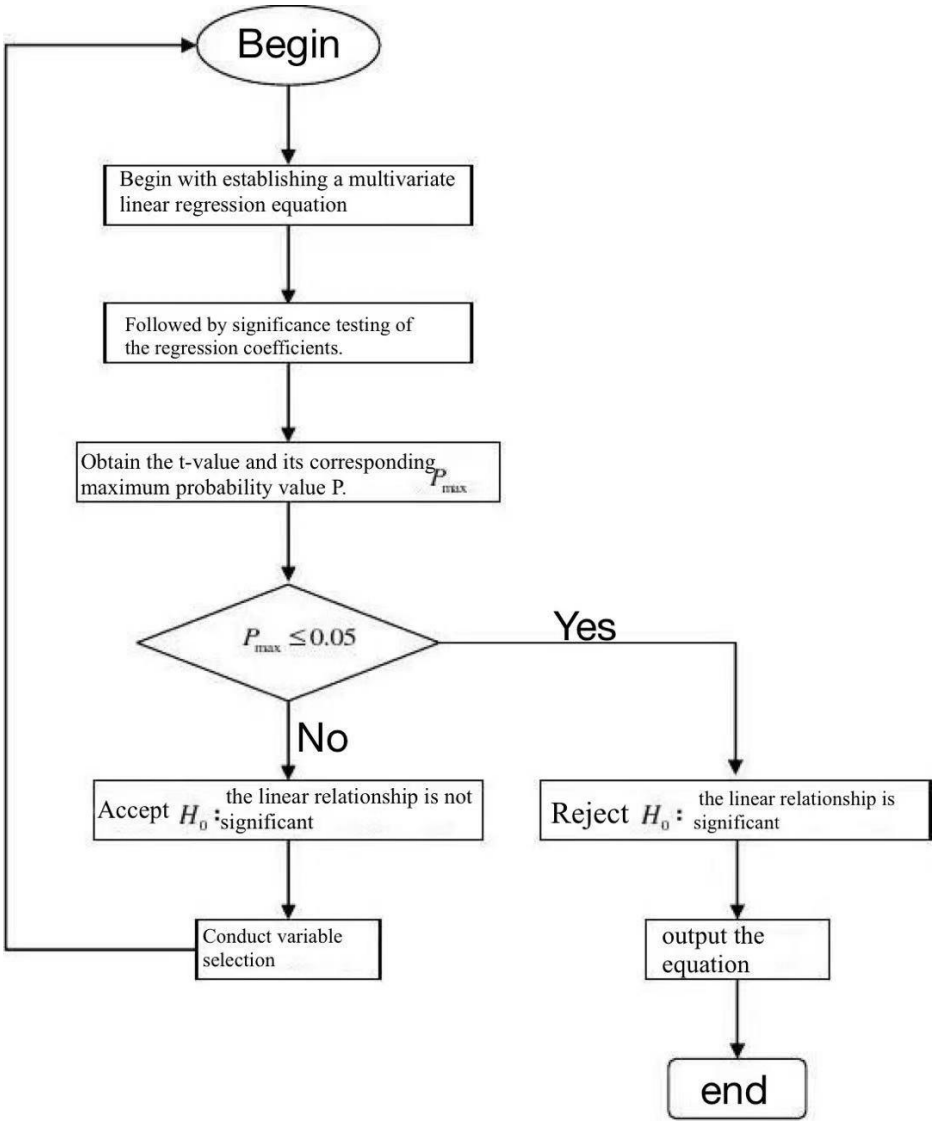
4. Arrange the eigenvectors according to their eigenvalues from largest to smallest, and take the first k rows to form a matrix P.

5. The data Y after dimensionality reduction is obtained by calculating Y = PX. Use the following formula to calculate the contribution rate V_i of each feature root:

$V_i = \frac{x_i}{\left(x_1 + x_2 + ...\right)}$. These eigenroots and eigenvectors explain the physical meaning of the principal components.



（2）Multiple Regression Analysis

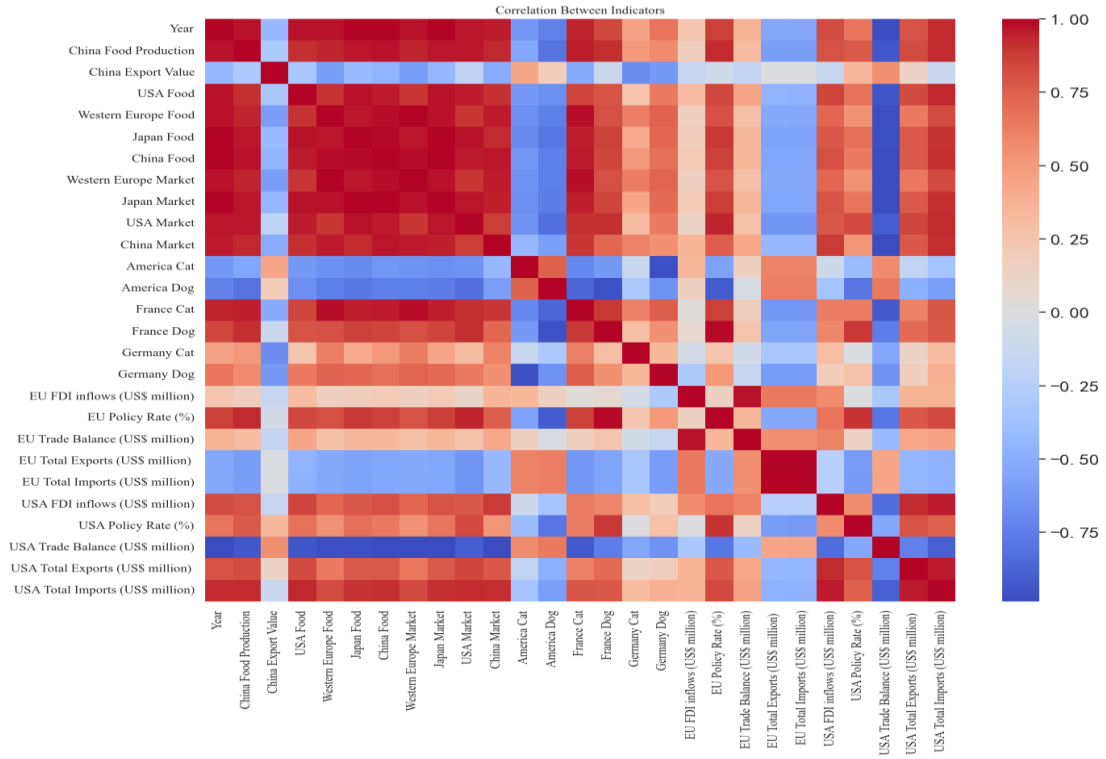


3.4.3 Model solving and analysis

3.4.3.1 Model substitution solving

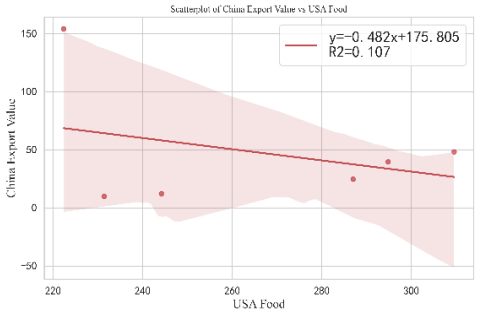
	Year	China Food Production	China Export Value	USA Food	Western Europe Food	Japan Food	China Food	Western Europe Market	Japan Market	USA Market	China Market	America Cat	America Dog	France Cat	France Dog	Germany Cat	Germany Dog
0	2019	440.7	154.1	222.4	144	31	68.4	270.9	97.3	529.2	221.2	9420	8970	1300	740	1470	1010
1	2020	727.3	9.8	231.4	152.6	31.3	81.5	287.8	98.5	547.2	295.3	6500	8500	1490	775	1570	1070
2	2021	1554	12.2	244.2	155.7	31.4	92.1	294.8	99.1	577.2	494.2	9420	8970	1510	750	1670	1030
3	2022	1508	24.7	287	156.5	31.7	99.3	295.8	100.2	675	493.6	7380	8970	1490	760	1520	1060
4	2023	2793	39.6	294.8341671	161.2885527	31.95700987	111.1986762	305.6114219	101.1533982	786.283172	564.8423825	7380	8010	1660	990	1570	1050
5	2024	3409.326805	48.14963873	309.5756045	165.2767642	32.21469885	122.1622508	313.6855467	102.119575	889.7426186	628.5281991	5839.833864	7852.508163	1765.496877	998.69905	1620.466314	1083.614189
6	2025	4205.318639	63.34809655	324.3170419	169.2649757	32.47238783	133.1258254	321.7596715	103.0857518	993.2020652	692.2140158	4619.284711	7557.088375	1876.322041	1049.134658	1661.891937	1107.931962
7	2026	5001.310473	78.54655438	339.0584793	173.2531872	32.73007681	144.0894	329.8337963	104.0519286	1096.661512	755.8998324	3398.735559	7261.668586	1987.147205	1099.570266	1703.31756	1132.249736

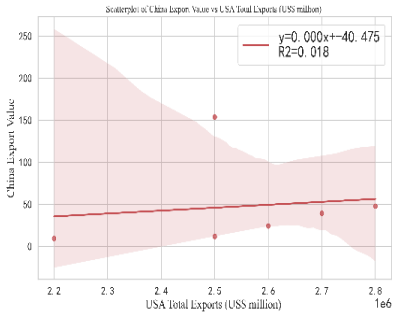
The correlation coefficient matrix obtained from some existing relevant data in the figure above is shown in the figure below：

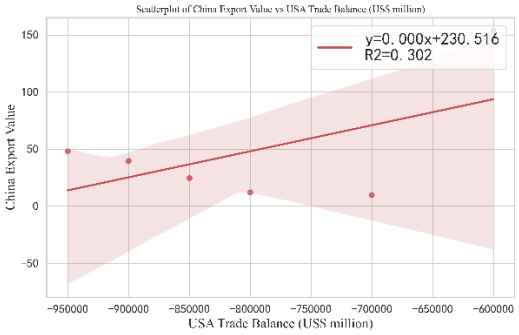


It can be seen from the figure that some foreign economic policies of the United States have a great impact on China's exports, which are shown in dark blue and dark red in the figure. More importantly, the trade difference between China's exports and the United States shows a significant positive correlation, which indicates that the foreign economic policies of the United States have a significant impact on China. In

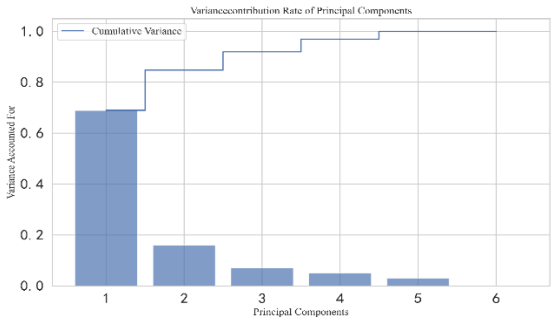
order to understand the impact of each index on China's exports more specifically, we analyzed it by establishing OLS regression model. However, due to too little data and too many features, we first used PCA for data dimensionality reduction. In the process of dimensionality reduction, we need to focus on the factors after dimensionality reduction to determine which factors represent which features. Then, the regression analysis of each influence factor and China's export volume is carried out to observe the influence of a single factor on China's export volume. (Due to the large number of factors, only a few are listed here.)



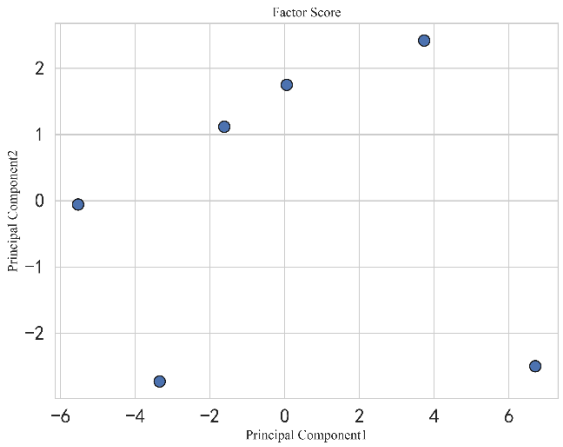




The cumulative variance contribution rate is calculated, and 6 principal components with the largest contribution rate are obtained:



To calculate the score of the principal component factor:



Then calculate the PCA component matrix, that is, the eigenweights corresponding to each principal component:

	Feature_1	Feature_2	Feature_3	Feature_4	Feature_5	Feature_6	Feature_7	Feature_8	Feature_9	Feature_10	Feature_11	Feature_12	Feature_13	Feature_14	Feature_15	Feature_16	Feature_17	Feature_18	Feature_19	Feature_20	Feature_21	Feature_22	Feature_23	Feature_24	Feature_25
Principal_Component_1	0.237982	0.228997	0.231117	0.238882	0.238294	0.231103	0.238321	0.23718	0.223672	-0.1569137	-0.1968757	0.23130256	0.21760373	0.10885221	0.16237085	0.03466823	0.22078076	0.06856541	-0.1476825	-0.1482851	0.18839756	0.17969659	-0.2294253	0.19255773	0.220863
Principal_Component_2	0.023845	0.087821	0.000773	0.013572	0.032806	0.003708	0.028127	-0.006158	0.113987	0.26030405	0.15728967	-0.0698225	-0.0338526	-0.0623585	-0.240588	0.49005994	0.01463441	0.45981569	0.34908474	0.34801639	0.25984291	-0.0353185	-0.0841876	0.16119332	0.14041297
Principal_Component_3	0.026591	0.013143	-0.210383	-0.046742	-0.093099	-0.209546	-0.06953	0.119208	-0.169181	0.12415287	-0.1156831	-0.1379245	0.17807597	-0.4679754	-0.2805655	-0.0635594	0.21306219	-0.0236692	-0.1489771	-0.1492128	0.06886956	0.48317501	0.17850325	0.30713173	0.12780097
Principal_Component_4	0.079301	-0.086759	0.00106	-0.064919	0.014754	0.014633	-0.043988	-0.036011	0.165391	0.44053891	0.26033384	-0.0232553	-0.1641059	0.43289721	-0.3064158	-0.1166157	-0.1452843	-0.2456143	-0.2595408	-0.2591323	0.25413866	-0.0354163	-0.0333074	0.23565584	0.16826287
Principal_Component_5	0.128284	-0.27104	0.03454	-0.093311	-0.043393	0.041621	-0.095862	-0.067869	-0.035697	0.21475505	-0.4051965	0.1782488	0.34280045	0.45363331	-0.2760621	0.02188467	0.2616698	0.04723728	0.14361906	0.14374846	-0.1997147	0.17336889	0.09431389	-0.1645117	-0.1467641
Principal_Component_6	0.696437	-0.173065	-0.251159	-0.285883	0.213232	-0.12753	0.159546	0.047453	-0.234736	-0.183022	-0.0543193	-0.0804075	-0.2136433	0.04574069	0.05485644	0.03156501	0.0352728	0.0933956	-0.0089604	-0.0350537	0.20726119	-0.1286434	0.13214371	-0.1080909	0.03305367

The score of each principal component is used for regression：

OLS Regression Results

Dep. Variable:	y	R-squared:	1.000
Model:	OLS	Adj. R-squared:	nan
Method:	Least Squares	F-statistic:	nan
Date:	Sun, 24 Nov 2024	Prob (F-statistic):	nan
Time:	11:43:58	Log-Likelihood:	172.85
No. Observations:	6	AIC:	-333.7
Df Residuals:	0	BIC:	-334.9
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	48.0916	inf	0	nan	nan	nan
x1	-4.1849	inf	-0	nan	nan	nan
x2	-0.1395	inf	-0	nan	nan	nan
x3	34.3312	inf	0	nan	nan	nan
x4	5.7560	inf	0	nan	nan	nan
x5	-1.1503	inf	-0	nan	nan	nan
x6	1.518e-14	inf	0	nan	nan	nan

Omnibus:	nan	Durbin-Watson:	1.174
Prob(Omnibus):	nan	Jarque-Bera (JB):	2.541
Skew:	1.545	Prob(JB):	0.281
Kurtosis:	3.788	Cond. No.	4.76

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The input rank is higher than the number of observations.

As shown in the figure above, regression did not work and we tried Gaussian noise perturbations to expand the data, the amplified data are as follows:

	Principal_Component_1	Principal_Component_2	Principal_Component_3	Principal_Component_4	Principal_Component_5	Principal_Component_6	y
0	-5.532147282	-0.057747793	2.35181943	0.363161451	-0.015387873	3.15579E-16	154.1
1	-3.346214322	-2.728824704	-1.302909393	-1.313514669	0.334547426	3.15579E-16	9.8
2	-1.619708719	1.116958885	-1.52886986	1.853720434	0.605244282	3.15579E-16	12.2
3	0.052952861	1.7494642	-0.646829634	-0.465525255	-1.703845745	3.15579E-16	24.7
4	3.734492344	2.418690925	0.454528308	-1.202598339	1.050242154	3.15579E-16	39.6
5	6.710625118	-2.498541513	0.672261149	0.764756379	-0.270800245	3.15579E-16	48.14963873
6	-3.337147992	-2.721874192	-1.313825641	-1.319356277	0.311908807	0.001452167	9.801285725
7	-1.620253372	1.110453348	-1.518552495	1.850202663	0.623725786	-0.013706424	12.21626624
8	0.050302402	1.749173163	-0.649935061	-0.460704307	-1.709970717	-0.001555996	24.70817586
9	-1.612892097	1.109464162	-1.537346392	1.856923779	0.60912201	-0.01701276	12.1973304
10	3.733173268	2.414085749	0.455280264	-1.196308846	1.058383344	-0.015585072	39.60281967
11	-5.529003845	-0.062612326	2.354306028	0.357342941	-0.010780107	0.007451581	154.100781
12	-1.633602638	1.134118873	-1.541387022	1.859135131	0.599569132	-0.013851174	12.20874305
13	6.714210958	-2.496040865	0.676268613	0.773993403	-0.267468875	-0.010591448	48.15326603
14	3.73719468	2.417466228	0.459899414	-1.210219468	1.055158836	-0.000327819	39.58432717
15	0.050291739	1.739038981	-0.63566776	-0.470083322	-1.685775561	0.019600504	24.68596431
16	6.689599374	-2.486602952	0.687113746	0.783217175	-0.283983091	-0.001542181	48.1405107
17	3.719190053	2.41483927	0.450287036	-1.193984754	1.043763177	-0.000339783	39.58808801
18	0.060331642	1.745522615	-0.652513104	-0.461787685	-1.700356794	0.007621753	24.69438685
19	3.747974078	2.412569175	0.469091964	-1.203129156	1.048339723	-0.019758368	39.60927994
20	-3.330085422	-2.720712089	-1.317574259	-1.311865848	0.333093237	0.01916927	9.794449169
21	0.055320601	1.763069584	-0.612718347	-0.450130342	-1.708201457	0.011152798	24.70686961
22	-3.353420023	-2.696068649	-1.318778617	-1.297220701	0.32752108	0.00064272	9.798732848
23	-1.612780559	1.101819734	-1.536824388	1.869860169	0.600032484	0.004604225	12.18282171
24	0.063702169	1.740323402	-0.668534611	-0.475862005	-1.696141339	-3.05752E-05	24.7055105
25	0.031846441	1.752876086	-0.642722745	-0.450696752	-1.71956134	-0.005888755	24.70479364
26	-1.621258482	1.142236937	-1.536571551	1.84956405	0.591569029	0.008781307	12.20363704
27	6.710919769	-2.493586638	0.670565158	0.760157363	-0.278821547	-0.005314814	48.14574398
28	3.742737843	2.427741394	0.453084478	-1.20737541	1.041206811	0.005986361	39.5912325
29	6.726200951	-2.521453557	0.670645107	0.776429243	-0.265947753	-0.010482133	48.15102208
30	6.726550883	-2.490387453	0.65375491	0.762942913	-0.29566262	0.000456933	48.14456623
31	6.697817953	-2.495824223	0.68690945	0.741728995	-0.269404938	0.001024851	48.14558781
32	-3.337678483	-2.720056966	-1.310141094	-1.321584802	0.313370201	-0.017230765	9.814753221
33	-1.620807201	1.115477034	-1.538715362	1.844134594	0.607739078	0.005455481	12.2013121
34	3.734526885	2.425480471	0.449474007	-1.212429589	1.037497496	-0.001231714	39.62154127
35	0.051292563	1.754824976	-0.641045007	-0.460000089	-1.69553329	-0.010076612	24.68282252

Perform the regression calculation again:

OLS Regression Results

Dep. Variable:	y	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	1.212e+05
Date:	Sun, 24 Nov 2024	Prob (F-statistic):	2.08e-62
Time:	11:43:58	Log-Likelihood:	-8.8579
No. Observations:	36	AIC:	31.72
Df Residuals:	29	BIC:	42.80
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	48.1419	0.059	812.398	0.000	48.021	48.263
x1	-4.1866	0.013	-314.902	0.000	-4.214	-4.159
x2	-0.1389	0.029	-4.860	0.000	-0.197	-0.080
x3	34.3471	0.047	735.586	0.000	34.252	34.443
x4	5.7406	0.060	95.122	0.000	5.617	5.864
x5	-1.1311	0.068	-16.528	0.000	-1.271	-0.991
x6	-2.3551	6.248	-0.377	0.709	-15.134	10.423

Omnibus:	0.585	Durbin-Watson:	2.011
Prob(Omnibus):	0.747	Jarque-Bera (JB):	0.145
Skew:	-0.142	Prob(JB):	0.930
Kurtosis:	3.124	Cond. No.	481.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

It can be seen that all the other principal components except principal component 6 have significant impacts on the effect of the model, and finally a multiple linear regression model containing 6 variables is obtained:

$$y = coef_1 x_1 + coef_2 x_2 + ... + coef_6 x_6 + const$$

3.4.3.2 Result analysis and interpretation

According to the relevant data and literature collected by our team and the model analysis, we believe that the economic policies of the United States have a great impact on the export of pet food in China. Therefore, to build a good trade relationship between China and the United States and promote the development of bilateral trade are effective measures to promote the export of pet food in China, strengthen the economic exchanges between the two sides, and jointly promote the development of the world pet

industry.

Ⅳ. Model evaluation and promotion

4.1Advantages and disadvantages of the model

★ We visualize the development factors and other parameters of the pet industry through charts, and visually show their impact and future development trends.

★ We used the ARIMA time series model for forecasting, which can give stable prediction results when predicting the scale and investment of China's pet industry, two stable time series data with linear trend flat.

★ After observing the approximately linear change of the number of pet cats and dogs, we chose the Holter linear trend method, which is relatively simple, easy to understand and implement, and has a good effect in processing linear trend data, and better captures the future linear development trend of the predicted data.

★ We used a combination of principal component analysis and multiple linear regression in the process of solving problem 4. Principal component analysis (PCA) is able to extract key information from multiple variables, simplifying the data structure through dimensionality reduction, and helping to identify key metrics more clearly. Multiple regression analysis can reveal the quantitative relationship between the independent and dependent variables, and predict changes in the dependent variables by constructing regression models. The two complement each other, alleviate the problem of multicollinearity to a certain extent, and help to improve the stability and reliability of multiple regression models.

4.2Improvement and generalization of the model

★ Not applicable to non-linear trend data: For data with non-absolute linear trends or complex fluctuations, the Holter linear trend method may not be as effective as a prediction. For this, we can use a combination of nonlinear time series forecasting methods, such as neural networks, support vector machines, etc., to deal with nonlinear trends.

★ Problems 1 to 3 are that in the process of making time series model predictions, we have too little raw data, and the time interval is one year, which is too long. As a result, we can collect more raw data or shorten the time interval.

Ⅴ. Citations

[1] https://www.jianshu.com/p/39d22980dd61“**PCA(principal component analysis,主成分分析)**”

[2] https://zhuanlan.zhihu.com/p/24263575 “**统计学之美（二）：多元回归分析**”

[3] https://blog.csdn.net/weixin_39753819/article/details/136100732“**时间序列预测——霍尔特线性趋势法与 Holt-Winters 方法**”

[4] https://zhuanlan.zhihu.com/p/24263575“时间序列(ARIMA)案例超详细讲解”

[5] https://blog.csdn.net/FrankieHello/article/details/86766625“**时间序列分析之 ADF 检验**”

[6] https://blog.csdn.net/buracag_mc/article/details/89107240“**AIC 和 BIC 相关知识**”

[7] https://ec.europa.eu/eurostat

[8] 2021 年中国宠物行业市场发展现状分析 宠物犬饲养数量下降原因何在？ 行业研究报告 - 前瞻网

[9] https://www.bea.gov/

[10] 《中国金融》丨疫情下的国际经济金融形势分析_腾讯新闻