Analysis of Correlation Between China's Gross Domestic Product and Fresh Food E-Commerce Transactions Scale 2014-2022 by Using Simple Linear Regression Model

Keru Xu

International Department of Beijing No.80 High School, Beijing 100020, P.R. China

ABSTRACT

Along with the gradual increase of China's Gross Domestic Product (GDP) and the technological advances in the era of big data, the grain production has reached a new high and the sales scale of fresh food e-commerce transactions has expanded. This paper will use the parameter estimation of linear regression model and model test to analyze. The relationship between China's GDP and the sales scale of fresh food e-commerce transactions in recent years is analyzed by using simple linear regression model, and forecasts are made for the future. A linear equation for China's GDP as well as the sales scale of fresh food e-commerce transactions is fitted using parameter estimation. Homoscedasticity, independence of observations and normality were performed on this linear equation. The results show that the use of simple linear regression models can be used to analyze and predict the future trends of China's GDP and Fresh Food E-Commerce Transaction Scale.

KEYWORDS

Simple linear regression; GDP; Fresh Food E-Commerce Scale

DOI: 10.47297/taposatWSP2633-456907.20230402

1 Introduction

Fresh food e-commerce, which refers to the use of e-commerce means to sell fresh products directly on the Internet, such as fresh fruits, vegetables, fresh meat and so on.[1] Cold chain transportation is a crucial component of the fresh food e-commerce industry, as it must guarantee product freshness and quality throughout lengthy transport processes. The rapid growth of China's population is contributing to an increase in consumption and the number of consumers, which is leading to an increase in gross domestic product. As a result, the market demand for cold chain logistics and transportation has increased dramatically year by year, leading the fresh food e-commerce transactions scale to increase year by year under the chain reaction. In this paper, in order to study the future development trend of fresh food e-commerce Transactions Scale, this paper will need to establish an appropriate mathematical model, use SPSS software to predict and analyze the regression model, test the fit of the model through the regression estimation results, and predict the future fresh food transactions according to the regression model with a better fit after verification.

2 Literature Review

In this paper, regression analysis will be used as the main analytical method to predict the scale of fresh food e-commerce transactions with simple linear regression model. And in the actual analysis of

the future development trend of fresh food e-commerce transactions scale, the previous researchers mainly discussed the types and diversified development of fresh food e-commerce, and almost did not analyze the two variables from a mathematical perspective and with mathematical methods. Moreover, some researchers have conducted in-depth research on linear regression method in practical applications.

Yu Lu solely analyzed the development of fresh e-commerce food, that is the development history, development status and classification of fresh food e-commerce. Furthermore, the author explored the development advantages of different types of fresh food e-commerce, and combined the current characteristics to analyze that ensuring the efficient operation of the supply chain is a significant role in the development of fresh food e-commerce.[2]

Jiulong Li, Ling Li, Ruimin Liu, Xu Sun, Wei Xue, Teng Xu and Yi Yang established simple linear regression model and analyzed the error of the model, so as to fit and analyze the pressure sensor calibration data in metrological calibration work. This is an important reference value for metrological inspectors to study the fitting method of other similar inspection data.[3]

Song Qianqian and Zhang Yanling based on the simple linear regression method to revise the ECMWF gust forecasts in Henan Province and to test and evaluate them.[4]

Park, J. Y., Kim, S. H., Jang, C. Y., Kim, J. H., & Lee, S. B. designated monthly electricity consumption as the dependent variable and average outdoor temperature data as the independent variable. They applied regression analysis to estimate the building's energy performance profile. The findings indicate that the predicted model consumes 23,608 kWh of total electricity, which is higher than the actual value. The findings indicate that the predicted model consumes 23,608 kWh of total electricity, which is higher than the actual value. The findings indicate that the predicted model consumes 23,608 kWh of total electricity, which is higher than the actual value. However, the modified model demonstrates greater energy savings compared to the previous one.[5]

This paper applies a simple linear regression analysis to establish the mathematical relationship that exists between the Gross Domestic Product and the scale of Fresh Food E-Commerce Transactions.

3 The Introduction of Simple Linear Regression

Regression is a statistical term used for describing models that estimate the relationships among variables. Linear regression models study the relationship between a single dependent variable Y and one or more independent variables, denoted by X.[6] A simple linear regression equation can be used to show how a dependent variable Y is adjusted conditional on changes in the independent variable X:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Y is the dependent variable in the simple linear regression equation, known as the predicted value of the sample;

X is the independent variable in the regression model, known as the eigenvalue of the sample;

ε is the error term in the regression model, i.e., variability that cannot be accounted for by the linear relationship between x and y;

β_0 is the constant;

β_1 is the regression coefficient.

The simple linear regression model can be used for analysis only when the data meet the following

conditions: There is a theoretical causal relationship between the independent variable and the dependent variable; The dependent variable is continuous and the independent variables can be categorical or ordinal variables;The relationship between the variables shows a linear trend; The residuals satisfy homoscedasticity, independence of observations and normality.

4 Inspection of Measurement Data

(1) Evolution of gross domestic product and fresh food e-commerce transaction scale in the period of 2014-2022(2023)

The ISV (Independent Software Vendors) e-commerce database is used to list the GDP and fresh food e-commerce turnover of China for the period of 2014-2022, and the one-way regression parameters of the two variables are counted. This information can be summarized in the Table No.1..

Table 1. Table of Gross Domestic Product and Fresh Food E-Commence Transaction Scale for the period of 2014-2022

Year	GDP (Trillion Yuan)	Fresh Food E-Commerce Transaction Scale (Trillion Yuan)
2014	64.36	0.029
2015	68.89	0.0542
2016	74.64	0.0914
2017	83.2	0.14028
2018	91.93	0.195
2019	98.65	0.25545
2020	101.36	0.36413
2021	114.92	0.46581
2022	121.02	0.56014

(2) Analysis of correlation between GDP and Fresh Food E-Commerce Transaction Scale by means of simple linear regression

(Assuming that China's GDP in the following data tables are all variables X and fresh food e-commerce transactions are all variables Y)

1) A characteristic analysis of the relationship between China's GDP and the scale of fresh e-commerce transactions

Based on the statistical data, the linear relationship between China's domestic GDP and the size of fresh food e-commerce transactions is constructed using the Excel software application, and linear regression analysis is used to find a straight line corresponding to a set of observations using the least squares method:

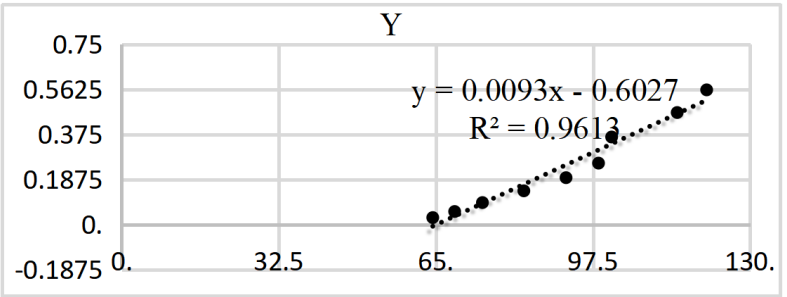


Figure 1 The relationship between China's GDP and Fresh Food E-Commerce Transactions Scale

This scatterplot presents a linear form and is able to present a straight linear relationship between

China's GDP - total value of fresh food e-commerce transactions. This shows that the linear regression model can be very successful in describing the one-factor relationship between the two indicators analyzed.

The straight line calculated by the least square method can be obtained with GDP as the dependent variable and the fresh e-commerce transaction scale as the independent variable.

2) Independence

Table 2 The Model Summary of Regression

R	R Square	Adjusted R Square	Durbin-Watson
.980[a]	0.961	0.956	1.019
a. Predictors: (Constant), Gross Domestic Product(Trillion Yuan)			
b. Dependent Variable: Fresh Food E-Commerce Transaction Scale(Trillion Yuan)			

Through Table No.2 to determine whether the variable Y obeys independence. When the value is generally distributed between 0 and 4 and the value is close to the Durbin-Watson test value of 2, the more likely the two variables are independent of each other. From the chart analysis, it can be seen that the Durbin-Watson test value is 1.019 in this case, which can be considered to meet the condition of independence of linear regression.

3) Normality

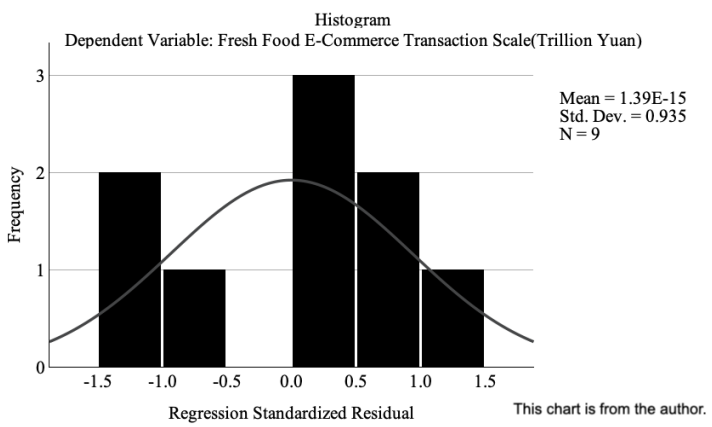


Figure 2. The Histogram of Regression

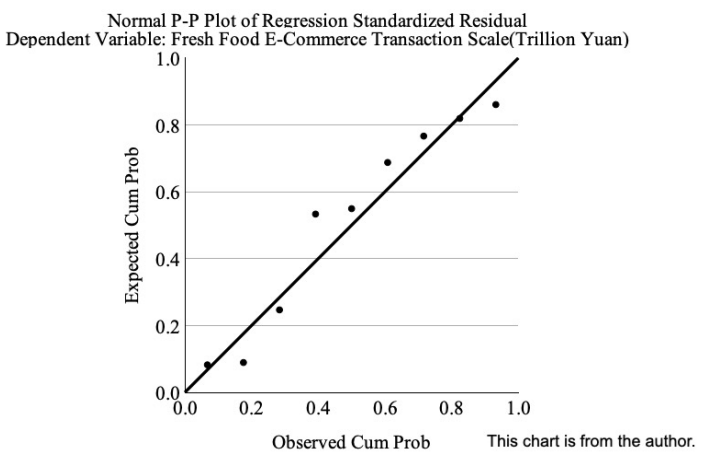
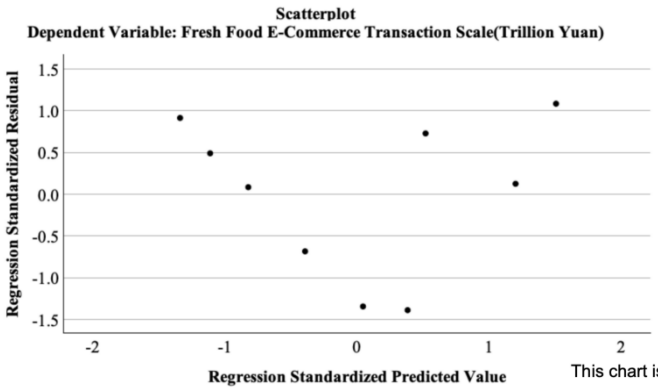


Figure 3. The Normal P-P Plot

The normality is determined by the histograms of the standardized residuals and the P-P plots of

the residuals from Figure No. 2 and Figure No. 3. From the histogram of Figure No.2, it can be seen that the standardized residuals approximately obey a normal distribution. The situation where the data is close to a normal distribution can be observed by observing whether the distribution of points in the P-P plot is close to the diagonal. If the points fall completely on the diagonal in the P-P plot, the data is normally distributed. According to Figure No.3, the points of P-P plot distribute along the diagonal line indicates that the regression residuals are approximately normally distributed. Therefore, by combining the standardized residual histogram and the P-P plot, it is evident that the residuals follow an approximately normal distribution.

4) Homoscedasticity



This chart is from the author.

Figure 4. Scatterplot of Homoscedasticity

This data can be judged for homoscedasticity by means of a scatter plot of the standardized predicted values. If the predicted values and the corresponding residuals are more evenly distributed and the data are distributed within ±2 standard deviations, the residuals are satisfied with isotropic congruence. According to Figure No.4, all the data in the graph are distributed within ±2 standard deviations and are evenly distributed, so the one-way linear regression model satisfies the residuals' chi-square.

5 Simple Linear Regression Analysis

(1) The degree of fit of model

Table 3. The Model Summary of Regression

R	R Square	Adjusted R Square	Durbin-Watson
0.980[a]	0.961	0.956	1.019
a. Predictors: (Constant), Gross Domestic Product(Trillion Yuan)			
b. Dependent Variable: Fresh Food E-Commerce Transaction Scale(Trillion Yuan)			

Table No.3 lists the degree of fit of the model. R is the complex correlation coefficient, and the complex correlation coefficient of this paper is R=0.980. R^2 is the coefficient of determination, and the coefficient of determination of this paper is R^2=0.961, which represents that the independent variable, known as Gross Domestic Product, can account for 96.1% of the variance in the dependent variable which is Fresh Food E-Commerce Transactions Scale. Adjusted R^2 is generally used to correct the effect of R^2 by the number of independent variables, and in this paper, the adjusted R^2=0.956, close to 1, which represents that a better model fit means that the independent variables do a better job of explaining the dependent variables.

(2) Analysis of variance

Table 4. The ANOVA of Regression

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	0.271	1	0.271	173.792	<.001[b]
Residual	0.011	7	0.002		
Total	0.282	8			
a. Dependent Variable: Fresh Food E-Commerce Transaction Scale(Trillion Yuan)					
b. Predictors: (Constant), Gross Domestic Product(Trillion Yuan)					

Table No.4 shows the consequences of analysis of variance, which shows F=173.792, P<0.001, indicating that the model is statistically significant. Conversely, if P>0.05, the regression model is not statistically significant.

(3) Modelling data

Table 5. The Coefficients of Regression

Model	Unstandardized B	Coefficients Std. Error	standardized Coefficients Beta	t	Sig.
(Constant)	-0.603	0.065		-9.240	<.001
Gross Domestic Product	0.009	0.001	0.980	13.183	<.001
a. Dependent Variable: Fresh Food E-Commerce Transaction Scale(Trillion Yuan)					

The table represents the interpretation of the regression coefficients in Table No.5. The value of unstandardized B is the regression coefficient. In this case the coefficients of the constant and variable Gross Domestic Product are -6027.023 and 0.009 respectively. Finally, the regression analysis equation can be obtained:

$$\mathrm{FreshFoodE-CommerceTransactionScale} = -603 + 0.009 \times \text{Gross Domestic Product}$$

6 Conclusion

In this study, we utilize a straightforward linear regression model to forecast the future growth pattern of e-commerce transactions in the fresh food industry, with China's gross domestic product serving as the independent variable. After conducting a significance test of the regression equation, analyzing residuals, and estimating regression coefficients' intervals, it has been determined that a linear relationship exists between the two variables. With the simple linear model, predictive analysis is possible, and ultimately, the regression equation is derived and confirmed.

The study also has several limitations. Firstly, the amount of research data is relatively small, and GDP data along with fresh food e-commerce transactions before 2014 will be continuously collected and substituted into the research model to improve prediction accuracy for the future scale of development of fresh food and e-commerce using the simple linear regression equation. Secondly, studying the effect of GDP solely on the scale of e-commerce transactions for fresh food has limitations because the scale of such transactions can be significantly influenced by various factors like climate and technology. Therefore, under certain circumstances, a multiple linear regression model can be employed to enhance the accuracy of forecasting fresh food transaction scale.

About the Author

Keru Xu is student of Beijing No.80 High School, and her research field is development economics of China.

References

[1] Wang, X. (2017). Study on the profit distribution of fresh electricity supplier pricing and logistics system. Journal of Beijing Jiaotong University, 83-84.

[2] Yu, L. (2019). Research on the development status and trend of Fresh E-commerce in China. Logistic Engineering and Management, 41: 113-15.

[3] Li, J., Li, L., Liu, R., et al. (2022). Application of univariate linear regression analysis in data fitting of pressure sensor. Metrology and Measurement Technology, 42: 40-49.

[4] Song, Q., Zhang, Y. (2023). Revised evaluation of ECMWF fine grid gust forecast products based on unitary linear regression. Journal of Shaanxi Meteorology, 17-22.

[5] Park, J.Y., Kim, S.H., Jang, C.Y., et al. (2015). The audit method of cooling energy performance in office building using the simple linear regression analysis model. KIEAE Journal, 15: 13-20.

[6] Bangdiwala, SI. (2018). Regression: Simple linear. International Journal of Injury Control and Safety, 25: 113-15.