Frontiers in Economics and Management Volume 4 Issue 11, 2023
ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0010

Least Squares Matrix Analysis Assisted by SPSS

-- Taobao's "Double 11" Commodity Sales Forecast

Yaoxin Hu

Jinan University Guangzhou 510000, China

Abstract

In economics, as a method of sales forecasting, this paper introduces the expression of the least squares method in linear algebra, emphasizes the core role of the normal equation in linear fitting, and establishes a regression linear prediction model with a slight combination of calculus and statistics knowledge, and uses SPSS software to theoretically predict the sales volume of Taobao Double 11.

Keywords

Least Square Method; SPSS; Linear Fitting; Sales Forecast.

1. Introduction

Finding projection is not common in practical problems, but there is some equivalence between projection and curve fitting or linear regression in statistical situations, which is why projection is so important. Projection problems include two aspects: one side is geometric orthogonality, which is the origin of normal equation, and the error is orthogonal to projection space; On the other hand, it is the extremum problem. Projection is the vector with the minimum error between the projection space and the original vector. Of course, these two aspects are unified in essence. In this article, we will reveal how to use linear fitting to predict and estimate problems in economics.

2. Literature Review

Yue Lingshui, Zhao Guigui (1997), Chen Qiuling, Chen Zhong (2012) established a linear regression prediction model using the least square method, and selected a suitable fitting curve to predict the trend of automobile sales.

In the CSDN forum, a lot of linear algebra theory knowledge related to machine learning was involved, such as Jin Liang (2016).

Two expressions of the least squares method and their internal relations are given respectively, while Liyiersan(2022) describes the least squares method from the perspective of integral and function construction, which further paves the way for economic prediction in this paper.

3. Basic Principles

The least square method is a data fitting method.

First, the matrix formula of the least squares solution is given:

From the normal equation:

$$A^T Ax = A^T b \tag{1}$$

Derive:

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0010

$$x = (A^T A)^{-1} A^T b \tag{2}$$

(Matrix A must be a column full rank matrix, otherwise (it does not exist) (Note: if A is a column full rank, columns of A are independent and square, so it is reversible).

Of course, this formula can also be understood from the perspective of calculus (partial derivative of error function), so we will not prove it here.

However, we are more familiar with multivariate linear functions in statistics:

$$Y_t(X) = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + \cdots + a_n X_n \tag{3}$$

So we need to change the matrix form here:

$$Y = Xt \tag{4}$$

Change into matrix form:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} \tag{5}$$

(Where$a_{0,}a_1 \ldots a_n$are undetermined constants, also called regression coefficients).

For this equation, it can be regarded as a system of linear equations (assuming that all equations are linearly independent). If the number of samples m is less than the characteristic dimension n, then this system of equations has infinite solutions. If m=n, there is a unique solution. If m is greater than n, there is no solution (i.e., there is a contradictory solution). The least squares method is generally used when m is greater than n, and the solution obtained is the optimal approximate solution.

The loss function is expressed as:

$$L(t) = \sum_{j=1}^{m} (Y|_{x=x_m} - y_m)2 = (At - Y)^T (At - Y) \tag{6}$$

Then use both sides of the equation to calculate the partial derivative at the same time:

$$\frac{\partial L(t)}{\partial t} = 2X^T (Xt - Y) \tag{7}$$

By making the derivative result equal to 0 matrix, we can reverse the normal equation:

$$X^T Xt = X^T Y \quad \Rightarrow \quad t = (X^T X)^{-1} X^T \tag{8}$$

Using least square method, we can substitute each group of observations into (5), and get:

$$Y_t(X) = a_0 + a_1 X_{i1} + a_2 X_{i2} + a_3 X_{i3} + \cdots + a_n X_{in} \qquad (i = 1,2 \cdots n) \tag{9}$$

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0010

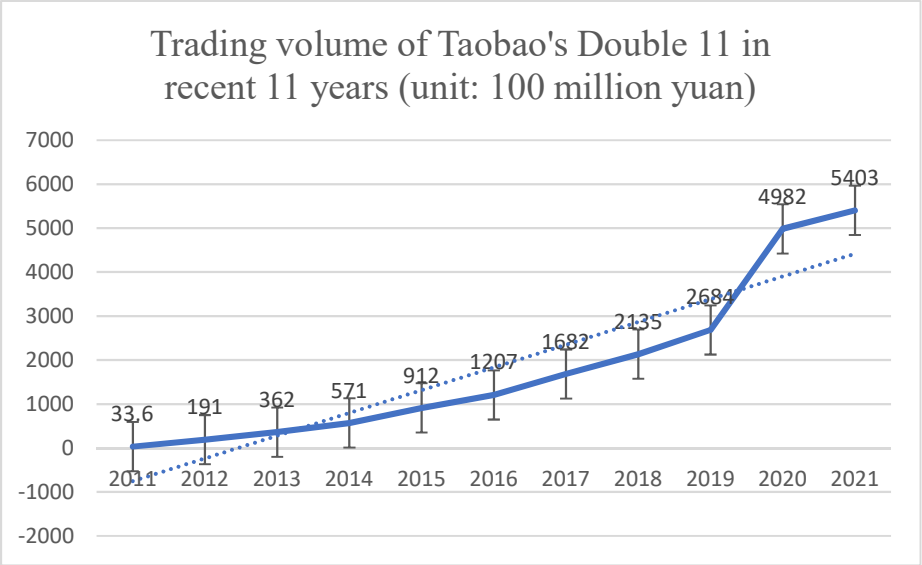
In particular, if there are only two variables in (3), we call it univariate linear regression:

$$Y_t(X) = a_0 + a_1 X_1 \tag{10}$$

In the next practical example, we will use the linear regression based on the least square method to analyze and forecast the trading volume of Taobao's "Double 11".

4. Background: Analysis and Forecast of Taobao's Double 11 Trading Volume

After data collection, the sales volume from 2011 to 2021 is as follows:



(Data source: Internet)

Figure 1. Total sales volume

We draw a line chart through Excel to see that although the sales volume fluctuates up and down, it generally shows a linear development trend, so we can give the equation:

$$b = C + Dt \tag{11}$$

The data of 11 years are regarded as 11 points, and the years are marked again with serial numbers 1, 2... 11:

First, suppose that all 11 points(1,33.6), (2,191) ⋯ (11,5403) are on a straight line:

$$\begin{cases} C + D \cdot 1 = 33.6 \\ C + D \cdot 2 = 191 \\ \vdots \\ C + D \cdot 11 = 5403 \end{cases} \tag{12}$$

$$\text{If} \qquad A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & 11 \end{bmatrix} \quad x = \begin{bmatrix} C \\ D \end{bmatrix} \quad b = \begin{bmatrix} 33.6 \\ 191 \\ \vdots \\ 5403 \end{bmatrix} \tag{13}$$

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0010

It is easy to find that the formula $Ax = b$ has no solution, so find a nearest solution;
Also,this solution should make $\left||b - Ax|\right|^2$ as small as possible:

$$\text{To solve} \qquad x = \begin{bmatrix} C \\ D \end{bmatrix} \tag{14}$$

$$\text{We have} \qquad A^T Ax = A^T b \tag{15}$$

$$\mathrm{A^T A} = \begin{bmatrix} 11 & 66 \\ 66 & 506 \end{bmatrix} \quad A^T b = \begin{bmatrix} 20162.6 \\ 177850.6 \end{bmatrix} \tag{16}$$

Use Gaussian elimination method:

$$\text{Get} \qquad \left[\begin{array}{cc|c} 11 & 66 & 20162.6 \\ 0 & 110 & 56875 \end{array}\right] \tag{17}$$

$$\text{So} \qquad \begin{cases} 11C + 66D = 20162.6 \\ 110D = 56875 \end{cases} \Rightarrow \begin{cases} C = -1269.6 \\ D = 517.1 \end{cases} \tag{18}$$

$$\text{That is, the fitting curve is:} \quad b = -1269.6 + (517.1)t \tag{19}$$

We can also use another method to verify the feasibility of the above methods:

$$y = bx + a \tag{20}$$

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{21}$$

$$a = \bar{y} - b\bar{x} \tag{22}$$

After calculation, it can be concluded that:

$\boldsymbol{\bar{x} = 6}$	$\boldsymbol{\sum_{i=1}^{n} x_i^2 = 506}$	$\boldsymbol{\sum_{i=1}^{n} y_i = 20162.6}$
$\sum_{i=1}^{n} x_i = 66$	$\left(\sum_{i=1}^{n} x_i\right)^2 = 4356$	$\sum_{i=1}^{n} x_i y_i = 177850.6$

$$b = \frac{11 \cdot 177850.6 - 66 \cdot 20162.6}{11 \cdot 506 - 4356} \tag{23}$$

$$\begin{cases} b = 517.1 \\ a = -1269.6 \end{cases} \tag{24}$$

The results of the two methods are the same, so in this case, the statistical estimation made by using SPSS is completely in line with our requirements.

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0010

The linear fitting of SPSS is used below. R2 of this straight line is 0.930, close to 1, which is of certain relevance, so we can use it to forecast the trading volume of the following years.

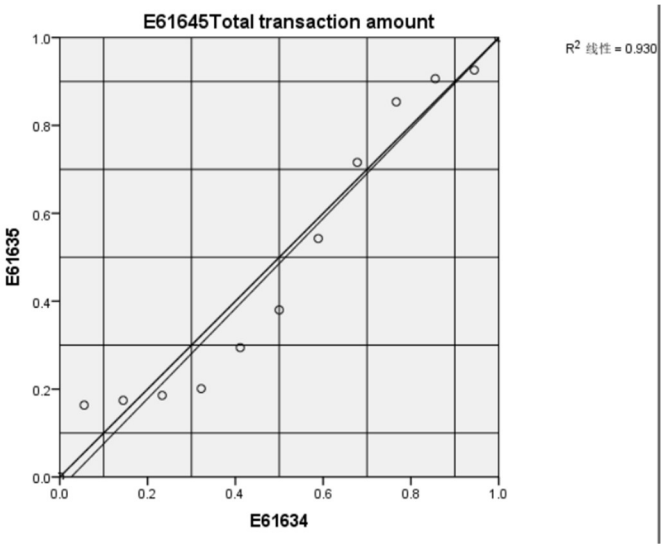


Figure 2. Fit lines drawn using SPSS

Forecast of transaction volume in 2022-2024:

$$y_{12} = -1269.6 + (517.1) \cdot 12 = 4935.6$$
$$y_{13} = -1269.6 + (517.1) \cdot 13 = 5452.7$$
$$y_{14} = -1269.6 + (517.1) \cdot 14 = 5969.8$$

In this year's "Double 11", Taobao and JD successively released war reports and hid GMV for the first time. Taobao implicitly said that it was "the same as last year"; JD said that it "exceeded the growth rate of the industry", so we have reason to believe that the predicted results are close to the actual situation.

5. Deficiencies

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.924[a]	.854	.838	747.06486	.854	52.691	1	9	.000	.719

a. Predictors: (Constant), Time

b. Dependent Variable: Total transaction amount

Figure 3. Data Model Overview

D.W statistics is used to test whether the residual distribution is normal distribution, because regression estimation using the least squares method assumes that the model residual is normal distribution. Therefore, if the residual is not subject to normal distribution, the model will be biased, that is, the model's interpretation ability is not strong.

D. The W statistic is about 2, which means that the residual is normally distributed. Here, the Durbin Watson value is too small to be convincing. The reason may be that the sample size is less than 30.

ISSN: 2692-7608 DOI: 10.6981/FEM.202311_4(11).0010

6. Conclusion

At the macro level, the current "Double 11" is a symbol to some extent. Through the "Double 11" activities, residents' strong consumption will and higher consumption capacity are displayed, which provides a positive signal for stimulating domestic demand and improving market confidence. The article takes the minimum multiplication method as the basic framework, combines the basic knowledge of linear algebra, and the application method is simple and clear, which is more suitable for these enterprises to carry out pre planning.

It can be used as a reference when measuring the product output, so as to maximize the profit forecast results for the enterprise as much as possible, and provide a basis for the enterprise to make business decisions and business plans.

References

[1] Yue Lingshui, Zhao Guigui. The application of least square method in the prediction of commodity sales [J]. Geological Technology and Economic Management, 1997 (01): 59-63.

[2] Jinliang; Principle derivation and code implementation of the general form and matrix form of the least squares method, CSDN.

[3] liyiersan123; Derivation and principle introduction of least squares method of linear regression, CSDN.

[4] Chen Qiuling, Chen Zhong. Application of Least Square Method in Automobile Sales Volume Forecast [J]. Cooperative Economy and Technology, 2012 (10).