Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 131 (2018) 825–831

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Congress of Information and Communication Technology (ICICT-2018)

Aero-Material Consumption Prediction Based on Linear Regression Model

Yanming Yang*, Lulu Sun, Chaoran Guo

Naval Aviation University Qingdao Campus, Qingdao, 266041, PR China

Abstract

It is indispensable to scientifically predict the consumption of aero-material spare parts and to make scientific decisions on aviation equipment maintenance resources and make full use of existing resources to improve maintenance capability. In this paper, the mathematic model and calculation method of linear regression model are introduced. And the parameter estimation and model test method of linear regression model is discussed. A linearization method is proposed for nonlinear problems. The application of linear regression model in forecasting the consumption of aero-material spare parts is analyzed by examples. Finally, the regression equation is analyzed for significance analysis, variance analysis and residual analysis. According to the analysis results, the regression equation is modified as necessary to further improve the prediction accuracy. The results show that the linear regression model is feasible and effective for the prediction of aero-material spare parts consumption.

© 2018 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Selection and peer-review under responsibility of the scientific committee of the 8th International Congress of Information and Communication Technology.

Keywords: aero-material consumption, linear regression model, parameter estimation, model test, residual analysis, prediction;

1. Introduction

The aero-material spare parts are essential material basis for aviation equipment maintenance engineering. With the development of aviation equipment, more and more complex equipment, maintenance of spare parts required for the variety and quantity are more and more spare parts financing, supply and storage process is also more complex. In this paper, we use linear regression method to forecast the consumption of spare parts [1]. The regression analysis and forecasting method is based on the analysis of the correlation between independent variables and dependent variables, the regression model between variables is built, and the regression model is used as the forecasting method. There are a lot of ways of regression analysis. Depending on the number of independent variables in the relationship,

* Corresponding author. Tel.: +86-13969715598.
*E-mail address:*yymqd@126.com

1877-0509 © 2018 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Selection and peer-review under responsibility of the scientific committee of the 8th International Congress of Information and Communication Technology
10.1016/j.procs.2018.04.271

the regression models can be divided into simple regression analysis and multivariate regression analysis. Depending on the correlation between independent variables and dependent variables, the regression models can be divided into linear regression forecasting and nonlinear regression forecasting. This paper focuses on simple linear regression prediction.

2. The principle of simple linear regression model

The linear regression is a linear method used to simulate the relationship between one dependent variable and many explanatory variables. The case of one explanatory variable is called simple linear regression model, while the case of multiple explanatory variables is called multivariate linear regression model [2-4]. As a commonly used statistical methods and for its principles is clear, model is simple and easy to use, classical linear regression model has been a very wide range of applications in the aviation equipment maintenance and support.

2.1. Regression model

The simple linear regression model is based on the approximate linear relationship between an independent variable and a dependent variable, and is fitted with a linear equation to predict the linear equation. A simple linear regression model is [5]:

$$y = a + bx \tag{1}$$

where y is the forecast object, known as dependent variables or explanatory variables; x is the influencing factor, known as independent variables or explanatory variables; a, b for the pending regression coefficient.

2.2. Parameter estimation

Estimating the parameters a, b in the model, from the point of view of curve fitting, least square method can be adopted. Suppose you have collected n pairs of data that predict the target y and the influencing factor x: (x_i, y_i) $(i = 1, \dots, n)$. After analyzing the historical data, assuming a linear relationship between y and x, you can use the regression model of equation (1). Applying the least squares method:

$$\begin{cases} b = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \\ a = \overline{y} - b\overline{x} \end{cases} \tag{2}$$

where $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$; $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

2.3. Model test

After the regression model is established, whether the model can be used for prediction or not, the model test is also needed. Common statistical tests are standard deviation test and correlation coefficient test.

(1) Standard deviation test. Standard deviation s, used to test the accuracy of regression prediction model, the formula is:

$$s = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{3}$$

where $\hat{y}_i$ is the predicted object actual value of the estimated value (or analog value).

The standard deviation s reflects the average error between the estimated and actual values obtained by the regression prediction model, so the smaller the value of s, the better. In general, the following requirements should be met:

$$\frac{s}{\overline{y}} < 10\% \sim 15\% \tag{4}$$

(2) Correlation coefficient test. The correlation coefficient is used to test the significance of the linear correlation between two variables, which is calculated as:

$$r = \sqrt{1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}_i\right)^2}} \tag{5}$$

It is easy to see that when r=l, the actual value completely falls on the regression line, and y is completely related to x; when r=0, y and x are completely uncorrelated; when 0<r<l, y and x has a certain correlation. Generally only when r is close to 1, we can describe the relationship between y and x using a linear regression model. To what extent is r, regression prediction model has practical significance? The actual test is through the critical correlation coefficient r_{α} (usually take a significant level $\alpha = 0.05$) to determine, this process is called correlation test.

2.4. Linearization of nonlinear problems

In practical problems, sometimes the relationship between y and x is not necessarily a linear relationship, but a certain curve relationship. For such problems, we should generally use curvilinear regression to describe them, but directly finding regression curves is rather difficult. For some special cases, it can be treated by variable replacement for linear regression problems.

To a nonlinear regression problem into linear regression, we must first determine (or approximate) type of nonlinear functions, and then see if you can use the linear variable replacement, in general; determine the type of nonlinear function is not easy, but some problems can be determined with professional knowledge and experience, or by mathematical method estimated. If the scatter plot of measured data is roughly around the following curve, that is, roughly showing a certain curve, it can take the corresponding replacement and turn it into a linear regression problem. The linearization methods of several common curves are as follows.

The power function:

$$y = a + x^b \tag{6}$$

Take the logarithm on both sides of the equation, then $\ln y = b\ln x + \ln a$; let $y' = \ln y$, $a' = \ln a$, $x' = \ln x$, then linear equations is $y' = a' + bx'$.

The exponential curve function:

$$y = ae^{bx} \tag{7}$$

Take the logarithm on both sides of the equation, then $\ln y = bx + \ln a$; let $y' = \ln y$, $a' = \ln a$, then linear equations is $y' = a' + bx$.

The logarithmic curve function:

$$y = a + b\lg x \tag{8}$$

Let $x' = \lg x$, then linear equations is $y = a + bx'$.

The hyperbolic function:

$$\frac{1}{y} = a + \frac{b}{x} \tag{9}$$

Let $y' = \frac{1}{y}$, $x' = \frac{1}{x}$, then linear equations is $y' = a + bx'$.

The *s* curve function:

$$y = \frac{1}{a + bc^{-x}} \tag{10}$$

Let $y' = \frac{1}{y}$, $x' = c^{-x}$, then linear equations is $y' = a + bx'$.

3. Application example analysis

3.1. Problem description

The number of aircraft landing and landing and the main tire consumption in 2012-2016 are shown in Table 1. Try to establish a simple linear regression model to predict the main tire consumption.

Table 1. Aircraft landing frequency and main tire consumption statistics in 2012-2016.

Year	Aircraft landing frequency				Main tire consumption			
	First quarter	Second quarter	Third quarter	Fourth quarter	First quarter	Second quarter	Third quarter	Fourth quarter
2012	623	239	289	302	70	26	36	33
2013	504	656	405	462	54	54	48	51
2014	745	543	648	432	83	76	60	44
2015	346	294	448	316	48	39	49	33
2016	672	592	503	572	78	62	53	54

3.2. Construct regression prediction model

Assuming that *y* is the main tire consumption, *x* is the number of aircraft landing, according to the equation (2) to calculate the regression coefficient: *b*=0.0934, *a*=7.776. Then get a linear regression model:

$$y = 7.776 + 0.0934x \tag{11}$$

Using Minitab software regression analysis function to get the following fitted line plot, as shown in Fig.1.

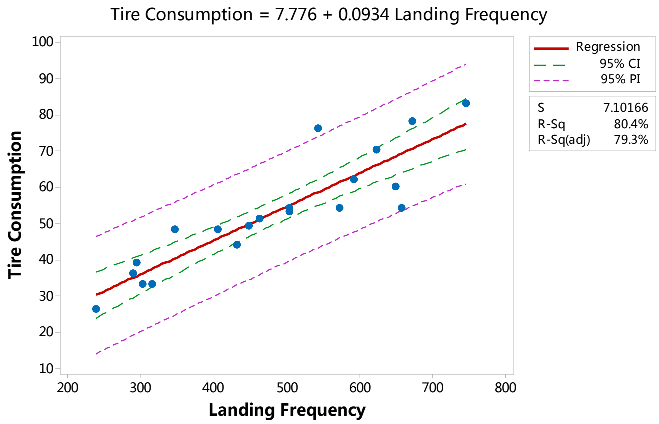


Fig. 1. Fitted line plot of aircraft landing frequency and main tire consumption.

3.3. Regression model test and analysis

(1) Significance test of regression equation. According to Table 2, analyze the results in the ANOVA table first. The *P*-value corresponding to 73.94 of the *F*-Value is 0.000 <0.05, so judging the regression equation as a whole is remarkably effective.

(2) A measure of the total effect of the regression equation. According to Table 3, R-Sq is very close to R-Sq (adjustment), and R-Sq (adjustment) is 79.3%, indicating that the fitted model can explain 79.3% of the variation in tire consumption y, so the model fitting has a better overall effect.

(3) Significance test of regression coefficient. According to Table 4, the independent variable x coefficient $p = 0 < 0.05$, indicating that the independent variable x is a significant factor. But the constant $p=0.170>0.05$ shows that the constant is not significant, and the regression equation should not contain the constant. Therefore, the equation (11) necessary amendments to obtain a new regression prediction model:

$$y = 0.108x \tag{12}$$

The model test and analysis are carried out for the equation (12). The test results of the linear regression model are shown in Table 5. The total effect measure of the regression model is shown in Table 6, and the significance test of the regression coefficient is shown in Table 7. According to the above model test and analysis results, all the tests of equation (12) passed. Therefore, regression equation (12) is more appropriate as a prediction model.

Table 2: Variance analysis of regression model (11)

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	3729.1	3729.14	73.94	0.000
Landing frequency	1	3729.1	3729.14	73.94	0.000
Error	18	907.8	50.43		
Total	19	4637.0			

Table 3: Model summary of regression model (11)

S	R-sq	R-sq(adj)	R-sq(pred)
7.10166	80.42%	79.33%	75.77%

Table 4: Coefficients of regression model (11)

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	7.776	5.44	1.43	0.170	
Landing frequency	0.0934	0.0109	8.60	0.000	1.0

Table 5: Variance analysis of regression model (12)

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	58856	58856.3	1106.41	0.000
Landing frequency	1	58856	58856.3	1106.41	0.000
Error	19	1011	53.2		
Total	20	59867			

Table 6: Model summary of regression model (12)

S	R-sq	R-sq(adj)	R-sq(pred)
7.29353	98.31%	98.22%	98.07%

Table 7: Coefficients of regression model (12)

Term	Coef	SE Coef	T-Value	P-Value	VIF
Landing frequency	0.108	0.00325	33.26	0.000	1.0

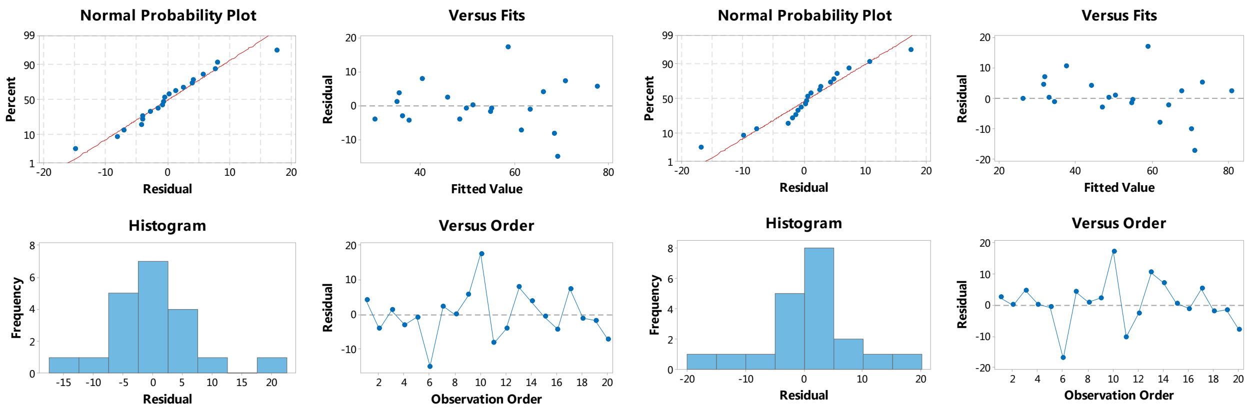
3.4. Residual plots analysis

Residual analysis was used to test the goodness of fit of the model in regression and analysis of variance. Checking the residual plot helps determine if the ordinary least-squares assumption is fulfilled. If these assumptions are satisfied, the ordinary least squares regression will yield an unbiased estimate of the smallest variance. [6]. Minitab software provides the following residual plots. The residual plot of the regression equation (11) is shown in Fig. 2 (a). And the residual plot of the regression equation (12) is shown in Fig. 2 (b).

(1) Residuals versus order of data. To observe the scatter plot of the residuals in the order of the observational order as the transverse axis, we focus on the random fluctuation of the residual value in the map in the horizontal axis. If random fluctuations, the residual values are mutually independent. In this example, the residuals fluctuate randomly and are independent of each other.

(2) Residuals versus fitted values. The emphasis is on whether the residual difference in the scatter plot is maintained or not, that is, whether there is a "funnel-shaped" or a "horn shape". If there is an obvious funnel shape or a horn shape, it is necessary to do some kind of transformation for the response variable *y*, and then fit the model after the transformation, and the fitting effect will be better. In this example, the graph is normal and the residual is equal variance.

(3) Normal probability plot of residuals. The normal probability graph of the residual is observed to see if the residual value is in the normal distribution. If the residuals are in a normal distribution, the point in the graph will generally form a straight line. In this example, the points are basically in a straight line, and the residuals can be regarded as normal distributions. The residual histogram in the lower left corner can be used to check the distribution of residuals. If you have one or two bars that are farther away from other bars, these may be outliers.

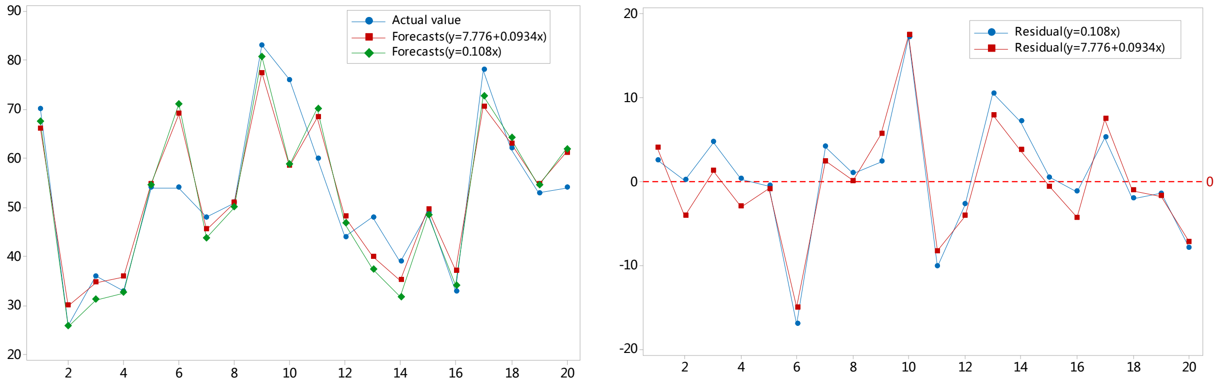


(a) (b)

Fig. 2. (a) residual plot of the regression equation (11); (b) residual plot of the regression equation (12).

3.5. Comparative analysis of prediction results

The regression equations (11) and (12) are respectively used to predict the main tire consumption of the aircraft. The predicted value and the true value are shown in Fig. 3 (a). The residuals of the two models are compared as shown in Fig. 3 (b). As can be seen from the figure, the regression model better fitted the actual value. At the same time, the residual values randomly fluctuate irregularly above and below the horizontal axis and are independent of each other with no abnormalities. Therefore, it is feasible and effective to use the linear regression model to forecast the aero-material consumption.



(a) (b)

Fig. 3. (a) comparison of prediction results and actual values; (b) comparison of residuals between two regression equations.

4. Conclusions

The regression analysis and prediction method is a classical and practical forecasting scheme. It is precisely because of its classic, so it is also mature, and it is easier to understand, and more widely used. In the actual use process, if we choose specific methods and models, we can analyze the data in detail, and observe and analyze the scatters, which can also be more detailed and the prediction results will be satisfactory. When the regression prediction method is applied, it is necessary to determine whether there is a correlation between the variables. If there is no correlation between the variables, the regression prediction method of these variables will result in the wrong result. When the regression analysis is applied correctly, we should pay attention to the qualitative analysis of the relationship between the phenomena, avoid the extrapolation of the regression prediction and apply the appropriate data. At the same time, the linear regression prediction method proposed in this paper not only applies to the prediction of aero-material spare parts consumption, but also to other equipment indexes or parameters, which provides a scientific method and means for equipment support forecast.

References

1. Yang, Yanming, W. Wang, and C. Guo. "Forecasting for Air Material Consumption Based on Winters Exponential Smoothing Model." International Conference on Automation, Mechanical Control and Computational Engineering
2. David A. Freedman. Statistical Models: Theory and Practice. Cambridge University Press. p. 26. (2009).
3. Rencher, Alvin C.; Christensen, William F., "Chapter 10, Multivariate regression – Section 10.1, Introduction", Methods of Multivariate Analysis, Wiley Series in Probability and Statistics, 709 (3rd ed.), John Wiley & Sons, p. 19, (2012).
4. Information on https://en.wikipedia.org/wiki/Linear_regression.
5. Yanming YANG, Yang YANG, Ruili ZHANG. Statistical Analysis and Application of Quality Management. Beijing: Tsinghua University Press, 2015.
6. Yang, Yan Ming, H. Yu, and Z. Sun. "Aircraft failure rate forecasting method based on Holt-Winters seasonal model." IEEE, International Conference on Cloud Computing and Big Data Analysis IEEE, 2017.