Research on Prediction of Equipment Maintenance Outlay Based on Linear Regression

FENG Zhang \(^1\), BING Chen \(^1\), SHOUquan Wang \(^2\), GAOlun Cui \(^2\) and WEibo Guo \(^2\)

\(^1\)Institute of Complexity Science Qingdao University, China
\(^2\)Naval Aviation University Qingdao Branch, China

\(^a\) Corresponding author: 18766286110@qq.com

Abstract. The equipment maintenance outlay is analyzed and predicted with R language statistics analysis software based on linear regression. The analysis results accord with the objective reality. Thus the method is proved practicably.

1. Introduction
Predication helps to guide analysis work of aviation maintenance. Pushing scientific maintenance and enhancing maintenance prediction have great importance on reducing the consumption of maintenance support resource and advancing the level of aviation maintenance management and the ability of supporting.

Regression analysis is an important method applied widely in statistics\([1-5]\). The correlation relations of variables are researched by means of this method. The philosophical elements of regression analysis are the mutual relations and object rule of the thing. The certain corresponding relations are functional relations but the relations do not exist frequently in practice. The regression analysis method can solve the problem of uncertain relations. The mathematical expressions of mutual relations are provided depending on regression analysis method. In practice the mathematical expressions is called as experience formulas. The experience formulas need validate test by means of relate knowledge of probability statistics. Moreover, it can be predicted and controlled variable at some extent, and the precision degree is calculated. In order to practice making decision and controlling work better, an important and secondary variables are distinguished depending on the influence degree of prediction variable over response variable.

With the high-speed development of information technology, some experts adore statistics analysis softwares such as R\([6]\], Excel, Minitab\([7]\), JMP\([8]\) and SPSS. The advantages of them are high-efficiency and high-stimulation precision. Because of open access and free virtue, R language is an widely adopted method in data analysis and visualization platform. Windows, Mac OS and Linux are convenient system for running R. In other words, the high efficiency of the statistics analysis software to some extent promotes the further development of statistics analysis and provides a big advantage for the research experts at analysing information.
2. Linear regression method

2.1 Principle of method
The question about adopting the expression of \( y = a + bx \) to estimate expectation of \( y \) is called simple linear regression. In the expression \( y \) is called response variable, and \( x \) is called prediction variable, and \( a \) and \( b \) are undetermined coefficient. These parameters are estimated by least square method. The expressions are listed as follows

\[
\hat{b} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \\
\sum (x_i - \bar{x})(y_i - \bar{y}) \\
= \sum (x_i - \bar{x})^2 \\
\hat{a} = \bar{y} - \hat{b}\bar{x}
\]

(1)

If the response variable \( y \) are related to prediction variables such as \( x_1, x_2, \ldots, x_n \), the question is called multiple regression. The expectation of \( y \) is estimated by the expression

\[ y = a + b_1x_1 + b_2x_2 + \ldots + b_nx_n. \]

To some extent the regression formula reflects inner rule of prediction variable and response variable. At the same time, the regression effect needs to validate after determining the regression formula.

\[ t - \text{value test method:} \]
Assumption \( H_0: b_i = 0, i=1,2, \ldots, m. \)
Statistics \( t_i = \frac{b_i}{\sqrt{c_{ii}s}} \)

\[ C = L^t = (c_{ij}) \]
\[ l_y = l_{ij} = (x_j - \bar{x})(y_j - \bar{y}) \]

In upper expression \( x_{ij} \) is \( k \) observation of \( x_j \)

Residual standard deviation \( s = \sqrt{\frac{Q}{n-m-1}} \)

Sum of squares of residuals \( Q = \sum (y_i - \hat{y}_i)^2 \)

\[ p - \text{value test method:} \]
\[ p(|t_i|) \] is the probability of random variable bigger than \( |t_i| \) following \( t \) distribution that degree free is \( n-2. \) If the expression \( p(|t_i|) \leq \alpha \) is true, the assumption \( H_0i \) will be rejected.

2.2 Procedure of method
There is a set procedure for achieving linear regression as follows.

(1) Estimating the function expression about multiple regression
The function expression about multiple regression is deduced according to experience and major knowledge. Moreover depending on scatter plot to choose is a very good method in practice. The regression formula is determined according to the determined regression coefficient and variance.

(2) Significant test about linear assumption
The regression effect is tested by means of t test method or p test method.
(3) Regression diagnosis
Depending on R language the diagnosis graph is produced to validate the effect. The indices such as standard deviation and deviation are major in testing effect.

3. Stimulation example
In view of protecting secret, stimulation data of open delivered literature is adopted[9]. In order to predict the relation of equipment quantity and equipment maintenance outlay, ten samples are taken out from different arm. Table 1. show the data information.

| NO | Equipment Quantity /Ten Thousand Pieces | Maintenance Outlay / Ten Thousand Yuan |
|----|----------------------------------------|----------------------------------------|
| 1  | 0.2                                    | 5.5                                    |
| 2  | 0.6                                    | 6.5                                    |
| 3  | 0.8                                    | 12                                     |
| 4  | 1                                      | 10                                     |
| 5  | 1.2                                    | 13                                     |
| 6  | 1.6                                    | 15                                     |
| 7  | 2                                      | 20                                     |
| 8  | 2.2                                    | 18                                     |
| 9  | 2.4                                    | 21                                     |
| 10 | 2.8                                    | 28                                     |

Procedure of analysis:
(1) Judging the form of regression function
The scatter plot graph of response variable and prediction variable is drawed by R statistics analysis software and is shown in the Fig.1. From this graph we can draw conclusion that the relation of equipment quantity and maintenance outlay is linear. The prediction work is conducted by linear regression method.
Figure 1. Relation of equipment quantity and maintenance outlay

Figure 2. Deviation and fitted graph
Figure 3. Normal Q-Q graph

Figure 4. Scale-Location graph
From Table 2, the regression formula is \( y = 3.1788 + 7.9197 \times x \). \( x \) and \( y \) stand for equipment quantity and maintenance outlay respectively.

(2) Significant test about linear assumption

From Table 2 we can understand the regression coefficients are significant at the level of \( p \) under 0.001. The results indicate the relation of equipment quantity and maintenance outlay is linear. In practice the relation of equipment quantity and maintenance outlay is linear indeed. In summary linear assumption is true.

(3) Regression diagnosis

Depending on R language the diagnosis graph is produced to validate the effect. In Fig.2 clear liner indicate that there no systematic conjunction between residual deviation value and fitted value. The fact that dependent variable and independent variable is linear correlation meet linear correlation assumption. In Fig.3 all points lie the 45 degree line, so the normality is true. In Fig.4 all points distribute around the horizontal line, and the assumption of the same variance is certain. In Fig.5 there no outlier, leverage point or strong influence point. In conclusion regression module meet statistical assumption, and is available and reasonable.

4. Conclusion

The linear regression analysis method is a mature method accepted widely at various fields. The advantage is easy and comprehensible. The R statistics analysis software is a convenient tool to
analyse data, and shorten miscellaneous operations, and construct very beautiful pictures. The researcher can make decision quickly by R software. The linear regression method may be introduced in aviation maintenance information analysis field to achieve the making decision goal of depending on data. Applying linear regression method to predicate the equipment maintenance outlay is important parts of fine support.

References
[1] LI ZHE, BAI Lin, LI Xingqian, ZHANG Hao, QU Xi, YU Hongmiao. Analysis method of on-orbit noise in manned spacecraft pressurized cabin based on non-linear regression[J]. SPACECRAFT ENVIRONMENT ENGINEERING, 2017, 34(1):63-69
[2] Kumar Hemant, Deb Debasis, ChaKravarty D. Design of crown pillar thickness using finite element method and multivariate regression analysis[J]. International Journal of Mining Science and Technology, 2017, (27):955-964
[3] Ajitanshu Vedtrtnam, Gyanendra Singh, Ankit Kumar. Optimizing submerged arc welding using response surface methodology, regression analysis, and genetic algorithm[J]. Defence Technology, 2018, (14):204-212
[4] Farhadian Hadi, Katibeh Homayoon. New empirical model to evaluate groundwater flow into circular tunnel using multiple regression analysis[J]. International Journal of Mining Science and Technology, 2017, (27):415-421
[5] GUI Bin, GUI Zhan-ji, CHEN LAN-sun. Statistical Analysis of Population Growth Model in Hainan[J]. MATHEMATICS IN PRACTICE AND THEORY, 2017, 47(12):166-171
[6] WANG Shu-Dong, ZHANG Shan-Qiang, HE Si-Cheng. Application of Mutual Information Network in Detection and Analysis of Breast Cancer Susceptibility Genes Using R Language[J]. Computer Systems & Applications, 2018, 27(1):143-148
[7] HOU Jia-Dong, YAN Feng, Li Ming. Reliability Analysis to Life Limit Parts of General Aviation Engine Based on Minitab
[8] LEI Xiang, JIANG Ling, ZHANG Shao-hua, ZHOU Run-ze. Research On The Secondary Development Methods Of JMPcSOFTWARE, 2014, 35(2):48-49
[9] YANG Yan-ming. Statistics Analysis and Application About Quality Management[M]. Beijing: tsinghua publishing company, 2015