Research on Consume Law Prediction of Plane Tyre Based on Principal Component Regression

FENG Zhang¹, BING Chen², SHOUquan Wang³, GAOlun Cui⁴, WEIbo Guo⁵

¹Complexity Science Qingdao University
Qingdao, China
18766286110, 86

²Complexity Science Qingdao University
Qingdao, China
13708989984, 86

³Naval Aviation University Qingdao Branch
Qingdao, China
13793235036, 86

⁴Naval Aviation University Qingdao Branch
Qingdao, China
18661627769, 86

⁵Naval Aviation University Qingdao Branch
Qingdao, China
15053273430, 86

¹8766286110@qq.com

ABSTRACT: Considering the problem of collinearity, the consume law of plane tyre is analyzed and predicted with R language statistics analysis soft based on Linear Regression. The analysis results accord with the real condition. Thus the method is proved practicably.

1. INTRODUCTION
The variables are orthogonality or non linear correlational when there are no linear relation completely. In the course of applying multielement regression method the condition that there is no strong linear relation between the prediction variables is assumed. When analysing the regression results, the regression coefficients are illustrated as the variable value of the response variable on the assumption of other prediction variables being constant. In fact, the power of persuasion is not adequate, considering that there is strong linear relation between the prediction variables. In this case, the regression results have some problem such as non-stabilization, low precision, low reliable regression module. The problem of strong linear relation between the prediction variables is called collinearity. Deleting some prediction variables is a method to eliminate or shorten collinearity, but it is not good method and is not working even sometimes.

The principal component method can find some several mutual disjoint comprehensive variables that reflect their information from prediction variables, and attain the goal of simplify variables. The method based on this principle is called principal component regression method in statistics.
2. PRINCIPAL COMPONENT METHOD

The variables such as $X_1, X_2, ..., X_p$ possess correlative relation from the same collective. The mutual orthogonal variables such as $C_1, C_2, ..., C_m$ can reflect information among the prediction variables and is acquired by combining the prediction variables linearly. In order to balance the role of variables exactly, it is necessary to implement the work of centralization and normalization. Every variable $C_j$ is the linearity combination of standardized variables. The equation is

$$C_j = v_{j1} \tilde{X}_1 + v_{j2} \tilde{X}_2 + \cdots + v_{jp} \tilde{X}_p$$

where $j=1,2, ..., p$.

Let $\lambda_j$ be the $j$th largest eigenvalue of the correlation matrix of $p$ variables, and the coefficient in the linear combination is the eigenvector of the corresponding eigenvalue of the correlation matrix,

$$\mathbf{v} = (v_{11}, v_{12}, \cdots, v_{1p}, v_{21}, v_{22}, \cdots, v_{2p}, \cdots, v_{p1}, v_{p2}, \cdots, v_{pp})$$

where $\text{Var}(C_j) = \lambda_j$.

The principal component covariance matrix is

$$\begin{bmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_p
\end{bmatrix}$$

Where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$, when a eigenvalue is 0, there is a strict linear relationship between the original variables. This is a special case. More commonly, one of the eigenvalue is much smaller than other eigenvalues.

The conventional regression analysis model is

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$$

The standardized regression analysis model is

$$\tilde{Y} = \theta \tilde{X}_1 + \cdots + \theta \tilde{X}_p + \epsilon$$

$\overline{Y}$ and $\overline{X}_j$ represent the mean of $Y$ and $X_j$ respectively.

$s_y$ and $s_j$ represent the standard deviation of $Y$ and $X_j$ respectively.

Standardized response variable is $\tilde{Y} = (y_j - \overline{Y}) / s_y$.

Standardized prediction variable is $\tilde{X}_j = (x_j - \overline{X}_j) / s_j$.

The relationship between the regression coefficients estimates of the above two models is

$$\tilde{\beta}_j = s_j \beta_j / s_y$$

$$\tilde{\beta}_j = s_j \beta_j / s_y$$

Substituting principal components for standardized predictors, then

$$\tilde{Y} = \alpha_1 C_1 + \alpha_2 C_2 + \cdots + \alpha_m C_m + \epsilon$$

where $\theta = v_1 \alpha_1 + v_2 \alpha_2 + \cdots + v_p \alpha_p$.

The purpose of principal component analysis is to replace as few primary components as possible with the original variables, and to ensure that data information is not lost. The Kaiser-Harris criterion is used to preserve the principal components with eigenvalues greater than one. The principal components with eigenvalues less than 1 have fewer variances than those contained in a single variable, so try not to use them.
3. PLANE TIRE
In the aviation maintenance process, the problem of spare parts consumption is a very important issue in the maintenance support process. The consumption of spare parts is related to many factors, but the influence degree of various factors on the consumption of spare parts is different. Therefore, the most important factors should be selected for analysis.

Plane tyre are made of plane rubber, nylon thread and steel. They work in a harsh environment. They are not only subjected to hot temperatures, but also against runway foreign objects and friction damage, and they also need to bear huge loads. Once the plane tyre fail, the consequences are very serious. Therefore, the protection of plane tyre is essential to ensure flight safety. The actual consumption of tyre is related to many factors such as the number of planes, flight hours, flight and landing, runway, weather, pilot operation, etc. The relevant experts agree that the most important factors affecting tire consumption are the number of planes, flight hours, flight and landing factors.

4. INSTANCE SIMULATION
The statistics of the number of tire consumption and the number of planes, flight hours, and flight and landing of a station from 1997 to 2012 are shown in Table 1. A multivariate regression model is established to predict the amount of tire consumption. The data is shown in Table 1.

| Year | Number of planes | Flight hours | Flight and landing | Consumption |
|------|------------------|--------------|--------------------|-------------|
| 1997 | 28               | 2596         | 1982               | 112         |
| 1998 | 30               | 2798         | 1996               | 116         |
| 1999 | 30               | 3033         | 2005               | 118         |
| 2000 | 31               | 3053         | 2169               | 124         |
| 2001 | 31               | 3088         | 2246               | 126         |
| 2002 | 32               | 3106         | 2277               | 128         |
| 2003 | 34               | 3256         | 2436               | 135         |
| 2004 | 33               | 3189         | 2383               | 133         |
| 2005 | 34               | 3246         | 2442               | 136         |
| 2006 | 35               | 3321         | 2515               | 139         |
| 2007 | 38               | 3546         | 2616               | 145         |
| 2008 | 38               | 3496         | 2596               | 143         |
| 2009 | 39               | 3683         | 2768               | 151         |
| 2010 | 38               | 3607         | 2722               | 149         |
| 2011 | 40               | 3802         | 2824               | 155         |
| 2012 | 42               | 3895         | 2923               | 159         |

| Year | Number of planes | Flight hours | Flight and landing | Consumption |
|------|------------------|--------------|--------------------|-------------|
| 1997 | -1.576           | -1.950       | -1.489             | -1.656      |
| 1998 | -1.096           | -1.386       | -1.443             | -1.375      |
| 1999 | -1.096           | -0.730       | -1.413             | -1.234      |
| 2000 | -0.856           | -0.675       | -0.869             | -0.812      |
| 2001 | -0.856           | -0.577       | -0.614             | -0.672      |
The correlation matrix of the three predictors is
\[
R = \begin{bmatrix}
1.000 & 0.980 & 0.980 \\
0.980 & 1.000 & 0.971 \\
0.980 & 0.097 & 1.000
\end{bmatrix}
\]

The corresponding eigenvalue is
\[
\lambda_1 = 2.954, \lambda_2 = 0.029, \lambda_3 = 0.017
\]

The corresponding eigenvector is
\[
V_1 = \begin{bmatrix} 0.578 \\ 0.577 \\ 0.577 \end{bmatrix}, \quad V_2 = \begin{bmatrix} 0.000 \\ -0.707 \\ 0.707 \end{bmatrix}, \quad V_3 = \begin{bmatrix} 0.816 \\ -0.409 \\ -0.409 \end{bmatrix}
\]

In this way, the main component of the plane tire consumption data is
\[
C_1 = 0.578 \hat{X}_1 + 0.577 \hat{X}_2 + 0.577 \hat{X}_3
\]
\[
C_2 = 0.000 \hat{X}_1 - 0.707 \hat{X}_2 + 0.707 \hat{X}_3
\]
\[
C_3 = 0.816 \hat{X}_1 - 0.409 \hat{X}_2 - 0.409 \hat{X}_3
\]

Table 3. Three main components of consumption data

| Year | $C_1$ | $C_2$ | $C_3$ |
|------|-------|-------|-------|
| 1997 | -2.896 | 0.326 | 0.120 |
| 1998 | -2.266 | -0.04 | 0.263 |
| 1999 | -1.870 | -0.483 | -0.018 |
| 2000 | -1.386 | -0.138 | -0.067 |
| 2001 | -1.182 | -0.026 | -0.211 |
| 2002 | -0.955 | 0.011 | -0.078 |
| 2003 | -0.131 | 0.087 | -0.073 |
| 2004 | -0.479 | 0.095 | -0.120 |
| 2005 | -0.136 | 0.121 | -0.069 |
| 2006 | 0.263 | 0.144 | -0.058 |
| 2007 | 1.235 | -0.063 | 0.136 |
| 2008 | 1.117 | -0.011 | 0.221 |
| 2009 | 1.886 | 0.023 | -0.030 |
| 2010 | 1.536 | 0.065 | -0.077 |
| 2011 | 2.323 | -0.080 | -0.046 |
| 2012 | 2.940 | -0.032 | 0.106 |

The regression results using conventional regression analysis methods are shown in Table 4.
Table 4. Regression results using conventional regression analysis method

| Variable | Coefficients | Estimate | Std.Error | t-value | Significance level |
|----------|--------------|----------|-----------|---------|--------------------|
| $\hat{X}_1$ | 0.055 | 0.055 | 1.012 | 1.000 |                      |
| $\hat{X}_2$ | 0.260 | 0.046 | 5.707 | 0.000 | ***                |
| $\hat{X}_3$ | 0.690 | 0.046 | 15.126 | 0.000 | ***                |

$n=16$ $R^2=0.991$ degree freedom=12

*** : (0,0.001) ** : (0.001,0.01) * : (0.01,0.05) : (0.05,0.1) blank : (0.1,1)

The regression results using principal component regression analysis are shown in Table 5.

Table 5. Regression results using principal component regression analysis method

| Variable | Coefficients | Estimate | Std.Error | t-value | Significance level |
|----------|--------------|----------|-----------|---------|--------------------|
| $C_1$ | 0.580 | 0.005 | 114.566 | 0.000 | ***                |
| $C_2$ | 0.304 | 0.051 | 5.906 | 0.000 | ***                |
| $C_3$ | -0.343 | 0.067 | -5.133 | 0.000 | ***                |

$n=16$ $R^2=0.999$ degree freedom =12

*** : (0,0.001) ** : (0.001,0.01) * : (0.01,0.05) : (0.05,0.1) blank : (0.1,1)

Table 6. Estimates of three principal component regression coefficients

| Variable | First principal component | First and second principal component | All principal component |
|----------|---------------------------|-------------------------------------|------------------------|
|          | normal | Original | normal | Original | normal | Original | normal | Original |
| Constant | 0       | 13.766   | 0      | 17.237   | 0      | 15.857   |        |            |
| number of planes | 0.335   | 1.146   | 0.335  | 1.146    | 0.055  | 0.189    |        |            |
| flight hours | 0.335   | 0.013   | 0.120  | 0.013    | 0.260  | 0.010    |        |            |
| flight and landing | 0.335   | 0.016   | 0.550  | 0.016    | 0.690  | 0.033    |        |            |
| $R^2$     | 0.997   | 0.997   | 0.997  | 0.999    |        |            |        |            |

Comparing the above two models, it is known that the significance is good, and $R^2$ is also increased, the regression effect is better, and the principal components are completely orthogonal, which has a computational advantage.

When three principal components are used completely, the principal component analysis method is equivalent to the least-squares method of constant regression. In order to eliminate the collinearity problem in the data, it is necessary to determine the number of principal components. In the plane tire consumption data, the sample variance of the principal component is, because the sum is relatively small, and the linear combination of the corresponding predictors is approximately 0, which is the root of the data collinearity. Since the original principal coefficients corresponding to the first principal component and the first principal component are substantially unchanged, and $R^2$ remains substantially unchanged. Therefore, considering the choice of a principal component for prediction, the corresponding regression equation is

$$Y = 13.766 + 1.146X_1 + 0.013X_2 + 0.016X_3.$$  

The regression equation corresponding to the conventional regression analysis method is
The results are different due to the collinearity problem.

5. CONCLUSION

The linear regression analysis method is used to predict the number of plane tyre consumed for different plane numbers, flight hours and flight movements, so that the storage can be “sense in mind” and better secure the aviation materials. In reality, the number of plane, flight hours, flight and landing are inevitably linearly related to a certain degree, that is, the overlap of information, resulting a different linear regression result.

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 Vedrtnam, 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] Ding CAI hong, HUANG Hao, YANG YANzhu. Description and Classification of Leather Defects Based on Principal Component Analysis[J/OL]. Journal of Donghua University(English Edition):1-7[2018-12-25]. http://kns.cnki.net/kcms/detail/31.1920.N.20181220.1744.010.html.

[7] HE Tong, XIONG Fengguang, HAN Xie, ZHANG Yuan. A Feature Curve Extraction Algorithm for Point Cloud Based on Covariance Matrix[J]. Computer Engineering, 2018, 44(3) :275-280

[8] WANG Rui. The Application of Ridge Regression in Solving the Problem of Collinearity of Economic Data[J]. Economic Research Guide, 2018, 0(22):144-147

[9] ZHAO Bin, WANG Chaoguang, LI Xiaolong, YU Ancai, LI Yang, ZHANG Jian. Design and Simulation of Aircraft Tires Testbench Based on ADAMS[J]. Machine Tool & Hydraulics, 2017, 45(10):15-18

[10] YANG Yan-ming. Statistics Analysis and Application About Quality Management[M]. Beijing: tsinghua publishing company, 2015