Estimation of single equipment standard in equipment maintenance expenditure based on cluster analysis

Wang Weijie\textsuperscript{1*}, Xiao Bin\textsuperscript{1}

\textsuperscript{1} Department of Management Engineering and Equipment Economics, Navy University of Engineering, Wuhan, Hubei, 430000, China

*Corresponding author’s e-mail: wangyiyun_wwj@163.com

Abstract. The single equipment standard in equipment maintenance expenditure is an important basis for applying for the expenditure needed for equipment daily maintenance. The estimation of single equipment standard is the main problem in the current equipment maintenance expenditure management, but the existing research results lack the quantitative analysis method. Aiming at this problem, first of all, this paper analyses the data estimated by all unit, uses clustering to simplify data size and improve data quality, then introduces the concept of expectation standard and establishes the unit satisfaction function about the expectation standard, and finally uses the particle swarm algorithm to obtain the optimal solution under the constraint of the total amount of expenditure. This paper provides a new idea of quantitative analysis for single equipment standard estimation.

1. Introduction

The equipment maintenance standard expenditure as an important part of the equipment maintenance expenditure is the main source of expenditure for the troops at all levels to perform the daily maintenance and support functions of the equipment and is usually obtained by multiplying the equipment strength by the standard [1]. At present, our military equipment maintenance expenditure standards are mainly divided into three parts: single equipment standard, unit standard and subsidy standard. The single equipment standard is the main basis for the daily maintenance of the equipment. It involves the largest number of categories, the largest amount of data, and the most complicated estimation. In addition, the single equipment standard is not affected by factors such as unit type and location, it is only related to equipment type, so it has a certain degree of independence and can be studied separately.

The previous single equipment standards of our military have generally been obtained in the form of inviting experts to organize meetings and decision-making by the authorities. This method is mainly based on the management experience of institutions and experts, is relatively simple and is greatly influenced by human factors. At present, in order to adapt to the pace of modernization of our military, and to strengthen the standardization management of equipment maintenance expenditure, the single equipment standard in equipment maintenance expenditure is in urgent need of large-scale adjustment. Faced with the complex and large-scale raw data, how to reasonably analyse the raw data, scientifically estimate the single equipment standard, and improve the accuracy and effectiveness of the standard, becomes an urgent problem to be solved in the management of military equipment maintenance expenditure.

Checking the relevant literature, most experts and scholars focus on the construction of expenditure standard system or the research on standard expenditure management mode, and less research on
standard estimation methods. In the literature [2], for the estimation of logistics expenditure standards such as heating expenditure, it was proposed to combine the longitudinal analysis of consumption data per unit over the years and the horizontal analysis of the same type of units with the annual consumption data, and to determine the standard by considering price factors in different regions. Literature [3] summarized the five steps in the estimation of expenditure standards and pointed out the importance of raw data for the estimation of expenditure standards.

Although there are some solutions to the problem of estimating expenditure standards, the research results are more general, lack of detailed discussion, and the proposed estimation method cannot better meet the new requirements of the military reform.

In view of this, this paper takes the actual needs of the troops as the starting point, uses clustering as the main analysis tool, analyses the data of the unit measurement data, and introduces the expectation standard and satisfaction function, and finally provides a theoretical reference for the estimation of the single equipment standard.

2. Cluster analysis
Cluster analysis is a multivariate statistical analysis method that classifies objects according to their similarity. It divides objects into several categories without prior knowledge, originating from taxonomy [4]. The basic idea of general clustering is to give a data set with \( n \) objects, constructing \( k \) groups according to the similarity size, and each group represents a cluster. The general steps are: firstly, determining \( k \) cluster centres, calculating the similarity between the remaining objects and each centre, and dividing the objects into different classes according to the principle of maximum similarity, secondly, re-determining the cluster centre, finally, repeat the previous steps until there is almost no difference between the data in the class.

Cluster analysis was first proposed by American scholar Macqueen [5]. It is currently used in many fields. Its general role is to mine deep information in the data and summarize the characteristics of each category in order to make corresponding decisions. For example, in the commercial field, the company divides customers by studying the consumption characteristics of different customers for similar products, and develops different marketing combinations for different types of customers, thereby increasing the sales of goods.

Considering that the general clustering method is greatly affected by the outliers, for the outliers, this paper uses the method of quadratic clustering to cluster the data, finds the outliers in the initial clustering. If an object does not strongly belong to any cluster, the object is an outlier that belongs to the cluster [6]. By diagnosing outliers, deleting noise information, retaining useful information, and ultimately improving the quality of data in the second clustering.

The main functions of clustering in this paper are:
- Diagnosing outliers, improving the quality of raw data;
- Combining units with similar actual conditions, reducing data complexity and simplifying data size.

3. Expectation standard and satisfaction function
In order to quantify the actual demand of units for the standard, this paper introduces the concept of expectation standard and satisfaction function, assumes that the unit has stable and well-organized preferences, and can objectively and truly reflect its real needs.

The expected standard \( b^0 \) is defined as the expected value of a unit for the single equipment standard about equipment \( p \). Satisfaction \( u(b_p) \) reflects the unit’s satisfaction level about the single equipment standard \( b_p \) when the unit expectation standard is \( b^0_p \), its maximum value is set to 100.

When the single equipment standard is less than the unit's expected standard, the unit believes that the normal maintenance and repair of the equipment is not effectively guaranteed, and the urgency to raise the standard is greater, and as the standard increases, the unit satisfaction rate rises faster; when
the single equipment standard is meeting or exceeding the unit's expectations, the rate of increase in satisfaction will slow down, and as the standard increases, unit's satisfaction should gradually approach the maximum of 100.

Therefore, referring to the logistic equation [7], this paper constructs the satisfaction function of the unit about the single equipment standard.

Its differential form is:

\[
\frac{du}{db_p} = \frac{ru(100-u)}{100}
\]  

(1)

Integration:

\[
u = \frac{100}{1 + e^{a-b_p}}
\]  

(2)

In the above formula, \( a = rb_p^0 \), \( r \) is a constant.

The function image is shown below:

![Figure 1. Unit's satisfaction function image.](image)

4. Steps and models of estimating single equipment standard
The main idea of single equipment standard estimation method based on cluster analysis is to use cluster analysis to merge the expectation standards of similar units, reduce the data complexity, and then obtain unit's satisfaction \( U \) to the standard combination \([b_1, b_2, \ldots, b_m]\) according to the satisfaction function, and strive to maximize \( U \) under the conditions of meeting total expenditure constraints.

The basic steps are as follows:

**Step 1: Clustering**

Assume that the total number of equipment categories that need to be estimated is \( m \). After collecting the unit estimation data, the units are sequentially grouped into \( k \) categories according to different equipment. For any equipment \( p \), if the total number of units that actually owning the equipment is \( n \), and the number of object variables is \( t \), then the object matrix of the cluster is:

\[
X_p = \begin{bmatrix}
x_{11} & x_{12} & \cdots & x_{1t} \\
x_{21} & x_{22} & \cdots & x_{2t} \\
\vdots & \vdots & \ddots & \vdots \\
x_{n1} & x_{n2} & \cdots & x_{nt}
\end{bmatrix}
\]

Then choose the similarity algorithm between the objects. Objects are divided into different categories by comparing the similarities between different objects.

The number of categories \( k \) can be determined by comparing the average contour value corresponding to the value of different \( k \). The larger the average contour value, the more obvious the
distinction between different classes [8]. The principle of determining the number of categories is: on the one hand, there is a clear distinction between classes, and on the other hand, the larger the number of categories, the better. The number of categories satisfying these two conditions is more appropriate. The average contour value is defined as:

$$S(i) = \frac{\min(b) - a}{\max[a, \min(b)]}$$ (3)

In the above formula, \(a\) is the average distance between object \(i\) and other objects in the same class, \(b\) is the average distance between object \(i\) and other objects in the different classes.

**Step 2: Diagnose outliers**

An outlier is a small number of data objects that behave abnormally in the dataset, either as noise information or as useful information. Referring to the methods proposed by other scholars to diagnose outliers [9-12], based on the initial clustering, this paper screens the outliers by calculating the outlier factors of each object. The outlier factor \(e\) is defined as the weighted average of the distance between each object and the various cluster centres.

Suppose the data set is \(D\) and is divided into \(k\) categories, such as \(\{D_1, \cdots, D_k\}\), the distance between object \(i\) and the cluster centre of \(D_j\) is \(d_{ij}\). Then, the outlier factor of object \(i\) is:

$$e_i = \sum_{j=1}^{k} \frac{|D_j|}{|D|}d_{ij}$$ (4)

Calculate the mean and standard deviation of all object’s outlier factors, refer to statistical knowledge, object \(i\) that satisfies the condition \(e_i \geq \mu + \beta \sigma, (1 \leq \beta \leq 2)\) is judged as an outlier.

Further analysis of outliers to determine whether to retain or reject. Repeat step 1 to perform secondary clustering on the processed data set and divide the units into several classes.

**Step 3: Standard Satisfaction Calculation for Single Equipment**

After the previous clustering, all units have been divided into several categories. At this time, similar units have a high degree of similarity, and a uniform standard can be used to indicate the expected standard of all units in the class. Considering that the units with more equipment have generally calculated the standard more reasonably, they are given higher weight when integrating the expected standards, and the weighted average of the expected standard of each unit in the class is taken as the expected standard of the same unit.

If the total number of class \(i\) units is \(l_i\), and the expected standard of unit \(j\) in the class is \(x_j\), the number of equipment it owns is \(c_j\), then the expected standard for equipment \(p\) in class \(i\) units \(b_{pi}^0\) is:

$$b_{pi}^0 = \sum_{j=1}^{l_i} \frac{c_j}{\sum_{j=1}^{l_i}c_j}x_j$$ (5)

According to the satisfaction function, the satisfaction of the class \(i\) units on the single equipment standard \(b_p\) of equipment \(p\) is:

$$u_i = \frac{100}{1 + e^{a-b_p}}$$ (6)

In the above formula, \(a = rb_p^0, r\) is a constant.
Next, the paper calculates the overall satisfaction of units with equipment $p$ on the standard $b_p$ by integrating the satisfaction of various units. If the number of unit classes is $k$, the ratio of the number of equipment owned by a certain unit to the total is $d_{pi}$, then the overall satisfaction $u(b_p)$ is:

$$u(b_p) = \sum_{i=1}^{k} d_{pi} u_i$$  \hspace{1cm} (7)

**Step 4: Build the single equipment standard estimation model**

According to the above steps, for the single equipment standard $b_p$ of any equipment $p$, there is a corresponding unit satisfaction value.

Increase the type of equipment and consider the importance of the equipment. Assume that the importance of each equipment is $\omega_1, \omega_2, \ldots, \omega_m (\sum_{i=1}^{m} \omega_i = 1)$, then units' satisfaction $U$ to the standard combination $[h_1, h_2, \ldots, h_m]$ is calculated as follows:

$$U = \omega_1 u(h_1) + \omega_2 u(h_2) + \cdots + \omega_m u(h_m)$$  \hspace{1cm} (8)

The greater the satisfaction $U$, the more the standard combination of single equipment meets the unit's requirements.

Consider the total amount of expenditure constraints and form a constraint function. If the total standard expenditure under the original standard level is $S_0$, the limit ratio of the expenditure is $\theta$, and the total number of each equipment is $c_1, c_2, \ldots, c_m$, then the final set of standard combinations must meet:

$$\sum_{i=1}^{m} h_i c_i \leq \theta S_0$$  \hspace{1cm} (9)

The standard combination that satisfies the constraints and has the highest units' satisfaction is the finalized single equipment standard.

**Step 5: Use the particle swarm optimization algorithm to find the optimal solution**

After the previous processing, the problem becomes: under the constraint condition, seek the optimal solution of the objective function. At this time, the particle swarm algorithm can be used to solve [13]. Due to space limitations, this paper only provides a brief description.

The decimal real number is used to encode the single equipment standard combination $[h_1, h_2, \ldots, h_m]$, and the position range is the upper and lower bounds of the various expected standards of the corresponding equipment.

Fitness function: $U = \omega_1 u(h_1) + \omega_2 u(h_2) + \cdots + \omega_m u(h_m)$. In particular, when the total amount of expenditure exceeds expectations, units' satisfaction $U$ is equal to 0.

Boundary conditions: The position or velocity of the particle exceeds the set value, regenerating the position and velocity of the particle.

Termination condition: The accuracy or number of iterations is met.

**5. Example analysis**

This paper selects three kinds of equipment for single equipment standard estimation, which are equipment 1, equipment 2, equipment 3, and the importance of equipment is $\omega_1, \omega_2, \omega_3$.

Some unit estimation data is shown below (Z-score standardization has been done, two decimal places are reserved):
Table 1. Partial raw data

| Unit   | Standard | Strength |
|--------|----------|----------|
| 1      | -0.97    | -0.52    |
| 2      | -0.42    | -0.52    |
| 3      | -1.08    | -0.72    |
| 4      | -0.48    | 2.93     |
| 5      | 0.14     | -0.72    |
| 6      | -0.24    | 0.09     |
| 7      | -0.49    | 2.12     |
| 8      |          |          |

Regarding equipment 1, the total number of units actually equipped with the equipment 1 is 27, let \( k \in [2,10] \), calculate the average contour value corresponding to different \( k \) values, and draw the line graph with the number of categories as the abscissa and the average contour value as the ordinate. The figure is shown below:

![Figure 2. Number of categories and average contour values](image)

As can be seen from figure 2, when the number of categories is 3, the average contour value is the largest, and the number of categories is moderate. Therefore, it is appropriate to take 3 as the number of categories. Then, do cluster analysis and calculate the size of the outliers of each unit. The mean of the outliers is 1.4163 and the standard deviation is 0.76844, let \( \beta = 1 \), then the threshold is (retaining two decimal places): \( \mu \sigma + \beta \sigma = 2.18 \). Comparing the threshold and the outliers of each unit, the units belonging to the outliers have units of 4, 7, 8, 24, and 25.

Further analysis shows that units 4, 7, and 8 are due to the fact that the number of possessed equipment is significantly higher than that of other units, resulting in large outlier factors, and their data should be taken seriously and still retained, while units 24 and 25 are significantly higher than other units, their data will have a large adverse effect on the measurement criteria, so their data should be eliminated.

Perform secondary clustering after renormalizing the processed data set. After comparing the average contour values, it is still appropriate to take the number of categories as 3. Calculate the ratio
of equipment strength and total equipment in various units and the expected standards. The results are shown in the following table:

Table 3. Basic information of various units (equipment 1)

| Cluster | Proportion of equipment owned | Expectation standard |
|---------|-------------------------------|----------------------|
| 1       | 10.08%                        | 14333.33             |
| 2       | 38.66%                        | 5591.30              |
| 3       | 51.26%                        | 5147.54              |

Characteristics of various clusters of units: The expectation standard of the first cluster of units are high, the units of the second cluster contain more units, and the units of the third cluster have more equipment.

For the same reason, calculate the expectation standards of various units of equipment 2 and equipment 3. Regarding equipment 2, according to the same principle, data of units 22 and 54 are eliminated. Regarding equipment 3, data of units 11, 32, and 50 are excluded. The calculation results are as follows:

Table 4. Basic information of various units (equipment 2)

| Cluster | Proportion of equipment owned | Expectation standard |
|---------|-------------------------------|----------------------|
| 1       | 10.93%                        | 15370.37             |
| 2       | 19.84%                        | 1738.57              |
| 3       | 36.44%                        | 3060.00              |
| 4       | 32.79%                        | 4501.23              |

Table 5. Basic information of various units (equipment 3)

| Cluster | Proportion of equipment owned | Expectation standard |
|---------|-------------------------------|----------------------|
| 1       | 24.02%                        | 1068.95              |
| 2       | 15.08%                        | 3648.15              |
| 3       | 60.90%                        | 1173.39              |

Refer to the expert opinion to determine the importance of the equipment as \( \omega_1=0.2 \), \( \omega_2=0.3 \), \( \omega_3=0.5 \). Consider the size of the standard data, the constant \( r \) is equal to 0.001.

About equipment 1: \( u(b_1) = \sum_{i=1}^{3} d_{i1}u_{i1} \); About equipment 2: \( u(b_2) = \sum_{i=1}^{4} d_{i2}u_{i2} \); About equipment 3: \( u(b_3) = \sum_{i=1}^{3} d_{i3}u_{i3} \). In the above formula, \( u_p = \frac{100}{1+e^{\theta b_p}} \).

Finally, using the particle swarm optimization algorithm to find the optimal solution, let the initial population number of particles be 100, the particle dimension be 3, the maximum speed be 100, the minimum value be -100, the learning factor be 1.5, and the inertia weight be 0.8.

Fitness function: \( U = 0.2u(b_1) + 0.3u(b_2) + 0.5u(b_3) \); Funding constraints: \( \sum_{i=1}^{3} b_iC_i \leq \theta S_0 \).

Query historical data, the original standard levels of the three types of equipment are 4000, 3000, 1500, then \( S_0 = 4000c_1 + 3000c_2 + 1500c_3 \).

When the proportion of restrictions are 1.2, 1.5, and 1.8, the calculation results are as follows (two decimal places are reserved):

Table 6. Calculation results

| \( \theta \) | Equipment 1   | Equipment 2   | Equipment 3   | \( U \) |
|-----------|--------------|--------------|--------------|------|
| 1.2       | 7134.50      | 1739.97      | 2307.22      | 55.17|
| 1.5       | 6922.05      | 3423.75      | 2686.06      | 65.91|
| 1.8       | 7249.37      | 4683.39      | 3178.01      | 76.23|
As can be seen from the experimental results, increasing the total amount of standard expenditure when the proportion of restrictions is small, in order to obtain higher satisfaction, the extent of the increase of the single equipment standard about less important equipment 1 should be appropriately reduced, and the extent of the increase of the single equipment standard about greater important equipment 2 and 3 should be improved; when the proportion of restrictions is large, the single equipment standard of equipment with higher importance shouldn’t increase significantly, while the single equipment standard of each equipment should be increased together to obtain the maximum satisfaction.

6. Conclusion
The quality of the single equipment standard directly affects the effectiveness of the equipment maintenance management expenditure and the effectiveness of equipment maintenance management. In this paper, the method of clustering analysis is adopted. By introducing the concept of expectation standard and unit satisfaction function, a quantitative analysis method suitable for single-standard standard measurement is constructed. Finally, the feasibility of the method is verified by an example. The research results show that the method is suitable for the analysis of large-scale measured data, and has good expansibility. At the same time, it can determine the single equipment standard combination according to the importance of equipment, and provides a new idea of quantitative analysis for single equipment standard estimation.

References
[1] Zhang Shujun. (2017) Research on management of equipment maintenance management expenditure. Journal of Naval University Of Engineering, 14: 78–81.
[2] Lv Lichang, Ming Zhaochun, Zhang Tong. (2009) Construction of the "Three Fees" Integrated Standard System. Military economic research, 1: 49-51.
[3] Zeng Dejun, Xu Hangjun. (2014) Discussion on the formulation of military funding standards. Modern trade industry, 22: 113-114.
[4] Sun Jigui, Liu Jie, Zhao Lianyu. (2008) Clustering Algorithms Research. Journal of Software, 19: 48–61.
[5] J.macqueen. (1967) some methods for classification and analysis of multivariate observations. FIFTH BERKELEY SYMPOSIUM, 1: 281-297.
[6] Jiawei Han (2012) Data mining concepts and technologies. Beijing Machinery Industry Press, Beijing.
[7] Cheng Maolin. (2003) Several Extension Models and Applications of Logistic Curves. Journal of Operations and Management, 12: 86–88.
[8] Zhou Ying, Zhuo Jinwu, Bian Yueqing (2016) Big Data Mining System Method and Case Analysis. Beijing Machinery Industry Press, Beijing.
[9] Ni Weiwei, Lu Jieping, Chen Geng, and Sun Zhihui. (2006) An Efficient Data Stream Outliers Detection Algorithm Based on k-Means Parti-tioning. Journal of Computer Research and Development, 43: 1639-1643.
[10] Huang Hongyu, Lin Jiaxiang, Chen Chongcheng. (2006) Review of Outlier Detection. Application Research of Computers, 8: 8-13.
[11] Xue Anrong, Yao Lin, Ju Shiguang. (2008) Survey of Outlier Mining. Computer science, 35: 13-18.
[12] Yang Bin. (2008) Research on Clustering Based Anomaly Detection Technology. Central South University, Hunan.
[13] Bao Ziyang. (2018) Intelligent Optimization Algorithm and Its MATLAB Example. Electronic Industry Press, Beijing.