https://doi.org/10.1051/e3sconf/202124503072

Study on safety evaluation of hazardous chemicals transportation industry in Capital city based on principal component analysis and cluster analysis

Zekai Lv[1], Shuran Lv[2]

[1] Management engineering department, Capital University of Economics and Business, Fengtai district, Beijing, 100070, China
[2] Management engineering department, Capital University of Economics and Business, Fengtai district, Beijing, 100070, China

Abstract. With the reform of chemical transportation industry accelerating, the safety evaluation of chemical transportation industry in China is becoming more and more standardized. Based on the theory of big data analysis and SPSS software as a platform, first of all, of the Beijing 91 the system construction and management of dangerous chemicals transportation enterprises, safety culture, safety production standardization 24 safety evaluation indexes such as principal component analysis, it is concluded that safety training basic requirements, loading and unloading operation, parking lot, civil air defense, vehicle dynamic supervision, system construction and management, safety production standardization, pilot eight kinds of factors as the main indexes for safety assessment. After using the eight indexes of 91 clustering analysis for transport enterprises, the enterprise is divided into five clusters of reasonable results, through the summary of common characteristics within the same cluster enterprises, it is concluded that the handling operation index plays an important role in classification for transport enterprises such as conclusion, can provide reference for other areas. At the same time, the conclusion can be used as a reference for the follow-up safety evaluation of hazardous chemical transportation enterprises, so as to achieve the purpose of targeted safety inspection.

1 Introduction

Hazardous chemicals (hereinafter referred to as "hazardous chemicals") refer to chemicals that are explosive, flammable, toxic, corrosive, radioactive and other in nature, and that are liable to cause personal casualties and property damage in the process of production, transportation, use, loading, unloading, and storage, and therefore require special protection[1]. The occurrence of hazardous chemical accidents seriously threatens the safety of people's lives and property, causes serious pollution to the environment, and reduces the city's ability to resist disasters[2]. According to statistics, 80% of the domestic hazardous chemicals needs to be transported by highway, and the total amount of dangerous goods transported by road every year is 300 million tons. With the rapid development of economy, the quantity of dangerous chemical cargo is also on the rise. According to relevant experiments, hazardous chemicals in transit are dynamic and dangerous. Once accidents occur, they will cause great harm to the lives and property of surrounding residents[3]. In recent years, chemical transport accidents occur frequently, such as the "8•26" accident of Baomao expressway, the "7•19" especially serious road traffic explosion accident of chemical products, and the "11•1" typical transport accident of Fuquan Expressway, etc., which all sound the alarm for people[4]. The background and operation status of hazardous chemical transportation enterprises in Beijing are different, and the current evaluation of hazardous chemical transportation enterprises is not targeted. The purpose of this paper is to classify and summarize the characteristics of various types of enterprises and identify the weak links of accidents.

At present, there are many researches on hazardous chemical transportation industry in China. For example, **Luo Dan** et al. [1] analyzed the characteristics and causes of hazardous chemical transportation accidents from the time and province of the accidents through statistics of 356 hazardous chemical transportation accidents occurred in China from 2013 to 2017, and proposed countermeasures from multiple perspectives. **Zhao Jiangping** et al.[5] applied disturbance origin theory (P theory) and explanatory structure model (ISM) to the construction of scenario model of accident emergency management, and used this model to carry out scenario analysis of events related to emergency management in event concentration. The above studies have analyzed the important factors of accidents in the transport of dangerous chemicals from different perspectives, but it is not enough to analyze them from the perspective of different links of safety evaluation in the transport of dangerous chemicals.

Combining with Beijing for transportation safety evaluation report and related data, using SPSS software for the Beijing 91 the system construction and management of dangerous chemical enterprises, safety

*Corresponding author's e-mail: 1227005186@qq.com

culture, safety production standardization 24 safety evaluation data such as principal component analysis and cluster analysis, which can extract the most critical of Beijing for transport enterprises safety evaluation indexes, and indicators identified in Beijing for the different types of transport enterprises, in order to summarizes situation of existing for transportation industry in Beijing, and provide reference for other regions[6].

2 Theoretical models and data sources

2.1 Principal component analysis

Principal component analysis (PCA) uses the idea of dimensionality reduction to transform multiple indexes in complex problems into a few comprehensive indexes, and improves the efficiency on the basis of ensuring the accuracy of the results[7]. Specific steps are as follows:

(1) Significance test. Testing is the premise of analysis, and only the results of qualified analysis are meaningful. Significance test refers to KMO value and Bartlett significance coefficient to judge the correlation between the analyzed objects.

(2) Extract the variance of common factors. Through variance, we can know the degree of retention of each index after analysis so as to prove whether the index after analysis is still representative.

(3) Statistic the cumulative contribution rate of each component. From the contribution rate, we can get the different importance degree of each index to the overall evaluation. Firstly, the correlation coefficient matrix R is determined, and the similarity measurement of index variables is carried out with the correlation coefficient. The correlation coefficient matrix R= (rjk) m×m, where

$$r_{jk} = \frac{\sum_{i=1}^{n}\left(\overline{x_{ij}} * \overline{x_{ik}}\right)}{n-1} \quad (j, k = 1,2,...,m) \quad (1)$$

In the formula: rjj =1, rjk = rkj, rjk is the correlation coefficient between the jth index and the kth index.

Then, compute the eigenvalues and eigenvectors. Calculate the eigenvalue of the correlation coefficient matrix R, λ1⩾λ2⩾... ⩾λM ⩾0, and the corresponding eigenvectors μ1, μ2... , μm, where μj =(μ1j, μ2j... , μ nj)T, and m new index variables are composed of eigenvectors.

$$\begin{cases} y_1 = \mu_{11}x_1 + \mu_{21}x_2 + ... + \mu_{n1}x_n \\ y_2 = \mu_{12}x_1 + \mu_{22}x_2 + ... + \mu_{n2}x_n \\ ... \\ y_m = \mu_{1m}x_1 + \mu_{2m}x_2 + ... + \mu_{nm}x_n \end{cases} \quad (2)$$

In the formula:y1, y2... ym is number 1,2, and... , m principal component[8].

2.2 Cluster analysis

As a multivariate statistical method in the theory of multivariate statistical techniques, cluster analysis, also known as group analysis and point group analysis, contains two categories, one is system clustering method, the other is K-means clustering method[4]. The idea of the system clustering method is to treat each sample as a class, combine the most similar classes according to the distance or similarity degree between the classes, and then calculate the similarity degree between the new class and other classes, until all the classes are combined into a class[9].

The research content of this paper is more appropriate to choose the system clustering method, mainly for the following reasons: first, the system clustering method performs clustering under uncertain conditions, so as to show the possibility of all clustering, and make the clustering result more rigorous; second, the system clustering method is more accurate. There is no artificial setting of the number of clusters so as to reduce the influence of subjectivity. Second, due to the small size of the sample data, it effectively avoids the shortcoming of low efficiency of system clustering in practical operation for large quantities of security statistical data[10]. The specific process is as follows:

(1) Feature extraction. According to the input of the original sample to determine which features to characterize the essential nature and structure of the sample.

(2) Execute the clustering algorithm to obtain the clustering genealogy graph (tree graph). Each sample in the input matrix is assumed to be a point in the characteristic variable space, without the participation of domain experts, and only geometric knowledge is used to obtain the most essential "clustering" property that can reflect the sample points in the X-dimension space. The clustering pedigree graph output by the system clustering method can reflect the situation of all classifications from rough to fine. The fast clustering method gives the specific classification scheme directly, including several categories in total, and the specific samples contained in each category.

(3) Select an appropriate classification threshold. According to the distance between each item, SPSS judged the appropriate threshold, so that the classification scheme can be directly seen on the clustering pedigree diagram.

3 Example application and analysis

3.1 Application of principal component analysis

Before conducting principal component analysis on the 91 hazardous chemical transportation enterprises mentioned above, significance test should be conducted first, as shown in Table 1.

Table 1. KMO and Bartlett tests

Kaiser-meyer-olkin measurement with sufficient sampling	The approximate chi-square	Bartlett's sphericity test (df)	Bartlett's sphericity test （Sig）
0.813	958.727	276	0.000

 https://doi.org/10.1051/e3sconf/202124503072

As can be seen from table 1, KMO value =0.813 > 0.6, bartlett significance coefficient sig=0.00 < 0.05, indicating that the data obey normal distribution and the variables have a strong correlation, indicating that principal component analysis is applicable.

Then, the statistical analysis software SPSS was used to conduct principal component analysis on 24 safety evaluation indexes of hazardous chemicals transportation enterprises in Beijing, and the common factor variance table of each index was obtained, as shown in Table 2.

Table 2. Common factor variance table

factors	The initial	extract
Driver management	1.000	.736
Vehicle dynamic monitoring personnel	1.000	.655
Vehicle basic requirements	1.000	.644
Vehicle dynamic monitoring	1.000	.756
Handling operation	1.000	.770
Basic parking requirements	1.000	.758
Technology protection	1.000	.595
Personnel protective	1.000	.756
Management organization and personnel	1.000	.654
Safety responsibility system	1.000	.561
System construction and management	1.000	.741
Input into production safety and use of funds	1.000	.703
Work safety meeting	1.000	.614
Safety training	1.000	.775
Safety culture	1.000	.675
Safety rewards and punishments	1.000	.616
Safety inspection	1.000	.725
Emergency management	1.000	.612
Contract management	1.000	.568
Traffic violation and accident management	1.000	.592
Occupational health management	1.000	.712
Personnel assessment	1.000	.495
Information management	1.000	.694
Safety production standardization	1.000	.736

The extracted values in Table 2 indicate the extent to which each index can express the original data after principal component analysis. At the same time, we can also get the cumulative contribution rate of variance of each component, as shown in Table 3.

Table 3. Cumulative contribution rate of difference

ingredients	Initial eigenvalue		
	A combined	% of the variance	Cumulative %
1	7.694	32.059	32.059
2	2.013	8.386	40.445
3	1.667	6.945	47.390
4	1.399	5.828	53.218
5	1.193	4.972	58.190
6	1.145	4.769	62.959
7	1.033	4.302	67.262

By table 3 can be concluded that when take 7 principal components, can express the original data by about 67%, more than 60%, we can accept is the extraction of the scope of the principal component, so should be the size of the values in table 2, the largest in the first seven indicators as the main ingredient, at a time when we find the extraction of 7 and 8 values are the same, so at this time will principal component extraction for eight, about 71% of the original data expression.

To sum up, it can be determined that there are 8 main indicators affecting the safety evaluation, which in order of influence degree are: safety training, loading and unloading operations, basic requirements of parking lot, civil air defense, vehicle dynamic supervision, system construction and management, safety production standardization, and driver management.

3.2 Application of cluster analysis

For the 24 samples in Table 1, 7 principal components obtained by principal component analysis were used as indicators to classify enterprises, and a cluster analysis pedigree was obtained, as shown in Figure 1.

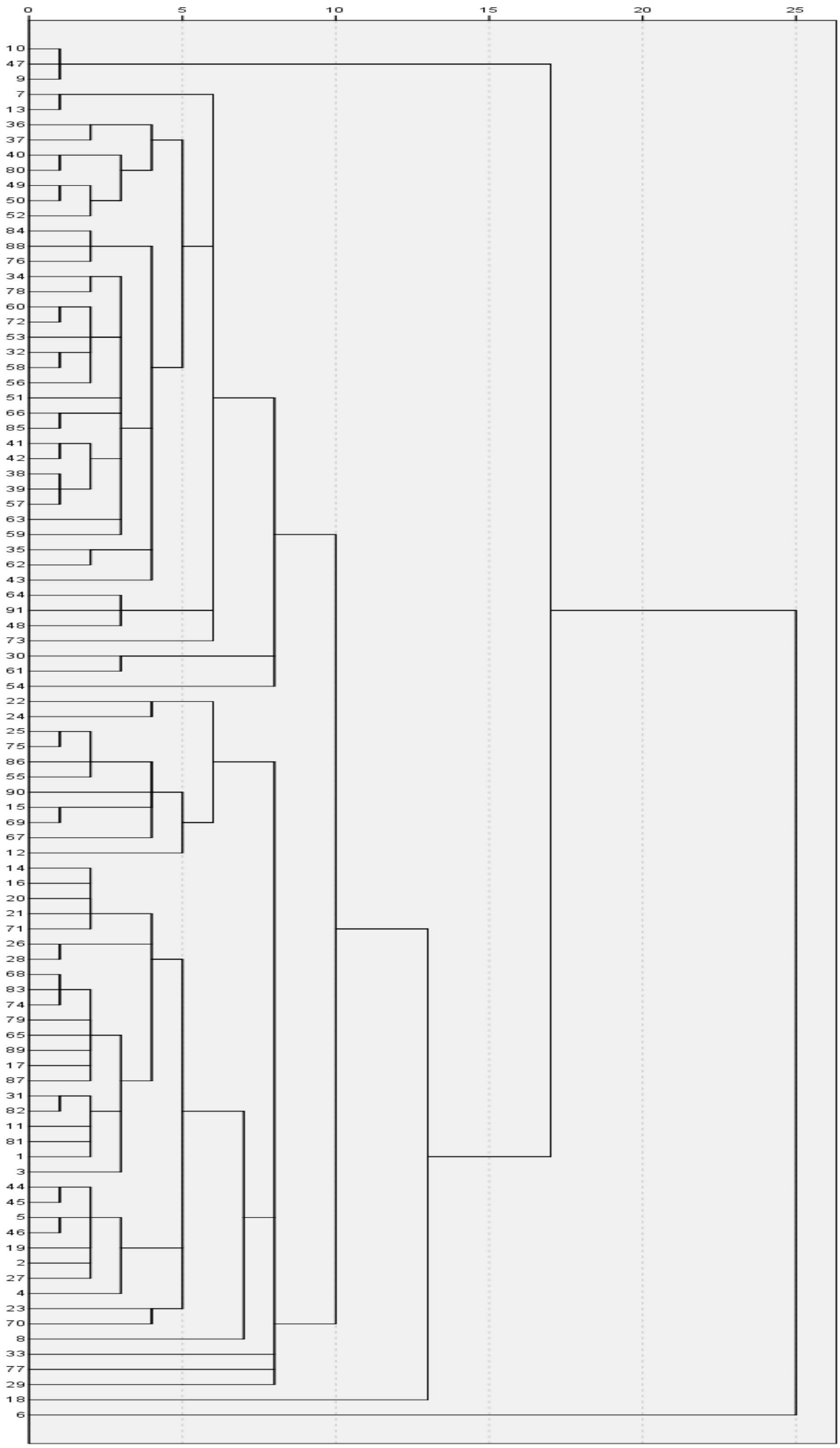

Figure 1. Cluster analysis pedigree diagram

It can be seen from Figure 1 that it is more appropriate to divide the enterprises into five clusters, which are found to have the following characteristics: the business scope of the first type of enterprises does not include loading and unloading operations, and it mainly focuses on natural gas transportation. These enterprises are all state-owned enterprises, and the overall safety management level is high, especially in the system construction and management, safety production standardization performance is outstanding. The second category of enterprises does not include loading and unloading operations, but the overall level of safety management has

https://doi.org/10.1051/e3sconf/202124503072

yet to be improved, and most of them are private enterprises. The third category of enterprises also does not include loading and unloading operations, but the overall level of safety management is poor, but the number of such enterprises is very small. The fourth category covers loading and unloading operations, and most have problems with "safety training" and "vehicle dynamic monitoring". The fifth type of enterprise is an individual case. In the sample of this study, there is only one enterprise, which has the qualification to transport hazardous chemicals, but is not engaged in relevant business at present.

4 Conclusion

Through principal component analysis of various safety evaluation indexes, this study finds that safety training, loading and unloading operations, basic requirements of parking lots, civil air defense, dynamic vehicle supervision, system construction and management, safety production standardization, and driver management are the main factors affecting the safety evaluation results of hazardous chemicals transportation[6]. Then, based on cluster analysis of composition indexes of some hazardous chemical transportation business owners in Beijing, the following conclusions are drawn:

1. The indicators of "loading and unloading operations" in the eight indicators play a crucial role in distinguishing enterprise types, and are one of the indicators to be paid special attention to in safety assessment.

2. Among the enterprises excluding the business scope of loading and unloading operations, state-owned enterprises with natural gas transportation as the main business have a higher overall management level, especially in system construction and management, and production safety standardization; However, the vast majority of private enterprises excluding the scope of loading and unloading operations still need to improve the overall level of safety, safety awareness needs to be strengthened, and even a few enterprises have poor safety management status, continue to rectify.

3. Among the enterprises covering the business scope of loading and unloading operations, most of them still have big problems in the two indexes of "safety training" and "vehicle dynamic supervision".

4. At present, there are still very few enterprises in Beijing that have the qualification to transport hazardous chemicals but have not engaged in relevant businesses, which deserve the attention of relevant departments.

To sum up, "loading and unloading" is a weak link in the transport industry of hazardous chemicals, which should be paid close attention to in the safety evaluation, and whether "loading and unloading" can be regarded as an important standard to distinguish enterprise types. The safety management and safety training level of hazardous chemical transportation in private enterprises are insufficient, and more attention should be paid to private enterprises in the future safety evaluation.

References

1. Luo Dan, Liao Chen-juan, YU Xinrou. Statistical Analysis and Prevention Countermeasures for Transportation accidents of hazardous chemicals in China from 2013 to 2017 [J]. Public Security in China (Academic edition),2019(01):28-32.
2. Li Na, Chen Jianhong. Statistical Analysis of Hazardous chemicals in China from 2013 to 2019 [J]. Applied chemical engineering,2020,49(05):1261-1265.
3. Chen Yue, Zhang Yuling. Research on Risk Assessment and Management of Road Transport of Dangerous Goods Based on index Model [J]. Road Traffic Science and Technology, 2008,35(03):143-150.
4. Qi Caijuan, Yu Xiaobao, Han Yiming, Dang Dongsheng, GUO Ning. Cluster Analysis of county-level Power supply enterprises based on internal and external factors [J]. Enterprise Reform and Management,2019(01):217-218.
5. Zhao Jiangping, Zhao Tengfei, Yang Zhen, Dong Shu, Fan Chunlei, Ren Xingmin. Research on emergency management of chemical road transport accidents based on p-ism [J]. Logistics science and technology,2020,43(03):39-43+69.
6. Shi Xuefeng, Ding Fangran, Cheng Wei. Health supervision and evaluation of disinfection products based on principal component and cluster analysis [J]. Chinese journal of preventive medicine,2017,18(04):277-280.
7. Zhang Ru, ZHANG Xueyang, Lu Hongguang, Liu Qiang. Evaluation of Urban Air Quality based on AhP and PCA -- A Case study of Xuzhou city [J]. Safety and environmental engineering,2017,24(03):103-107.
8. Yan Chengyong, Zhang Wenjun, Yin Jianchuan, Wang Pinglin, Xue Zongyao. Comprehensive evaluation of ship maneuverability based on principal component Analysis [J]. Journal of Shanghai Maritime University,20,41(02):11-15.
9. Wang Jingyu, Li Cuiping, Li Zhongxue. Prediction of water inrush risk from coal floor based on principal component clustering analysis [J]. Chinese journal of safety science,2013,23(08):120-125.
10. Qin Su Lin, Zhou Jun. Application of Cluster Analysis based on Fuzzy Equivalence matrix in the Assessment of production staff in textile enterprises [J]. Industrial control computer,2015,28(09):119-121.