3δ principle and boxplot method are used to estimate the cost interval of tower replacement technology

Jianqing Li¹, Xiaohu Zhu¹, Chen Zhao²*, Jian Kang¹
1State Grid Anhui Electric Power Company, Hefei, China;
2China energy engineering group Anhui electric power design institute, Hefei, China;
*Corresponding author e-mail: zcncepu@163.com

Abstract. The replacement of tower is an important part of the technical reform of power grid lines. To analyze the rationality of the cost of the technical reform of tower, it is necessary to deal with the abnormal engineering data based on the cost data of previous projects. In this paper, the unit cost of tower replacement is sorted out by using 3δ principle and boxplot method, and the reasonable cost range is obtained to provide a basis for project investment.

1. Sample information
This time, 138 items of data of the transmission tower replacement project from 2014 to 2017 were sorted out. Before the sorting, the cost sample range of each voltage level was 500kV: 409,700 yuan per base - 904,100 yuan per base, 220kV: 35,500 yuan per base - 407,000 yuan per base, 110kV: 22,100 yuan per base - 246,600 yuan per base, and 35kV: 101,100 yuan per base - 120,000 yuan per base. As the sample of tower with voltage grade of 500kV is no more than 5, no interval sorting is done for it. This time, only 220kV, 110kV and 35kV tower projects are sorted and analyzed. [1]

Table 1. Table of engineering samples of various types of pole tower (tower body) replacement

| The serial number | Voltage grade | type | unit | Data year interval | Sample size | The average | Sample data interval |
|-------------------|---------------|------|------|--------------------|-------------|-------------|---------------------|
| 1                 | 500 kV        | Angle steel tower | Ten thousand yuan per base | 2014-2017. | 3 | 58.42 | [40.97, 90.41] |
| 2                 | 220 kV        | Angle tower/steel tube tower | Ten thousand yuan per base | 2014-2017. | 77 | 13.5 | [3.55, 40.70] |
| 3                 | 110 kV        | Angle tower/steel pole | Ten thousand yuan per base | 2014-2017. | 33 | 8.76 | [2.21, 24.06] |
| 4                 | 35 kV         | Angle tower/steel pole | Ten thousand yuan per base | 2014-2017. | 25 | 4.78 | [0.11, 12.00] |

Sample description statistics with voltage grade of 220kV:
Table 2. 220kV sample description statistics (before sorting out)

| Replace the tower (tower body) 220kV | statistic | Standard error of |
|-------------------------------------|-----------|------------------|
| The median                          | 11.3400   |                  |
| The variance                        | 56.036    |                  |
| The standard deviation              | 7.48575   |                  |
| minimum                             | 3.55      |                  |
| The maximum                         | 40.70     |                  |
| The scope of                        | 37.15     |                  |
| The interquartile range             | 8.94      |                  |
| Partial degrees                     | 1.760     | 274.             |
| kurtosis                            | 3.501     | 541.             |

Data scatter diagram:

![Scatter diagram of unit cost of 220kV tower pole (tower body) project](image)

Figure 1. Scatter diagram of unit cost of 220kV tower pole (tower body) project

2. Organizing method
The data analysis of all technical transformation cost data is mainly divided into three parts: [2]

2.1. descriptive statistics of samples
Mean, median, variance, standard deviation, minimum, maximum, range, quartile distance, skewness, kurtosis, scatter plot, etc. [3]

Software implementation: SPSS software is used for descriptive statistical analysis.

2.2. outlier removal
Common methods to screen outliers are 3δ principle and boxplot method.

3δ principle. [4]

First, use the sample data to plot the histogram of frequency distribution.

Secondly, the probability density curve is fitted according to the display result of frequency distribution histogram, and the corresponding probability density function is obtained. When data is normally distributed, mu is the mean of sample data and delta is the deviation of sample data according to the definition of normal distribution. The probability of \( P(|x-\mu|>3\delta) \) less than 0.003δ is a very small probability event, and by default we can assume that there are no samples outside the mean of 3δ.
Therefore, when the average distance from the sample data is greater than $3\delta$, the sample data is considered to be an outlier. [5] When data does not obey normal distribution, it can be determined by the standard deviation of the number of times away from the average distance, and the value of the number of times needs to be determined according to experience and actual situation, as shown in figure 2-1. [6]

![Figure 2. 3δ principle corresponding to the normal distribution](image)

Data in this report are sorted out by boxplot method. The boxplot consists of five parts, which are the minimum value, median, maximum number and two quartiles Q1 and Q2. The abnormal value is the value less than $q1-1.5IQR$ or greater than $Q2+1.5IQR$ in the sample. [7]

The first step is to calculate Q1, the first quartile of the sample, which is equal to the 25% of all the Numbers in the sample from the smallest to the largest.

The second step is to calculate the median $F$ of the sample, which is equal to the 50% of all the Numbers in the sample arranged from smallest to largest. [8]

The third step is to calculate Q2 of the sample, which is equal to the 75% of all the Numbers in the sample arranged from smallest to largest.

Step 4, calculate the sample quartile distance $IQR$

$$IQR = Q2 - Q1$$

Finally, the values of $q1-1.5IQR$ and $Q2+1.5IQR$ were calculated, and the outliers outside the interval were eliminated. [9]

SPSS, python and matlab software implementation: used for drawing and calculation, boxplot the two ends of the beard is not the exact length of 1.5 times the box, but no more than the value of the length of the longest, so using the python/matlab calculation scope of outliers, because the cost is positive, the data for the lower limit of the negative to the area between the minimum value, abnormal value ultimately determine the SPSS output boxplot results shall prevail.

After the outliers are selected, it is necessary to analyze the outliers according to specific projects, and analyze the causes of the outliers, so as to reduce the outliers in the data filling of subsequent technical renovation cost.

2.3. overall interval estimation

Interval estimation from sampling point estimate and standard error, according to the given probability value contains to estimate parameters of interval is established. With the given probability value is called the confidence level or confidence level, the built contains's called the confidence interval estimation function, refers to the overall parameter values fall in the sample statistic probability of a certain area. [10]

Since the sample size of many data is small and the overall distribution is unknown, Walsh average order statistics are used to construct the confidence interval of the center of symmetry. This method is
independent of the overall distribution and can estimate the confidence interval more accurately. Set $x_1, x_2, \ldots, x_n$

Is the original data, is a continuous symmetric independent identically distributed the symcenter of random sample, calculate walsh average

$$W(\cdot) = \frac{x_i + x_j}{2} \quad 1 \leq i \leq j \leq n$$

(2)

Exponential order of Walsh mean, denoted as $W_1(\cdot)$ cars only $W_2(\cdot)$ ... $W_N(\cdot)$ cars only $N = \frac{1}{2} n(n + 1)$ cars only

If you give confidence $1 - \alpha$, it is estimated that the $\theta$ Location interval $(\widehat{\theta}_L, \widehat{\theta}_u)$, determined by the following equation.

$$\widehat{\theta}_L = W(K), \quad \widehat{\theta}_u = W(N - K + 1)$$

(3)

Among them:

$$K \approx \frac{N}{2} - \frac{Z_{\alpha/2}/\sqrt{n}}{2/\sqrt{3}}$$

(4)

Here, $Z_{\alpha/2}$ is a normal random variable $\alpha/2$ Critical value, if $\Phi(\cdot)$ is the standard normal distribution function, then

$$1 - \Phi(Z_{\alpha/2}) = \frac{\alpha}{2}$$

(5)

The calculation shall be carried out in the following steps:

by $\alpha$, if you look at the normal distribution table, you get $Z_{\alpha/2}$;

K and $n - k + 1$ are calculated by formula (4).

The $x_1, x_2, \ldots, x_n$ The KTH value $W(K)$ of (2) is calculated from smallest to largest, and the KTH value $W(n - k + 1)$ is calculated from largest to smallest.

From equation (3), we can get $\widehat{\theta}_L, \widehat{\theta}_u$ cars only

Software implementation: use R software to calculate interval estimation results. As the sample size of some voltages is small, R cannot calculate the grade of 95% or 99% confidence interval and the scheme does not give the result of interval estimation.

3. Finishing process

3.1. outlier removal

Draw the boxplot of 220kV tower (tower body) and calculate that the normal value range is 35,500 yuan per base to 303,000 yuan per base, and there are 4 abnormal values.
The unit cost of 220kV bar erection (tower body) project is estimated by interval, and the calculated result of software is 90,300 yuan per base - 134,400 yuan per base (99%).

### 3.3. abnormal data analysis

Abnormal data 1-4. The unit cost of the 4 sample data is relatively high, which is due to the large weight of the tower per unit and other high costs. The unit cost reaches 30.67-407,000 yuan per base, and the deviation from the mean is 135,000 yuan per base. In the actual project, there is the possibility that these two data are abnormal, and other costs are too high, which are related to demolition. However, in the process of data sorting, common situations need to be considered, and some samples with high deviation from the normal value will be eliminated in this sorting.

### 4. Sample collation results

This time, 73 items of unit cost data of 220kV tower project were changed from 2014 to 2017, 6 items of abnormal data (4 items of 220kV tower (tower body) samples) were eliminated, and 69 items of sample data were retained. After finishing, the sample range of unit cost of 220kV tower (tower body) project is: 35,500 yuan per base - 279,100 yuan per base.

| The serial number | Voltage grade | type | unit | Data year interval | Sample size | The average | Sample data interval |
|-------------------|---------------|------|------|--------------------|-------------|--------------|---------------------|
| 1                 | 220 kV        | Angle tower/steel tube | Ten thousand yuan per base | 2014-2017 | 73           | 12.26        | [3.55, 27.91]       |
| 2                 | 110 kV        | Angle tower/steel pole | Ten thousand yuan per base | 2014-2017 | 32           | 8.28         | [2.21, 15.4]        |
| 3                 | 35 kV         | Angle tower/steel pole | Ten thousand yuan per base | 2014-2017 | 24           | 4.48         | [0.11, 9.74]        |

Statistical results of sample description of 220kV after sorting:
Table 4. Description statistics of 220kV samples (after sorting)

| Replace the tower (tower body) 220kV | statistic | Standard error of |
|-------------------------------------|-----------|-------------------|
| The median                          | 11.1000   |                   |
| The variance                        | 27.844    |                   |
| The standard deviation              | 5.27670   |                   |
| minimum                             | 3.55      |                   |
| The maximum                         | 27.91     |                   |
| The scope of                        | 24.36     |                   |
| The interquartile range             | 7.99      |                   |
| Partial degrees                     | 994.      | 281.              |
| kurtosis                            | 646.      | 555.              |

The overall interval estimation results of 220kV are shown as follows:

![Scatter diagram of unit cost interval estimation of 220kV bar erection (tower body) project](image)

The probability that the true value of the sample after outliers are removed falls between 90,300 yuan/basis and 134,400 yuan/basis is 99%.

References

[1] GU Jifa, "The System Methodology of Principles of Matter-human-affair", Transportation System Engineering and Information, no. 3, pp. 25-28, 1995.
[2] XU Changfu, "As is ought to and to do: a metaphysical approach to the questions of explaining regulating and transforming the world", New trends of Philosophy, no. 10, pp. 6-9, 2001.
[3] Sajad Sabzi, Hossein Javadikia, Juan Ignacio Arribas. A three-variety automatic and non-intrusive computer vision system for the estimation of orange fruit pH value[J]. Measurement, 2019.
[4] Luzzi Sabino, Maestro Mattia Del, Elia Angela, Vincitore Francesca, Perna Giuseppe Di, Zenga Francesco, Garbossa Diego, Elbaba Samer K, Galzio Renato. Morphometric and Radiomorphometric Study of the Correlation Between the Foramen Magnum Region and the Anterior and Posterolateral Approaches to Ventral Intradural Lesions.[J]. Turkish neurosurgery, 2019, 29(6).
[5] Sajad Sabzi, Hossein Javadikia, Juan Ignacio Arribas. A three-variety automatic and non-intrusive computer vision system for the estimation of orange fruit pH value[J]. Measurement, 2019.
[6] Alison L. Armstrong, Harvinder P. Singh. Clinical and cost comparison of manipulation under anaesthetic and steroid injection for frozen shoulder (stage II) with arthroscopic interval release [J]. Journal of Arthroscopy and Joint Surgery, 2019.

[7] Catherine Qiu, Carina Lott, Patrick Cahill, Jason Anari. Is the Juice Worth the Squeeze: Additional Spine Length with Shorter Distraction Interval, but at What Cost? [J]. Spine Deformity, 2019, 7(6).

[8] González-Mohíno Fernando, Santos-Concejero Jordan, Yustres Inmaculada, González-Ravé José M. The Effects of Interval and Continuous Training on the Oxygen Cost of Running in Recreational Runners: A Systematic Review and Meta-analysis. [J]. Sports medicine (Auckland, N.Z.), 2019.

[9] Anonymous. Suitable replacement for balloon light towers [J]. Quality Progress, 2018, 51(3).

[10] Anonymous. PHARMACEUTICAL COOLING TOWER REPLACEMENT DESIGN-BUILD PROJECT [J]. Engineered Systems, 2017, 34(11).