Discussion on the Application of Mathematical Statistics in Construction Quality Control

Shuxi Yan

School of Mathematics and Statistics, Yulin University, Yulin, 719000, China

Abstract: generally, sampling is used to control the quality of engineering construction. If the data obtained by sampling is able to reflect the actual condition of the construction should be verified through mathematical statistics. The idea of mathematical statistics focuses on random sampling, i.e. randomly select a small portion of sample from the whole studying objects for experimental verification and obtaining corresponding data, and then analyze these data to conclude the whole objects. Sampling inspection is also used to control the quality of the specific construction practice. This paper summarized the collection methods of mathematical statistics, introduced the application methods of mathematical statistics in construction quality control, as well as discussed the application of mathematical statistics in actual construction quality control by instances.

Key words: mathematical statistics, construction, quality control.

Construction quality is the key link of building engineering construction. For the construction enterprises, construction quality control is the priority of the productive process. The quality has a direction relation to the economic benefits and reputation of the enterprises. At present, the overall construction controlling quality in our country is poorer than that in the western developed countries. Therefore, we must introduce mathematical statistics and apply it to the construction quality control, scientifically verify the data obtained from the construction process, so as to find out the major causes influencing the construction quality and the internal rules for taking corresponding measures.

I.   Collection methods of mathematical statistics

Firstly, implement one hundred percent inspection. It refers to observe, inspect, count and register all individuals of the whole one by one, by which a conclusive method for evaluating the overall quality can be obtained. Secondly, implement random sampling inspection. It refers to select several individuals as samples from the whole on the basis of the random sampling principle, and then infer the method for evaluating the overall quality in accordance with the inspection results of the samples. Samples for inspection are not subject to the inspectors' subjective wishes and each individual's rate of being selected is exactly the same. In this way, the samples in the whole are well distributed, which are not only of fair representativeness but also can be used in the destructive inspection and construction quality control. Hereby, the tasks that cannot be completed by the one hundred percent inspection can be finished accordingly; and its application space can be further extended.

II.   The application methods of mathematical statistics in construction quality control

Random sampling is generally used in the process of actual construction quality control. Now, take the quality control of the asphalt mixture used in construction as the studying point, the

grading analysis on the finished materials and the analysis on the asphalt contents mainly include many random samples, while the number of the samples is directly related to the reliability of the test results. The following problems many occur when processing the measured data:

Firstly, select n_i samples from the whole for testing. Work out the accuracy range of the n_i samples' average value $\overline{x}$ for reflecting the average value of the whole, or work out the necessary number of the tests required by the accuracy range after completed n_i samples. It is a problem involving the number of the samples and the average value in the process of unknown variability. For statistics, this problem can be solved by the following methods. After completed n tests, the possible changing interval within the assurance rate required by the average value $\overline{x}$ can be obtained on the basis of the principle of random variable subjecting to t distribution. Once the average value of the accuracy is given and n_i samples are selected, the number of extra tests can be obtained; thus, the given accuracy can be met, i.e. the total number of the tests can be calculated accordingly to the formula: $n_i+n_2= n_1F (R/D)$. In this formula, n_1 refers to the initial number of tests, while n_2 refers to the number of the increased tests. R refers to the difference between the maximum and minimum initial trial values, i.e. the range; D refers to the allowable deviation of the average value; F refers to the correlation coefficient between the initial number of tests and the confidence probability.

Secondly, once n tests have been done on the individual content of the whole finished material set, such as asphalt content and the average value and standard deviation have been obtained, the possible deviation of this content in the whole batch of finished materials have to be worked out. Hereby, it is the maximum deviation that may be caused by a certain random variable obtained through the limited samples. The limited samples may cause deviation between overall teat expectation value μ and variance α^2, and between the estimated value without deviation $\overline{x}$ and squared value with standard deviation s^2; therefore, it is improper to replace μ and α^2 with $\overline{x}$ when n-the number of tests is too low. According to the relevant theory of statistics, the maximum deviation obtained from the following formula shall be more proper: x = $\overline{x}$ ±k s, among which k refers to the correlation coefficient between the number of tests and the confidence probability; k can be selected from the table1.

Table 1: the relationship between the coefficient k and n

The number of tests: n	2	3	4	5	6	7	8	9	10	11
Coefficient: k	86.622	22.869	11.266	7.914	6.381	5.511	4.952	4.563	4.275	4.503
The number of tests: n	12	13	14	15	16	17	18	19	20	21
Coefficient: k	3.877	3.733	3.614	3.513	3.425	3.351	3.283	3.224	3.172	3.124
The number of tests: n	22	23	24	25	26	27	28	29	30	31
Coefficient: k	3.082	3.043	3.007	2.975	2.944	2.917	2.891	2.867	2.844	2.823

Thirdly, it can be found from the test results that a few data is significantly large or small; whether the above situation should be attributed to operation error, mixing equipment suffering from instant loss of control, misread during the tests and other abnormal factors? It relates to abnormal data selection. Statistical principles can be used to decide whether to eliminate the abnormal data under suspicion. The most common method is applying Grubbs for calculation and making $T = | x_d - \overline{x} | / s$. In this formula, T refers to the determination value as comparing with $G(\alpha, n)$, while $G(\alpha, n)$ is the Grubbs coefficient, x_d is the data under suspicion, $\overline{x}$ is the average test data containing x_d, s is the standard deviation of test data including x_d. When $T > (\alpha, n)$, it is bound to be the abnormal data and should be excluded, otherwise be kept.

III. The application of mathematical statistics in actual construction quality control

A certain suspension bridge project possesses main span of 610m and tower height of 142.6m; the compressive strength of its major tower tip can be analyzed through frequency distribution histogram. Sum up and draw up the scattered distribution diagram of quality deviation composing of several column diagrams according to the data implementation frequency, so as to work out the quality characteristic values and further determine the construction capacity.

i. data collection

Firstly, collected the report of the C40 concrete compression strength of the bridge project, and found the maximum and minimum value of all data, and the maximum and minimum are 53.9MPa and 48.8 MPa respectively, with range of $R = X_{maz} - X_{min} = 53.9 - 48.8 = 5.1MPa$。

ii. Determine the class interval and the number of classes

The class interval is subject to the required accuracy of the measured data; and the number of classes should be determined by the total number of collected data. If collected 50~100 data, then 10~20 classes are acceptable, 10 classes are generally appropriate. If collected 20~50 data, then 5~10 classes are acceptable. The class interval is expressed by h, and the number of classes is expressed by k.

iii. Determine the interval value of individual class

After classified all data, the interval value of individual class should be kept in succession, while any interruption shall result in difficulties in data statistics. How to determine the value of the lower bound of the first interval is critical in that the values of the upper and lower bounds of all other intervals vary with the first one. To effectively prevent the data from being on the interval, the following formula can be used to calculate the first interval value: value of the lower bound of the first interval$= X_{min} - h/2 = 48.8 - 1.02/2 = 48.29MPa$; the value of the upper bound of the first interval$= X_{min} + h/2 = 48.8 + 1.02/2 = 49.31MPa$。

iv. Statistical frequency

Table 2: statistical table for frequency distribution

No.	Class Interval	Frequency	Relative Frequency
1	48.29-49.31	3	0.1
2	49.31-50.33	3	0.1
3	50.33-51.35	4	0.133
4	51.35-52.37	15	0.5
5	52.37-53.39	3	0.1
6	53.39-54.41	2	0.067
Total		30	1.000

The distribution of the whole data can be mastered from the above statistical table for frequency distribution. For further understanding of the quality control status, the histogram of the frequency distribution should be drawn (refer to figure 1). It is a coordinate graph, in which the abscissa represents the classes while the ordinate represents the frequency of individual interval value. The histogram is drawn to observe the stability of the quality and the distribution situation. Applying histogram to construction quality control makes it possible to estimate the likely percent defective, judge the distribution situation of the quality and the construction capacity of the enterprise.
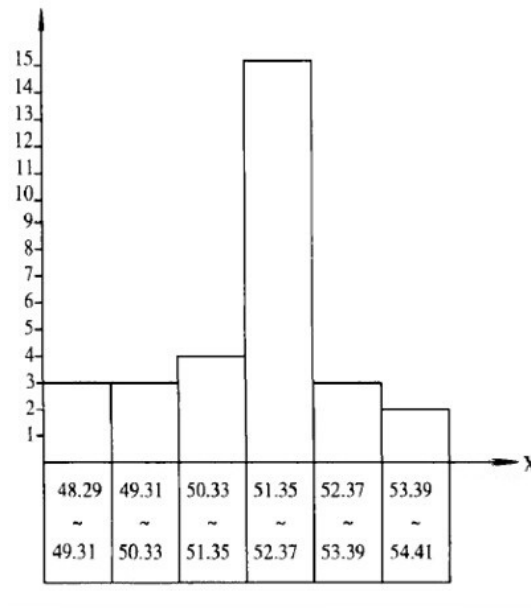

Figure 1: histogram of concrete compression strength frequency distribution

Firstly, evaluate the likely percent defective. The quality evaluation standard generally includes two major boundary values-upper bound value and lower bound value, i.e. T_u and T_L respectively. Therefore, the percent defective is divided into super upper bound percent defective P_u and super lower bound percent defective P_L, and total percent defective is $P=P_u+P_L$. In order to calculate P_u+P_L, introduce coefficient: $K_u=|T_u-x|/S$, $K_L=|T_L-\overline{x}|/S$. refer to the normal probability coefficient distribution table according to K_u and K_L, i.e. determine proper P_u and P_L. there is only the standard lower bound value in this case: $T_1= 40 \times 0.85=34MPa$, $K_L=|34-51.5|/2.4 =7.28$. If it exceed the bound, the percent defective $PL = 0$.

Secondly, observe the process capability. The process capability refers to the capability of producing acceptable products under the stable processing condition, which is generally expressed by the capability index of process, i.e. Cp. To be specific, it refers to the ratio between the quality standard range and the production accuracy of this process. The calculation methods can be divided into: in case of the quality standard center coinciding with the quality distribution center, $Cp=T/6S= Tu-TL)/6S$; in case of the quality standard center differing from the quality distribution center, $C_{pK}= (T_u-T_L)/6S (1-K)$. When the quality center of the construction only possesses lower bound or upper bound, $Cp=（X-T）/3S$. In this case, the quality center only possesses lower bound, thus $C_p=(X-T_L)/3S= (51.5-34)/(3\times2.4)=2.43>1.33$. Its process capability is able to fully meet the requirements. Due to the excessive capability, the process is not very economical, thus the mix proportion should be adjusted appropriately.

Thirdly, analyze the distribution situation of the construction quality. If the production condition is in normal state, the histogram will present a trend of higher in the middle and lower in the two sides, and the left and right will approach mutual symmetry. Such a trend is normal. Histogram with any abnormal trend should be deeply analyzed and corrected by taking corresponding measures immediately. If the trend is in the shape of folded teeth and the histogram is full of unevenness, which is mainly caused by unreasonable classification and improper class interval. If the trend is in the shape of lonely island, and the histogram presents as a small isolated histogram, which is mainly caused by a few of disqualified materials or unskilled operation in short term. If there is a double peak in the histogram, i.e. two peaks in the graph, which is generally caused by mutually mixed data from two classes with different production condition. If the trend is in the shape of slight slope and the histogram presents a slight slope to the right or left, i.e. the average value significantly incline towards left or right, which is mainly caused by excessive controlling on the upper or lower bound during the construction process. If the trend is in the shape of a bold cliff, and the distribution center of the histogram incline towards one side, which is generally attributed to the subjective factor of the operator or caused by abnormal data collection, or man-made interference during the process inspection, which hereby should be solved by re-statistics or re-inspection as regulated. In terms of this case, it can be found from the histogram that the quality distribution is in good condition.

vi. Conclusion analysis

It can be found from the data analysis on this case that the compressive strength and process capability of the main tower is too excessive. Although the quality distribution is in good condition, it is not economic; thus it suggests adjusting and optimizing the concrete composition reasonably based on guaranteed project quality.

V. Summary

To sum up, applying mathematical statistics to the construction quality control system is significant for determining the number of the samples, ensuring the reliability of the test data, selecting individual data, etc. At present, China mainly chooses to refer to overseas data in the construction quality control. Therefore, how to reasonably apply mathematical statistics to determine the sample value and the number of the samples, and further improve the construction quality is subject to the study of the professional statistics personnel in their future work.

Bibliography:

[1] Yan Shijian. *Probability Theory and Mathematical Statistics* [M]. Shanghai: East China Normal University Press, 2003.

[2] He Ping. *Mathematical Statistics and Multivariate Statistics* [M]. Chengdu: Southwest Jiaotong University Press, 2004.

[3] Tang Xun. *The Application Methods of the Mathematical Statistics in Construction Quality Inspection* [J]. Journal of Xiaogan University, 2004 (6).

[4] Wang Lihua, Chen Lida. *Analysis and Evaluation on the Application of Mathematical Statistics in the Emergent Beijiang Dike Reinforcement Project Quality Test* [J]. Guangdong Water Resources and Hydropower, 2006 (6).