International Journal of Science Academic Research
Vol. 06, Issue 01, pp.9007-9013, January, 2025
Available online at http://www.scienceijsar.com

ISSN: 2582-6425

Research Article

COMPARISON OF STATISTICAL METHODS FOR OUTLIER DETECTION IN PROFICIENCY TESTING DATA ANALYSIS OF FINENESS IN ORDINARY PORTLAND CEMENT

***Shwet Vashishtha**

R.P.P.T. Laboratory, National Test House, Ghaziabad, India

Received 09th November 2024; **Accepted** 11th December 2024; **Published online** 15th January 2025

Abstract

The study evaluates the performance of participating laboratories in a Proficiency Testing Program for Ordinary Portland Cement (OPC) fineness analysis using various statistical methods. The study found that nIQR, which considers both within- and between-laboratory variance, is most suitable for outlier detection in the dataset. The analysis showed that different statistical approaches can significantly affect the evaluation of laboratory performance, emphasizing the importance of selecting an appropriate method for proficiency assessment in OPC fineness testing. The study compared Cochran's, Grubb's, Hampel's, Dixon's, Mandel's, nIQR, robust algorithm-A, and classical z-score methods for outlier detection and performance assessment. The z-score indicates laboratory performance relative to others, with values between ±2 and ±3 considered questionable and those outside ±3 considered unsatisfactory. The study highlights the importance of an effective evaluation system and the significance of selecting the appropriate statistical method for proficiency assessment.

Keywords: Proficiency Testing, z-score, Statistical Techniques, Outliers, OPC.

INTRODUCTION

Proficiency testing programs are essential for evaluating the analytical performance of laboratories, ensuring the validity of experimental assays, and identifying analytical issues. These programs adhere to international standards such as ISO 5725, ISO/TR 22971, ISO/IEC 17043, and ISO 13528, which provide frameworks for measurement control, statistical analysis, and proficiency testing scheme requirements [1-9]. In the context of Ordinary Portland Cement (OPC) fineness analysis, selecting appropriate statistical methods for outlier detection is crucial for accurately evaluating laboratory performance. A study compared various statistical methods, including Cochran's, Grubb's, Hampel's, Dixon's, Mandel's, nIQR, robust algorithm-A, and classical z-score methods for outlier detection and performance assessment in OPC fineness testing [1]. The nIQR method, which considers both within- and between-laboratory variance, was identified as the most suitable for outlier detection in the dataset, highlighting the importance of selecting the correct statistical approach for proficiency assessment [1]. The z-score was used to indicate laboratory performance relative to others, with values beyond ±3 indicating unsatisfactory performance [1]. Different statistical methods, such as Grubb's, Dixon's, Hampel's, Cochran's, Mandel's, nIQR, robust algorithm-A, and classical z-score, were utilized for outlier detection and performance evaluation [1]. The nIQR method, considering both within- and between-laboratory variance, proved effective in detecting outliers in the dataset, underscoring its suitability for proficiency assessment in OPC fineness testing [1]. The z-score values obtained through various statistical methods provided insights into laboratory performance, with values outside the range of ±3 indicating problematic measurements [1]. The study emphasized the importance of an effective evaluation system and the significance of selecting appropriate statistical methods for proficiency assessment in OPC fineness testing [1].

Corresponding Author:** ***Shwet Vashishtha,
R.P.P.T. Laboratory, National Test House, Ghaziabad, India.

By comparing different statistical approaches, the research aimed to enhance the statistical methodology used in proficiency testing programs, ensuring the accurate evaluation of laboratory performance and outlier detection [1]. The results obtained through robust statistical methods like nIQR and robust algorithm-A highlighted the importance of outlier detection and data validation in ensuring the reliability and accuracy of laboratory testing [1]. The study underscored the critical role of statistical methods in outlier detection and performance assessment in proficiency testing programs for OPC fineness analysis. By employing robust statistical approaches like nIQR and robust algorithm-A, laboratories can effectively identify outliers, enhance data quality, and improve the reliability of their testing procedures. The findings emphasize the importance of meticulous data analysis, robust statistical methods, and quality assurance measures in ensuring accurate proficiency assessment and outlier detection in laboratory testing scenarios [1]. Proficiency testing programs are designed to evaluate the analytical performance of participating laboratories, making it possible to accomplish a critical evaluation of the validity of routinely carried out experimental assays, to identify analytical problems, and to facilitate the implementation of necessary corrective actions. A Proficiency Testing Program provider (PT scheme) is responsible for conducting the statistical analysis and supplying an indicator of the performance of all participants. The most important international standard for measurement control in laboratories is ISO 5725 with the group name "Accuracy (trueness and precision) of measurement methods and results" (6 parts) [10–15].The international standard ISO/TR 22971 [16] provides users with practical guidance for the use of ISO 5725-2 [11] and presents simplified step-by-step procedures for the design, implementation, and statistical analysis of interlaboratory studies to assess the variability of a standard measurement method and for the determination of repeatability and reproducibility of data obtained in interlaboratory testing. The standard ISO/IEC 17043 [17] specifies general requirements for the competence of providers of proficiency testing schemes and for the development and

operation of proficiency testing schemes. These requirements are intended to be general for all types of proficiency-testing schemes, and can be used as a basis for specific technical requirements for particular fields of application. The international standard ISO 13528 [18] contains detailed descriptions of statistical methods that are used for data analysis, which were obtained by means of charts of qualification verification, and provides recommendations concerning the use of such charts and authorized persons. It is added to the standard ISO/IEC 17043 concerning the verification of abilities by means of inter-laboratory comparisons.

The primary objective of organizing this proficiency testing is to assess the laboratory's technical competence to perform measurements and fulfil the requirements of ILAC/APLAC with regard to the compatibility of results submitted by these laboratories. Participation in the Proficiency Testing Programme/Interlaboratory Comparison is mandatory for NABL-accredited cement-testing laboratories.The laboratory carried out physical test parameters: Blaine's air permeability method [19] for Ordinary Portland Cement. Initially, z-scores were evaluated for the results of the participating laboratories based on the median and NIQR. The NIQR was determined to be 0.714 × IQR (IQR is the difference between the 3rd and 1st quartiles).The z-scores demonstrated the laboratory's ability to perform the above-mentioned analyses competently. Proficiency testing programs are statistical quality assurance programs that enable laboratories to assess their performance in conducting tests within their laboratories when their data are compared to those of other laboratories that participate in the same program. Proficiency testing aims to independently assess the competence of participating laboratories. Together with the use of validated methods, proficiency testing is an essential element of laboratory quality assurance. The PT programs conducted by the National Test House are independent schemes arranged by an independent section in the PTP Division.PT is designed for laboratories to ensure the performance of individual laboratories for specific tests or measurements, and to monitor the continuing performance of laboratories. It provides laboratories with an objective means of assessing the reliability and confidence of the data they produce. It also complies with the requirements of ISO/IEC 17025:2017:7.7 – Ensuring the validity of results – clause 7.7.2 (a) "participation in proficiency testing' [20].

The assessment of different statistical methods is essential for evaluating the performance of each laboratory in the PT schemes. Certainly, PT scheme providers and participants need to know whether there are any significant differences in the evaluation results when different performance statistical methods are applied. Based on these considerations, this study compared the suitability of different statistical approaches for determining the assigned value and standard deviation for proficiency assessment. This study also aimed to improve the statistical approach currently used in this proficiency-testing program. The results were evaluated using different statistical approaches. Cochran's test, Grubb's test, Hampel's test, Dixon's test, Mandel's Method, nIQR method, robust algorithm-A statistical method and classical z-score to establish the best method to present the performance of the participant along with the outlier detection. When the amount of data is high and there is a significant difference between the laboratory results, the detection of outliers is a difficult task. Thus, in the present study, emphasis was placed on establishing a suitable method to determine outliers in such a way that the extreme values provided by some PT participants would not influence the results of the participants close to the reference value.

The Z-score is the performance score of a laboratory compared with that of other laboratories. The quality of the measurements increased with a decrease in z-score. In this study, laboratories with z-score values between ± 2 and ± 3 were considered questionable and advised to closely examine their results to rectify the fault. The cutoff value of the z-score evaluated by any method was outside the range of ± 3, which indicates that there is a problem with the measurement.

Statistical Techniques

The statistical evaluation of the data obtained from the participants is illustrated in this paper using both numerical techniques such as Cochran's, Grubb's, Hampel's, and Dixon's statistics, and graphical techniques such as classical z-score, Median & NIQR, Robust Algorithm-A, and statistical methods.

Grubb's Method

Grubbs were used to detect single outliers in the univariate dataset. The dataset follows an approximately normal distribution. Grubbs' test was based on the following two hypotheses:

H0: There is no outlier in the data set
H1: There is at least single outlier in the data set

The general formula for Grubbs' test statistic is defined as

$$G = \max \left| yi - y \right| / s$$

Where yi is the element of the data set, y and sdenoting the sample mean and standard deviation and the test statistic is the largest absolute deviation from the sample mean in units of the sample standard deviation. The calculated value of G was compared with the critical value obtained from the Grubb's test. When the calculated value is higher or lower than the critical value, the calculated value can be accepted as an outlier when choosing statistical significance. The statistical significance (α) describes the maximum mistake level which a person searching for outlier can accept.

Dixon's Method

The test developed by Dixon and used in this study was appropriate for small sample sizes. The test has some limitations for n≤30, and was later extended to n≤40.The test is the first step for organizing the data in ascending order, and the next step is to count the parameter R.

The test has various statistics. Suppose that to test a large set of elements as outliers, the sample is arranged in ascending order X1≤X2≤.≤Xn, implying that a large sample element is given by Xn. Dixon proposed the following test statistics defined as

$R_{10}=x_n-x_{n-1}/x_n-x_1$, $3\leq n\leq 7$
$R_{11}=x_n-x_{n-1}/x_n-x_2$, $8\leq n\leq 10$
$R_{21}=x_n-x_{n-2}/x_n-x_2$, $11\leq n\leq 13$
$R_{22}=x_n-x_{n-2}/\ x_n-x_3$, $14\leq n\leq 30$

For testing the smallest sample element to be an outlier, the sample is ordered in descending order implying that the smallest sample element is labeledXn. The selection of the test statistics depends on Dixon's criteria.

The variable $X_n$is marked as an outlier, when the corresponding statistic $_{(n)}$ exceeds a critical value, which depends on the selected significance level α.

The calculated value of the parameter R was compared using Dixon's test critical value to determine statistical significance. When the calculated value of parameter R is larger than the critical value, it is possible to accept the data from the dataset as an outlier.

Hampel's Method

Statistical tables are not necessary to calculate Hampel's test. Theoretically, this method is resistant, which means that it is insensitive to outliers. It also has no restrictions on the abundance of the dataset. Hampel's test was performed using the following steps for the datasets:

- Compute the median (Me) of the entire dataset. The median is described as a numeric value, separating the higher half of the dataset from the lower half.
- Compute the value of the deviation ri from the median value; this calculation should be done for all elements from the data set:

$$r_i=(x_i–Me)$$

where, x− simple data from the data set, i− belongs to the set for 1 to n, n− number of all element of the set and Me− median

- iii. Calculate the median for deviation $Me_{|ri|}$
- iv. Check the conditions: $|r_i|\geq 4.5Me_{|ri|}$

If this condition is executed, the value from the dataset can be accepted as an outlier.

Cochran's Method

According to ISO 5725-2, Cochran's test is recommended for detecting outliers in a given set of interlaboratory variability tests. It is a one–sided outlier test that only examines the greatest standard deviation and eliminates problematic results with within-laboratory reproducibility and repeatability. Cochran's statistic C is calculated using the following formula

$$C = SD^2_{max}/\Sigma^{p}_{j=1}SD_j^2$$

where SD_{max} is the maximum standard deviation among the investigated laboratories

SD_j is the standard deviation of data from the laboratory
p is the number of participating laboratories

The calculated C value can be compared with the critical value for a given n value, that is, the number of results given by each laboratory.

Mandel's k/h Method

Mandel's k and h consistency test statistics are discussed in the ASTM E691 [21] standards for inter-laboratory analysis. 'k' value is a measure of within-laboratory consistency in repeatability and the 'h' test statistic is used to examine the consistency of interlaboratory data, confirming if any laboratory data is an outlier. In other words, it indicates the accuracy of the laboratory results compared to those reported by others. 'h' test statistic value reflects the deviation of a single laboratory's mean test results from the overall mean results obtained from all participating laboratories. 'h' and h_{crit} are a measure of seriousness in a lab's inaccuracy and define as;

$$h_j = d_j / S_x$$

Where, d_j is deviation of mean result of a lab (j) from the overall mean and S_x is the standard deviation

$$h_{crit} = t\,(p – 1)/\sqrt{\{p\,(t^2 + (p\text{-}2)\}}$$

Where, t is the student's distribution with degree of freedom v=p-2 and $\alpha = 0.05$
p is the number of participating laboratories

When the 'h' value is larger than h_{crit}, it is concluded that the mean result given by the laboratory is not accurate or reliable.

Normalized interquartile range (*nIQR)* Method

The performance evaluation process of the participating laboratories is illustrated by the z-score values of the results obtained in the PT program in accordance with the ISO 13528 guidelines. The values of standardized sum (S_i) and standardized difference (D_i) between two results of laboratory "i" have been calculated using equations.

$$S_i = (A_i + B_i) / \sqrt{2}$$
$$D_i = (A_i - B_i) / \sqrt{2} \text{ if median } (A_i) > \text{median } (B_i)$$

Whereas $A_i$and $B_i$are the two measurement values of the laboratory 'i'

Using the values of $S_i$and D_i, the values of both Z-scores, that is, The Z-score between laboratories (Zbi) and within laboratories (Zwi) can be calculated using equations.

Zbi = Si – Median (Si)/IQR (Si) x 0.7413 (Variation between the laboratory: Reproducibility)
Zwi = Di – Median (Di)/IQR (Di) x 0.7413 (Variation within the laboratory: Repeatability)

The interquartile range (IQR) was defined as the difference between the lower and upper quartiles of data. The lower quartile (Q1) is the value below, which is a quarter of the result, and the upper quartile (Q3) is the value above, which is a quarter of the result. Quartiles were calculated analogous to the median and interquartile range (IQR) = Q3–Q1.The term "Normalized IQR" is comparable to the standard deviation and is equal to IQR × 0.7413.Factor 0.7413 is derived from the standard normal distribution, which has a mean of zero and a standard deviation equal to one. The widths of the interquartile ranges of these distributions were 1.34898 and 1/1.34898 = 0.7413, respectively. z-score values Zbi and Zwi of laboratory "i" are the robust z-scores of the *ith S* and *ith D* values, respectively. Finally, each participating laboratory was assigned two z-scores based on their results. Generally, laboratories with z-score values outside the range of ± 3 are considered outliers.

Robust Algorithm-A Method

Robust statistics can be calculated according to ISO 13528: robust analysis algorithm The robust average (x*) and standard deviation value (s*) of the participants' results could be achieved by an iterative calculation, as described in ISO 13528, and is not affected by the results far from the reference value.Performance is evaluated by calculating the z-score or z′-score (z prime) in the given expression as the uncertainty of the assigned value $u(x_{pt}) < 0.3\sigma_{pt}$.The$Z'$ score was calculated as follows:

$$Z' = (x_i - x_{pt})/ \sqrt{\sigma^2_{pt} + u^2(x_{pt})}$$

where, x_i is the test result from the participant laboratory, x_{pt} is the assigned value and σ_{pt} is the standard deviation for proficiency assessment (SDPA).

$$u(x_{pt}) = 1.25s^*/\sqrt{p}$$

where, $u(x_{pt})$ is the uncertainty of assigned values, s* is robust standard deviation and p is number of participants

Classical Z-Score Method

The z-score is the score given to participants according to their performance. The classical z-score can be calculated as follows

$$z = (X_{lab} - X_{mean})/SD$$

where X_{lab} is the result of an individual laboratory, X_{mean} is the mean value of the analyte obtained from the participants, and SD is the standard deviation of the data.

SAMPLE DETAILS OF PT

The ordinary Portland cement (53 grade) provided to the laboratories is the ISI-marked cement sample, which confirms the requirement of IS 269:2015.It was ensured to make further homogeneous PT items by thoroughly mixing, coning, and quartering on the line of IS 3535-1986 under dry conditions and to prepare representative samples. In this study, two test parameters were considered and the results of the participants were evaluated using different statistical methods. The PT items were sent to 13 laboratories. All participating laboratories provided results in duplicate. The details of the results are given in Table 1. The participant laboratories used IS 4031 (Part-2):1988 for the determination of Fineness by Blaine's Air Permeability Method.

Table-1. Details of laboratory results

S.NO.	Lab Code	Fineness (m^2/kg)	
		Result-A	Result-B
1	A	331	329
2	B	333	333
3	C	333	334
4	D	331	329
5	E	322	321
6	F	322.5	318.1
7	G	329	332
8	H	322	313
9	I	304	306
10	J	327.5	332.1
11	K	296	319
12	L	325	323
13	M	320	327

RESULT AND DISCUSSION

In proficiency testing programs, the outcomes acquired by participating laboratories are typically conveyed in the form of a Z-score, which serves as a representation of their performance. The Z-score of each laboratory signifies the extent to which the reported outcome deviates from the designated reference value, thereby classifying the results as satisfactory, questionable, or unsatisfactory. This assessment of laboratory performance involves the conversion of data obtained from participating laboratories into Z-scores through the use of statistical techniques. Based on the conventional Z-score approach, outlier results affect the mean and standard deviation of the dataset. Under the classical framework, a substantial majority of laboratories indicate that their Z-scores fall within an acceptable range, specifically within the confines of a Z-score ± 2. The Z-scores, as determined through the implementation of the classical, robust, and nIQR methods, are presented in Table 3.On the other hand, ISO 13528 employs robust algorithmic analysis to yield resilient estimates.

First, the PT results were sorted in ascending order and the absolute deviation from the median was computed. Subsequently, the process converges through iterations, specifically by updating the robust average (x*) and robust standard deviation (s*) multiple times to ensure consistency up to the third decimal place between consecutive iterations. The robust average is regarded as the designated value, whereas the uncertainty value is obtained by substituting the robust standard deviation into the equation. Z-score was determined using robust estimates. The laboratories identified as outliers using this approach exhibit similarities to the laboratories identified in the context of the Grubb, Dixson, Mandel, and classical methods for density test parameters. Based on the analysis of the data, it is evident that this method is the most optimal for deriving the designated value when the reference value for PT is unknown.

The measurement quality of the data received from the participating laboratories was quantitatively evaluated using the nIQR method. To achieve this, PT data were processed to determine two types of Z-scores: the Z-score within the laboratory (Zwi) and the Z-score between the laboratories (Zbi).The intra-laboratory z-score value indicates the variability in data generated within the same laboratory, whereas the inter-laboratory z-score indicates the variability in data generated by different participating laboratories. The quality of the measurements improved as the Z-score decreased. The values of both the intra-laboratory Z-score and inter-laboratory Z-score were calculated for the parameters and plotted against the laboratory code (Figure 1a and 1b).By referring to the Z-score results chart, each laboratory can easily compare its performance with that of other laboratories. The analysis of cement fineness across multiple laboratories revealed intriguing insights into the variability of test results. Fineness values, measured in m2/kg, ranged from 296 to 333, indicating a diverse spectrum of cement quality among the tested samples. Two sets of results were obtained for each sample, labelled as Result-A and Result-B, providing additional layers of data for comparison. Upon closer examination, it became evident that certain laboratories consistently yielded higher or lower fineness values than the others.

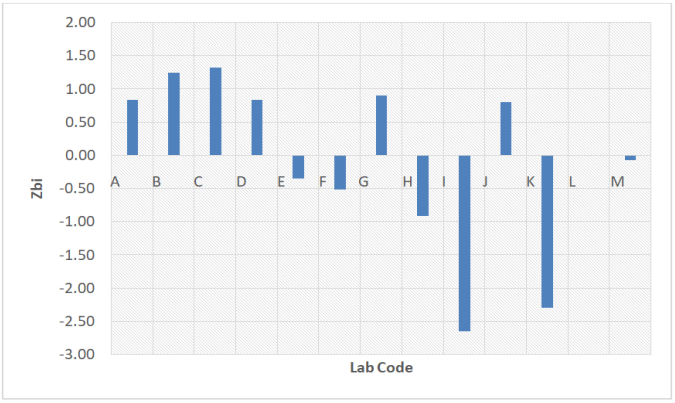

Fig. 1(a). Z-score within the laboratory (Zwi) obtained by nIQR Method

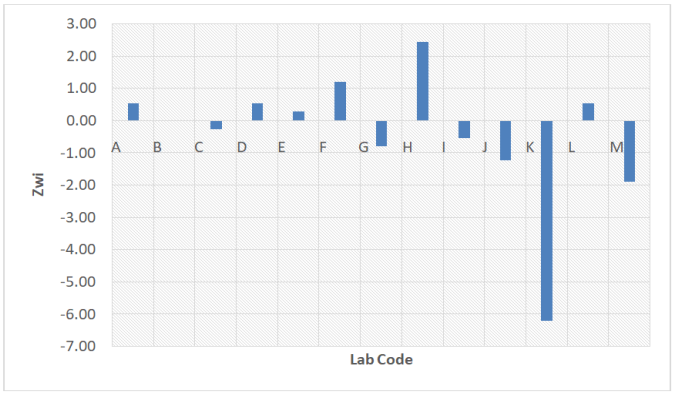

Fig. 1(b). Z-score between the laboratory (Zbi) obtained by nIQR Method

For instance, laboratories I, K, and M consistently exhibited values at the extremes of the spectrum, suggesting potential discrepancies in the testing methodologies or equipment calibration. These outliers underscore the importance of implementing robust quality control measures to ensure the reliability and accuracy of cement fineness tests. Furthermore, statistical analysis revealed the presence of outliers using various methods such as Grubb's, Dixson's, Mandel's, nIQR, Robust Algorithm-A, and Classical Method. Laboratories with consistently divergent results were flagged as outliers across multiple statistical techniques, indicating the need for further investigation of the root causes of discrepancies. The variability observed in cement fineness values highlights the challenges associated with standardizing testing procedures across different laboratories. Factors such as equipment calibration, sample preparation techniques, and operator proficiency can significantly influence the test results, leading to inconsistencies in data interpretation.

Addressing these challenges requires collaborative efforts among stakeholders to establish standardized protocols and best practices for cement testing. Furthermore, the identification of outliers using statistical methods underscores the importance of data validation and quality assurance in research and testing. Laboratories with consistently divergent results should undergo rigorous evaluations to identify potential sources of error and implement corrective measures. The reliability and accuracy of cement testing can be enhanced by fostering a culture of continuous improvement and knowledge sharing, ultimately benefiting industries that rely on high-quality cement products. Moreover, the implications of outliers extend beyond the realm of cement testing to broader applications in the research and industry. Outliers can signal the underlying issues in data collection, analysis, or interpretation, highlighting the need for robust statistical techniques and quality control measures. By proactively and systematically addressing outliers, researchers and practitioners can enhance the integrity and reproducibility of their findings, ultimately advancing scientific knowledge and innovation. This comprehensive examination of the results and discussion emphasizes the importance of meticulous data analysis, robust statistical methods, and quality assurance measures in research and testing. By critically evaluating research findings and their implications, researchers and practitioners can enhance the reliability and validity of their work, thereby contributing to meaningful advancements in their respective fields.

Table 2. Outliers detected by various statistical approaches

Statistical Methods	Lab Code
Grubb's Method	I
Dixson's Method	I
Hampel's Method	Nil
Cochran's Method	K
Mandel's Method	I
nIQR Method	
➢ Zbi	I and K
➢ Zwi	H and K
Robust Algorithm-A Method	I
Classical Method	I

The table-2 provides information on the outliers detected by various statistical methods for different laboratory codes. Outliers were identified using Grubb's Method, Dixson's Method, Cochran's Method, Mandel's Method, nIQR Method (including Zbi and Zwi), Robust Algorithm-A Method, and Classical Method. Notably, Hampel's method did not detect any outliers in specified laboratory codes. Table-3 displays data from various laboratory samples along with the corresponding z-scores calculated using the classical, robust, and nIQR (Zwi and Zbi) methods.

Table 3. Z-score obtained by Classical, Robust and nIQRMethods

Lab Code	Value (m^2/kg)	z-score classical	z-score Robust	z-score by nIQR (Zwi)	z-score by nIQR (Zbi)
A	330.0	0.703	0.618	0.54	0.83
B	333.0	1.029	0.945	0.00	1.25
C	333.5	1.084	0.999	-0.27	1.32
D	330.0	0.703	0.618	0.54	0.83
E	321.5	-0.223	-0.307	0.27	-0.35
F	320.3	-0.353	-0.438	1.19	-0.51
G	330.5	0.757	0.673	-0.81	0.90
H	317.5	-0.658	-0.743	2.43	-0.90
I	305.0	-2.019	-2.104	-0.54	-2.64
J	329.8	0.681	0.596	-1.24	0.81
K	307.5	-1.747	-1.832	-6.21	-2.29
L	324.0	0.049	-0.035	0.54	0.00
M	323.5	-0.005	-0.090	-1.89	-0.07

The z-scores indicate the deviation of each sample's value from the mean, in terms of standard deviations. Positive z-scores indicate values above the mean, whereas negative z-scores indicate values below the mean. In this dataset, Lab Code B, C, G, and J exhibited positive z-scores across all methods, indicating values above the mean, whereas Lab Code E, F, H, I, K, and L showed negative z-scores, suggesting values below the mean. Lab-I stands out, with the most negative z-scores across all methods, indicating a significant deviation from the mean. Compared to the classical, robust, and nIQR methods, we observed slight variations in the calculated z-scores for each sample. While the classical method provides a standard approach to calculating z-scores, the robust method offers resistance to outliers, resulting in slightly different values, especially in samples with extreme values, such as I and K. The nIQR method, specifically Zbi, incorporates the interquartile range to account for variability, resulting in z-scores that might differ slightly from those of the classical method. Overall, the z-scores provide insights into the distribution of values within the dataset and the extent of each sample's deviation from the mean, highlighting potential outliers and variability in the measurement methods. Further analysis may involve investigating the reasons behind outliers and assessing the reliability of different z-score calculation methods. The results and discussion of cement fineness analysis provide valuable insights into the complexities of testing methodologies and data interpretation. By identifying outliers and addressing potential sources of error, researchers and practitioners can improve the reliability and accuracy of cement testing, thereby contributing to advancements in construction materials and infrastructure development. Collaborative efforts and standardized protocols will be essential to ensure consistency and reproducibility in cement testing practices, ultimately benefiting industries and communities worldwide.

Conclusion

The performance of the laboratories was expressed by z-score and the laboratories having $|z| \leq 2$ are classified as satisfactory, $2 < |z| < 3$ are classified as questionable and $|z| \geq 3$ are considered as unsatisfactory. Among all the methods, the highest number of outliers was obtained by nIQR statistical analysis. As the nIQR method considers variance both within and between laboratory results, it seems to be the most suitable method for outlier detection in the dataset evaluated in this study. The presence of outliers in the data analysis poses challenges and opportunities for deeper insights. Researchers can effectively detect and treat outliers by employing various statistical methods and data visualization techniques. Common approaches, such as analyzing interquartile ranges (IQRs), provide valuable insights into the distribution of data and help identify potential anomalies. Understanding the causes of outliers, including measurement and sampling errors, is crucial for interpreting their impacts on statistical analyses. Through careful consideration and appropriate treatment of outliers, researchers can enhance the robustness and accuracy of their analyses, thereby leading to more reliable conclusions. Moreover, addressing outliers facilitates the identification of underlying patterns, trends, and relationships within the data, ultimately contributing to advancements in the research and decision-making processes. Therefore, a comprehensive understanding of outlier detection and treatment methodologies is essential to ensure the validity and reliability of statistical analyses in various fields.

Acknowledgment

The PTP Division of NTH (WR) expresses gratitude to the Director General, National Test House, for his valuable support and continued encouragement. The PTP Division also expresses thanks to the Director, National Test House (WR), Mumbai, for his valuable guidance and the facilities provided for the successful completion of the work throughout the project.The PTP Division sincerely expresses thanks to the Civil Engineering Laboratory for rendering the service for this program.We express our gratitude to all participating laboratories for joining the scheme.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

OPC: Ordinary Portland Cement
IQR: Interquartile Range
nIQR: Normalized Interquartile Range
PT: Proficiency Testing
PTP: Proficiency Testing Program
SD: Standard Deviation
SDPA: Standard Deviation for Proficiency Assessment
Z' Score: Z-prime Score

REFERENCES

1. Hodge, V. and Austin, J. (2004). A survey of outlier detection methodologies. Artificial Intelligence Review, 22(2), 85-126. https://doi.org/10.1023/b:aire.0000045502.10941.a9
2. Inazumi, S., Shinsaka, T., Nontananandh, S. & Chaiprakaikeow, S. (2021). Applicability of additives for ground improvement utilizing fine powder of waste glass. Materials, 14(18), 5169. https://doi.org/10.3390/ma14185169
3. Jackson, A., Kanak, M., Grishman, E., Chaussabel, D., Levy, M., & Naziruddin, B. (2012). Gene expression changes in human islets exposed to type 1 diabetic serum. Islets, 4(4), 312-319. https://doi.org/10.4161/isl.21510
4. Lee, E., Parkin, N., Jennings, C., Brumme, C., Enns, E., Casadellà, M. & Ji, H. (2020). Performance comparison of next generation sequencing analysis pipelines for hiv-1 drug resistance testing. *Scientific Reports,* 10(1). https://doi.org/10.1038/s41598-020-58544-z
5. Mohamed, M. and Tran, D. (2022). Exploring the relationships between project complexity and quality management approaches in highway construction projects. *Transportation Research Record Journal of the Transportation Research Board,* 036119812211313. https://doi.org/10.1177/03611981221131308
6. Ullmann, S., Nordsiek, S., &Lowke, D. (2022). Significance or scatter—statistical evaluation of rapid chloride migration test results for sprayable cement mortar. Materials, 15(17), 6050. https://doi.org/10.3390/ma15176050
7. Verma, S. (1997). Sixteen statistical tests for outlier detection and rejection in evaluation of international geochemical reference materials: example of microgabbro pm-s. *Geostandards and Geoanalytical Research,* 21(1), 59-75. https://doi.org/10.1111/j.1751-908x.1997.tb00532.x
8. Verma, S., Orduña-Galván, L., & Guevara, M. (1998). Sipvade: a new computer programme with seventeen

statistical tests for outlier detection in evaluation of international geochemical reference materials and its application to whin sill dolerite ws-e from england and soil-5 from peru. Geostandards and Geoanalytical Research, 22(2), 209-234. https://doi.org/10.1111/j.1751-908x.1998.tb00695.x

9. Zhu, W., Huang, L., Mao, L., & Esmaeili-Falak, M. (2021). Predicting the uniaxial compressive strength of oil palm shell lightweight aggregate concrete using artificial intelligence-based algorithms. Structural Concrete, 23(6), 3631-3650. https://doi.org/10.1002/suco.202100656
10. ISO 5725-1:1994. Accuracy (trueness and precision) of measurement methods and results. Part 1: General principles and definitions.
11. ISO 5725-2:1994. Accuracy (trueness and precision) of measurement methods and results. Part 2: Basic methods for repeatability and reproducibility of standard measurement methods.
12. ISO 5725-3:1994. Accuracy (trueness and precision) of measurement methods and results. Part 3: Intermediate measures of precision of the standard measurement method.
13. ISO 5725-4:1994. Accuracy (trueness and precision) of measurement methods and results. Part 4: Basic methods for determining the trueness of a standard measurement method.
14. ISO 5725-5:1994. Accuracy (trueness and precision) of measurement methods and results. Part 5: Alternative methods for determining the precision of a standard measurement method.
15. ISO 5725-6:1994. Accuracy (trueness and precision) of measurement methods and results.Part 6: Use accuracy values in practice.
16. ISO/TR 22971:2005. Accuracy (trueness and precision) of measurement methods and results: Practical guidance for the use of ISO 5725-2:1994 in designing, implementing, and statistically analyzing inter-laboratory repeatability and reproducibility results.
17. ISO/IEC 17043:2010. Conformity assessment. General requirements for proficiency testing.
18. ISO 13528:2015 Statistical methods for use in proficiency testing by inter-laboratory comparison
19. IS 4031 (Part 2):1988, Methods of Physical Tests for Hydraulic Cement- Determination of Fineness by Blain Air Permeability Method.
20. ISO/IEC 17025:2017, General requirements for the competence of testing and calibration laboratories
21. ASTM E691, Standard practice for conducting an inter-laboratory study to determine the precision of a test method

****