Journal of Building Engineering 95 (2024) 110276

Contents lists available at ScienceDirect

Journal of Building Engineering

journal homepage: www.elsevier.com/locate/jobe

Comparison of outlier detection approaches for compressive strength of cement-based mortars

Nara Cangussu[a,*], Paula Milheiro-Oliveira[b,c], Ana Mafalda Matos[a], Farhad Aslani[d,e], Lino Maia[a,f]

[a] CONSTRUCT-LABEST, Faculty of Engineering, University of Porto, 4200-465, Porto, Portugal
[b] Faculty of Engineering, University of Porto, 4200-465, Porto, Portugal
[c] CMUP, Faculty of Sciences, University of Porto, 4200-465, Porto, Portugal
[d] Materials and Structures Innovation Group, School of Engineering, The University of Western Australia, WA, Australia
[e] School of Engineering, Edith Cowan University, Joondalup, WA 6027, Australia
[f] Faculty of Exact Sciences and Engineering, Campus da Penteada, University of Madeira, 9020-105, Funchal, Portugal

ARTICLE INFO

Keywords:
Compressive strength
Cement
Mortar
Outlier
Test statistics

ABSTRACT

EN 196-1 states that the compressive strength result of testing of standard mortar specimens corresponds to the average of the six individual measurements. If one of the six individual measurements varies more than ±10 % from the average, it must be excluded, and a new average is calculated with the five remaining values. This methodology might be interesting for quality control or characterization of cement; however, it may not be the best approach for research or scientific purposes. Applying EN 196-1 to non-standard cement-based mortars with multiple and even unconventional supplementary cementitious materials may induce even more variation in the result. Also, mixtures with higher mechanical strength generally present the highest absolute variation. The main goal of the present work is to compare different methodologies, such as Tails of Normal Distribution, Thompson-Tau, Box-plot, T-test, Dixon Criteria, and other methods provided by engineering standards, to evaluate the existence of discrepant values and the effect of their exclusion (or not) on compressive strength experimental values, namely for science purposes. A sample of compressive strength of 338 mortar specimens was used The main findings revealed that excluding data points as a result of applying 8 of the 13 methods under study did not affect the mechanical strength result by more than ±5 %. The method described in the Brazilian standard that guides preparation, control, receipt, and acceptance for Portland cement concrete, would invalidate almost 46 % of all mixtures produced and did not prove to be reliable. Simplified methods of excluding outliers in relative or absolute percentages of 6 % around the average value can be considered an interesting option for professionals involved in laboratory testing of building materials. Otherwise, the authors would recommend applying the either Dixon Criteria or T test of the ASTM E178 for scientific purposes.

1. Introduction

According to ASTM E178, an "outlier" is one that appears to deviate markedly from other members of the sample in which it occurs

* Corresponding author.
E-mail address: naracan@gmail.com (N. Cangussu).

https://doi.org/10.1016/j.jobe.2024.110276
Received 9 December 2023; Received in revised form 9 June 2024; Accepted 23 July 2024
Available online 23 July 2024

[1]. Other authors define outlier 'as a data point which is significantly different from the remaining data' [2] or as 'an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism' [3].

The ASTM E178 advises adopting a cautious attitude to identify outliers and comes with the purpose of orientation relative to the overall problem of experimentation. The American Standard is precise in the position on the matter of screening samples for outlying observations when the "physical reason known or discovered for the outlier(s)". It recommends actions in this order: 1. reject observation(s); 2. correct observation(s) on physical grounds; 3. reject it (them) and possibly take additional observation(s). However, the Standard recalls that the statistical test may always be used to support a judgment that a physical reason does actually exist for an outlier, or the statistical criterion may be used routinely as a basis to initiate action in order to find a physical cause. In the case of "physical reason unknown" the use of statistical tests is recommended [1].

It is known that any experimental data set has an associated error, which may be systematic, random, or, in most cases, both. Correct use of experimental data requires a careful analysis of the associated errors, which indicates the reproducibility, representativeness, and reliability of the data obtained. There are several procedures for the treatment and analysis of experimental data sets, based on statistics or on the physics of the process, but these procedures are difficult to apply [4].

Even more complex analyses, such as Data Mining in Knowledge Discovery in Databases (KDD), include data clearings, such as handling missing values and removal of noise or outliers. Several methods are explained in the handbook [5] which go from doing nothing to becoming the major part (in terms of time consumed) of a KDD process in certain projects. The analysis may involve complex statistical methods, or using specific Data Mining algorithms in this context.

Outlier detection methods have been suggested for numerous applications, for instance in credit card fraud detection, clinical trials, voting irregularity analysis, data cleansing, network intrusion, severe weather prediction, geographic information systems, athlete performance analysis, and other data-mining tasks [6]. In the context of civil engineering research and practice, there are many options for dealing with outliers. With the wide range of options, the researchers adopt the most varied methods for detecting and disregarding outliers, which, consequently, influence the results of data analysis [7–30]. In all of the criteria for testing outliers, the doubtful observation is included in the calculation of the numerical value of the statistic obtained on the observed sample, which is then compared with a critical value based on the theory of random sampling to determine whether the doubtful observation is to be retained or tagged as an outlier [31].

Some standards rely on rather simplified criteria for detecting outliers, in order to facilitate faster results. Among them, the European Standard EN 206 [32] and the Brazilian Standard NBR 12655 [33], both for cement-based mixtures can be mentioned. The EN 206 guides that when two or more specimens are manufactured from a sample and the range of variation of the test results is greater than 15 % of the mean, these results shall be disregarded, unless investigation shows that there is an acceptable reason justifying the elimination of an individual test value. The NBR 12655 advises that the greatest of the two values obtained in the compressive strength test should be taken as the strength of the sample. Although both standards adopt simplified procedures, the action to be undertaken when one or more records are statistical outliers is a matter of engineering judgment, and any established method for assessing statistical outliers must be permitted [34].

Besides the aforementioned European and Brazilian standards, the methods used in the present study were: the Box-plot representation; the Grubbs test, T-test, and Dixon Criteria from ASTM E178 [1]. The statistical analyses proposed in this work aim to enhance data analysis for civil engineering researchers. By assessing the difference between the lowest or highest test result and the distribution of all the test results, it is possible to determine if the lowest or the highest result is a statistical outlier, according to each method. In fact, many standards do not care if a measured value is an outlier or not, they just define an approach to exclude measured values.

Portland cement clinker is the most widely used binder for concrete production, and it is expected to be dominant for the foreseeable future. Last forecast of the International Energy Agency Cement Roadmap projects cement demand growth by 12%–23 % by 2050 [35] and the Organisation for Economic Co-operation and Development admits that its demand might double by 2060 [36]. Efforts have been made worldwide to reduce the environmental footprint of Portland cement, namely by using supplementary cementitious materials as partial Portland cement replacements as well as the optimisation Portland cement dosage of in mortars and concrete product-based [37]. Concrete nowadays is a multifarious material and its properties are significantly affected by the individual properties of its constituents. Concrete ingredients have recently changed a lot by introducing various materials and admixtures that either added before or during mixing. As such, the mix design of concrete materials has become more complex due to the number of constituent raw materials and from different sources. Thus, the mixture design usually starts at paste or mortar levels, which are easier and less timer and resources are needed. One of the most widely used tests to check cement based material's performance is mechanical strength according to EN 196-1 [38]. This standard describes a methodology to determine the compressive strength of standard mortars. The standard mortar presents a mix ratio of 1:3:0.5 in mass, cement, aggregate and water, respectively. The sand is standardized by CEN. EN 196-1 establishes that the compressive strength result corresponds to the average of the six individual measurements. However, if one of the six individual results varies more than ± 10 % from the average, it must be excluded and a new average is calculated with the five remaining values. Even though this methodology is very easy to apply, it is not scientific and does not come with evidence of statistical support. Thus, the methodology might be interesting for quality control or characterization of cement mortars, however, it may not be the best approach for research or scientific purposes. The application of EN 196-1 to non-standard cement based mortars with multiple and even unconventional supplementary cementitious materials may induce more variation in the result. Also, mixtures with higher mechanical strength generally present a larger absolute variation.

2. Research significance and objectives

Being aware that different standards define distinct methodologies to detect and exclude values, here a general analysis is made for the specific case of determining the compressive strength of nonstandard cement-based mortars. Even though, Standards for civil engineering practice establish procedures to disregard measurements when testing specimens for concrete strength, there was a gap in the literature regarding the investigation of the different approaches that these standards describe as well as the comparison of these procedures with other established statistical methods for outlier detection. The present study intents to investigate the set of methods found. Thus, the main goal of the present work is to compare methods like Tails of Normal Distribution, Thompson-Tau, Box-plot, T-test, Dixon Criteria, and other methodologies contained in some engineering standards (EN 13791, EN 196-1, NBR 7215, NBR 13279, EN 206, NBR 12655, NBR 7584) to calculate the representative value based on possible result exclusions and recommend the best method for research purposes. As such, a compressive strength data set of 338 mortar specimens, performed in Laboratory of Building Materials in Faculty of Engineering of the University of Porto, was used (a total of 2028 specimens). For each set of compressive strength tests, 13 outlier detection approaches were applied with the outcome value of the compressive strength of the set being determined accordingly.

The current work starts with a brief explanation of the data obtained in Section 3. Different outlier excluding methodologies are presented in Section 4 and the application of such methods in the data set is shown in Section 5. Finally, Section 6 presents the main conclusions.

3. Compressive strength data

3.1. Mechanical strength test

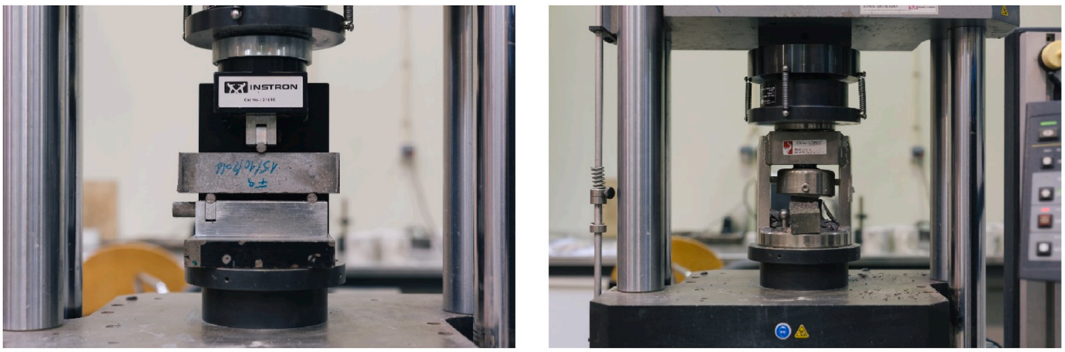
The compressive strength test of mortars in prismatic specimens 4 × 4 × 16 [cm^3] is one of the most widely performed tests in cement-based materials. In Europe, specifically in Portugal, the test method is described in EN 196-1 [38]. A few adaptations may occur from one country to another but, summarily, three prismatic specimens with dimensions of 4 × 4 × 16 [cm^3] are moulded with mortar. Afterward, specimens are cured at 20 °C under HR > 95 % until testing age.

Usually, the test is performed at 28 days, but cement types "R", according EN 196-1, require also 2 days compressive strength assessment. However, it can be performed at any age, depending on the objective of the work. As such, at a pre-defined testing age, the three specimens are firstly tested to the flexure strength (3-point bending, see Fig. 1-a). Six halves cross section of 4 × 4 cm^2 and with a length of approximately 8 cm result from this procedure. The six halves are, subsequently, tested for compressive strength with the load being applied in an area of 4 × 4 cm^2, as can be seen in Fig. 1-b. Thus, from this test method, the obtained raw data are the six compressive strength values. Typically, the users calculate the average value and display it as being the output from the test. Sometimes the standard deviation or the coefficient of variation is reported too. In fact, there are several standards worldwide that rule the compressive strength tests of mortars in prismatic specimens 4 × 4 × 16 [cm^3], even though, usually they do not explicitly define an approach to detect outliers. Typically, standards define an analysis to discharge values, but that is not an outlier analysis. When inexplicable values are observed, usually it is based on the user experience that the value is accepted or rejected, without any statistical support.

3.2. Tests results sample

A sample of 338 sets of six halves specimens, totalizing 2028 raw values for compressive strength was used and presented in Annex A. The individual measurements that resulting from testing (collected data) are used in the current study. The tests were carried out in the Laboratory of Building Materials of the University of Porto. For the outlier analysis, each mixture was analysed separately, since different sets correspond to different mixtures. The specimens were tested for research purposes, as such, evaluating the effect of using different cement types, supplementary cementitious materials (fly ash, metakaolin and silica fume) and admixtures.

Annex A presents the raw values obtained in the compressive strength tests which were used for the calculations, analyses and detection of outliers. Those results are also illustrated in Fig. 2. Each marker in Fig. 2 represents the average value of six compressive



a) b)

Fig. 1. Mechanical strength test of cement-based mortars according EN 196-1 [38]: a) Flexure; b) Compressive.

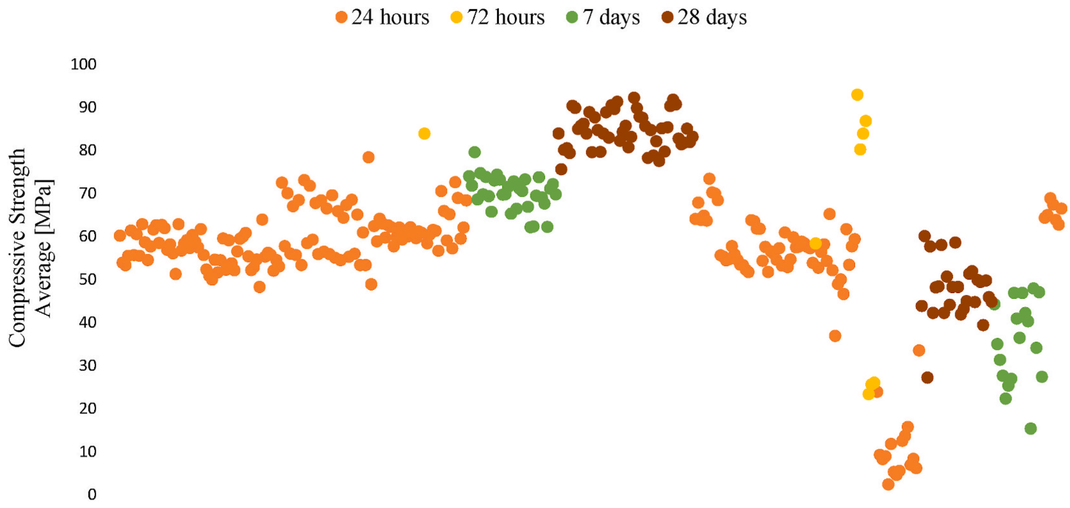


Fig. 2. Compressive strength data set test results for different ages.

strength results (according to test method explained in section 3.1). As can be perceived, the compressive strength test was performed at different ages, namely 24h, 72h, 7 and 28 days. Table 1 presents a statistic summary of the data set under study.

4. Approaches to obtain representative values

In many data analysis tasks, a large number of variables are being recorded or sampled. One of the first steps towards obtaining a coherent analysis is the detection of outlying observations [6], that is data that differ drastically from all other records (see Fig. 3). In other words, an outlier is a value that drastically deviates from remaining data and that can (and probably will) cause anomalies in the results obtained through algorithms and systems analysis. For all the methodologies studied in this work, the compressive strength representative value was the average of the remaining data.

One way to find outliers in data analysis is to find the statistical distribution that most closely matches the data and use statistical methods to detect outliers.

The ISO/TR 22971 conceptualizes outlier as an observation which is sufficiently different from all others to warrant further investigation. The standard states that, in tests used to determine the presence or absence of outliers, it is assumed that the results are distributed in a Gaussian manner (normal distribution) or at least follow a single unimodal distribution [40].

The methods described below were applied to the data set introduced in section 3.2.

4.1. Tails of the normal distribution and the QQ plot

Errors in scientific measurements are generally well approximated by a normal distribution. Karl Friedrich Gauss derived his equation from a study of errors in repeated measurements of the same quantity [41].

According to the normal distribution, it is expected that data that are around twice the standard deviation make up 95 % of all data; the outliers represent, in this analysis, 5 % of all, falling in the tails of the distribution [41]. In practice, however, in large majority of problems, the parameters μ (expected value) and σ (standard deviation) of the underlying Gaussian distribution are not known. The QQ-plot may help one find the outliers, but one should not rely on a test that simply uses the tails of the Gaussian distribution as if parameters were known.

4.2. Thompson-Tau method

The Thompson-tau technique is a statistical method of deciding whether to keep or discard suspected outliers in a sample of a single variable. The procedure considers the mean ($\overline{x}$) and standard deviation (s) of the sample of size n. For each data point, the absolute value of the deviation with respect to the mean is calculated, $\delta_i = |x_i - \overline{x}|$. The data point most suspected as a possible outlier is the data point with the maximum value of δ_i. Thompson τ is calculated from the T-Student distribution, as shown in Equation Eq. (1) [44].

Table 1
Statistic summary of compressive strength test data set.

Testing age	Min (MPa)	Max (MPa)	Overall Average (MPa)	Mean of Std Dev (MPa)	Mean of Coef. Variation (%)
24 h	1.80	80.34	55.43	1.34	2.49
72 h	23.10	97.15	62.39	1.85	2.52
7 days	14.40	81.56	57.70	1.58	2.59
28 days	26.40	99.90	71.80	2.14	2.87

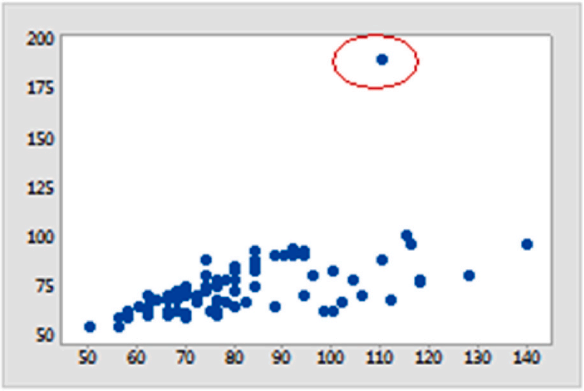


Fig. 3. Outlier Detection graphically.
Source [39]:

$$\tau = \frac{t_{\frac{\alpha}{2}} \cdot (n-1)}{\sqrt{n}\sqrt{n-2+t_{\frac{\alpha}{2}}^2}} \tag{1}$$

where $t_{\alpha/2}$ is the critical Student's T-value which corresponds to having an area under the curve to the left of point $t_{\alpha/2}$ equal to $\alpha_{/2}$. Very often one takes $\alpha = 0.05$. If $\delta_i > \tau \cdot s$ then the tested data should be considered an outlier; otherwise, it is not. With the Thompson-tau method, one can only detect a maximum deviation δ_i which gives data x_i as a suspected outlier. If δ_i is detected as abnormal, then x_i will be singled out, and then re-start a new round of search. This process continues until there are no abnormal data found [42].

4.3. Turkey's method (box-plot)

Boxplots are a popular and easy tool for identifying outliers. Values above (Q_3 + 1.5 x IQR) or below (Q_1 - 1.5 x IQR) can be considered outliers, where Q_1 and Q_3 are the first and third quartiles, respectively. IQR is the interquartile range (IQR = Q_3 - Q_1). Fig. 4 illustrates the box-plot representation.

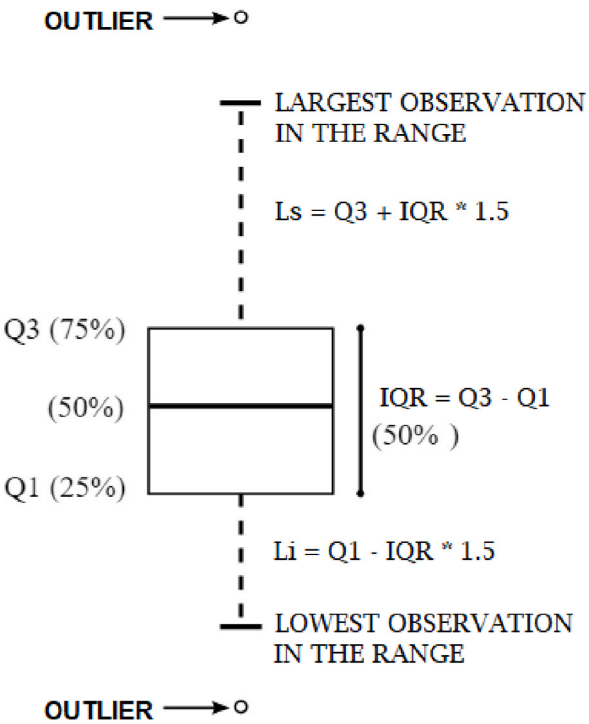


Fig. 4. Boxplot method to identify outliers [39].

4.4. ASTM E178 - Standard practice for dealing with outlying observations - Tstat

This method is presented in item 6.2 of ASTM E178 Standard [1]. The hypothesis is that all observations in the sample come from the same normal population. The work considered possible outliers occurring on either side, simultaneously. It uses the statistic shown in Equation Eq. (2).

$$T_n = \frac{(x_n - \overline{x})}{S} \text{ and } T_1 = \frac{(\overline{x}) - x_1}{S} \tag{2}$$

where $\overline{x}$ is the average and s is the estimate of the standard deviation, σ, based on the sample data, and x_1 and x_n are the smallest and the largest value in the data, respectively. The values obtained for the statistics T_n and T_1 are compared with the corresponding critical values, for a given significance level. In the present work, the sample size is 6 for each set/mixture. A level of significance of 5 % is set, which gives a critical value as red on Table 1 of ASTM E178 [1]. The largest value x_n is marked as an outlier if T_n overtakes the critical value and a similar reasoning is applied for x_1.

4.5. ASTM E178 - Standard practice for dealing with outlying observations - Dixon Criteria

The Dixon criteria is based entirely on ratios of differences between the observations and may be used in cases where avoiding calculation of the estimate of the standard deviation s is desirable or where quick judgment is required [1]. Table 2 from ASTM E178 gives the appropriate statistics and also the critical values for the 1, 5, and 10 % levels of significance [1]. The Standard reminds one that the T-statistic discussed above is the best to be used for the single-outlier case, and the final statistical judgment should be based on it.

In this work, the sample size is 6 for each set/mixture and a 5 % level of significance is considered. The outlier analysis was carried out according to Equation Eq. (3).

$$\begin{aligned} r_{10} &= \frac{x_2 - x_1}{x_n - x_1} \text{ \textit{if smallest value is suspected;}} \\ r_{10} &= \frac{x_n - x_{n-1}}{x_n - x_1} \text{ \textit{if largest value is suspected.}} \end{aligned} \tag{3}$$

If $r_{10} > r^*$ with r* given by Table 2 from ASTM E178 [1], then the suspected value should be considered as an outlier.

4.6. EN 13791 - assessment of in-situ compressive strength in structures and precast concrete components - Grubbs test

Grubbs' test [43] is used to detect a single outlier in a univariate data set that follows an approximately normal distribution. The Grubbs test is also known as the maximum normalized residual test. The standard EN 13791 alerts to the fact that in the case of possible outliers at both extremes, the described technique should be first applied to the value that deviates most from the mean. Each outlier shall be documented and evaluated individually. If more than two records are outliers, this may be an indication that the test sample comprises at least two concretes and this possibility should be examined [44].

The highest test value of n consecutive test values should be considered an outlier when the inequality (Equation Eq. (4)) holds true. The same occurs with the lowest test value (Equation Eq. (5)).

Table 2
Summary of the results of applying the different methods considered in this work.

Method/Standard	Type of specimen	Brief description of detection criteria	Number of mixtures with detected outliers	Number of outliers detected
Tails of the Normal Distribution	All	Tails of the Gaussian Curve (2 Sd)	2	2
Thompson-Tau	All	Tau Statistic	111	111
Turkey's procedure	All	Box-plot analysis	5	5
ASTM E178	All	*T Stat*	43	43
ASTM E178	All	Dixon Criteria	51	51
EN 13791	All	Grubbs test	9	9
EN 196-1 - Cement strength	Prismatic 4 x 4 × 16 cm	Max variation around the mean (10 %)	10	12
NBR 7215 - Cement strength	Cylindrical Ø5 x 10 cm	Mas relative deviation to the mean (6 %)	39	46
NBR 13279 - Mortar strength	Prismatic 4 x 4 × 16 cm	Max. variation around the mean (0.5 MPa)	317	1358
EN 206 - Concrete strength	Cubic 150 mm or cylindrical Ø15 cm × 30 cm	Max. variation around the mean (15 %)	10	12
NBR 12655 - Concrete strength	Cylindrical Ø10 × 20 cm (more usual)	Max. of the two specimens	338	1014
NBR 7584 - Hardened concrete - reflection hammer	In situ	Max. deviation around the mean (10 %)	10	12
EN 13791 - Prefabricated concrete strength	Cylindric or cubic cores, with length: diameter ratio of 2:1 or 1:1 and a diameter ≥75 mm	No outlier detection criteria	0	0

$$\frac{(fc, is, highest - fc, m(n)is)}{s} > \mathrm{Gp} \tag{4}$$

$$\frac{fc, m(n)is - fc, is, lowest}{s} > \mathrm{Gp} \tag{5}$$

where *fc,m(n),is* is the mean of the in-situ compressive strength; the sample standard deviation is represented by *s;* and G_p is the critical value of the test, given in EN 13791 [44], which corresponds to a significance level of 1 %.

4.7. EN 196-1 - methods of testing cement - Determination of strength

The EN 196-1 [38] establishes that the result of the compressive strength test will be the arithmetic mean of the six individual results, executed as presented in Section 2 of this work. If one of the six individual results deviates more than ±10 % from the mean, this result must be deleted and a new mean must be calculated with the five remaining values. The final result is the new average strength. If a result among the remaining five varies by more than ±10 % from the new mean, the entire series of results is discarded and the testing is repeated.

4.8. NBR 7215 - Portland cement - Determination of compressive strength

Brazilian standard NBR 7215 [45] establishes the calculation of the maximum relative deviation of the series of four results, dividing the absolute value of the difference between the average strength and the individual strength that is most distant from this average, by the average strength and multiplying the quotient by 100 %. When the maximum relative deviation is greater than 6 %, a new mean must be calculated, disregarding the outlier. The final result is the new average strength.

4.9. NBR 13279 - mortar for laying and coating walls and ceilings

The Brazilian standard, NBR 13279 [46], deals, specifically, with testing the tensile and compressive strength of mortars in prismatic specimens. The standard advises that when the maximum absolute deviation to the mean is greater than 0.5 MPa, a new average must be calculated, disregarding outliers. The test is considered valid only when the result consists of the average of at least four specimens.

4.10. NP EN 206 - concrete - Specification, performance, production, and compliance

EN 206 [32] directs that, when two or more specimens are manufactured from a sample and the range of variation of the individual test results is greater than 15 % of the average, the entire set must be discarded, unless an investigation shows that there is an acceptable reason that justifies the dismiss of an individual test value.

4.11. NBR 12655 - Portland cement concrete - preparation, control, receipt and acceptance - procedure

NBR 12655 (ABNT, 2015a), is a Brazilian Standard that specifies the method for determining the compressive strength of cylindrical concrete specimens, tested as performed in NBR 5739 [47]. According to NBR 12655, the samples must be collected randomly during the concreting operation and each sub-sample must consist of two cylindric specimens of the same mixture, according to ABNT NBR 5738 [48], for each failure age, moulded in the same act. The greater of the two values obtained in the compressive strength test is taken as representative of the strength of the sub-sample.

For the present work, each of the 3 prismatic specimen were considered to be a sample, and the two halves of the a prismatic specimen form a sub-sample. Therefore, the detected discrepant values were, from the engineering point of view, the smallest value in the sub-sample formed by each prismatic specimen, although such records would not be immediately classified as “outliers” from a statistical point of view.

4.12. EN 13791 - evaluation of the strength of concrete in structures and prefabricated products

EN 13791 [44]allows the assessment of the compressive strength class of concrete and permits to follow up the construction of a structure where there is doubt about the compressive strength based on results of standard tests or there is doubt about the quality of execution. The standard aims at estimating in situ characteristic compressive strength of a test region and/or in situ strength at specific locations (areas).

EN 13791 [44] prescribes that the number of cores to be extracted from a test area must be determined by the volume of concrete involved and the purpose of testing the cores. Each test area comprises a core at least. For the evaluation of in situ compressive strength, for statistical and safety reasons, as many cores as practicable should be used. An in situ compressive strength assessment for a particular test zone should be based on at least 3 cores. These standards do not consider a specific method for detecting outliers. All data are considered in the computation of the average value which stands as the definition of precast concrete strength.

4.13. NBR 7584 - Hardened concrete - evaluation of surface hardness by reflection hammer

To estimate the surface hardness of concrete using a reflection hammer, NBR 7584 [49] recommends calculating the mean of the (9–16) individual values of the hammer index corresponding to a single test area. Any individual value that is more than 10 % away from the mean value obtained is considered an outlier and discarded determining that a new mean must be calculated. The final average index must be obtained with at least 5 individual records and, when this is not possible, the test in that area must be abandoned. None of the remaining individual measurements should differ by more than 10 % from the final mean.

5. Identification of outliers

When carrying out the outlier detection tests, it should be stressed that an outlier must not be discarded or rejected purely from a statistical point of view. For each specimen, the reason why the suspicious result is different from all the others should be investigated and identified [40].

In addition to conventional statistical procedures (Tails and QQplot, Thompson-Tau method, Turkey's Box-plot method, T*stat* test, Dixon Criteria, Grubbs test), the other procedures studied here are specific for different materials, which have in common that they are cement-based composite materials. In these cases, the detection of discrepant values recommended in the standards consists on simplified cautions.

Comparing the two types of analyses (statistical and simplified), it can be seen that by applying statistical methods, no abnormal measurements were found in 227 mixtures (67.2 %), while simplified methods detected discrepant values in some of those. It is important to notice that, among the methods for obtaining final compressive strength results, the methodology described in EN 13791 [44] does not consider exclusions, only considers the average of the values found. Furthermore, the guidelines described in NBR 12655 always reject three of a total of six values found in the tests.

The procedures reviewed in the previous section were applied to the database described in section 2. Annex B presents the number of outliers found in each mixture as well as the variation of compressive strength (in percentage), when compared to the difference between the initial and final compressive strength values.

Table 2 shows differences among the methods investigated, such as type of specimen that the method/standard refers and a brief description of each detection criteria. The table also informs the number of outliers detected in all database, by each method and the number of mixtures that had some influence of detecting outliers.

With the results in hand, it is clear that the Thompson-Tau, NBR 13279, and NBR 12655 methods detected a larger number of outliers. On the other hand, the tails of the Normal distribution (using twice the standard deviation to define the tails), Turkey's method, Grubbs test, EN 206, NBR 7584, and EN 13791 were not very strict in detecting outliers. Three methods proved to be moderate in terms of the number of outliers: T Stat and Grubbs Test, from ASTM E178, and NBR 7215. In this study, 85 % of the outliers detected by both methods were coincident, that is, only 14 of the 94 outliers detected were different in the two methods. Despite the moderate number of outliers detected by the NBR 7215 method (maximum variation of 6 %), 59 % of the detected outliers do not coincide with the outliers found in at least one of the ASTM E178 methods (T Stat or Grubbs).

Table 3 presents the number of mixtures for which the lowest/highest value has been identified in this work as an abnormal value, as a function of the number of outliers detected. For example, 37 sets/mixtures (around 11 % of all 338 mixtures) saw 3 lowest values excluded. Also, 10 sets (3 %) had 3 highest values dismissed as discrepant values.

Of all excluded values, two were detected by a greater number of methods, among all methods used. The lowest value found in mixture #126, fc = 64.81 MPa (individual compressive strength value), was excluded 11 times, that means, only 2 methods did not detect it for exclusion: EN 206 [32] which excludes measurements situated 15 % beyond the mean, and EN 13791 [44], that does not look for any outlier. The coefficient of variation, Cv, of the mixture was 6 %. Also, the value fc = 58.14 MPa found in mixture #133 was excluded in 11 of the 13 analysed procedures. The Cv of this mixture was 8 %. Apart from the EN 13791 the method that failed to detect any outlier was Turkey's method (Box-plot).

Taking into account the differences between the compressive strength results before and after exclusion by each method, a calculation was performed to measure the variation in the final compressive strength result (Δfcm). Due to the fact that some discrepant value was not detected in many mixtures, as shown in, Table 2 the total average of the variations of the compressive strength results did not reach high values (see Table 4). However, fcm variations obtained reached 11 %, as can be seen on Table 4. It

Table 3
Coincident detection of outliers (either lowest or highest values).

Number of outliers that coincide in different methods	Number of mixtures			
	Outlier: lowest value		Outlier: highest value	
	Number	Proportion	Number	Proportion
0	0	0.0 %	21	6.2 %
1	15	4.4 %	251	74.3 %
2	252	74.6 %	38	11.2 %
Example: 3	37	10.9 %	10	3.0 %
4	8	2.4 %	10	3.0 %
5	8	2.4 %	6	1.8 %
6	8	2.4 %	1	0.3 %
7	4	1.2 %	0	0.0 %
8	0	0.0 %	1	0.3 %
9	2	0.6 %	0	0.0 %
10	2	0.6 %	0	0.0 %
11	0	0.0 %	0	0.0 %
12	2	0.6 %	0	0.0 %
13	0	0.0 %	0	0.0 %
Total	**338**	**100.0 %**	**338**	**100.0 %**

Table 4
Variation of the fcm result after excluding outliers in the 338 mixes.

Method/Standard	Δfcm	
	Mininum	Maximun
Tails of the Normal distribution	0.0 %	2.6 %
Thompson-Tau	0.0 %	7.1 %
Turkey's Box-Plot	0.0 %	2.6 %
ASTM E178 - T *Stat*	0.0 %	5.2 %
ASTM E178 - Dixon Criteria	0.0 %	5.2 %
EN 13791 - Grubbs test	0.0 %	3.2 %
NP 196-1 - Cement strength	0.0 %	3.8 %
NBR 7215 - Cement strength	0.0 %	3.8 %
NBR 13279 - Mortar strength	0.0 %	2.4 %
NP EN 206 - Concrete strength	0.0 %	6.6 %
NBR 12655 - Concrete strength	0.2 %	11.0 %
NBR 7584 - Hardened concrete - reflection hammer	0.0 %	3.8 %
NP EN 13791 - Prefabricated concrete strength	0.0 %	0.0 %

involved the method described in NBR 12655.

NBR 13279 [46] - Mortar for laying and coating walls and ceilings - Determination of tensile strength in flexion and compression - is the specific Brazilian standard for the test carried out in this research. This Brazilian standard takes into account an allowed maximum absolute deviation from the mean (difference between the mean and the individual compressive strength that is farthest from the mean) equal to 0.5 MPa. According to this Standard, the test will be considered valid only when the final result consists of the average of at least four test records. Otherwise, the procedure must be redone. In the present situation, considering the guidelines established in NBR 13279, it can be seen that 155 mixtures (46 %) would be considered invalid, and the tests would be redone. As seen before, the EN 13791 does not provide for the exclusion of abnormal values.

Although there were several outliers detected in the database by each method (see Table 2), compressive strength variations greater than 5 % were not frequent, except for NBR 13279 [46], as can be seen on Table 5.

Analysing the methods that detected the highest number of outliers, except for the previously mentioned NBR 13279, Thompson Tau-method (111 outliers), and NBR 12655 (1014 discrepant values), it was possible to notice that, even excluding a higher number of values, just 2 mixtures (0,6 %) exhibited variation of compressive strength (Δfcm) greater than 5 %.

Application of the Thompson Tau-method to the 2 mixtures just mentioned resulted in $\Delta fcm = -7.1$ % (compressive strength at 24h) and 5.2 % (28 d). For the NBR 12655, respectively $\Delta fcm = 11{,}0$ % (24h) and 6.4 % (7 d) were obtained. Apparently, the high variations are not related to the age of the specimen. As shown in Table 5, after excluding the outliers detected by each procedure, eight of the 13 procedures did not show variations in their strength results greater than 5 %.

This study, and taking into account the importance of detecting outliers or reject discrepant values, reveals to which extent procedures that disregard the existence of outliers are not recommended, as is the case of NP EN 13791. Scientific methods for detecting outliers, on the other hand, demand specific knowledge of statistical methodologies, using more robust formulas, non-elementary computations, etc.

The methodology set in NBR 13279, which restricts measurement values to 0.5 MPa around the average, puts some practical problems, since it resulted in almost 46 % of the mixtures invalid. On the other hand, working with a semi-amplitude equal to twice the standard deviation was very comprehensive and detected few outliers. The method described in NBR 12655 works differently from the others in that it excludes the highest value of each pair of samples. This method was the one that presented the greatest variation compared to the results of the other methods.

Table 5
Number of mixtures with Δfcm > ± 5 % for each method.

Method/Standard	Number of mixtures (and percentage) with Δfcm > ± 5 %
Tails of the Normal distribution	0 (0.0 %)
Thompson-Tau	2 (0.6 %)
Turkey's Box-Plot	0 (0.0 %)
ASTM E178 - T *Stat*	1 (0.3 %)
ASTM E178 - Dixon Criteria	1 (0.3 %)
EN 13791 - Grubbs test	0 (0.0 %)
NP 196-1 - Cement strength	0 (0.0 %)
NBR 7215 - Cement strength	0 (0.0 %)
NBR 13279 - Mortar strength	0 (0.0 %)[a]
NP EN 206 – Concrete strength	1 (0.3 %)
NBR 12655 – Concrete strength	2 (0.6 %)
NBR 7584 – Hardened concrete – reflection hammer	0 (0.0 %)
NP EN 13791 – Prefabricated concrete strength	0 (0.0 %)

Legend:
[a] 17.2 % of mixtures were not considered, because all 6 values were excluded, considering NBR 13279 [46].

Among the remaining simplified procedures for the exclusion of discrepant values recommended for use by engineers and professionals involved in laboratory tests for the analysis of cement mixtures, four of them (NP 196-1, NBR 7215, NP EN 206, and NBR 7584) involve tolerance percentages around the mean. From the methods under consideration, it was noticed that the limit of 10 % (or greater) around the mean (NP 196-1, NP EN 206, and NBR 7584) detects a small number of outliers, compared to the scientific methods included in ASTM E178. On the other hand, the results obtained when applying NBR 7215 were close to those obtained when applying the ASTM. The NBR 12655 simplified method proved to be reasonably interesting, given its ease of application.

Based on the previous findings one may propose to divide the evaluated methods in the following groups.

(a) Group A: methods that apparently dismiss a high number of measurements without statistical justification - NBR 13279 and NBR 12655.
(b) Group B: methods that apparently dismiss dozens of measurements, possibly to many - Thompson-Tau method, ASTM E178 - T Stat, ASTM E178 – Dixon Criteria, and NBR 7215.
(c) Group C: methods that apparently dismiss only a few measurements - EN 13791 – Grubbs test, EN 196-1, EN 206, NBR 7584.
(d) Group D: methods that apparently dismiss one or just a couple of measurements, possibly too few – Tails of the Normal distribution, Turkey's method, and EN 13791.

Looking at the results and into the outliers that were excluded, one would argue that, for scientific purposes, the methods in the Group B are the most appropriate. The methods of Group A are out of sense for scientific purposes. The methods in Group D are too permissive. Looking at Group B it is observed that the NBR 7215 shall stand out from the other methods. One also finds that the methods Thompson-Tau, ASTM E178 - T Stat, ASTM E178 – Dixon Criteria work reasonably well while being statistically based and adequate for scientific purposes. Among the methods investigated, the authors of this paper recommend the methods Dixon Criteria and T Stat of the ASTM E178 as being the most adequate.

6. Conclusions

This work sought to analyse different ways of detecting and excluding outliers to compressive strength tests of cementitious mixes tested according to the EN 196-1. Different methodologies were applied both from standards and the statistical literature of outlier detection. Main findings of the current work are drawn as follows.

- The testing age did not seem to influence the detection of outliers;
- After excluding the discrepant values, most of the procedures did not show significant variations (±5 %) around the initial average value;
- The method described in NBR 13279 tends to invalidate a large proportion of mixtures produced and does not prove to be reliable;
- The method of excluding the lowest value of each pair of specimens proved to be easy and reliable to some extent; however a large number of exclusions occurred. It also led to the greatest discrepancy in the final results;
- Simplified methods of excluding outliers, such as the exclusion of values when deviating more than 6 % of the average, can be considered an interesting option for professionals involved in laboratory testing of building materials for cementitious mixtures, who request fast repeating;
- For scientific purposes, the researchers would better use statistical methods to detect outliers. Based on the data set used in the current work, the authors would recommend applying the Dixon Criteria (or T Stat) of the ASTM E178.

Future work should investigate in more detail application of all these methods to sets of specimens which include a discrepant measurement previously and intentionally mixed with the other specimens. Different degrees of discrepancy should be simulated.

Ethical approval

Not applicable.

Consent to participate

Not applicable.

Consent to publish

Not applicable.

CRediT authorship contribution statement

Nara Cangussu: Writing – original draft, Resources, Formal analysis, Data curation, Conceptualization. **Paula Milheiro-Oliveira:** Writing – review & editing, Validation, Formal analysis. **Ana Mafalda Matos:** Writing – review & editing, Supervision, Funding acquisition. **Farhad Aslani:** Writing – review & editing, Visualization. **Lino Maia:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared the data at the Attached File step

Acknowledgments

This work is financially supported by: Base Funding – UIDB/04708/2020 of the CONSTRUCT – Instituto de I&D em Estruturas e Construções - funded by national funds through the FCT/MCTES (PIDDAC); by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., under the Scientific Employment Stimulus – Institutional Call – CEECINST/00049/2018, and through the individual Scientific Employment Stimulus 2021.01765.CEECIND. The 2nd author was partially supported by CMUP, member of LASI, which is financed by national funds through FCT - Fundação para a Ciência e a Tecnologia, I.P., under projects with references UIDB/00144/2020 and UIDP/00144/2020. Collaboration and materials supply Secil, Omya Comital and Sika is gratefully acknowledged. Technical support of Laboratory of Building Materials of FEUP as also acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jobe.2024.110276.

References

[1] ASTM, ASTM E178 - Dealing with Outlying Observations, vol. 14, August. 2008, pp. 1–17, https://doi.org/10.1520/E0178-08.
[2] C.C. Aggarwal, Outlier Analysis, vol. 1, Springer, New York, 2019, https://doi.org/10.1007/978-14614-6396-2.
[3] D.M. Hawkins, Identification of Outliers, 1980.
[4] M.C.V. Andrade, C.A.O. Nascimento, e S.C. P. Migliavacca, Detection of outliers in a gas centrifuge experimental data, Braz. J. Chem. Eng. 22 (3) (2005) 389–400, https://doi.org/10.1590/S0104-66322005000300008.
[5] O. Maimon e Rokach Lior, Introduction to Knowledge Discovery and DataMining, second ed., Springer, Israel, 2010 https://doi.org/10.1007/978-0-387-09823-4. Maimon Oded e Rokach Lior, Orgs.
[6] I. Ben-Gal, Outlier Detection", Em *Data Mining And Knowledge Discovery Handbook*, Springer US, 2009, pp. 117–130, https://doi.org/10.1007/978-0-387-09823-4_7.
[7] R. Alsini, A. Almakrab, A. Ibrahim, e X. Ma, Improving the outlier detection method in concrete mix design by combining the isolation forest and local outlier factor, Constr Build Mater 270 (2021) 121396, https://doi.org/10.1016/j.conbuildmat.2020.121396.
[8] K. Dunphy, A. Sadhu, e B. Banting, Experimental and numerical investigation of tensile properties of early-age masonry, Materials and Structures/Materiaux et Constructions 54 (1) (2021) 1–18, https://doi.org/10.1617/s11527-021-01635-8.
[9] G. Araya-Letelier, F.C. Antico, M. Carrasco, P. Rojas, e C.M. García-Herrera, Effectiveness of new natural fibers on damage-mechanical performance of mortar, Constr Build Mater 152 (2017) 672–682, https://doi.org/10.1016/j.conbuildmat.2017.07.072.
[10] M. Miró, J.N. Eiras, P. Poveda, M. Climent, e J. Ramis, Detecting cracks due to steel corrosion in reinforced cement mortar using intermodulation generation of ultrasonic waves, Constr Build Mater 286 (2021), https://doi.org/10.1016/j.conbuildmat.2021.122915.
[11] R. Hendrickx, Using the Karsten tube to estimate water transport parameters of porous building materials: the possibilities of analytical and numerical solutions, Materials and Structures/Materiaux et Constructions 46 (8) (2013) 1309–1320, https://doi.org/10.1617/s11527-012-9975-2.
[12] J. Li, R. Wang, e L. Li, Influence of cellulose ethers structure on mechanical strength of calcium sulphoaluminate cement mortar, Constr Build Mater 303 (April) (2021) 124514, https://doi.org/10.1016/j.conbuildmat.2021.124514.
[13] C. Moro, H. El Fil, V. Francioso, e M. Velay-Lizancos, Influence of water-to-binder ratio on the optimum percentage of nano-TiO2 addition in terms of compressive strength of mortars: a laboratory and virtual experimental study based on ANN model, Constr Build Mater 267 (2021) 120960, https://doi.org/10.1016/j.conbuildmat.2020.120960.
[14] N.A. Libre, R. Khoshnazar, e M. Shekarchi, Repeatability, responsiveness and relative cost analysis of SCC workability test methods, Materials and Structures/Materiaux et Constructions 45 (7) (2012) 1087–1100, https://doi.org/10.1617/s11527-011-9818-6.
[15] R.L.S. Ferreira, M.A.S. Anjos, C. Maia, L. Pinto, A.R.G. de Azevedo, e J. de Brito, Long-term analysis of the physical properties of the mixed recycled aggregate and their effect on the properties of mortars, Constr Build Mater 274 (2021) 121796, https://doi.org/10.1016/j.conbuildmat.2020.121796.
[16] R. Neves, F. Branco, e J. De Brito, About the statistical interpretation of air permeability assessment results, Materials and Structures/Materiaux et Constructions 45 (4) (2012) 529–539, https://doi.org/10.1617/s11527-011-9780-3.
[17] S. Jueyendah, M. Lezgy-Nazargah, H. Eskandari-Naddaf, e S.A. Emamian, Predicting the mechanical properties of cement mortar using the support vector machine approach, Constr Build Mater 291 (2021) 123396, https://doi.org/10.1016/j.conbuildmat.2021.123396.
[18] E. Del Monte, S. Boschi, e A. Vignoli, Prediction of compression strength of ancient mortars through in situ drilling resistance technique, Constr Build Mater 237 (2020) 117563, https://doi.org/10.1016/j.conbuildmat.2019.117563.
[19] M. Ramesh, M. Azenha, e P.B. Lourenço, Quantification of impact of lime on mechanical behaviour of lime cement blended mortars for bedding joints in masonry systems, Constr Build Mater 229 (2019) 116884, https://doi.org/10.1016/j.conbuildmat.2019.116884.
[20] G. De Schuttere, A.M. Poppe, Quantification of the water demand of sand in mortar, Constr Build Mater 18 (7) (2004) 517–521, https://doi.org/10.1016/j.conbuildmat.2004.04.004.
[21] L. Falchi, C. Varin, G. Toscano, e E. Zendri, Statistical analysis of the physical properties and durability of water-repellent mortars made with limestone cement, natural hydraulic lime and pozzolana-lime, Constr Build Mater 78 (2015) 260–270, https://doi.org/10.1016/j.conbuildmat.2014.12.109.
[22] Y. Xu, G. Xing, J. Zhao, e Y. Zhang, The effect of polypropylene fiber with different length and dosage on the performance of alkali-activated slag mortar, Constr Build Mater 307 (February) (2021) 124978, https://doi.org/10.1016/j.conbuildmat.2021.124978.
[23] A. Vollpracht, B. Lothenbach, R. Snellings, e J. Haufe, The pore solution of blended cements: a review, Materials and Structures/Materiaux et Constructions 49 (8) (2016) 3341–3367, https://doi.org/10.1617/s11527-015-0724-1.
[24] S. Weigel e D. Stephan, Modelling of rheological and ageing properties of bitumen based on its chemical structure, Materials and Structures/Materiaux et Constructions 50 (1) (2017) 1–15, https://doi.org/10.1617/s11527-016-0957-7.

[25] G. Bai, C. Zhu, C. Liu, e B. Liu, An evaluation of the recycled aggregate characteristics and the recycled aggregate concrete mechanical properties, Constr Build Mater 240 (2020) 117978, https://doi.org/10.1016/j.conbuildmat.2019.117978.
[26] T.H. Panzera, K. Strecker, J.D.S. Miranda, A.L. Christoforo, e P.H. R. Borges, Cement - steatite composites reinforced with carbon fibres: an alternative for restoration of Brazilian historical buildings, Mater. Res. 14 (1) (2011) 118–123, https://doi.org/10.1590/S1516-14392011005000007.
[27] I. Galobardes, R.P. Salvador, S.H.P. Cavalaro, A. Figueiredo, e C.I. Goodier, Adaptation of the standard EN 196-1 for mortar with accelerator, Constr Build Mater 127 (2016) 125–136, https://doi.org/10.1016/j.conbuildmat.2016.09.147.
[28] A.M. Matos, L. Maia, S. Nunes, e P. Milheiro-Oliveira, Design of self-compacting high-performance concrete: study of mortar phase, Constr Build Mater 167 (2018) 617–630, https://doi.org/10.1016/j.conbuildmat.2018.02.053.
[29] K.L. Aughenbaugh, T. Williamson, e M.C. G. Juenger, Critical evaluation of strength prediction methods for alkali-activated fly ash, Materials and Structures/ Materiaux et Constructions 48 (3) (2014) 607–620, https://doi.org/10.1617/s11527-014-0496-z.
[30] L.N. Koutas e D.A. Bournas, Confinement of masonry columns with textile-reinforced mortar jackets, Constr Build Mater 258 (2020) 120343, https://doi.org/10.1016/j.conbuildmat.2020.120343.
[31] ASTM, E178 - Dealing with Outlying Observations, vol. 14, August. 2008, pp. 1–17, https://doi.org/10.1520/E0178-08.
[32] CEN, EN 206: Concrete - Specification, Performance, Production and Conformity, 2013.
[33] ABNT, NBR 12655 - Portland Cement Concrete - Preparation, Control, Receipt and Acceptance - Procedure. Brazil: (Only Available in Portuguese), 2015.
[34] N.M. de O. Cangussu, L.V. Martins, e L. Maia, "Exploratory study about rice husk ash as partial substitute of cement in concrete applications", em *21st international multidisciplinary geoconference - sgem 2021*, SGEM WORLD SCIENCE (SWS) Scholarly Society 2021, Org., Albena, Bulgaria: STEF92 Technology (2021) 359–366, https://doi.org/10.5593/sgem2021/6.1.
[35] International Energy Agency, "Technology Roadmap - Low-Carbon Transition in the Cement Industry". [Online]. Available in: www.wbcsdcement.org.
[36] OCDE, Global Material Resources Outlook to 2060, OECD, 2019, https://doi.org/10.1787/9789264307452-EN fev.
[37] K.L. Scrivener, V.M. John, e E.M. Gartner, Eco-efficient cements: potential economically viable solutions for a low-CO2 cement-based materials industry, Cem Concr Res 114 (2018) 2–26, https://doi.org/10.1016/J.CEMCONRES.2018.03.015, dez.
[38] CEN, EN 196-1: Methods of Testing Cement. Part 1, Determination of strength, 2006, pp. 1–37.
[39] L. Minitab, "Suporte ao Minitab 21", Identificação de Outlier. Acessado: 8 de dezembro de 2023. [Online]. Available in: https://support.minitab.com/pt-br/minitab/21/help-and-how-to/graphs/general-graph-options/graph-editing-basics/identifying-outliers/.
[40] ISO, ISO/TR 22971 - Accuracy (Trueness and Precision) of Measurement Methods and Results, 2005, p. 40. Switzerland.
[41] Walpole Ronald, Raymond Myers, Myers Sharon, e Ye Keying, Probability & Statistics for Engineers & Scientists, ninth ed., 2011.
[42] Q. Shen e R. Yang, Thompson-tau outlier detection method for detecting abnormal data of listed pharmaceutical companies in China. Em *Proceedings - 2015 8th International Symposium On Computational Intelligence And Design, ISCID 2015*, Institute of Electrical and Electronics Engineers Inc., 2016, pp. 379–382, https://doi.org/10.1109/ISCID.2015.298, maio.
[43] Procedures for detecting outlying observations in samples, Technometrics 11 (1) (1969) 1–21, https://doi.org/10.1080/00401706.1969.10490657.
[44] CEN, EN 13791: assessment of in-situ compressive strength in structures and precast concrete components, European Standard 5 (2019) 18–22, https://doi.org/10.4324/9781315780320-29. +95.
[45] ABNT, *PNBR 7215: Ortland Cement - Determination Of Compressive Strength*. Brazil: (Only Available in Portuguese), 1996.
[46] ABNT, *NBR 13279: Mortar For Laying and Covering Walls and Ceilings – Determination Of Tensile Strength in Flexion and Compression.* Brazil: (Only Available in Portuguese), 2005.
[47] ABNT, *NBR 5739: Concrete - Compression Test of Cylindrical Specimens*. Brazil: (Only Available in Portuguese), 2018.
[48] ABNT, *NBR 5738: Concrete - Procedure For Molding and Curing Specimens*. Brasil: (Only Available in Portuguese), 2015.
[49] ABNT, *NBR 7584 - Hardened Concrete - Assessment Of Surface Hardness Using a Reflection Sclerometer*. Brazil: (Only Available in Portuguese), 2012.