Journal of Computer Science

Original Research Paper

Determination of Optimum Alloy Sample Size for Ball Mill Abrasion Test Through Automated t-Student Distribution Analysis

[1]Carlos Daniel Garcia Mendoza, [1]Lelly María Useche Castro, [1]Miguel Ángel Lapo Palacios, [2]Óscar Vargas Ortiz and [3]Ronny Javier Maldonado Macías

[1]*Department of Mathematics and Statistic, Institute of Basic Sciences, Technical University of Manabi, Ecuador*
[2]*Faculty of Postgraduate, Technological Business University of Guayaquil, Ecuador*
[3]*Faculty of Earth Sciences, Amazon State University, Ecuador*

Article history
Received: 01-04-2022
Revised: 26-08-2022
Accepted: 02-09-2022

Corresponding Author:
Carlos Daniel Garcia Mendoza
Department of Mathematics and Statistic, Institute of Basic Sciences, Technical University of Manabi, Ecuador
Email: daniel.garcia@utm.edu.ec

Abstract: In material wear experiments, such as the Ball Mill Abrasion Test (BMAT), it is crucial to know the number of specimens to be used in each experiment to produce accurate and reliable results. In BMAT, however, a structured way of determining the necessary number of balls per type of material to be included in the experiments has not been considered to date. This article shows, using statistical tools and adhering to the standards that govern these experiments, that the optimal number of balls per type of material to be included in each experiment is 6. In addition, it is shown that if the initial differences between the balls are reduced, it is possible to obtain more accurate and reliable results.

Keywords: Ball Mill Abrasion Test, t-Student Distribution, Sample Size Optimization

Introduction

It is crucial in experimental procedures to obtain accurate and reliable measurements of the parameter of interest. To reach this goal, scientists resort to the use of precise equipment and the collection of a high number of measurements (Antony, 2014). The first approach is obvious: The more precise the machine is, the more reliable and reproducible the results are. In regards to the number of measurements, on the other hand, it is clear that the more abundant and systematic they are, the more representative they become.

In ideal circumstances, scientists aim to take as many measurements as possible to get increasingly closer to the actual value of the parameter being studied. When this is possible, a normal-like distribution is usually obtained (Ross, 2017; Ibe, 2013). In reality, however, in some experimental procedures, it is either impossible or impractical to repeat experimental measurements numerous times and get a normal distribution. In this scenario, the t-student distribution has shown to be more appropriate (Brereton, 2015). Although it is a general source of discussion, this method is commonly applied when the sample size drops below 30, given that a normal-like distribution is obtained above this sample size (Ross, 2017; Kim, 2015).

It is of particular interest to examine the possibility of optimizing the number of balls needed per alloy in Ball Mill Abrasion Tests (BMAT) to yield reliable results, given that only a single measurement can be obtained from each ball after every experiment. BMAT is used to measure the resistance of materials under wear conditions similar to those in an industrial ball mill (Gates *et al*., 2018). The experiment consists of a rotating drum containing metallic balls of the materials to be tested and an abrasive medium. The interactions between balls and the abrasive material lead to the crushing of the abrasive and mass loss of the grinding media due to wear (Ali *et al*., 2019). The optimization of these experimental setups can facilitate the obtention of experimental results, which are crucial to the industry given the significant losses comminution processes represent (Massola *et al*., 2016).

In BMAT experiments, given the number of alloys generally tested, it could be both impractical and impossible to accommodate more than 30 balls per alloy in a single experiment, just to obtain more than 30 measurements and get a normal-like distribution. It could be impractical given that, for example, adding an extra ball per alloy extends the time taken to measure the weight loss of the entire alloy type. If, for instance, 10 different alloys are being tested and 30 balls of each alloy are included (just to approximate closer to the normal distribution) and considering that the time taken to measure each ball's weight loss is approximately 10 sec; the weight loss measuring step would take a total of 5 h. If having such an abundant sample size for each alloy would be necessary, the interest in this methodology and its results would rapidly diminish.

Additionally, it must be remembered that laboratory equipment is generally small compared to industry-level equipment. Therefore, in some circumstances, it may be essentially impossible to try to get a great amount of data at the same time by increasing the number of balls per alloy. In BMAT, for example, if the same number of balls abovementioned are considered, with an average mass of 130 g, the total volume occupied by them would be approximately 5.7 L. If the volume occupied by the minerals to be ground, water, and the make-up charge are added, it would rapidly be noticed that a relatively big mill will be needed. Moreover, it is important to mention that handling such an amount of material in a laboratory, may be dangerous for the operator and will increase the time taken for unloading the mill and cleaning the samples. The aim of this study is, therefore, to determine the optimum alloy sample size for appropriate material performance estimation in BMAT.

Even though the analysis of the optimum number of specimens per alloy is significantly important for wear tests and specifically for BMAT, it is not possible to find a systematic approach to this issue in the literature.

One of the earliest works on wear tests similar to BMAT and with great impact on this field is that of Albertin and Sinatora (2001), who loaded 12 balls of each alloy to the laboratory mill used. Similarly, Chenje *et al*. (2004) exposed 3 balls each of five different alloys to the tested simultaneously, to determine their microstructure-property-wear performance relationships. Gates *et al*. (2008), who in more detail explored the possibility of using BMAT as a proper wear performance test, used an average of 11 balls per alloy. Jankovic *et al*. (2016) used two ball mills of different sizes, for the larger mill 21 balls of each of the four grades of media were charged to the mill, while for the smaller mill 16 balls of three different grades were used. Most recent works have been done by Ali *et al*. (2019), who used various sets of balls, each of which contained an average of 11 balls.

Other authors who have worked on wear tests similar to BMAT, on the other hand, have not stated the exact amount of specimens implemented in the tests but mentioned throughout their analysis the importance of taking a statistically significant number of test balls for each experiment.

As observed, the sample size considered for this type of wear test fluctuates greatly, being as low as 3 balls in some cases and as high as 21 for others. The absence of evidence showing a justified choice of sample size for each material type considered in these wears experiments sets the basis to consider a structured approach to determine the sample size needed for BMAT. The question to be answered is, therefore, whether it is possible to find a minimum sample size for which accurate measurements of alloy performance are still obtained.

In general, statistical terms, many authors have worked on the determination of appropriate sample size for experimental procedures, which include: Miot (2011) who states that one should determine the significance level of the estimate and the maximum tolerable sample error to determine the sample size. Moreover, if the population standard deviation is unknown and no literature presents similar data, a pre-test should be conducted to know the behavior of the subgroup. Similarly, Bujang and Baharum (2016) claim that sample size guidelines should be guided by the determination of a sizeable effect size that researchers can accept or tolerate. Additionally, it is stated that sample size determination very much depends on the particular study objectives. Both authors highlight the importance of considering the particularities of the experiment to be performed. Favorably, numerous BMAT experimental results are available, which give a great insight into the behavior of the data yielded by this wear test.

Verma and Verma (2020) analyze both extremes of the sample size spectrum. For instance, using a sample smaller than the optimum increases the probability of rejecting the correct claim and can fail to detect a small effect. This, however, does not imply that large samples provide better results than usually believed. One of the main issues with a large sample size is the need for more financial and human resources than possibly required. The method proposed by the authors to determine the sample size consists of deciding the level of accepted error and finding an estimate of the population standard deviation. Van de Schoot and Miocevic (2020) on the other hand, propose that a small sample size solution could lie in the addition of classical experimental design elements such as Randomization, blocking, and replication. Awe *et al*. (2022), as the previously presented authors, have done, claim that sample size determination depends strongly on the aims, nature, and scope of the study and the expected results. Moreover, it is limited by the availability of resources such as time, manpower, money, etc. A crucial statement made by the authors is that using the same sample size as those of similar studies is not recommended, as it is possible to run the risk of repeating errors that were made in determining such sample size. It is for this reason that this study aims to explore the particularities of the BMAT experiment to determine, with great statistical significance, the appropriate sample size, given that no authors have previously considered such an important topic.

Finally, Lakens (2022) claims that one of the methods for determining the sample size consists in planning for results precision. In more detail, the sample size based on precision requires the collection of data to achieve the desired width of the confidence interval around the parameter estimate. This decision should be taken considering previous experimental results and the nature of the test itself. This approach will be considered in this study, taking into account the particular characteristics of BMAT results.

Preliminaries

In the scenario of normal distribution behavior, scientists can make use of the mean and sample standard deviation to estimate the real data characteristics. As a general convention, the 95% limits are used to generate a range of values inside which the author is confident to conclude that the real data mean is found (Ahsanullah *et al*., 2014).

The distribution to be used in this study is the t-student distribution, which better estimates the errors at small sample sizes (Ahsanullah *et al*., 2014). The expression to be used to calculate the 95% confidence interval is:

$$I_{t-student} = \bar{X} \pm \frac{^{t}(1-\alpha, n-1)^{\alpha}}{\sqrt{n}} \quad (1)$$

where, $t_{(1-\alpha, n-1)}$ corresponds to the t value from the student distribution at which the 95% confidence interval is obtained. Given this expression, it is possible to analyze the behavior of this interval as a multiple of the standard deviation when increasing n. Figure 1 shows the behavior of the 95% confidence interval as a multiple of standard deviation vs. several measurements. It is observed that the 95% intervals given as multiples of the sample standard deviation rapidly drop as the number of measurements increases. In experimental procedures where the sample size has to be estimated, as the ASTM E122-17 states, it is necessary to first state the precision desired. This standard covers the methods to determine the sample size required to estimate, with specified precision, a measure of the quality of a lot of material, or produced by a process. The authors of this study, based on the analysis of the data behavior from previous BMAT results, consider that if the 95% confidence interval is smaller than X ± σ, then the measurement is considered to be relevant and precise. This consideration, in the context of the ASTM E122-17 standard, is translated to a maximum acceptable difference between the true average and the sample average to be equal to one advance estimate of σ.

It is noticeable that at around 6 measurements the confidence interval can be given as a multiple of one sample standard deviation and beyond this point the interval decreases at a much lower rate, taking 11 extra measurements to reduce the confidence interval to half of that obtained at 6.

This study will examine whether this phenomenon will apply to BMAT tests. If it is possible to reduce the sample size of each alloy to 6, for example, then it would be possible to reduce, by a decent amount, the time taken to measure the weight losses of the grinding media after a wear experiment, considering that 10 or more balls are generally included for each material type.

Data Collection

Existing results from Ball Mill Abrasion Tests are used. The BMAT experiments were run on a 601 mm diameter laboratory mill, at a rotational speed equivalent to 25% of the critical speed (the speed at which centrifuging occurs). The abrasive used was a mixture of pure quartz (10 mm) and A.F.S. sand, the grinding media were metallic balls of approximately 32 mm in diameter. To obtain the desired mill fill, a makeup charge was used, which consisted of 40 mm diameter martensitic steel balls. The tests were run wet, i.e., water was added to reduce the risks associated with inhaling dust produced by the grinding process. For each alloy considered in this study, more than 10 balls were included. The mass loss experimented by each ball (from each alloy group) was registered, which is the information being processed and analyzed in this study. The results used are in units of $\frac{mg}{dm^2h}$, corresponding to the amount of mass loss per surface area per duration of the test. The quantities presented in this study (unless stated otherwise) correspond to these units. The characteristics of each alloy used in this study are shown in Table 1.

Table 1: Chemical composition of materials studied

Material name	C	Cr	Mo	Cu	Mn	Si	Ni
MAT 1	3.13	10.62	0.01	0.10	0.50	0.36	0.12
MAT 2	0.76	0.27	0.02	0.29	0.70	0.15	0.10
MAT 3	2.69	29.97	0.03	0.07	0.71	0.45	0.20

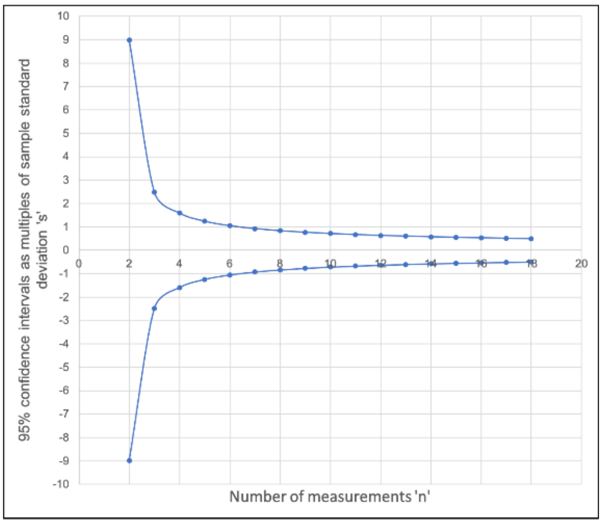

Fig. 1: 95% confidence intervals as multiples of standard deviation vs number of measurements

Materials and Methods

The aim of the methodology to be detailed hereunder is to determine the optimal number of measurements required for an alloy in a single BMAT experiment to achieve the precision stated previously. To achieve this, a set of measurements of N number of balls is provided from existing experimental results. The optimization method provides a combination of a smaller quantity of measurements n with the same reliability as that obtained when using all the measurements. In formal terms: Let N be the total number of specimens in a single alloy, then the number of unique combinations of n measurements is given by the following equation:

$$C_n^N = \frac{N!}{(N-n)!n!} \quad (2)$$

where each combination can be represented ${}^{i}B_n^N$ with 0≤ i ≤ C_n^N and can be thought of as a set of n artificial measurements done for the given alloy. The set of all possible combinations can be expressed as:

$$P_n = \left\{ {}^{i}B_n^N : 0 \le i \le C_n^N \right\} \quad (3)$$

It is then possible to obtain the sets of the averages and standard deviations of each combination set ${}^{j}B_n^N$, as follows:

$$P_n^{avg} = \left\{ \overline{{}^{i}B_n^N} : 0 \le i \le C_n^N \right\}$$
$$P_n^{stdv} = \left\{ s\left({}^{i}B_n^N\right) : 0 \le i \le C_n^N \right\} \quad (4)$$

These are sets with cardinality equal to C_n^N which, depending on the value of n, could be as high as 462 if N = 11 for instance. It is possible then to create histograms for these two variables P_n^{avg}, P_n^{stdv} which would illustrate the recurrence of each mean and standard deviation calculated from all combinations of n measurements as will be seen later. The behavior of these two variables is crucial for the optimization method.

For instance, $\overline{P_n^{stdv}}$ has a direct impact on the behavior of the confidence intervals at different n values. Given that the proposed method should yield an optimum combination of measurements ${}^{j}B_n^N$ that is representative of the materials' behavior, the first restriction imposed on the method is the following:

$$\overline{P_n^{avg}} - \frac{1.96\sigma}{\sqrt{C_n^N}} \le \overline{{}^{j}B_n^N} \le \overline{P_n^{avg}} + \frac{1.96\sigma}{\sqrt{C_n^N}} \quad (5)$$

where, σ is the standard deviation of the variable P_n^{avg}. Additionally, given that the confidence intervals are proportional to the standard deviation and that the goal is to minimize the width of this interval, it is also required for s $s\left({}^{j}B_n^N\right)$ to be minimal. By doing so, the method not only ensures that the optimal combination ${}^{j}B_n^N$ is representative of the real behavior of the material but also reduces the width of the confidence interval, providing more significance and reliability to the results yielded.

Results and Discussion

Analysis of Material MAT1

Evolution of Means and Standard Deviation Histograms

From the wear experiment considered it was possible to obtain 11 results corresponding to the alloy weight loss. As mentioned before, the proposed methodology can extract this information and create graphs of the recurrence of means and standard deviation from all possible combinations of n elements. The figures to be presented in this section have the intention of illustrating the evolution of the histograms of means and standard deviations as n increases, taking as particular cases n = 2,6,11, sufficient to understand the main changes produced by the increase in n. In the Appendix section, the reader can find the complete evolution of these histograms. Figure 2 shows the results from combining 2 measurements from the 11 pre-existing ones.

It is seen from Fig. 2 that the range of values of means is fairly widespread, starting at approximately 115 up to approximately 155. From this graph, it is not clear what the actual performance of the alloy is under this experimental condition, or at least it does not give the reader the confidence to state a single value that would represent the entire family of results. From the standard deviation histogram, it can be seen that the values range from around 0 to almost 25.

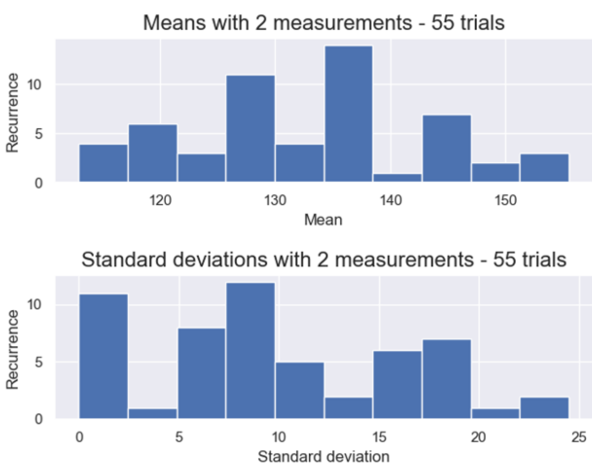

Fig. 2: Histograms of means and standard deviations using 2 samples (MAT 1)

DOI: 10.3844/jcssp.2022.841.851

From these results, it is clear that making only two measurements is simply inappropriate, as already suggested by some literature (Verma and Verma, 2020; Miot, 2011). There is of course a high chance to obtain results with very low standard deviation, although the same chances appear for high standard deviation. It is not appropriate then to make two measurements to predict the performance of an alloy in experimental conditions similar to these. Notice that Fig. 2 also shows the number of trials, corresponding to the total number of possible combinations of 2 elements from the list of pre-existing results, i.e., C_n^N .

If the number of measurements further increases to 6, Fig. 3 is obtained. When using 6 elements to form combinations, the histogram of mean values appears to resemble a normal-like distribution, provided the high number of trials, as predicted by the literature (Ross, 2017). In this case, the range of mean values starts at 120 and finalizes at approximately 145, with a clear concentration of data between 130 and 135, in contrast to the previous scenario where the mean values were spread across a large interval, making it particularly difficult to calculate the actual alloy performance. The results from combining 6 elements indicate that the actual mean value and therefore the actual alloy performance is likely to be inside the interval starting at 130 up to 135 (a smaller range than that provided by the previous case).

In terms of the standard deviation histogram, it is evident the values are spread over a relatively wide range, though most of the values are concentrated in the interval from 14 to 16. This standard deviation still represents around 11% of the weight loss value, which may not be considered acceptable for these experiments. However, notice as well how it is possible to produce combinations of 6 values that produce a relatively low standard deviation (approximately 10). If these combinations correspond to those producing an average performance inside the most recurrent mean values, it would be ideal to only use these specimen balls for future experiments. This will not only let the experiment produce data with low deviation but also with high accuracy in measuring the absolute performance of the alloy. In this scenario, the most convenient (and surprisingly easy) way to produce a more accurate and precise experiment setup will mostly depend on correctly choosing the right specimen balls. In future sections, this possibility will be further analyzed.

At last, it is possible to consider the combination of 11 elements, which is simply the average and standard deviation of the total amount of data points existing for the experiment, corresponding to 132.6 and 16.2 respectively. Since the pre-existing experiment results are by nature fairly spread, it is no surprise to have a relatively high standard deviation (Brereton, 2015). In this case, having numerous specimens, without considering their initial differences, appears to be an inappropriate approach, contradicting the general belief that simply having more specimens in an experiment will automatically produce better and more accurate results as also suggested by Verma and Verma (2020).

Averaged Standard Deviations

As mentioned in the methodology, the third type of graph can be obtained from the data, which shows the average standard deviation for each possible combination. This is done by calculating the mean standard deviation for every n number of measurements and plotting this data against n itself as shown in Fig. 4.

It can be observed that the average standard deviation increases as the number of measurements increases. As explained before, this is caused by the pre-existing differences in specimen balls. Therefore, it is expected to see the largest standard deviation in combinations where it is more likely to gather results further apart. For obvious reasons, this happens when the combinations are made with a larger number of elements and reach a maximum when the number of measurements equals the total number of data points collected, in this case when $n = 11$.

95% Confidence Intervals

The next step consists of analyzing whether this increase in standard deviation, apparently significant, can drastically change the 95% confidence intervals. It is important to recall that the theoretical analysis gave the number of measurements 6, needing 11 additional measurements to reduce the interval by a factor of two. However, the intervals were given as multiples of standard deviation, which we have seen an increase as the number of measurements increased. Figure 5 shows the 95% intervals (shifting the mean to zero) vs. the number of measurements.

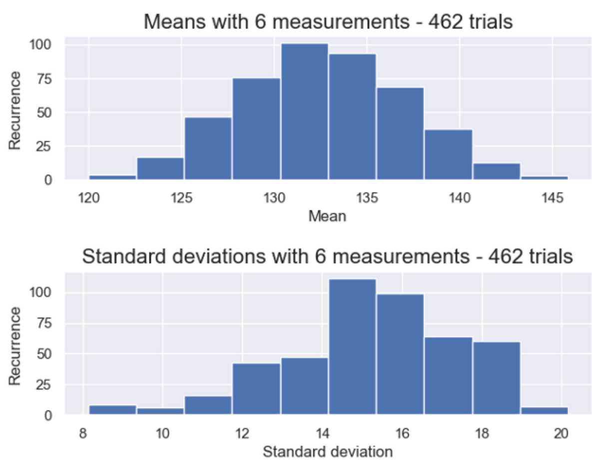

Fig. 3: Histograms of means and standard deviations using 6 samples (MAT 1)

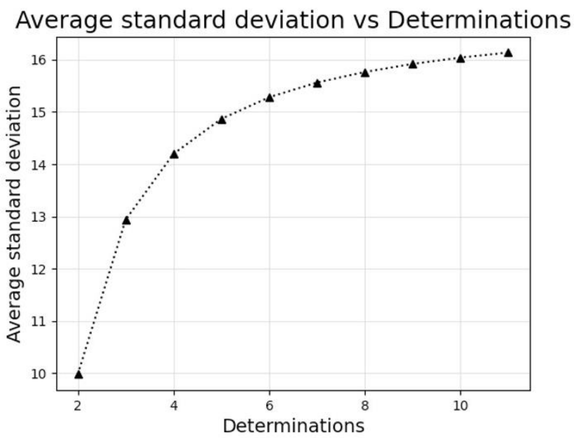

Fig. 4: Averaged standard deviation vs measurements (MAT 1)

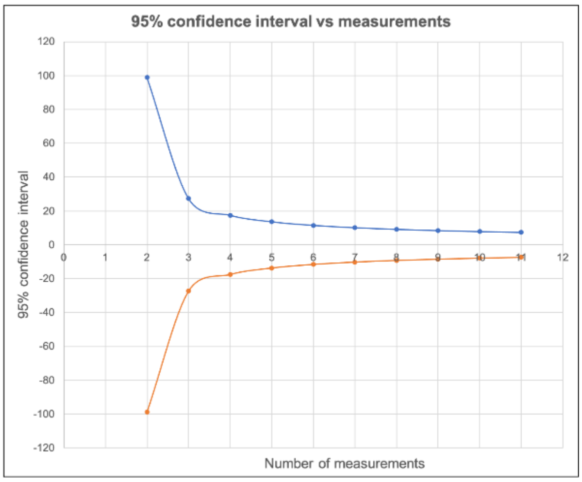

Fig. 5: 95% Confidence interval vs measurements (MAT1)

It is observed that the confidence interval rapidly decreases at the start, dropping from around 98 to approximately 20 in only 3 measurements. However, after around 6 measurements, the confidence interval decreases at a much lower rate. For example, increasing the measurements from 6 to 11 (almost twice the number of measurements) only drops the upper bound of the interval by approximately 5 points. Additionally, it is observed that the radius of the 95% confidence interval at 6 measurements is proportional to one standard deviation (the precision desired). Therefore, it is possible to state that a sample size of 6 balls is indeed an optimal amount.

It is important to recall that Fig. 5 is produced with averaged standard deviations for each determination (from all possible combinations of n values), therefore it could be possible to further decrease the confidence interval to a narrower range (for n values lower than 11) if, for example, the optimization process can find a combination of n values with a relatively low standard deviation, as explained in the methodology.

Optimizing the 95% Confidence Interval at 6 Measurements

The optimization method was able to choose those combinations which produced averages inside the 95% confidence interval of the variable P_n^{avg}, which is considered a reasonable range for the actual average performance of the alloy. This range was chosen based on the intervals with most data accumulation in the histogram shown in Fig. 3. The proposed method is then able to find those combinations which produce the lowest standard deviations. By doing this, it is guaranteed that the combination found not only falls in between a desired average value but also with a low enough deviation which could potentially reduce the 95% confidence interval at this number of measurements.

After running the optimization process, a combination of six values with an average of approximately 134 and a standard deviation of approximately 11.3 was found. This corresponds to the combination with the mean closest to the average of 11 values and the lowest standard deviation. By choosing this combination, it is possible to reduce the width of the 95% confidence interval to approximately 7. If only this value is optimized, it is possible to see a clear reduction in the 95% confidence interval at 6 measurements as shown in Fig. 6.

Therefore, it is possible to further reduce the 95% confidence interval by simply optimizing the specimens chosen for the experiment. In this case, the 95% interval width was reduced to a value that is only two points away from the interval generated with 11 measurements. Besides that, using optimized values, it was possible to reduce the standard deviation from 16.2 (using 11 measurements) to 11.3 (using 6 measurements). It is important to note as well that the standard deviation and 95% confidence interval at 11 measurements cannot be further reduced since the only available data consisted of 11 values. With this premise, it is unreasonable to use all 11 available specimen balls considering that carefully choosing 6 of them would lead to a significant decrease in standard deviation with similar confidence intervals.

Analysis of Material MAT 2

Evolution of Means and Standard Deviation Histograms

The same procedure seen above was applied for this material and the histograms of means and standard deviations from all possible combinations of n values were obtained, as shown in Fig. 7.

It is clear from Fig. 7, that the average data is widely spread over a large range of values, which confirms that using two samples is a scenario that has to be avoided if precise performance information is desired. Even though BMAT results can be reproducible with high precision, it is important to note that the significance of these results will be strongly linked to the number of samples used. For example, if certain combinations are used, it is possible to obtain a weight loss of approximately 157 while other combinations can produce a weight loss of 161. This difference may seem insignificant but when extrapolating these results to experiments with higher mass losses (e.g., higher rotational speeds), the gap between them gets larger. Additionally, for mining companies that lose significant amounts of resources during grinding media change, even the smallest benefit over the grinding media performance is desired. As with the previous material, it was possible to produce histograms considering all combinations of 6 values from the pre-existing data, generating Fig. 8.

The histogram of mean values shows a clear concentration of data around 158.0. In contrast with the previous case, using six values or measurements seems more reasonable, since it is more likely to fall between a range close to the actual alloy performance (and within the precision desired). In terms of the standard deviations, the data accumulates at relatively high values, however, as seen in MAT 1, there could be a combination of 6 values with a relatively good approximation of the mean performance with a low standard deviation. This would, in turn, make it possible for optimization to take place and reduce the 95% confidence interval at six measurements. Finally, the combination of the 11 samples produced a mean of 158.1 and a standard deviation of 1.8.

Averaged Standard Deviations

As with the first alloy, the average standard deviation increases as the number of measurements increases, reaching a maximum of 11, given that the total number of pre-existing results is 11. The range of values of average standard deviation is however not as wide as in the previous case. For this reason, it is expected to obtain a 95% confidence interval graph similar to that obtained previously.

Optimizing the 95% Confidence Interval at 6 Measurements

Following the previously mentioned method for optimizing the confidence intervals at 6 measurements only, Fig. 9 is obtained.

After applying the optimization process, the 95% confidence interval at 6 measurements could be reduced from 2.01 to 1.52. Further reducing the confidence interval not only warrants meeting the precision desired but also improves the quality of the population's behavior prediction (Verma and Verma, 2020). The optimized interval only differs by 0.5 units from the confidence interval obtained when using 11 measurements. Additionally, when using the optimized combination, the standard deviation could be reduced from 1.81 (using 11 measurements) to 1.37; further encouraging the use of carefully picked specimens from a larger but probably more deviated sample.

Analysis of Material MAT 3

Evolution of Means and Standard Deviation Histograms

The mean and standard deviation histograms from the last alloy considered in this analysis are shown below. Figure 10 shows the results from combining 2 sample data from the list.

As with the previous two alloys, the histogram of means when using 2 samples contains data spread over a wide range of values. This is also the case for the standard deviation histogram, confirming once again the risks associated with considering a low number of samples in BMAT. A more interesting result is obtained in Fig. 11.

In contrast to the two previous alloys, MAT 3 does not show a clear accumulation of mean values when using 6 samples. The histogram of mean values shows that the mean data is almost evenly spread over a large range of values, starting at around 100 up to approximately 125. One of the reasons for this to happen is the material characteristic. MAT 3, is a more resistant material in these conditions, observed as a lower mass loss compared to the previous two alloys. A natural consequence of this is that at experimental conditions such as the one considered now, which leads to relatively low mass losses (25% rotational speed), the more resistant materials show more data deviation.

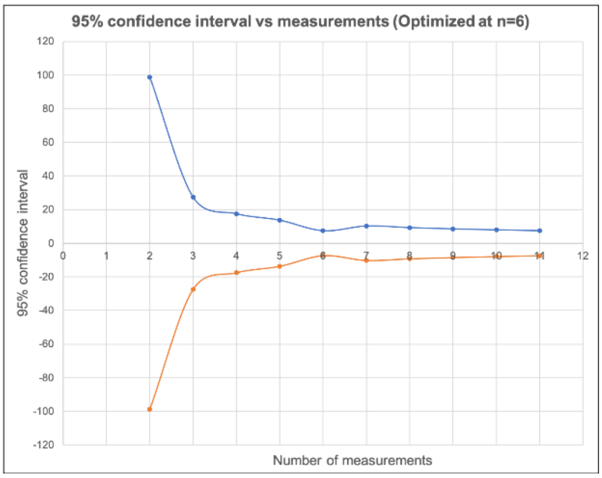

Fig. 6: Optimized 95% confidence interval at 6 measurements (MAT 1)

DOI: 10.3844/jcssp.2022.841.851

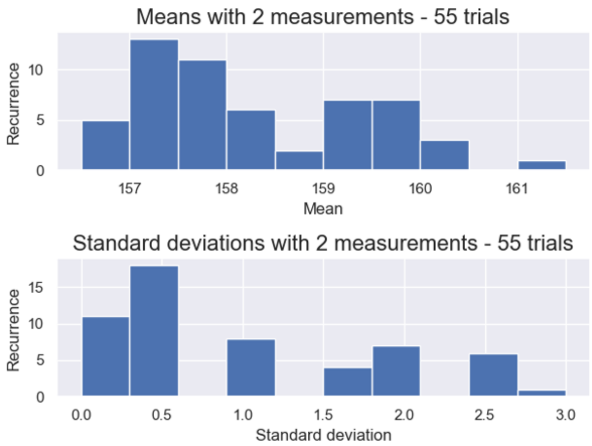

Fig. 7: Histograms of means and standard deviations using 2 samples (MAT 2)

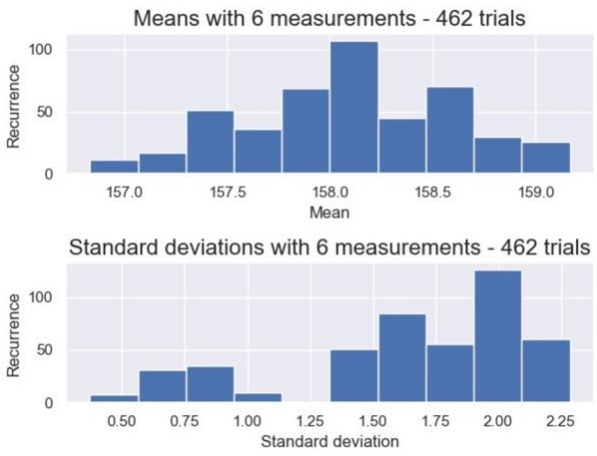

Fig. 8: Histograms of means and standard deviations using 6 samples (MAT 2)

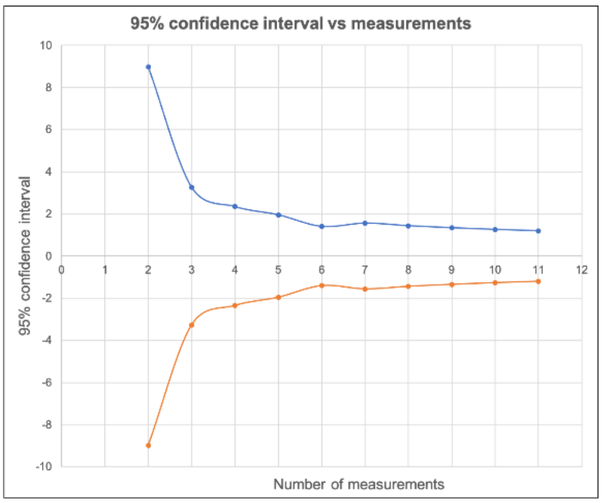

Fig. 9: Optimized 95% confidence interval at six measurements (MAT 2)

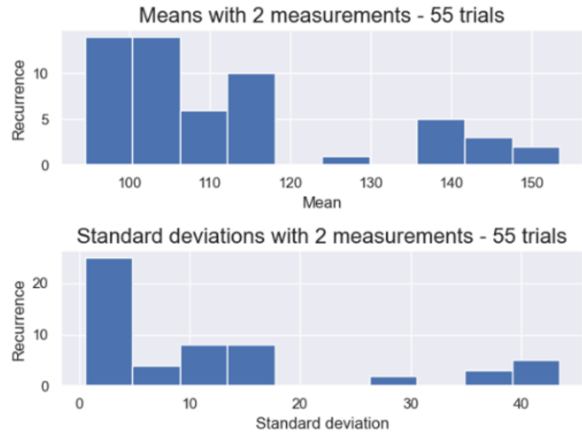

Fig. 10: Histograms of means and standard deviations using 2 samples (MAT 3)

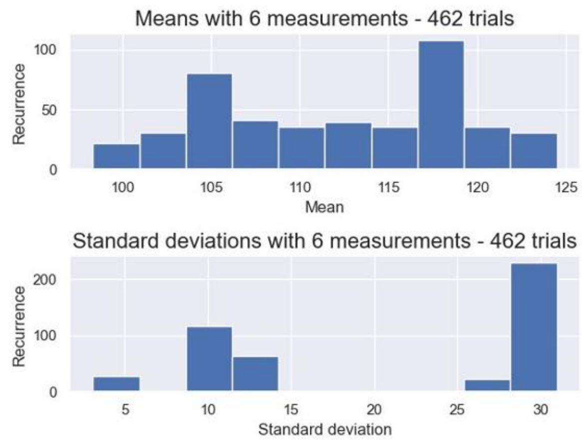

Fig. 11: Histograms of means and standard deviations using 6 samples (MAT 3)

The second reason for this to happen is the pre-existing differences inside this group before the experiments. The balls in this alloy group differ greatly in size and therefore initial mass and surface area and as previously mentioned, even after normalizing the weight losses, there is still a noticeable correlation between this parameter and ball mass and surface area (Gates *et al*., 2008). This result is a good way to show that it is necessary to set the basis for a standardized way to select the alloys and samples to be included in BMAT. As shown in previous cases, it is only necessary to carefully choose 6 samples for a given alloy to get accurate results. By doing this, it is possible to reduce the number of samples normally included in a wear test in BMAT and potentially accommodate more alloy types if needed.

The results from combining 11 samples resulted in a mean of 112.1 and a standard deviation of 24.0, which represents a large percent of the alloy's weight loss. This again results from the initial differences present before the experiment.

Optimizing the 95% Confidence Interval at 6 Measurements

Given that the mean values of mass loss at 6 measurements are spread over a wide range of values, it is not appropriate to conduct an optimization step for this material. Doing so would mean arbitrarily choosing a certain combination of 6 samples with a low standard deviation, without knowing whether this chosen combination is representative of the entire alloy group.

Conclusion

The methodology used in this study has shown that the optimum alloy sample size for Ball Mill Abrasion Tests is six. It was demonstrated that in BMAT it is not meritorious to include an abundant number of samples per material type if significant differences in size are present within it. Pre-existing size differences lead to highly deviated results, even if these are normalized. By using six samples per alloy, with similar initial sizes, it is possible to obtain accurate results while keeping the sample size relatively small. These results show the importance of setting the basis for a standardized way of alloy and sample selection for BMAT, which could significantly optimize the obtention of data from experimental procedures, critical to the mining industry. One of the limitations of this study is that only experiments run at 25% of the mill speed were considered. These findings can be further strengthened by considering more material types and experimental conditions, as those generally used in BMAT.

Acknowledgment

The authors would like to thank the support from the Institute of Basic Sciences-UTM in the production of this material.

Author’s Contributions

Carlos Daniel García Mendoza: Participation in experiments, data collection, and contribution to the writing of the manuscript.

Lelly María Useche Castro: Data analysis planification, the proposal of theoretical data processing method, and contribution to the revision of the manuscript.

Miguel Ángel Lapo Palacios: Proposal of Python script to be used for data analysis. Contribution to the design of data processing script.

Óscar Vargas Ortiz: Contribution to the revision of theoretical data analysis, the structure of the manuscript, and general organization of the data presented.

Ronny Javier Maldonado Macías: Contribution to the writing of the manuscript.

Ethics

This article is genuine and contains unpublished material. All authors have read it thoroughly and approved that it does not contain any material which is already published.

References

Ahsanullah, M., Kibria, B. M., & Shakil, M. (2014). Normal distribution. In *Normal and Student st Distributions and Their Applications* (pp. 7-50). Atlantis Press, Paris. https://link.springer.com/chapter/10.2991/978-94-6239-061-4_2

Albertin, E., & Sinatora, A. (2001). Effect of carbide fraction and matrix microstructure on the wear of cast iron balls tested in a laboratory ball mill. *Wear*, *250*(1-12), 492-501. https://doi.org/10.1016/S0043-1648(01)00664-0

Ali, Y., Garcia-Mendoza, C. D., & Gates, J. D. (2019). Effects of ‘impact’and abrasive particle size on the performance of white cast irons relative to low-alloy steels in laboratory ball mills. *Wear*, *426*, 83-100. https://doi.org/10.1016/j.wear.2019.01.048

Antony, J. (2014). *Design of experiments for engineers and scientists*. Elsevier.

Awe, O., Love, K., & Vance, E. (2022). Promoting statistical practice and collaboration in developing countries. Taylor and Francis Limited.

Brereton, R. G. (2015). The t-distribution and its relationship to the normal distribution. *Journal of Chemometrics*, *29*(9), 481-483. https://doi.org/10.1002/cem.2713

Bujang, M. A., & Baharum, N. (2016). Sample size guideline for correlation analysis. *World*, *3*(1), 37-46.

Chenje, T. W., Simbi, D. J., & Navara, E. (2004). Relationship between microstructure, hardness, impact toughness, and wear performance of selected grinding media for mineral ore milling operations. *Materials & Design*, *25*(1), 11-18. https://doi.org/10.1016/S0261-3069(03)00168-7

Gates, J. D., Bennet, P. J., McInnes, L. J., & Tunstall, B. R. (2018). The challenge of accurate prediction of industrial wear performance from laboratory tests. In *International Symposium on Wear Resistant Alloys for the Mining and Processing Industry*.

Gates, J. D., Dargusch, M. S., Walsh, J. J., Field, S. L., Hermand, M. P., Delaup, B. G., & Saad, J. R. (2008). Effect of abrasive mineral on alloy performance in the ball mill abrasion test. *Wear*, *265*(5-6), 865-870. https://doi.org/10.1016/j.wear.2008.01.008

Ibe, O. C. (2013). Basic concepts in probability. In Markov Processes for Stochastic Modeling, pages 1–27. Elsevier, Oxford.

Jankovic, A., Wills, T., & Dikmen, S. (2016). A comparison of wear rates of ball mill grinding media. *Journal of Mining and Metallurgy A: Mining*, *52*(1), 1-10. https://scindeks.ceon.rs/article.aspx?artid=1450-59591601001J

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1), 33267. https://doi.org/10.1525/collabra.33267

Massola, C. P., Chaves, A. P., & Albertin, E. (2016). A discussion on the measurement of grinding media wear. *Journal of Materials Research and Technology*, *5*(3), 282-288. https://doi.org/10.1016/j.jmrt.2015.12.003

Miot, H. A. (2011). Sample size in clinical and experimental trials. *Jornal Vascular Brasileiro*, *10*, 275-278. https://doi.org/10.1590/S1677-54492011000400001

Ross, S. M. (2017). *Introductory statistics*. Academic Press.

Van de Schoot, R., & Miocević, M. (2020). *Small sample size solutions: A guide for applied researchers and practitioners* (p. 284). Taylor & Francis. https://library.oapen.org/handle/20.500.12657/22385

Verma, J. P., & Verma, P. (2020). Determining Sample Size in Experimental Studies. In *Determining Sample Size and Power in Research Studies* (pp. 61-88). Springer, Singapore.

Appendix

This section shows the complete evolution of the histograms of means and standard deviations as the number of measurements is increased for MAT 1.

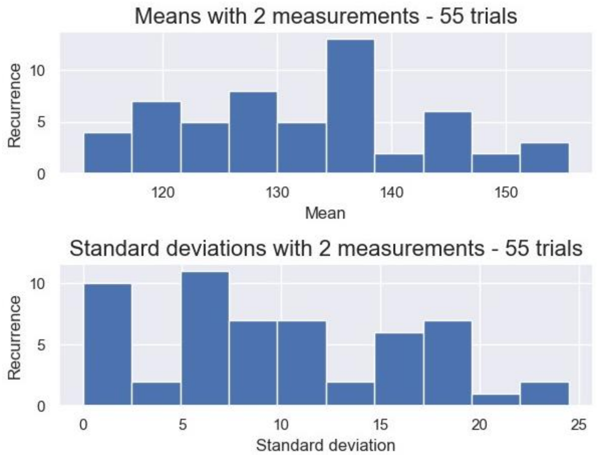

Fig. 12: Histograms of means and standard deviations using 2 measurements

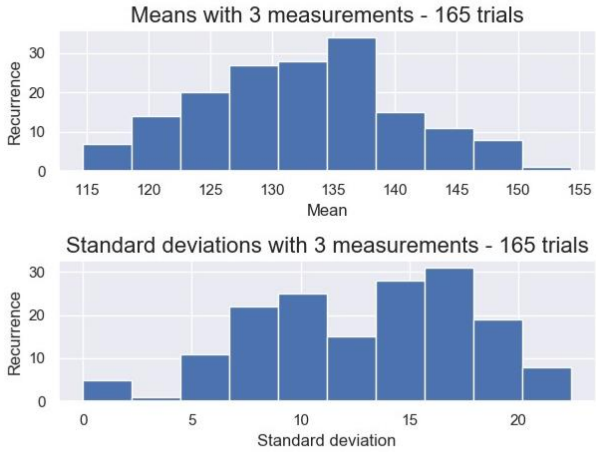

Fig. 13: Histograms of means and standard deviations using 3 measurements

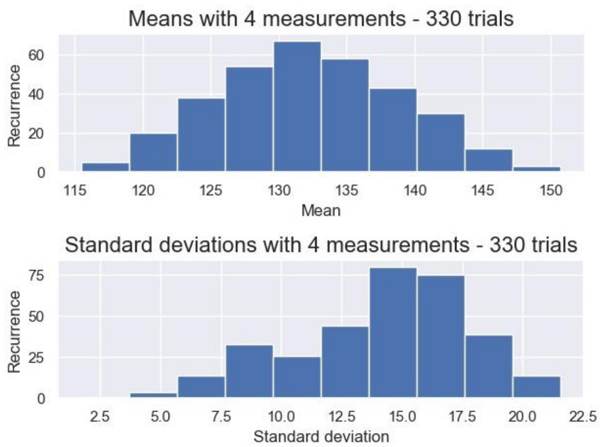

Fig. 14: Histograms of means and standard deviations using 4 measurements

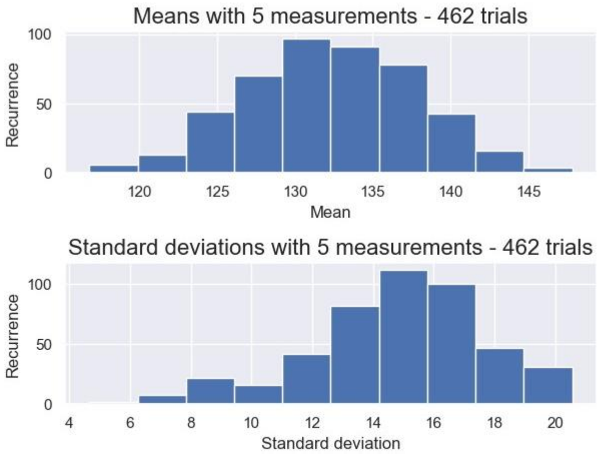

Fig. 15: Histograms of means and standard deviations using 5 measurements

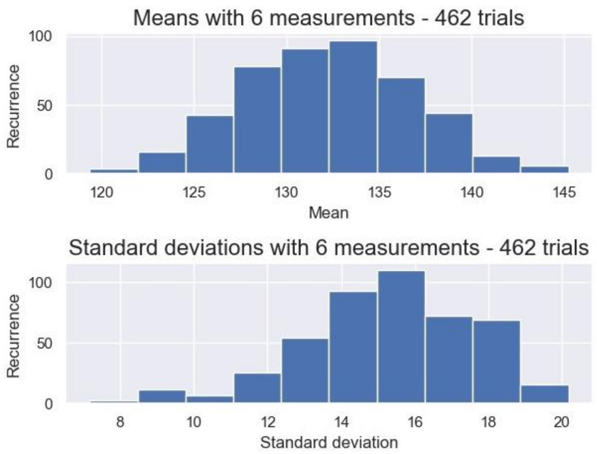

Fig. 16: Histograms of means and standard deviations using 6 measurements

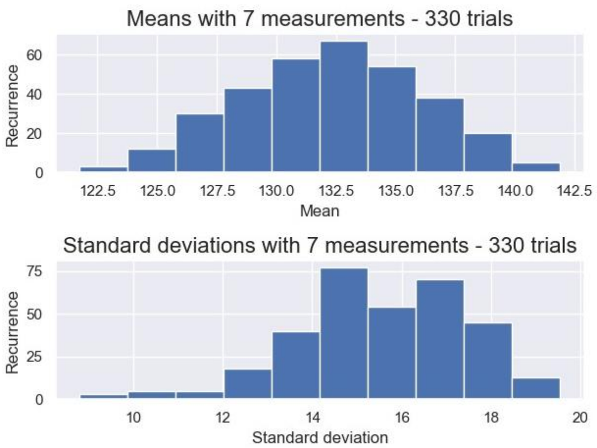

Fig. 17: Histograms of means and standard deviations using 7 measurements

Fig. 18 (a): Histograms of means and standard deviations using 8 measurements

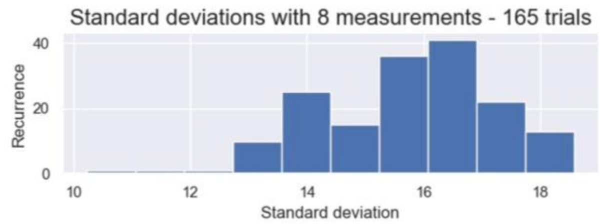

Fig. 18 (b): Histograms of means and standard deviations using 8 measurements

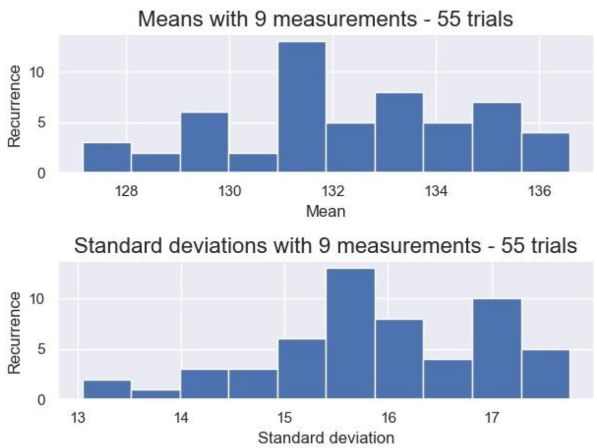

Fig. 19: Histograms of means and standard deviations using 9 measurements

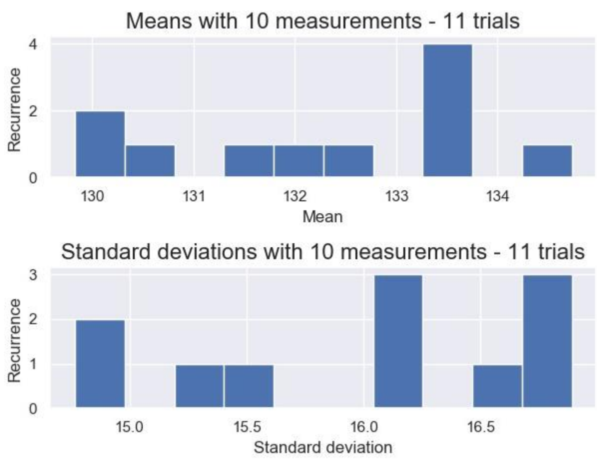

Fig. 20: Histograms of means and standard deviations using 10 measurements