Sample Size Determination to Estimate the Number of Units Shipped by a Foundry

(*Communicated to* MIIR *on 12 June 2023*)

Study Group: IX Brazilian Study Group with Industry of the Center for Mathematical Sciences Applied to Industry – CEPID-CeMEAI, February 06–10, 2023.
Communicated by: Francisco Louzada Neto and José Alberto Cuminato
Industrial Partner: Tecumseh do Brasil Ltda (https://www.tecumseh.com/)
Presenter: Gustavo Mazzei and Jean Menezes
Team Members: D.F. Silva, Institute of Mathematical Sciences and Computing, University of São Paulo, São Carlos, Brazil; E.S. Brito, Federal Institute of Education, Science and Technology of Goiás, University of São Paulo and Federal University of São Carlos; F.Louzada, Institute of Mathematical Sciences and Computing, University of São Paulo, São Carlos, Brazil; J.S. Leão, Federal University of Amazonas, University of São Paulo and Federal University of São Carlos; M.S.Lauretto, School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, Brazil; P. Moraes, Federal University of Paraná; P.H.B. Ribeiro, University of São Paulo and Federal University of São Carlos; V.L.Tomazella, Department of Statistics, Federal University of São Carlos, São Carlos, Brazil.

Industrial Sector: Data Analysis; Manufacturing.

Tools: Statistical sampling, quality control.

Key Words: Sample size; quantity estimation; quality control.

MSC2020 Codes: 62D05, 62P30.

Summary

The foundry industry involves pouring liquid metal or metal alloy into molds to create parts of specific shapes and measurements. The Tecumseh foundry is a notable player in this field, producing roughly 42,000 tons of metal each year. However, like many foundries in Brazil, Tecumseh faces challenges in accurately converting its productive volume into the number of pieces produced. Without a reliable and effective system for this conversion, the only option is manual counting. To address this issue, the group sought to estimate the number of pieces to be sampled while minimizing errors.

https://doi.org/10.33774/miir-2023-s27lz **ORCID:** https://orcid.org/0000-0002-6780-2089

1 Introduction

A foundry is a factory that produces metal castings, typically by melting metal and pouring it into molds to solidify. The casting process requires specialized equipment, such as furnaces, molding machines, and finishing tools. Foundries are used to manufacture a wide range of metal products, including parts for automobiles, machinery, and building structures. The choice of metal and casting method depends on factors such as the desired properties of the final product, the production volume required, and the specific application of the product. The casting process is highly specialized, and the final product's quality is influenced by various factors, including the casting method, metal selection and finishing processes.

Figure 1. Image of a foundry process.

The process for casting and sampling parts in a factory can vary depending on the type of product and casting process used, but in general, the flow is as follows:

- **Design and Modeling:** The design of the part to be cast is developed and a model or mold is created for the part;
- **Metal preparation:** The metal to be used in casting is prepared, including casting, bonding and mixing different types of metal if necessary;
- **Mold preparation:** The mold is primed with non-stick coating and is placed in position for casting;
- **Casting:** The metal is melted and poured into the mold, where it is allowed to solidify and harden;
- **Demoulding:** Once solidified, the mold is opened and the part is removed.
- Cleaning and finishing: The part is cleaned of mold residues and undergoes finishing work, such as grinding, grinding, cutting, folding and welding, if necessary;
- **Inspection:** The part is inspected to ensure it meets quality standards and specifications;
- **Sampling:** A representative sample of produced parts is selected for further inspection and testing;
- **Storage and distribution:** Parts that meet specifications are stored or distributed for use or sale.

https://doi.org/10.33774/miir-2023-s27lz **ORCID:** https://orcid.org/0000-0002-6780-2089

By the processes described, it is reasonable to expect that the parts produced may present variations in their weight that may occur as a result of the materials used to create the modes or even external factors such as temperature and air humidity. In this sense, any decision made based on the weight of the pieces produced will be subject to errors, therefore, it is necessary to analyze these variations and define effective strategies to deal with this type of problem.

1.1 Problem description

The problem presented in this work is located in the last two stages of the production process presented in the begging of the introduction, that is, in the stages of sampling and distribution of the produced parts. Currently, in the factory in question, these stages are manual and two scales are used: precision parts scale and box weight scale.

More specifically, these stages have the following steps as a process flow: Every empty package is weighed and its tare recorded (Figure 2-a); of the entire batch, 10 pieces are weighed and their average weight is recorded (Figure 2-b); and after the parts are packed inside the boxes, they go to the scale to perform the weighing (Figure 2-c).

(a) (b) (c)

Figure 2. Parts sampling and packaging process.

To estimate the number of pieces packed, the final weight of the box is utilized. Based on the average weight of a sample of 10 units information, the expected weight of a certain quantity of parts to be packed is determined, and the number of units that achieve this weight is packed accordingly. This approach is preferred because it is a more straightforward and cost-effective alternative to counting all the packed parts, particularly when dealing with large quantities.

However, it should be noted that this sampling process and weight-to-pieces conversion method are susceptible to errors, which can cause the estimated number of pieces to deviate from the actual number packed. As a result, the company may occasionally deliver a different number of parts than what was ordered, either more or less. Unfortunately, this problem is not unique to this particular foundry in Brazil, as there is currently no reliable or effective system for conversion other than manual counting.

The main challenge facing the company is to develop a reliable system for converting production volume (in Kg) into the number of pieces packed. One of the contributing factors to this issue is the lack of statistical criteria in the current sampling process, where 10 pieces are sampled regardless of their weight or type. Our proposed solution is to leverage well-established statistical principles to develop more accurate sampling methods, ultimately reducing packaging errors to acceptable levels for both the company and its customers. Furthermore, we suggest implementing quality control charts to enable operators to track historical averages of piece weights and detect potential anomalies.

To guide the resolution of this problem, a database comprising 2,000 records of average weights for 60 different types of parts produced over a one-year period was provided by the company, along with the corresponding weighing dates. This database has been continually updated over time and contains additional information, such as batch size and the number of parts per box, that were not incorporated into our current modeling efforts.

1.2 Theoretical Support

Sample size determination is the process of determining the number of observations or participants that should be included in a sample taken from a larger population. The sample size is an important factor in statistical analysis because it affects the accuracy and precision of the results. A larger sample size generally provides more information and leads to more precise estimates, but it also requires more resources and time to collect the data.

There are several methods for determining sample size, including:

(1) **Power analysis:** This method uses statistical analysis to determine the sample size needed to detect a specified effect size with a certain degree of confidence.

(2) **Margin of error:** This method uses the desired margin of error and the variability of the population to calculate the sample size needed.

(3) **Confidence intervals:** This method uses the desired level of confidence and the variability of the population to calculate the sample size needed.

(4) **Precision:** This method uses the desired level of precision and the variability of the population to calculate the sample size needed.

Ultimately, the appropriate sample size will depend on several factors, including the research question, the characteristics of the population, the desired level of precision, and the resources available for data collection.

1.3 Contributions

Tecumseh Company needs to develop new strategies for calculating the number of parts to be packaged. To achieve this, we propose a statistical method based on the company's database, which includes information on the average weight of samples of parts. Using established concepts and results from the statistical literature, we calculate the sample size and establish a relationship between the weight of the pieces produced and the number of pieces to be packaged. Additionally, we suggest creating quality control charts to monitor the historical behavior of the average weight of parts and detect batches with high weight variation.

Our contribution is significant as we establish criteria based on theories that allow interested parties to define maximum levels of tolerable error in the parts packing process and monitor the average weight of produced parts over time. Our proposed methods enable the calculation of the sample size based on a given tolerable error and the quick identification of anomalies in the production process.

1.4 Organization

This technical report is organized as follows. Section 2 describes the problem formulation and the methods for sample size determination. In Section 3, we selected three parts in the database to illustrate the application of the proposed methods, estimating the sample sizes needed to satisfy certain levels of significance and maximum tolerable error. Furthermore, in Section 3, we present a prototype app as a minimum viable product (MVP) for the Tecumseh company. In Section 4, we present the general comments on the work developed and reinforce some future proposals. Finally, in Appendix A we provide a tutorial in Brazilian Portuguese for using the developed MVP, which will be delivered to Tecumseh company employees.

2 Methods

In this section, we present the methodology for the sample size determination. More specifically, we give the methodology to determine how many units of a specific item must be picked from a production batch and weighed, such that estimates of the total number of units in a package is close enough of the specification. First we state the formal description of the problem; then, we present the method for sample size computation; finally, we present the estimation of parameters and error specification.

2.1 Problem formulation and notation

As previously mentioned in Section 1, the company needs to fill a container with N_i units of an item i. The challenge is that N_i is usually large (on the order of hundreds or thousands) and the company does not have a reliable and cost-effective system to individually count each unit in the container. In other words, it is not possible to guarantee that the container contains exactly N_i units.

To overcome this issue, the current procedure adopted by the company consists of estimating the *expected net weight* of the container if it had N_i units, here denoted NW_i. For this, a sample of n_i units is drawn from the current production batch and its average weight per unit is measured using a precision scale. Based on this average weight (here denoted AW_i), the best estimate for NW_i is $NW_i = N_i\, AW_i$. So, the procedure is to fill the container until its net weight reaches NW_i.

Due to the variability among units weights, this procedure does not guarantee a perfect filling of the container. More specifically, the estimated number of units placed in the container yielded by this procedure,

$$\hat{N}_i = NW_i \,/\, AW_i \ , \tag{2.1}$$

https://doi.org/10.33774/miir-2023-s27lz **ORCID:** https://orcid.org/0000-0002-6780-2089

may be different of the required quantity, N_i. Although this difference is intrinsic to the sample weight estimation procedure and cannot be completely eliminated, its magnitude can be controlled by the casting process (in order to reduce the variability between unit weights) or by an adequate specification of the sample size to calculate AW_i . The focus of this work is in the last aspect.

2.2 Sample size specification

The central problem is to determine the sample size n_i such that, with high probability, the error in estimating the number of units placed in the box ($\hat{N}_i$) is within a maximum tolerated limit, e.

We denote W_i the individual weight of any unit of item i and TW_i the total net weight of a container with N_i units of item i. We also assume that W_i follows a normal distribution with mean μ_i and standard deviation σ_i. Hence, TW_i follows a normal distribution with mean and standard deviations [2, Theorem 5.6.4]:

$$\mu_{TW_i} = N_i \, \mu_i \tag{2.2}$$

$$\sigma_{TW_i} = N_i \, \sigma_i \tag{2.3}$$

Under the assumptions above, the sampling size n_i is computed by [1, Section 8.2]:

$$n_i = \left(\frac{z_\gamma \; \sigma_{TW_i}}{e} \right)^2 , \tag{2.4}$$

where z_γ is the quantile of the standard normal distribution with confidence level γ and e is the maximum tolerated error, that is, the maximum absolute difference between the *actual* and *required* number of items placed in the container.

The interpretation of Equation (2.4) is as follows: suppose $\gamma = 0.95$ and $e = 0.05$. Then, n_i is the sample size necessary so that, with a probability of 95%, the filling error does not exceed 5% of the total required units. Or, in other words, the probability of filling a container with an error greater than 5% of the required total is 5%. These examples are merely illustrative; the maximum tolerated error and confidence level can be customized according to the item, customer, etc.

2.3 Parameters estimation and error specification

Equations (2.2), (2.3) and (2.4) depend on μ_i and σ_i, which are unknown. These parameters were estimated based on a database of samples previously collected by the company, for various items. Each record contains the average weight of a sample of 10 units. To provide more robustness for the estimates, only items with at leas 20 samples were considered.

Let k_i be the number of samples previously collected for item i, and $m_{i,j}$ be the average weight of item i measured in sample j ($j = 1, \ldots, k_i$). The parameters μ_i and σ_i

are estimated by:

$$\hat{\mu}_i = \frac{1}{k_i}\sum_{j=1}^{k_i} m_{i,j} \quad (2.5)$$

$$s_i = \sqrt{\frac{1}{k_i - 1}\sum_{j=1}^{k_i}(m_{i,j} - \hat{\mu}_i)^2} \quad (2.6)$$

$$\hat{\sigma}_i = \sqrt{10}\ s_i \quad (2.7)$$

Regarding the tolerable error e, it is important mentioning that Equation (2.4) concerns the total weight of the container, not the number of units placed in it. So, the tolerable error cannot be specified in terms of the number of total units. The solution is using a surrogate specification for e in terms of the error rate in the total container's weight if it was filled with exactly N_i items:

$$e = \alpha\ N_i\ \hat{\mu}_i, \quad (2.8)$$

where $\alpha \in (0, 1)$ is the maximum tolerable error.

2.4 Quality Control

As the sample size estimation proposal is based on the history of sample averages collected by the company, this is subject to errors that may result from the variation in weight inherent to the production of parts or sampling errors. To alert company operators about atypical cases, we propose the use of quality control charts.

Quality control charts are graphical tools used to monitor and analyze a process or product quality over time. These charts display data points plotted against control limits, which are calculated based on statistical analysis of historical data.

The purpose of quality control charts is to detect any variation or deviation from the expected process or product quality, and to take corrective actions before the situation becomes critical. In that sense, they are an important part of quality management and can help improve process efficiency, customer satisfaction and company profitability.

To construct the quality control chart, we propose to use the mean and standard deviation of the average measurements. The sample means should be plotted on a chart over time, and the control limits should be drown based on the mean and standard deviation. The upper and lower control limits are set at three standard deviations from the mean, and represent the acceptable range of variation for the process (more details in [3]). The equations for determining these values are given below:

$$\begin{aligned} CL_i &= \hat{\mu}_i \\ UCL_i &= \hat{\mu}_i + 3s_i/\sqrt{10} \\ LCL_i &= \hat{\mu}_i - 3s_i/\sqrt{10}, \end{aligned} \quad (2.9)$$

where, for the $i-$th part, CL_i is the *center line* that represents the average value of the quality characteristic, UCL_i is the *upper control limit*, LCL_i is the *lower control limit* and $\hat{\mu}_i$ and s_i are given respectively by the equations (2.5) and (2.6).

https://doi.org/10.33774/miir-2023-s27lz **ORCID:** https://orcid.org/0000-0002-6780-2089 Content not peer-reviewed by MIIR. **License:** CC BY-NC-ND 4.0

Table 1. Parameter estimates and number of observations for each Part.

Part i	k_i	μ_i	σ_i
Part 1	222	2163.62	52.48
Part 2	118	561.72	9.98
Part 3	32	109.50	4.61

3 Results and Discussion

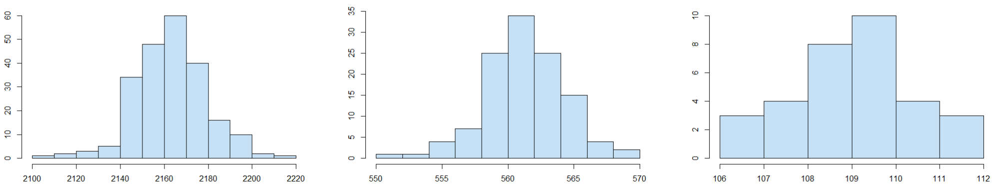
In this section we present some discussions about the results obtained in Section 2 and use the data provided by Tecumseh company to illustrate the procedures presented.

In the original database there is information on sample means referring to 60 different types of pieces, however, only 30 of them were used in our application, since there are few records of sample means for the other pieces. As previously mentioned, the reference parameter used to decide the types of pieces used in the application is to have 20 or more observations of sample averages recorded.

In addition, in the 30 types of parts with 20 or more records, normality tests and other statistical checks were performed to ensure that these samples met the theoretical assumptions shown in Section 2.

3.1 Sample size Estimation

To exemplify the estimation of the sample size, we selected three types of parts with different magnitudes and sample mean sizes, which will be denoted simply by Part 1, Part 2 and Part 3. Figure 1 below shows the histogram of the sample means for each of these pieces.



(a) Part 1 (b) Part 2 (c) Part 3

Figure 3. Histograms of average weights recorded for the three chosen parts.

Table 1 presents summary information about these three selected pieces. The table shows the number of samples previously recorded (k_i) for each of the three types of parts. In addition, the mean of the recorded means (μ_i) and their respective standard deviation (σ_i), calculated by the expressions (2.5) and (2.7), respectively, are also displayed.

The information in Table 1 is essential for estimating the sample size of parts needed for weighing. But in addition, there are two variable pieces of information depending on

https://doi.org/10.33774/miir-2023-s27lz **ORCID:** https://orcid.org/0000-0002-6780-2089 Content not peer-reviewed by MIIR.

Table 2. Parameter estimates and number of observations for each Part.

Part i	N_i	α	γ	n_i
Part 1	$N_1 = 250$	$\alpha = 1.6\%$	$\gamma = 0.95$	$n_1 = 9$ parts
	$N_1 = 500$	$\alpha = 1\%$	$\gamma = 0.90$	$n_1 = 16$ parts
	$N_1 = 1000$	$\alpha = 1\%$	$\gamma = 0.95$	$n_1 = 23$ parts
Part 2	$N_2 = 1000$	$\alpha = 1\%$	$\gamma = 0.95$	$n_2 = 12$ parts
	$N_2 = 2000$	$\alpha = 0.5\%$	$\gamma = 0.90$	$n_2 = 34$ parts
	$N_2 = 2000$	$\alpha = 1\%$	$\gamma = 0.90$	$n_2 = 9$ parts
Part 3	$N_3 = 1300$	$\alpha = 1\%$	$\gamma = 0.90$	$n_3 = 48$ parts
	$N_3 = 1300$	$\alpha = 2\%$	$\gamma = 0.90$	$n_3 = 12$ parts
	$N_3 = 1300$	$\alpha = 2\%$	$\gamma = 0.95$	$n_3 = 17$ parts

the situation desired by the company's operators: the number of pieces to be boxed and the maximum acceptable error.

For example, suppose that for Part it is desired to pack a batch of one thousand parts ($N_1 = 1000$) with a maximum error rate $\alpha = 1\%$, at a significance level of 10% ($\gamma = 90\%$). So, the mean and standard deviation for the total net weight of a container (TW_i) are respectively calculated by the expressions (2.2) and (2.3) and given by

$$\mu_{TW_1} = N_1\,\mu_1 = 1000 \times 2163.62 = 2163620$$
$$\sigma_{TW_1} = N_1\,\sigma_1 = 1000 \times 52.46 = 52460,$$

and the maximum tolerated error e is calculated by the expression (2.8) and given by

$$e = \alpha\ N_1\ \hat{\mu}_1 = 0.01 \times 1000 \times 2163.62 = 21636.2.$$

Then, the desired sample size is finally calculated by the expression (2.4), so that

$$n_i = \left(\frac{z_{0.90}\ \sigma_{TW_1}}{e}\right)^2 = \left(\frac{1.64 \times 52460}{21636.2}\right)^2 \approx 16,$$

and therefore, the weighed sample will consist of 16 units.

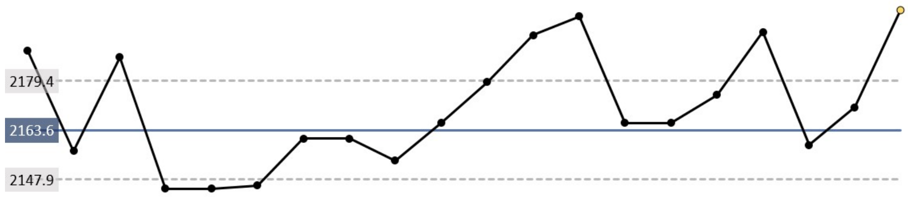
In Table 2 we show some sample size results obtained for weighing the three types of parts selected, varying the size of the batch to be packaged and the maximum tolerable error rate.

3.2 Quality Control Charts

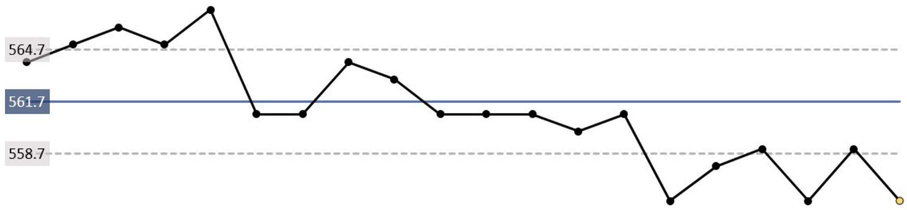
As said before, the proposed sample size estimation proposal is subject to errors and it is necessary to realize a historical follow-up to capture these errors and verify whether they are due to the production process or the process of sampling parts.

Complementing the sampling process described in the previous section, the proposed monitoring idea is as follows: If a new weight measurement falls within the upper and lower control limits, no alert is generated and the company operator can continue with the parts packing process. However, if a new average weight falls outside of these limits, an alert is generated for the operator. In this case, it is recommended that the operator take another sample and obtain a new average weight measurement. If this new measurement falls within the control limits, the operator can continue with the packing process. If the new measurement is still outside of the control limits, this suggests that there may be an issue with the weight of parts being produced in this batch, and further investigation may be necessary.

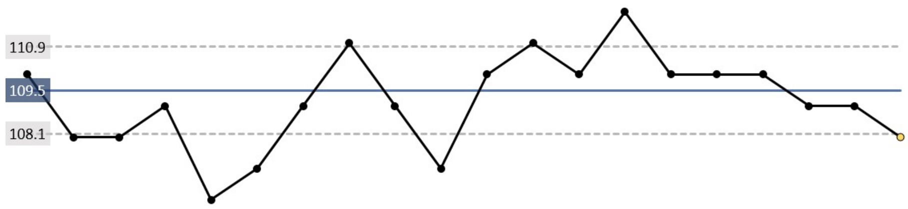
Based on historical weight averages, the information shown in Table 1, the ideas previously discussed and the equations in (2.9), Figure 4 below shows the quality control charts for the three parts chosen as examples, considering the last 20 samples observed.



(a) Part 1



(b) Part 2



(c) Part 3

Figure 4. Quality control charts of the three chosen parts.

3.3 Minimum Viable Product - An app to predict reliability

A Minimum Viable Product (MVP) was implemented in Excel®, and incorporates the methods described in this work. It allows the user to select the desired item, the required number of units to be packed, the significance level and and the tolerated error percent-

https://doi.org/10.33774/miir-2023-s27lz **ORCID:** https://orcid.org/0000-0002-6780-2089

age. Based on these parameters, the program returns the sample size and the quality control chart.

In Figure 5 below, we have a print of the calculator application's initial screen. Note that the initial screen is composed of two regions, where region A returns the sample size of parts to be weighed based on the desired parameters entered by the user, while region B returns the quality control chart for the type of part indicated. In addition, the application tabs are indicated in C, namely the database, the registration of new sample weight averages and the calculator on the initial screen.

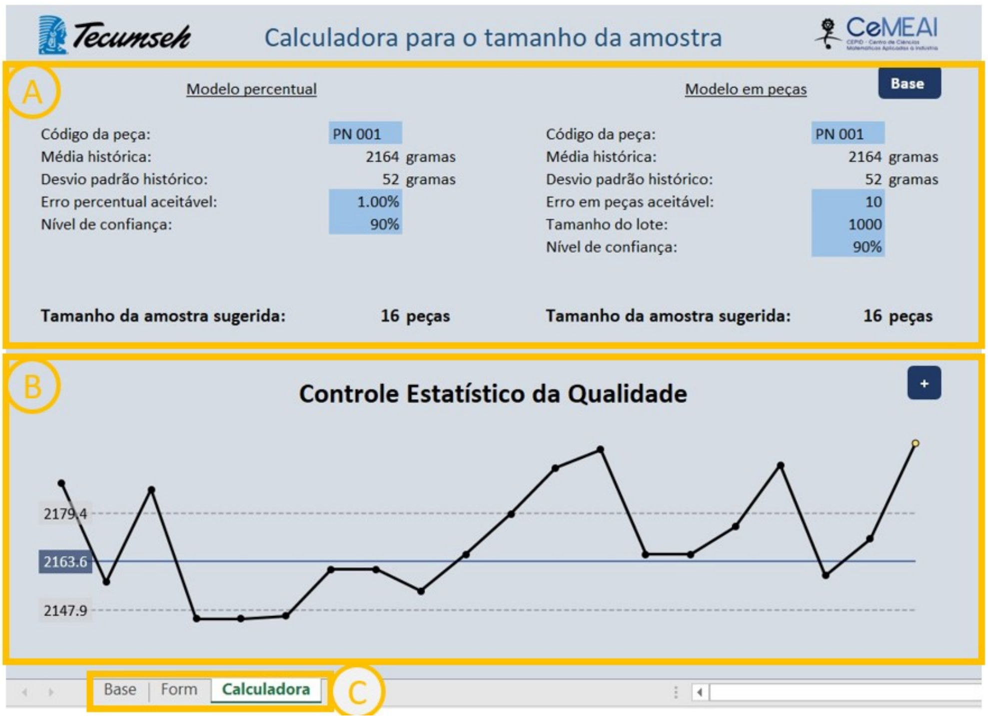


Figure 5. Screenshot of the calculator's home screen.

In Figure 6 we present some highlights for the sample calculator region. As previously mentioned, obtaining the sample size depends directly on the maximum tolerated error defined by the company's operators. Thus, we propose two different ways to obtain the sample size: in region A, the user can directly indicate the maximum percentage error tolerated, while, in region B, the user can inform the size of the lot to be packed and the maximum number of parts count error. For both situations, highlight 1 indicates the inputs that will be made by the user, namely the part code, the confidence level and the tolerated error (in A) or the batch size and the tolerated error of the number of pieces packed (in B). Finally, highlight 2 indicates the return of the implemented operation, that is, it returns to the user the size of the sample that should be taken from the lot to calculate the average of the weights.

Figure 7 shows the quality control chart for the selected part and is automatically generated when the user selects the part type from the calculator region (Figure 6,

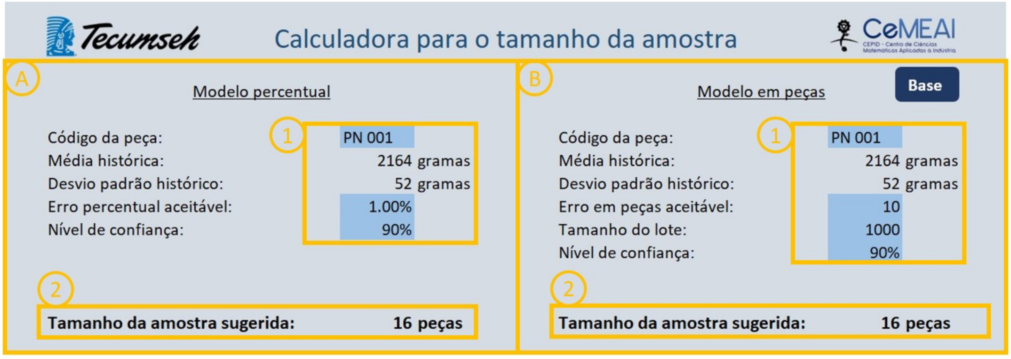


Figure 6. Screenshot of sample size calculator region.

highlight 1). When taking a new sample of parts from a batch and obtaining the weight average, the user can record this new information by accessing the tab for *recording new averages* using the button highlighted in A. The new record will be automatically recorded in the database tab and the new point will be included in the quality control chart with a different color from the others, as highlighted in B.

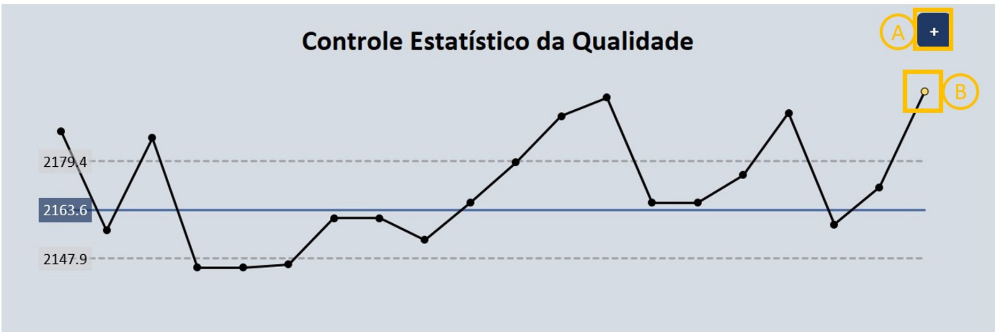


Figure 7. Screenshot of quality control charts region

Finally, the tab for recording new averages can be accessed from the home screen (Figure 5, highlight C) or via the previously mentioned button in the quality control region (Figure 7, highlight A). Figure 8 below shows this tab, where the user can select the type of part he is handling and register the weight obtained from the sample, the weight of the package that will be used and the number of parts that must be packed. As previously mentioned, when saving this information, a new line is added to the database and this information can be used for future analysis.

Despite being an initial and simplified version, the proposed calculator application is already a great advance for immediate use by the company, since currently there is no statistical sampling criterion being used. In this sense, the expectation is that errors in the number of parts packed tend to decrease with the help of the proposed calculator. In addition, quality control charts can also help company operators to monitor the homogeneity of produced parts and quickly identify possible anomalies that may

https://doi.org/10.33774/miir-2023-s27lz **ORCID:** https://orcid.org/0000-0002-6780-2089

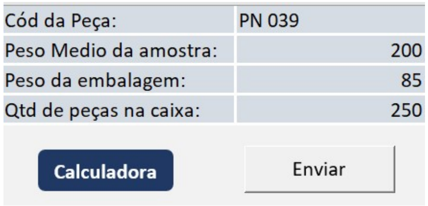


Figure 8. Screenshot of the tab for recording new averages.

occur. In this way, this application can be a first step towards reducing the company's losses and increasing the safety of customers when purchasing their products.

3.4 Future work

There are numerous work possibilities that can be explored based on the preliminary findings presented thus far. It is important to note that these results are based on the data provided by Tecumseh during a one-year period and represent our first interaction with the company.

Moving forward, a range of new possibilities can be explored once the company implements our proposed sampling methodology. By collecting new types of data, we can refine and evaluate the methodology to further improve its effectiveness. Additionally, incorporating new sample data into the company's history could enhance quality control monitoring by analyzing the moving average of sample weights and variations between sequential samples.

We have also developed a minimum viable product (MVP) for calculating sample sizes based on the data collected thus far. Improving this application should be a key focus for future work, with an emphasis on automating data collection through systems that connect to the company's scales. This would enable the automatic updating of the database, quality control charts, and estimates, resulting in a more streamlined process.

4 Conclusion

In this study, we presented an initial proposal for solving a casting problem faced by a Brazilian company. The challenge was to develop a sampling methodology that could reduce the percentage error in the number of packed parts to an acceptable level.

Our proposed model was built on established theories of statistical inference and sampling, relating the total number of parts to be packaged to the expected total weight. We developed an estimation technique based on small samples taken from the batch in question, using the sample mean to calculate the total number of pieces.

Additionally, we proposed the construction of quality control charts using the sample mean and variance information to help operators identify sampling errors or anomalies in the production process. Finally, we created an MVP that automatically calculates the number of pieces to be packed based on the average weight of a batch sample and inserts the new measurements in the quality control charts to detect high variations.

We suggested some improvements to the data recording process, such as automating data collection through computational integration between the scales. The developed

https://doi.org/10.33774/miir-2023-s27lz **ORCID:** https://orcid.org/0000-0002-6780-2089 Content not peer-reviewed by MIIR. **License:** CC BY-NC-ND 4.0

model is a potential solution for improving the sampling methods adopted by the company and will aid in evaluating new results and proposals, as outlined in Section 3.4.

The methodologies and MVP proposed in this study have significant potential for application and respond to the company's initial demand. We hope that implementing these solutions will help reduce losses and improve the company's relationship with its customers.

Acknowledgements

This research was supported by CEPID-CeMEAI – Center for Mathematical Sciences Applied to Industry (grant 2013/07375-0, São Paulo Research Foundation–FAPESP).

References

[1] Bluman, A. G. *Elementary statistics: A step by step approach*, 3 ed. McGraw-Hill, Boston, 1997.

[2] DeGroot, M. H., and Schervish, M. J. *Probability and statistics*, 4 ed. Addison-Wesley, Boston, 2012.

[3] Montgomery, D. C. *Introduction to statistical quality control.* John Wiley & Sons, 2020.

Appendix A User's Guide to the MVP

In this Appendix we present a tutorial on the use of the developed MVP, to be delivered to Tecumseh company employees. The MVP was called "Calculadora de Amostras" (Sample Calculator) and the tutorial was written in Brazilian Portuguese so that its use and understanding are facilitated among the company's operators.

9º Workshop de Soluções Matemáticas para Problemas Industriais

Tutorial - Calculadora de Amostras v1.0

Francisco Louzada Neto[1]
Vera Lúcia Damasceno Tomazella[2]
Marcelo de Souza Lauretto[3]
Diego Furtado Silva[1]
Jeremias Leão[1,4]
Éder Silva de Brito[1,2,5]
Paulo Henrique Brasil Ribeiro[1,2]
Pablo Moraes[6]

[1]Instituto de Ciências Matemáticas e Computação - Universidade de São Paulo (ICMC-USP)
[2]Departamento de Estatística - Universidade Federal de São Carlos (DEs-UFSCar)
[3]Escola de Artes, Ciências e Humanidades - Universidade de São Paulo (EACH-USP)
[4]Departamento de Estatística - Universidade Federal do Amazonas (DE-UFAM)
[5]Instituto Federal de Educação, Ciência e Tecnologia de Goiás - Campus Anápolis (IFG)
[6]Departamento de Matemática - Universidade Federal do Paraná (DM-UFPR)

6 a 10 de fevereiro de 2023

https://doi.org/10.33774/miir-2023-s27lz ORCID: https://orcid.org/0000-0002-6780-2089 Content not peer-reviewed by MIIR.

Apresentação

Apresentamos ao usuário da *Calculadora de Amostras* um tutorial de uso da versão *v1.0* da ferramenta. Esse tutorial está dividido em quatro subseções: na primeira apresentamos cada uma das três telas iniciais que compõem a ferramenta e nas demais detalhamos o uso e funcionalidade de cada uma delas.

Ao seguir este tutorial, o usuário fará um *tour* guiado pelas telas disponíveis na ferramenta, onde poderá conhecer suas funcionalidades e a sequência de como utilizá-las. As imagens são capturas de um uso real da ferramenta, o que possibilita ao usuário visualizar os resultados retornados após cada passo ou cada clique.

Para melhor entendimento da ferramenta e de suas possibilidades, os destaques nas capturas de tela deste tutorial estão divididos em três categorias:

Destaques em vermelho: Constituem o passo-a-passo guiado para que o usuário utilize a ferramenta. Esses destaques indicam ao usuário o local e a respectiva ação que deverá ser executada para uso da ferramenta.

Destaques em verde: Constituem os resultados fornecidos pela ferramenta em cada tela. Esses resultados são automaticamente obtidos após o usuário realizar qualquer ação na ferramenta.

Destaques em laranja: Constituem apenas observações de *layout* da ferramenta para deixar explícito ao usuário algumas diferenças de funcionalidade.

1 Tutorial

A ferramenta **Calculadora de Amostras** é composta por três telas divididas em abas do Excel: *Calculadora*, *Form* e *Base*. Cada uma dessas planilhas possui uma finalidade específica e podem ser acessadas por atalhos criados na ferramenta ou simplesmente pela seleção de abas do próprio Excel.

A seguir apresentaremos uma visão geral de cada uma dessas telas e especificaremos as funcionalidades e resultados relacionados a cada uma delas.

1.1 Telas iniciais da ferramenta

Antes de tratar das funcionalidades da ferramenta, apresentaremos a visão geral dessas três telas iniciais.

1.1.1 Tela Calculadora

A tela *Calculadora* é a principal tela da ferramenta. Nela encontram-se as calculadoras de tamanho de amostra desejada e os gráficos de controle de qualidade plotados a partir dos resultados calculados. A figura a seguir mostra uma visão geral dessa tela.

Para acessar essa tela e as demais, basta que o usuário siga a descrição dos destaques na figura, explicados logo abaixo.

Ⓐ O usuário pode selecionar qual tela da ferramenta quer acessar diretamente pela seleção da aba correspondente.

Ⓑ Da tela *Calculadora* o usuário também pode acessar a tela *Form* clicando no botão "+" indicado.

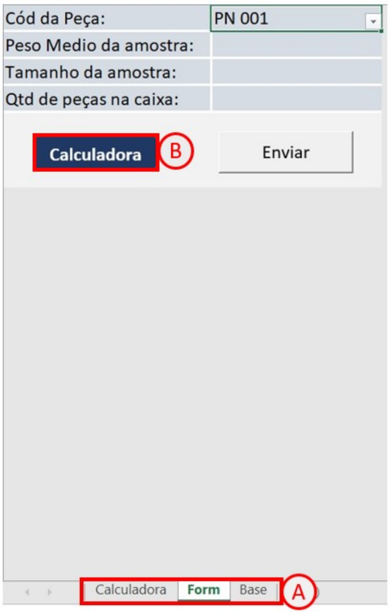
Ⓒ Da tela *Calculadora* o usuário também pode acessar a tela *Base* clicando no botão "Base" indicado.

https://doi.org/10.33774/miir-2023-s27lz **ORCID:** https://orcid.org/0000-0002-6780-2089 Content not peer-reviewed by MIIR. **License:** CC BY-NC-ND 4.0

1.1.2 Tela Form

A tela *Form* é a tela onde o usuário poderá registrar e salvar novas médias de pesos e o tamanho de uma amostra selecionada. A ideia é que a base de dados possa ser constantemente alimentada a partir daqui.

Para acessar essa tela e as demais, basta que o usuário siga a descrição dos destaques na figura, explicados logo abaixo.



Ⓐ O usuário pode selecionar qual tela da ferramenta quer acessar diretamente pela seleção da aba correspondente.

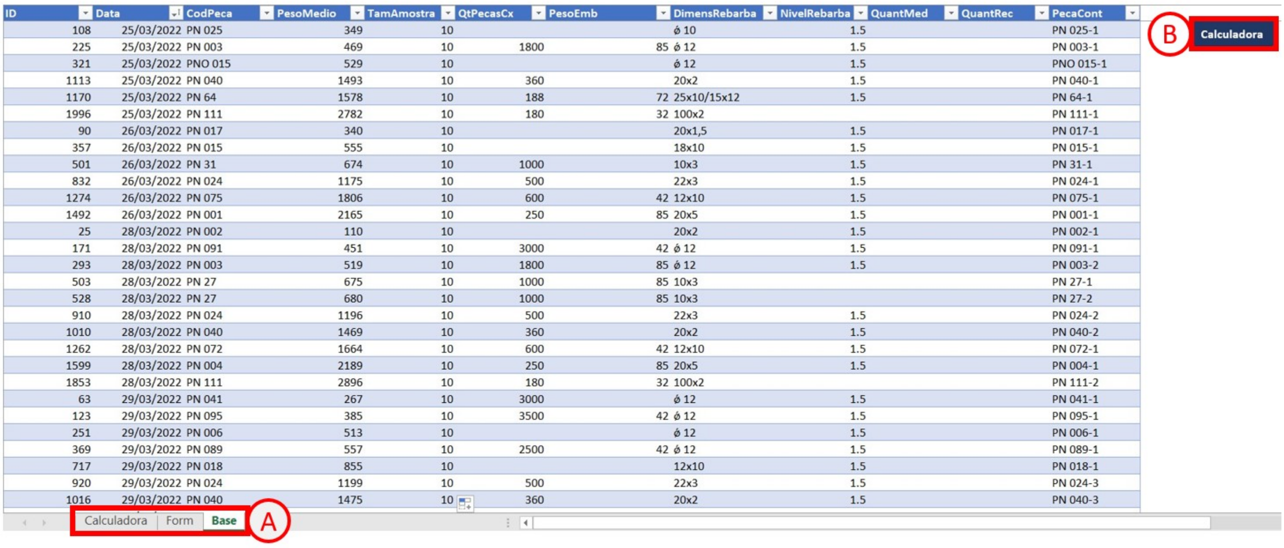
Ⓑ Da tela *Form* o usuário também pode acessar a tela *Calculadora* clicando no botão "Calculadora" indicado.

1.1.3 Tela Base

A tela *Base* corresponde a base de dados disponibilizada pela Tecumseh para o trabalho do grupo no Workshop, contendo 2000 linhas de informações referentes aos 60 tipos de peças. Nessa tela os dados estão formatados como tabela, o que possibilita o fácil manuseio do usuário em ferramentas de seleção de códigos de peça, data de registro, etc.

Além disso, a ferramenta permite a constante atualização dessa base de dados a medida que o usuário registre novas médias de pesos de amostras. Os novos valores serão salvos nas linhas dessa tabela de forma sequencial. Por esse motivo, foi acrescentada na base de dados a coluna "TamAmostra" que refere-se ao tamanho da amostra selecionada, já que a ideia é que o valor seja diferente para os diversos tipos de peças.

Para acessar essa tela e as demais, basta que o usuário siga a descrição dos destaques na figura, explicados logo abaixo.



ID	Data	CodPeca	PesoMedio	TamAmostra	QtPecasCx	PesoEmb	DimensRebarba	NivelRebarba	QuantMed	QuantRec	PecaCont
108	25/03/2022	PN 025	349	10			ø 10	1.5			PN 025-1
225	25/03/2022	PN 003	469	10	1800	85	ø 12	1.5			PN 003-1
321	25/03/2022	PNO 015	529	10			ø 12	1.5			PNO 015-1
1113	25/03/2022	PN 040	1493	10	360		20x2	1.5			PN 040-1
1170	25/03/2022	PN 64	1578	10	188	72	25x10/15x12	1.5			PN 64-1
1996	25/03/2022	PN 111	2782	10	180	32	100x2				PN 111-1
90	26/03/2022	PN 017	340	10			20x1,5	1.5			PN 017-1
357	26/03/2022	PN 015	555	10			18x10	1.5			PN 015-1
501	26/03/2022	PN 31	674	10	1000		10x3	1.5			PN 31-1
832	26/03/2022	PN 024	1175	10	500		22x3	1.5			PN 024-1
1274	26/03/2022	PN 075	1806	10	600	42	12x10	1.5			PN 075-1
1492	26/03/2022	PN 001	2165	10	250	85	20x5	1.5			PN 001-1
25	28/03/2022	PN 002	110	10			20x2	1.5			PN 002-1
171	28/03/2022	PN 091	451	10	3000	42	ø 12	1.5			PN 091-1
293	28/03/2022	PN 003	519	10	1800	85	ø 12	1.5			PN 003-2
503	28/03/2022	PN 27	675	10	1000	85	10x3				PN 27-1
528	28/03/2022	PN 27	680	10	1000	85	10x3				PN 27-2
910	28/03/2022	PN 024	1196	10	500		22x3	1.5			PN 024-2
1010	28/03/2022	PN 040	1469	10	360		20x2	1.5			PN 040-2
1262	28/03/2022	PN 072	1664	10	600	42	12x10	1.5			PN 072-1
1599	28/03/2022	PN 004	2189	10	250	85	20x5	1.5			PN 004-1
1853	28/03/2022	PN 111	2896	10	180	32	100x2				PN 111-2
63	29/03/2022	PN 041	267	10	3000		ø 12	1.5			PN 041-1
123	29/03/2022	PN 095	385	10	3500	42	ø 12	1.5			PN 095-1
251	29/03/2022	PN 006	513	10			ø 12	1.5			PN 006-1
369	29/03/2022	PN 089	557	10	2500	42	ø 12	1.5			PN 089-1
717	29/03/2022	PN 018	855	10			12x10	1.5			PN 018-1
920	29/03/2022	PN 024	1199	10	500		22x3	1.5			PN 024-3
1016	29/03/2022	PN 040	1475	10	360		20x2	1.5			PN 040-3

- (A) O usuário pode selecionar qual tela da ferramenta quer acessar diretamente pela seleção da aba correspondente.
- (B) Da tela *Base* o usuário também pode acessar a tela *Calculadora* clicando no botão "Calculadora" indicado.

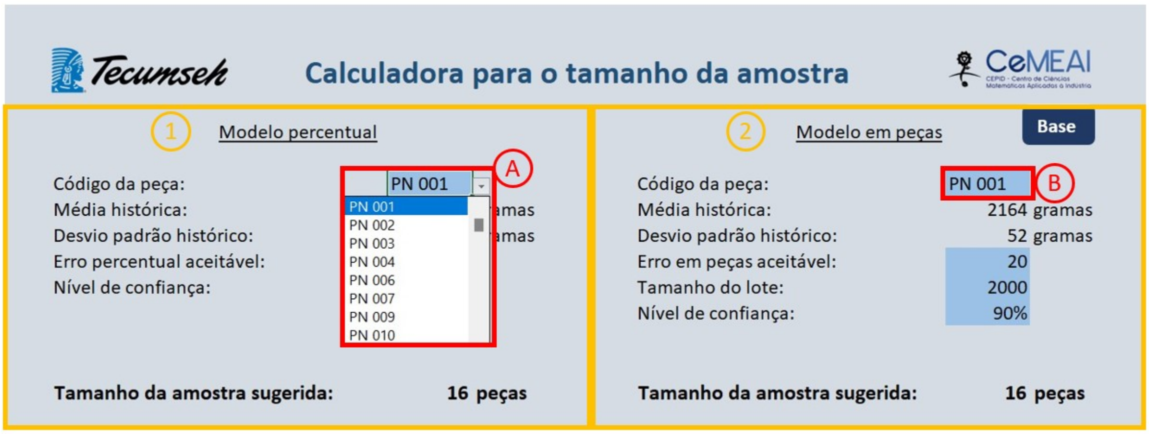
1.2 Calculadora de Amostras e Gráficos de Controle de Qualidade

Na tela *Calculadora* o usuário poderá calcular o tamanho da amostra que deverá ser tomada para um determinado erro máximo tolerável e nível estatístico de confiança desejado. Para isso, basta seguir os passos indicados na figuras a seguir.

Antes de tudo, a calculadora de amostras está dividida em duas partes (destacadas em laranja na figura abaixo) que se diferenciam pelo modo como o usuário quer definir o **erro tolerável** na amostra, em percentual ou em número de peças:

(1) Na região *Modelo percentual* o usuário deverá indicar o **percentual** de erro aceitável para empacotamento no lote desejado.

(2) Na região *Modelo em peças* o usuário deverá indicar o **número de peças** aceitável como erro (para mais ou para menos) no empacotamento no lote desejado.



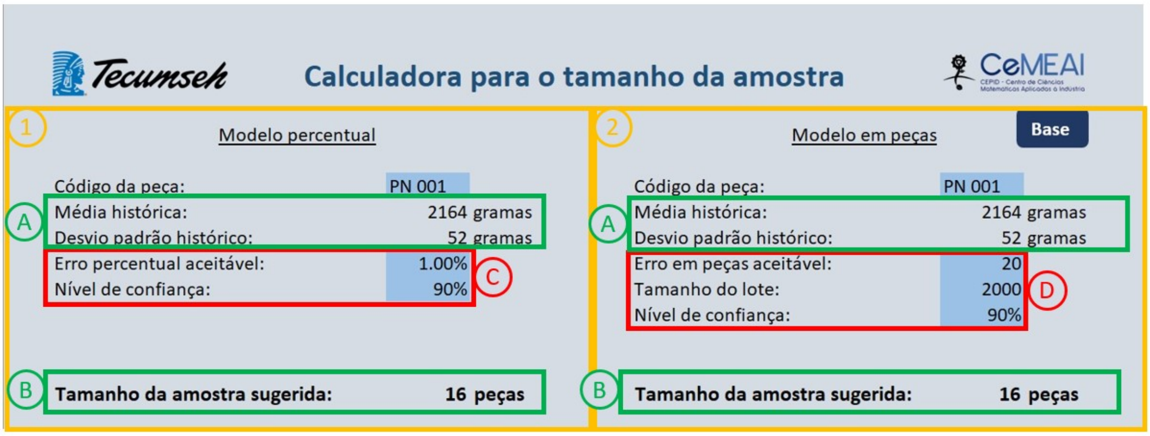
(A) O usuário deve selecionar o *Código da peça* que deseja tomar a amostra. Todas as 60 peças do banco de dados estão relacionadas, basta que o usuário selecione a desejada.

Ao escolher a peça na seleção (A), automaticamente a mesma peça será selecionada na região (B).

(B) Pode ser utilizado para seleção do tipo de peça caso deseje-se calcular o tamanho da amostra com base no erro por número de peças empacotadas.

Observação Importante: Na atual versão da Calculadora de Amostras, recomendamos que o usuário selecione a peça pela seleção (A). Apenas essa seleção atualiza os gráficos de controle de qualidade para a peça selecionada.

https://doi.org/10.33774/miir-2023-s27lz **ORCID:** https://orcid.org/0000-0002-6780-2089 Content not peer-reviewed by MIIR.



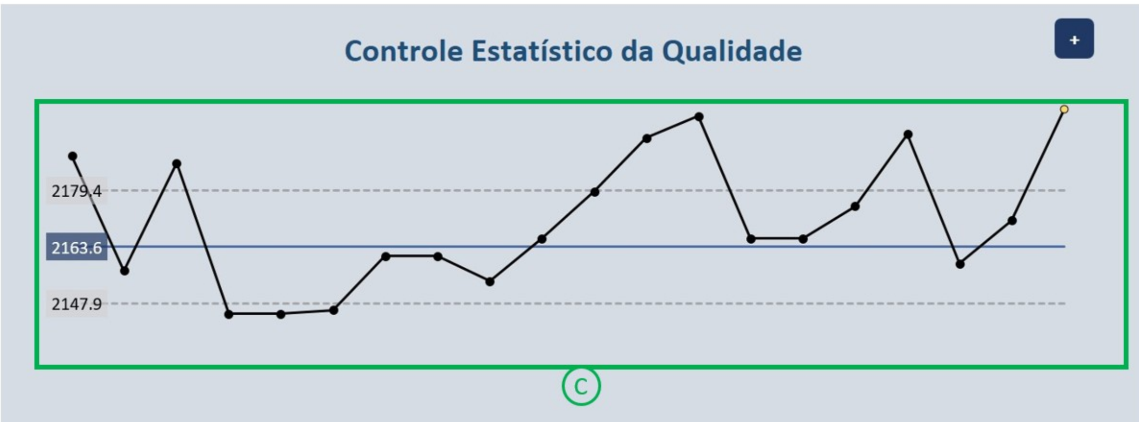
(A) Após selecionar o código da peça, para ambas as regiões (1) ou (2) são retornados automaticamente a **Média histórica** e o **desvio padrão histórico** dos pesos dessa peça.

(C) Se o usuário optar pelo *Modelo percentual* (1) , deverá digitar nas células destacadas o **erro percentual aceitável** para esse lote e o **nível de confiança** desejado.

(D) Se o usuário optar pelo *Modelo em peças* (2) , deverá digitar nas células destacadas o **erro em peças aceitável** e o **tamanho do lote** a ser empacotado, bem como o **nível de confiança** desejado.

(B) Independente da escolha do modelo (1) ou (2), após preencher as células com as informações da peça, do erro e do nível de confiança, automaticamente será retornado o **tamanho da amostra sugerida** pelo modelo.

Ao indicar o código da peça analisada no passo (A) acima, automaticamente o **Gráfico de Controle de Qualidade** com o histórico dessa peça é gerado, conforme a figura abaixo:

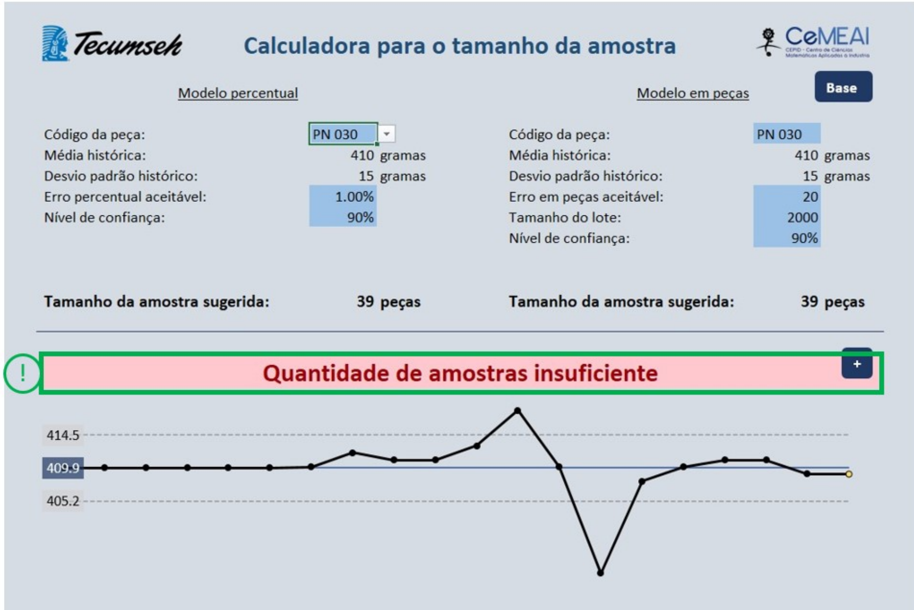


(C) Gráfico de controle de qualidade gerado a partir das últimas 20 observações de médias amostrais do tipo de peça escolhido. Algumas características desse gráfico:

- A linha central azul representa o valor médio das médias de peso registradas nas amostra da peça escolhida;
- As linhas pontilhadas representam o *limite superior de controle* (LSC) e o *limite inferior de controle* (LIC)
- O último ponto registrado está propositalmente colorido em outra cor, para destaque.

(!) **Observação importante:** A teoria utilizada para o cálculo de tamanho amostral tem como pressuposto a normalidade da distribuição das médias de pesos registradas. Esse pressuposto não é garantido para amostras pequenas. Dessa forma, o modelo é válido para as peças cujo o histórico de registro possui mais do que 20 observações.

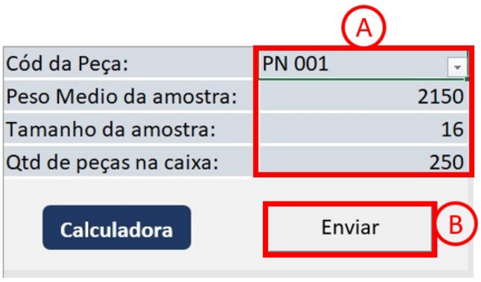
Caso o usuário selecione uma peça com histórico menor que 20 registros, um aviso é retornado ao usuário, conforme indicado pelo marcador (!) na figura abaixo:



Nesse caso, sugerimos que o usuário decida entre seguir o tamanho amostral indicado pela ferramenta ou tomar apenas 10 unidades conforme procedimento padrão anterior.

1.3 Registro de Novas Médias e Tamanhos Amostrais

Após tomar uma amostra e observar o peso médio das peças amostradas, o usuário pode registrar esse valor para atualizar o banco de dados, bem como o tamanho da amostra observada. Isso será feito na tela *Form*, seguindo os passos destacados nas figuras a seguir:



(A) Selecionar o código da peça observadas e digitar as demais informações: peso médio da amostra, **tamanho da amostra** e **quantidade de peças na caixa**.

(B) Após informar os dados, clicar no botão "Enviar" para inserir essas informações no banco de dados.

Após enviar o novo valor, o usuário pode retornar à tela *calculadora* e atualizá-la para incluir esse novo valor médio no gráfio de controle de qualidade

(C) Em qualquer tela, selecionar a opção "Dados" do Excel.

(D) Clicar na opção "Atualizar Tudo".

Obs: Ou simplesmente apertar o botão "F5" ao invés dos passos (C) e (D).

(A) O Gráfico de Controle de Qualidade é atualizado com o novo valor, sendo este o novo último ponto, destacado em amarelo. O usuário pode avaliar se esse ponto está entre os limites de controle. Se não tiver, recomendamos que se tome uma nova amostra e calcule novamente a média de pesos para verificar se há alguma variação no lote de peças ou se foi apenas uma variação decorrente do processo de amostragem.

https://doi.org/10.33774/miir-2023-s27lz **ORCID:** https://orcid.org/0000-0002-6780-2089 Content not peer-reviewed by MIIR.

1.4 Banco de Dados

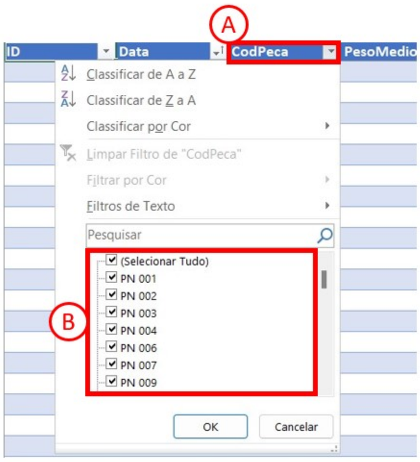
O banco de dados está alocado na tela *Dados* e conta inicialmente com as 2000 observações fornecidas pela Tecumseh ao grupo de trabalho do 9º Workshop de Soluções Matemáticas para Problemas Industriais (CeMEAI-USP). Esse banco está formatado como tabela e permite que o usuário realize seleções direcionadas ao seu interesse.

Para além das funcionalidades padrão do Excel, a ferramenta possibilita a retroalimentação do banco de dados com as novas observações de médias de pesos obtida de novas amostras. Cada novo registro é inserido na tabela como uma nova linha, conforme ilustrado na figura a seguir:

ID	Data	CodPeca	PesoMedio	TamAmostra	QtPecasCx
1456	18/01/2023	PN 001	2159	10	250
1533	19/01/2023	PN 001	2171	10	250
1651	23/01/2023	PN 001	2202	10	250
2001	24/03/2023	PN 001	2150	16	250

Ⓐ Linha inserida no banco de dados após o usuário utilizar a tela *Form* para enviar esse registro.

Observação importante: Ao utilizar a tela *Form* o usuário pode acidentalmente clicar mais de uma vez a tecla "Enviar" e registrar a média de peso de uma única amostra duas ou mais vezes. Se isso acontecer, o usuário pode ir até o banco de dados excluir as linhas adicionadas desnecessariamente, seguindo os seguintes passos:



Ⓐ Clicar no título da coluna "Código da Peça".

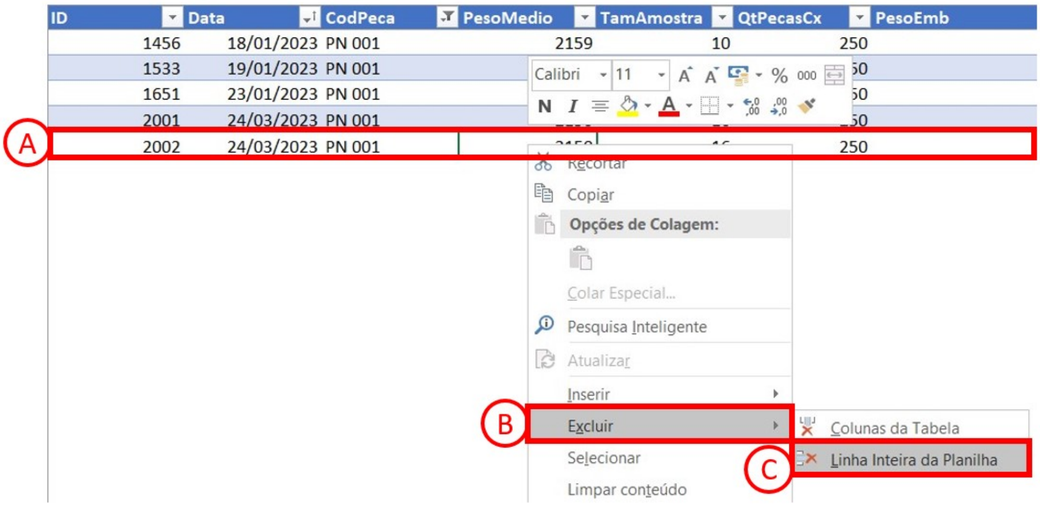
Ⓑ Na seleção aberta, deixar marcado apenas o código da peça desejada.

ID	Data	CodPeca	PesoMedio	TamAmostra	QtPecasCx
1456	18/01/2023	PN 001	2159	10	250
1533	19/01/2023	PN 001	2171	10	250
1651	23/01/2023	PN 001	2202	10	250
2001	24/03/2023	PN 001	2150	16	250
2002	24/03/2023	PN 001	2150	16	250

Ⓑ Seguindo até o final da tabela é possível verificar facilmente as linhas excedentes inseridas.

https://doi.org/10.33774/miir-2023-s27lz **ORCID:** https://orcid.org/0000-0002-6780-2089 Content not peer-reviewed by MIIR. **License:** CC BY-NC-ND 4.0

Basta que os seguintes passos sejam seguidos para excluir as linhas desnecessárias:



(A) Clicar com o botão direito do mouse sobre a linha que deseja-se excluir.

(B) Na seleção aberta, clicar na opção “Excluir”.

(B) Na nova seleção aberta, clicar na opção “Excluir linha inteira”.

https://doi.org/10.33774/miir-2023-s27lz **ORCID:** https://orcid.org/0000-0002-6780-2089