

Application of the Monte Carlo Simulation method to assess the overall productivity of a bromatological laboratory

Nadia Cárdenas-Escobar M.Sc.¹, Stalin Pinargote-Zavala M.Sc.¹, Sandra García-Bustos¹
Ph.D., Jorge Quichimbo M. Sc.²

¹ Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias Naturales y Matemáticas
Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador,
ncardena@espol.edu.ec; spinargo@espol.edu.ec; slgarcia@espol.edu.ec

² Universidad de Guayaquil, Facultad de Ciencias Químicas, Ciudadela Universitaria Salvador
Allende, Guayaquil, Ecuador, jorge.quichimbom@ug.edu.ec

Abstract— *The initial step of the production chain in a food processing factory in Ecuador is the qualification of raw materials, which is overseen by the bromatological laboratory of a company that expects reliable and timely results. The laboratory process starts from the reception of the raw materials until the sending of the results to the areas involved. However, some stages of the process are more elaborated than others, causing delays in the delivery of the raw material qualification results, which subsequently continue into the production chain process. For this reason, productivity at each stage of the analytical process must be improved to reduce results delivery time. This work used the Monte Carlo simulation method to find the best scenario for increasing productivity in the bromatology laboratory. The findings show that productivity could increase from 23% to 56% as long as the causes of delays in the analysis are solved.*

Keywords— Monte Carlo, simulation, bromatological laboratory, productivity

I. INTRODUCTION

According to the National Chamber of Aquaculture (NCA) [1], in Ecuador, between January and November 2018, 506 thousand MT of shrimp were exported, representing an increase of 15%, compared to the same period in 2017 and equivalent to USD 3,234.8 million. From December 2017 to June 2018, USD 1.5 million were sold to Brazil, according to Pro Ecuador. At that time, the market was opened after a six-month court battle that overcame an 18-year sanitary barrier.

This sector faced other challenges in 2019, reopening sales to Mexico and South Korea, both with health barriers. In 2017, South Korea was the fifth-largest destination and had been growing at a rate of 15% to 20% per year [1].

So far this year (data up to September 2019), the number of exported shrimps is 472 thousand MT with an amount of USD 2,705.8 million. Due to the increase in previous years, the demand and the production of shrimp feed, which provides solutions to the shrimp sector, have risen [1].

Consequently, the amount of raw material received by the company under analysis has increased. This company's main activity focuses on producing and marketing feed for various aquaculture and agricultural species. It is located in Durán, a province of Guayas-Ecuador.

The bromatological laboratory where this simulation was applied analyzes raw materials for their qualification (approval - rejection) before entering the process. Hence, the functioning of this laboratory is the first part of the whole production process of the plant.

Description of the problem

For each analysis in the bromatological laboratory, it is necessary to follow certain protocols and processes to fulfill the required activities. Each of these processes takes more time than others, so it is essential to identify the ones of vital importance to implement improvements and create a model that helps increase productivity in the laboratory.

In recent months, as the demand for shrimp feed has risen, the amount of raw material received has also increased. Due to this, the results have been delayed- Therefore, an improvement is needed in the processes of receiving and analyzing the samples.

In this regard, the present work focused on the elaboration of a statistical model of simulation to reduce results delivery time and improve productivity in a bromatological laboratory.

According to [2], simulations are used in the industry to predict the consequences of making certain decisions such as inventory control, maintenance plans, resource location, sales or demand prediction, etc. In addition, simulations allow complex problems to be solved, although the results are an approximation of the solution.

Scope

This study only applies to a bromatological laboratory in Durán, province of Guayas, Ecuador.

The data for this study were obtained from the same site where the analysis processes are executed, and each stage of the analysis was observed for the measurement of times. The data were taken in September 2019. In addition, the causes of the delay problem were measured through cause-effect analysis and a Pareto chart indicating the frequencies of the causes of the delays. Finally, simulations were carried out using historical data under different scenarios to observe the degree of improvement that could be obtained in the process.

II. LITERATURE REVIEW

Simulation Stages

The stages of a simulation are described below:

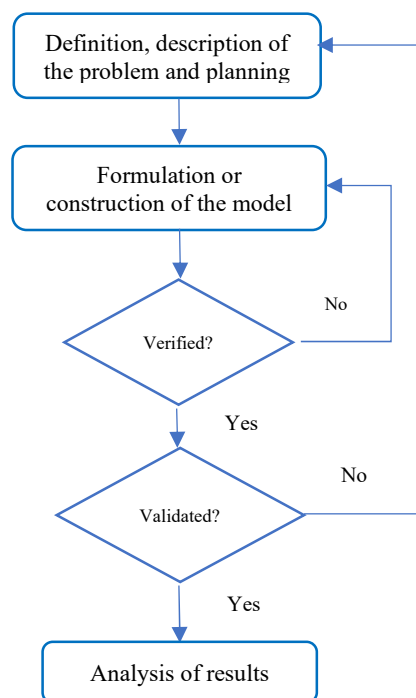


Fig. 1 Flowchart of a Simulation

Definition, description of the problem. Plan.

In this first stage, the objectives of the simulation must be identified. Also, the measurement indices have to be specified to determine the effectiveness of the system under study. Finally, specific objectives of the models and the systems to be modeled through the simulation must be defined.

Formulation or construction of the model.

Once the expected results of the study are set, it is necessary to define and build a model to obtain the desired

results. However, it must be considered that the purpose is not to develop the model itself but instead that it must solve the problem to reach the desired goal.

In this stage, the variables that are part of the model, their logical relationships, and the flow charts that fully explain the model must be defined.

A simulation model refers to a set of hypotheses that describe the functioning of a system expressed as mathematical and/or logical relationships between various elements that make up the system [3].

Model verification.

It is necessary to confirm that the simulated model meets the design requirements for which it was built. That is, it must be verified that the simulation model follows the specifications of the conceptual model.

Model validation.

During this stage, it is necessary to assess the differences between the simulation and the real system. Also, the correctness of the working hypotheses and the precision of the data must be checked. Furthermore, the structural assumptions established at the beginning and their statistical considerations have to be validated [14].

Analysis of results.

In this stage, the results generated by the simulation are interpreted, and decisions are made. Therefore, the results should collaborate with the decision support system to make semi-structured decisions.

Monte Carlo simulation.

Monte Carlo Simulation is a computerized mathematical technique that uses randomly generated numbers and inverse probability distributions [4]. The Monte Carlo simulation can show the extreme possibilities. It also allows observing all the possible scenarios of the decisions made and evaluating the impact of the risk. Therefore, it is possible to achieve the best decision-making under a certain uncertainty, even to analyze all the potential consequences of the possible scenarios during the simulation before making decisions.

Monte Carlo simulation performs risk analysis by building models of possible outcomes using a range of random values (a probability distribution) for each factor with inherent uncertainty.

In computing, the sequences of random numbers used are pseudorandom since they are generated by an algorithm that ensures that the sequence is sufficiently unpredictable and does not repeat itself in very short cycles. These algorithms use a seed or starting number as the starting point for generating the

sequence. Furthermore, this sequence of random numbers is usually constructed with a uniform distribution within the interval (0,1).

According to the recommendations made by [5] and [6], the pseudorandom numbers (iterations) that can be generated for the simulation must be determined. In this way, the simulation is carried out by the sample size, the classes, and the quantity of the set of pseudorandom numbers.

Generally, in statistics, random models are used to simulate phenomena that contain a certain random component. However, the Monte Carlo method considers a random or pseudorandom event to study the model. [7]

Next, some statistical concepts helpful in understanding these methods are defined. First, the probability distribution function (PDF) of a variable x ($p(x)$) is a function that contains the probability of success for each value of x . This function must not have negative values as a condition and must be normalized (transform to proportions without measurement units that allow elements of different variables and different measurement units to be compared.) within an interval (x_{min}, x_{max}) [18].

$$p(x) \geq 0 \quad \equiv \quad \int_{x_{min}}^{x_{max}} p(x) dx = 1 \quad (1)$$

The cumulative probability function (CPF) of a variable x is a function that contains the probability of success within an interval $[x_{min}, x]$. Therefore, it is a monotone increasing function with initial value $P(x_{min})=0$ y $P(x_{max})=1$ [18].

$$P(x) \equiv \int_{x_{min}}^x p(x') dx' \quad (2)$$

The generating function to be used as a probability distribution function (PDF) must present a form similar to the distribution of the elements of the sample for the classes to be calculated under predetermined conditions.

Steps to Build the Monte Carlo Simulation

The steps to follow for the application of the Monte Carlo simulation [9],[10],[12]; adapted for the construction of the model, are the following:

- Identify the probability distributions with their appropriate parameters for each stochastic variable. In this case, these are the analysis service volumes. This is explained in the identification of input and output variables.
- Calculate the correlations between the independent variables using a correlation matrix and other techniques. Include possible dependencies between variables.

- Build the model, checking all its assumptions.
- Generate random or pseudorandom numbers, which, as indicated above, are from 0 to 1.
- Feed the stochastic processes or the inverse probability distributions with the random numbers to determine to which class each probability belongs (random number).
- Calculate the number of simulations. The larger the number of simulations, the better, based on the law of large numbers. Each run is a possible and consistent scenario based on the assumptions established in the model. Therefore, each result must be saved. In this study, calculations of the number of simulations were performed.

The number of runs of a simulation depends a lot on the distribution to be simulated, in addition to the goodness of the random number generators $U(0.1)$ used and the conditions with which the simulation started. In general, the number of simulations is calculated using the formula:

$$n = \frac{\sigma^2 (Z_{\alpha/2})^2}{\varepsilon^2} \quad (3)$$

In which:

$Z \rightarrow$ Normal statistic for a given α

$\varepsilon \rightarrow$ Standard error or absolute deviation allowed on the sample of the distribution to be simulated

$\sigma^2 \rightarrow$ Variance of the distribution to simulate

In the event that the mean and variance of the distribution to be simulated are obtained from a population n_1 of 30 data points or less, then this calculation is modified, and the following equation is used:

$$n = \frac{s^2 (t_{n-1, \alpha/2})^2}{\varepsilon^2} \quad (4)$$

In which:

$t \rightarrow$ t-student distribution statistic

$\varepsilon \rightarrow$ Standard error or absolute deviation allowed on the sample of the distribution to be simulated

$s^2 \rightarrow$ Estimator of the variance of the distribution to be simulated

Both formulas can be used only when the information from which the estimators are obtained statistically follows a normal distribution [13]. If the analyzed data follow any other distribution, Chebyshev's theorem must be used so that the number of runs is calculated by the equation:

$$n = \frac{m^2}{\alpha}$$

In which:

m^2 → Maximum standard deviation allowed on the mean of the distribution to be simulated
 α → Allowable probability of error

The optimal number of runs of a model with some probability variables is obtained by executing the calculation for each one of them, and the largest is selected; then, this is the number of simulations of the model [2].

Recent research

Recent research on Monte Carlo simulations was reviewed to arrive at the desired results.

In the work of [17], an analysis of the Monte Carlo simulation was carried out as a tool to solve equations where the necessary variables are not known exactly; however, the ranges of the real data of these variables are known. The author used random number generation methods and frequency histograms. A spreadsheet was utilized for the Monte Carlo simulation in addition to a simulation called "MCP," which uses equations of great importance in the oil field. Using spreadsheets to perform a Monte Carlo simulation is a commonly used tool that allows any equation to be solved. The only problem is that many steps and checks for the correctness of the equations need to be performed before ensuring that the simulation is successful.

[7] focused on the Monte Carlo simulation applied to the inventories of an oil service company. The stages of the Monte Carlo simulation were defined, emphasizing the construction of a probability distribution and generation of random numbers. Also, the variables that influence the inventory were detailed. Consequently, it was determined that the Monte Carlo simulation is not a precise but a very close approximation to the real system. The more data generated, the greater its accuracy.

Description of the Work Area

The present work focused on the laboratory analysis activities explained in Fig. 2.

- Sample registration: Every sample admitted to the laboratory must be registered in a form to evidence its reception and further analysis.
- Sample homogenization: It is necessary to reduce the sample size and ensure that the resulting mixture represents the number of bags or bulk trucks previously sampled. Generally, the sample that arrives at the laboratory varies between 1 to 3 kg, and in the analysis, only approximately 10 g or less is needed.

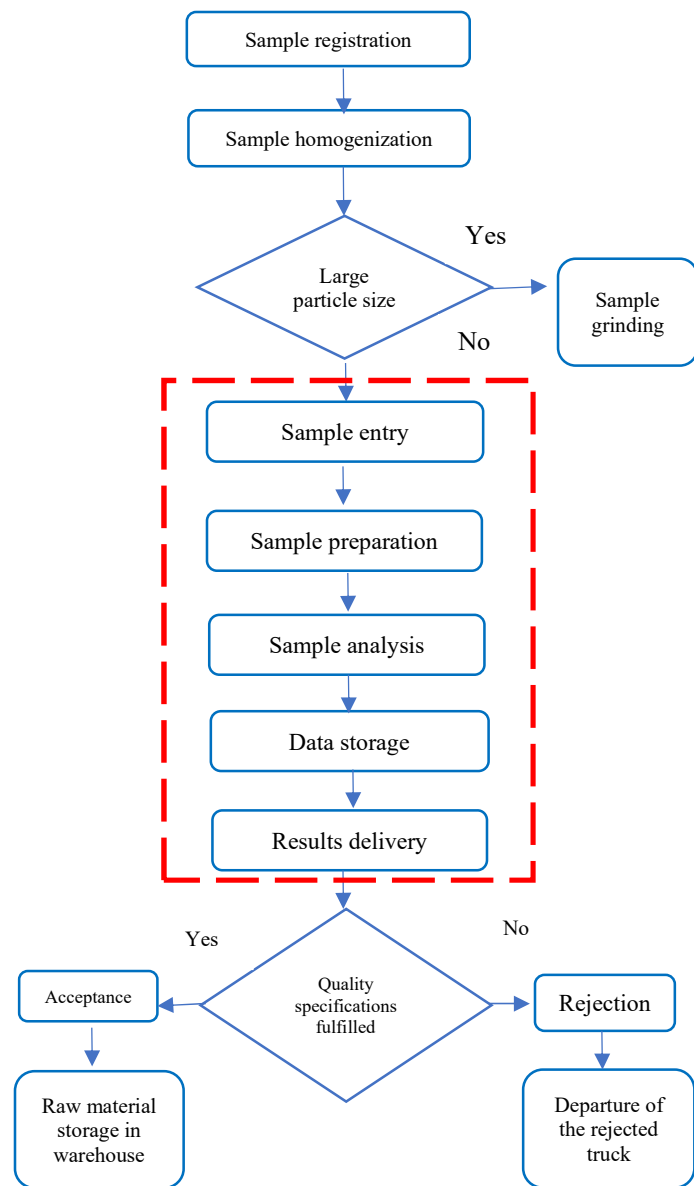


Fig. 2 Laboratory Analysis Process Diagram

- Sample grinding: some raw materials arrive in grains or are too coarse, so they must be ground to reduce their particle size and proceed with their respective analysis. If the samples are fine powders, grinding is not necessary.
- Sample entry: it is crucial to enter all the relevant information about the product, such as supplier, batch, production and expiration dates, and so on, into the respective system.
- Sample preparation: the instrument sample holders must be clean and in good condition to prepare the sample. It is also essential to place the correct amount of sample so that the NIR (Near Infrared Reflectance) analysis is accurate with

no prediction errors in the results. NIR technology involves a beam of light interacting with a material, where electromagnetic radiation occurs in the form of waves.

- **Sample analysis:** the sample holder is placed inside the NIR instrument and the instrument is allowed to run to obtain the results.
- **Data storage:** since the system of these analyzes is automatic; the results are expected to be stored in the respective database.
- **Results delivery:** the areas involved in this process are informed through email if the raw material meets the specifications and can be stored in the warehouses; or if the raw material is rejected and the truck containing the product is sent out of the plant.

This study focused on the processes starting from sample entry until the sending of the results (marked red in Figure 2), considering that no changes could be made in the previous processes since they are vital in the laboratory.

III. IMPLEMENTATION

The input variables and the output variable are identified as it follows:

Input variables

Sample entry time includes entering all the raw material information, such as supplier, production and expiration dates, etc., into the system, which is an online database platform. This is the initial stage of the analysis before starting sample preparation. This variable is identified as T_{in} .

Four distributions were tested to determine which one fits the behavior of the data of the variable T_{in} through the following hypothesis test:

$H_0: \geq 0.05$: The data follows an X Distribution.

$H_1: < 0.05$: The data does not follow an X Distribution.

TABLE I. P-VALUE OF THE VARIABLE DATA ENTRY TO THE SYSTEM

Distribution	P-value
Normal	0.025
Exponential	<0.003
Weibull	0.028
Gamma	<0.005

The null hypothesis was rejected according to the p-values shown in Table I. Therefore, it could be inferred that none of the proposed distributions fit the behavior of the data of the T_{in} variable. Consequently, it was calculated as an Empirical Distribution:

$$f(x) = \begin{cases} 0.15; 2.25 \leq x < 2.75 \\ 0.10; 2.75 \leq x < 3.25 \\ 0.05; 3.25 \leq x < 3.75 \\ 0; 3.75 \leq x < 4.25 \\ 0.35; 4.25 \leq x < 4.75 \\ 0.10; 4.75 \leq x < 5.25 \\ 0.15; 5.25 \leq x < 5.75 \\ 0.10; 5.75 \leq x < 6.25 \end{cases}$$

Sample preparation time refers to the time it takes for the analyst to place the sample in the sample holder cells. This stage is performed after entering the sample data into the system and before performing the NIR analysis. This variable is identified as T_{pr} .

TABLE II. P-VALUE OF THE SAMPLE PREPARATION TIME VARIABLE

Distribution	P-value
Normal	0,074
Exponential	0,067
Weibull	0,056
Gamma	0,045

According to the p-values from Table II, the behavior of the data could be adjusted to a Normal or an Exponential Distribution since they have values of >0.05 . However, the highest p-value was taken, so there was statistical evidence to reject the null hypothesis. Therefore, the data for the T_{pr} variable were considered to follow a Normal Distribution.

Sample analysis time corresponds to the time it takes for the equipment to get the results of the analyzed sample. The analysis, per se, is carried out after the preparation of the sample until the results are obtained. This variable is identified as T_{an} .

TABLE III. P-VALUE OF THE SAMPLE ANALYSIS VARIABLE

Distribution	P-value
Normal	0,78
Exponential	<0,003
Weibull	<0,003
Gamma	<0,005

According to the p-values from Table III, there was no statistical evidence to reject the null hypothesis. Therefore, the data for the T_{an} variable were considered to follow a Normal Distribution.

Data storage time is the time it takes for the system to capture the results of the analysis. This step is automatic, including from the moment the results are released to when they are stored in the database. This variable is identified as T_{al} .

TABLE IV. P-VALUE OF THE DATA STORAGE TIME VARIABLE IN THE SYSTEM

Distribution	P-value
Normal	0,689
Exponential	<0,003
Weibull	<0,003
Gamma	<0,005

According to the p-values shown in Table IV, the data of the T_{al} variable were considered to follow a Normal Distribution.

Results delivery time corresponds to the time it takes for the analyst to prepare and send a report of the results by email. It is measured from the time the results are stored in the database to when the email to all interested parties is sent. This variable is identified as T_{en} .

TABLE V. P- VALUE OF THE VARIABLE RESULTS DELIVERY TIME

Distribution	P-value
Normal	0,355
Exponential	<0,003
Weibull	0,251
Gamma	0,205

According to the p-values seen in Table V, the data could follow a Normal Distribution, the Weibull distribution, or the Gamma distribution. However, in this case, the highest p-value was taken, so there was no statistical evidence to reject the null hypothesis. Therefore, the data of the T_{en} variable were considered to follow a Normal Distribution.

Output variable

The data obtained from the measurements between the months of September and October 2019 are presented as mean values of the minutes taken to perform an analysis (minutes/analysis). This variable is identified as $Mnts$.

Subsequently, this output variable was transformed in terms of productivity (number of tests/day of work). It should be noted that the working hours are from 8:00 a.m. to 5:00 p.m. with a lunch break from 1:00 p.m. to 2:00 p.m., that is, eight working hours (480 minutes). The transformation was carried out using the following formula:

$$\frac{\text{Analysis}}{\text{day}} = \frac{480 \text{ min/day}}{n \text{ min/analysis}} = \frac{480}{n} \text{ analysis/day}$$

TABLE VI. P-VALUE OF THE OUTPUT VARIABLE MNTS (MINUTES/ANALYSIS)

Distribution	P-value
Normal	0,274
Exponential	<0,003
Weibull	0,219
Gamma	0,251

According to the p-values from Table VI, the data could follow a Normal distribution, the Weibull distribution, or the Gamma distribution. In this case, the highest p-value was taken, so there was no statistical evidence to reject the null hypothesis. Therefore, the data of the output variable $Mnts$ were considered to follow a normal distribution.

Calculation of correlations

With the data obtained through the measurements of the times, the model of the initial situation in the laboratory was constructed. As a first step, we elaborated a correlation matrix to determine any problem of multicollinearity between the variables:

```
> cor(Datos)
      Tin      Tpr      Tan      Tal      Ten      Mnts
Tin  1.00000000 -0.037247358  0.34913250 -0.2400806  0.328537444  0.6137862
Tpr  -0.03724736  1.000000000  0.05282465  0.1378738  0.003952431  0.1804886
Tan   0.34913250  0.052824655  1.00000000  -0.1167116  0.377432795  0.6087820
Tal  -0.24008060  0.137873844 -0.11671159  1.00000000 -0.192122955  0.3934099
Ten   0.32853744  0.003952431  0.37743280 -0.1921230  1.000000000  0.5262257
Mnts  0.61378624  0.180488561  0.60878203  0.3934099  0.526225654  1.0000000
```

Fig 1. Correlation matrix of the data of the initial situation

Since the values of the variables in the correlation matrix were not very high (Fig. 3), there were no multicollinearity problems. Hence, we could build the corresponding model by checking all its assumptions.

Model Construction

A proposed statistical model was carried out to possibly determine the initial situation of the laboratory. In this case, a variable selection method was applied through stepwise elimination, which starts without eliminating any variable (Start AIC); then, variables are eliminated or included until a model is obtained by selecting the best predictor variables with the Akaike measurement (AIC) of lower value.


```

> modelol<-lm(Datos$Mnts~DTin+DTpr+DTan+DTal+DTen,data=Datos)
> step(object=modelol,direction="both",trace=1)
Start: AIC=-58.24
Datos$Mnts ~ DTin + DTpr + DTan + DTal + DTen

    Df Sum of Sq    RSS   AIC
<none>                 0.5969 -58.236
- DTpr 1      0.2615  0.8584 -52.968
- DTen 1      2.6361  3.2330 -26.446
- DTan 1      3.1722  3.7691 -23.378
- DTin 1      6.6637  7.2605 -10.266
- DTal 1     10.2050 10.8019  -2.320

Call:
lm(formula = Datos$Mnts ~ DTin + DTpr + DTan + DTal + DTen, data = Datos)

Coefficients:
(Intercept)      DTin      DTpr      DTan      DTal      DTen
  2.5475      0.5551      0.4200      0.5774      0.5520      0.4129

```

Fig. 4 Akaike measurement to determine the significant variables in the model of the initial situation in the bromatological laboratory

According to Fig 4, in the Akaike measurement, the lowest AIC (-58.236) was found with all the initial variables (<none>), while eliminating any of the variables increased the AIC value. Hence, this indicates that the best model proposal adjusted to the reality of the system is the one with no eliminated variables. For this reason, we could preliminarily point out that the linear regression model proposed for the initial situation of the laboratory would be:

$$\text{Minutes/analysis} = \beta_0 + \beta_1 T_{in} + \beta_2 T_{pr} + \beta_3 T_{an} + \beta_4 T_{al} + \beta_5 T_{en}$$

With this background, we determined the coefficients β_0 , β_1 , β_2 , β_3 , β_4 y β_5 through the following hypothesis test with 0.05 significance:

$$H_0: \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

$$H_1: \beta_0 \neq \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4 \neq \beta_5 \neq 0$$

```

> summary(modelol)

Call:
lm(formula = Datos$Mnts ~ DTin + DTpr + DTan + DTal + DTen, data = Datos)

Residuals:
    Min       1Q   Median       3Q      Max
-0.22538 -0.13806 -0.01688  0.11909  0.33474

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.54752    0.30349   8.394 7.79e-07 ***
DTin         0.55505    0.04440  12.502 5.52e-09 ***
DTpr         0.42001    0.16958   2.477  0.0266 *
DTan         0.57742    0.06694   8.626 5.64e-07 ***
DTal         0.55198    0.03568  15.471 3.38e-10 ***
DTen         0.41288    0.05251   7.863 1.67e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2065 on 14 degrees of freedom
Multiple R-squared:  0.9799, Adjusted R-squared:  0.9728
F-statistic: 136.8 on 5 and 14 DF, p-value: 2.26e-11

```

Fig. 5 Proposed statistical model of the initial situation of the bromatological laboratory

The p-value of the model was significant (2.26e-11), so the null hypothesis was rejected. Therefore, we could indicate that at least one of the partial regression coefficients differed from 0. The coefficients for the proposed model could be observed as follows: $\beta_0=2.54752$, $\beta_1=0.55505$, $\beta_2=0.42001$, $\beta_3=0.57742$, $\beta_4=0.55198$, and $\beta_5=0.41288$.

The proposed linear regression model had an R^2 of 0.9799, so this model could explain 97.99% of the observed variable (minutes/analysis).

All the associated assumptions were tested, such as linearity, independence, homoscedasticity, normality of the residuals, and non-collinearity, to validate the regression model.

Simulation Verification

In this stage, a pilot test of 20 simulations was carried out. Furthermore, 20 random numbers considered initial data were generated, and the inverse probability distributions were fed with these generated random numbers to determine to which class each probability belongs (random number). This was performed for each variable in the model. The results were then obtained by applying the regression model to compare the output variable of the real system vs. the pilot simulation.

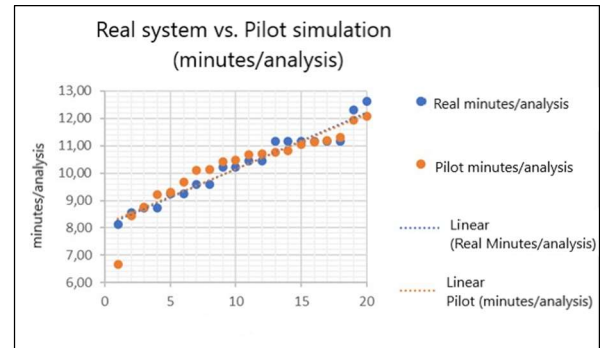


Fig. 6 Comparison of the Real System vs. Pilot Simulation (Minutes/analysis)

As seen in Figure 6, the slopes of the real system and the pilot simulation of the output variable (Minutes/analysis) were very similar, which indicates no difference between both. Additionally, it suggests the feasibility of carrying out a greater number of simulations.

Calculation of number of simulations

The number of simulations must be determined for each of the input variables used in the simulation model.

Since the data for the T_{in} variable follow an empirical distribution, the following equation was used:

$$n = \frac{m^2}{\alpha}$$

Where:

m : Standard deviation = 1.29

α : Allowable probability of error = 0.05

Then:

$$n = \frac{1.29^2}{0.05} = 33.282 \cong 33$$

Considering that the data for the T_{pr} variable follow a Normal distribution with an amount of data less than 30, the following equation was used:

$$n = \frac{s^2(t_{n-1,\alpha/2})^2}{\varepsilon^2}$$

Where:

s^2 : Variance = 0.34

ε : Standard error = 0.076

$t_{19, 0.05}$: t-student distribution statistic = 2.093

Then:

$$n = \frac{0.34^2(2.093)^2}{0.076^2} = 87.674 \cong 88$$

Given that the data for the T_{an} variable follow a Normal distribution with an amount of data less than 30, the following equation was used:

$$n = \frac{s^2(t_{n-1,\alpha/2})^2}{\varepsilon^2}$$

Where:

s^2 : Variance = 0.97

ε : Standard error = 0.217

$t_{19, 0.05}$: t-student distribution statistic = 2.093

then:

$$n = \frac{0.97^2(2.093)^2}{0.217^2} = 87.531 \cong 88$$

Considering that the data for the T_{al} variable follow a Normal distribution with an amount of data less than 30, the following equation was used:

$$n = \frac{s^2(t_{n-1,\alpha/2})^2}{\varepsilon^2}$$

Where:

s^2 : Variance = 1.31

ε : Standard error = 0.293

$t_{19, 0.05}$: t-student distribution statistic = 2.093

then:

$$n = \frac{1.31^2(2.093)^2}{0.293^2} = 87.568 \cong 88$$

Finally, given that the data for the T_{en} variable follow a Normal distribution with an amount of data less than 30 the following equation was used:

$$n = \frac{s^2(t_{n-1,\alpha/2})^2}{\varepsilon^2}$$

Where:

s^2 : Variance = 1.09

ε : Standard error = 0.244

$t_{19, 0.05}$: t-student distribution statistic = 2.093

Then:

$$n = \frac{1.09^2(2.093)^2}{0.244^2} = 87.420 \cong 87$$

So for this study, we need to run a total of 88 simulations..

Validation of the Initial Situation Model

Table VII shows the summary of the descriptive statistics of the response variable (minutes/analysis), as well as the transformation in terms of productivity of the data obtained from the simulation and the real measurements.

TABLE VII. STATISTICAL SUMMARY OF THE MINUTES PER ANALYSIS AND THE NUMBER OF DAILY ANALYSES OBTAINED THROUGH THE REAL DATA AND THE SIMULATION.

Statistics	Minutes/analysis		Analysis/day	
	Real	Simulation	Real	Simulation
μ	10,25	10,29	47,5	47,1
σ	1,25	1,05	5,9	4,9
Mín	8,14	8,00	38,0	37,5
Máx	12,63	12,37	59,0	55,1
Median	10,32	10,31	46,5	46,5
Q1	9,23	9,67	43,0	43,9
Q3	11,16	10,92	52,0	49,7

The decision variable in this study was the number of daily tests obtained from the response variable minutes per test. In table VII, the mean values varied very little. The median values were almost identical, while the standard deviation was somewhat lower in the simulation than in the real system.

Finally, a Chi-square goodness-of-fit test was performed to determine if the proposed model was described well enough by comparing the expected and observed frequencies in the simulation. The Chi-square test of goodness of fit is described below.

- H_0 : The proposed model fits the real system
 H_1 : The proposed model does not fit the real system

TABLE VIII. CHI-SQUARE GOODNESS-OF-FIT TEST

N	GL	Chi-cuad.	P-value
88	5	3,62879	0,604

According to the values obtained in Table VIII, the p-value was greater than 0.05, with no statistical evidence to reject H_0 . Therefore, we could indicate that the proposed model of the initial situation in the laboratory fits the real system.

IV. RESULTS

After applying the Cause-effect diagram and the Pareto chart, it was determined that the main cause of the productivity problem in the bromatological laboratory is the slow internet speed. Several departments share the network simultaneously, which causes slowness when performing the analyses. This delays the processes of entering data, analyzing samples, storing data, and sending the results by email.

It is important to mention that in experiments carried out in the laboratory, it has been possible to verify a reduction in the percentage of time of the activities with different scenarios. Each scenario shows the increment or decrement of the network speed according to the number of connected users.

For this reason, the simulation was carried out in three scenarios: most likely, optimistic, and pessimistic, which include, through historical data, changes in the values of sample entry time, sample analysis time, data storage time, and results delivery time variables (T_{in} , T_{an} , T_{al} y T_{en} , respectively). These times are directly related to the speed problem of the network used in the laboratory.

The values of the sample preparation time variable (T_{pr}) continue with the same characteristics since they cannot be changed under any circumstances.

TABLE IX. SUMMARY OF THE SIMULATIONS IN DIFFERENT SCENARIOS

Statistic	Most likely Scenario		Optimistic Scenario		Pessimistic scenario	
	Minutes/ analysis	Analysis / day	Minutes/ analysis	Analysis / day	Minutes/ analysis	Analysis / day
Mean	8,41	59	6,61	75	11,80	41
Increase	-18%	23%	-36%	56%	15%	-15%

Table IX shows the means and how they increase or decrease according to their behavior during the execution of the simulations:

Most Likely Scenario: with the proposed model, the values of the variables T_{in} , T_{an} , T_{al} , and T_{en} were reduced by 25% using historical data when the network was shared with some users of another department. Then, we checked if the simulated data of the T_{in} , T_{an} , T_{al} , and T_{en} variables followed the same distributions as the real data. All variables were found to fit the distributions of the original data.

Applying the model, the output variable shows that generally, in the analysis time, there can be a decrease by 18% when the

network works at "normal" speed, and there is not much demand for users. So, the number of daily tests could increase by 23%.

Optimistic Scenario: with the proposed model, the values of the T_{in} , T_{an} , T_{al} , and T_{en} variables were reduced by 50% using historical data when the network worked without users of another department. In addition, it was found that all the distributions of each of the T_{in} , T_{an} , T_{al} , and T_{en} variables fit the distributions of the original data correspondingly. Applying the model, the time per analysis can be reduced by 36% when the network works at a "high" speed, and there is no demand from other users. Thus, the number of daily tests could increase by 56%.

Pessimistic scenario: with the proposed model, the values of the T_{in} , T_{an} , T_{al} , and T_{en} variables increased by 20% using historical data when the network was slow and was shared with all users of another department. In addition, it was found that all the distributions of each of the T_{in} , T_{an} , T_{al} , and T_{en} variables followed the respective distributions of the original data shown above. In this scenario, the variable time per analysis can increase by 15% when the network works at a "slow" speed, and there is a high demand for users. Therefore, the number of daily tests could decrease by 15%.

According to the values obtained, the optimistic scenario is the one with the greatest increase in productivity, up to 56%, given that the minute/analysis variable decreased up to 36%. However, productivity could be reduced if the times of the variables used in the model are reduced by less than 50%, as in the most likely scenario. Therefore, we would have more minutes per analysis and consequently less daily analysis.

IV. CONCLUSIONS

By addressing the root cause of lab result delivery delays (slow network speed), productivity, in terms of daily tests, can increase by up to 56%, avoiding additional costs such as overtime paid to laboratory analysts. However, the network speed will depend on the users connected and the resources the company wants to invest in increasing its capacity.

Both the real system data and the simulations in each scenario are based on historical data. The input and output variables were verified to ensure they followed the same behavior as the data of the real system.

In the present study, the Monte Carlo simulation was designed by applying each of its stages to sufficiently describe the real system of the data gathered.

Productivity, in terms of daily tests, can increase from 23% to 56% if we consider the times of input variables of the most

likely scenario and the optimistic scenario, which is possible as long as the network is not shared with other departments so that its speed does not drop. Another possibility may be to have a dedicated network for the laboratory.

[18]López de Bermúdez, N. Y. Aspectos teóricos en la generación de variables aleatorias. Facultad de Ciencias Naturales y Matemática., 28-34, 2016

REFERENCES

- [1] CNA.Cámara Nacional de Acuacultura - Estadísticas. Obtained from Cámara Nacional de Acuacultura: <http://www.cna-ecuador.com/estadisticas/>, 2019
- [2] Rodríguez, I. Simulación Análisis Salidas Simulación. Obtained from SlideShare: <https://es.slideshare.net/inesarodriguez58/informe-simulacion-analisisalidassimulacion>, 2016
- [3] Sotou, J. *Simulación Método Monte Carlo*. Obtained from Wordpress: <https://jaimesotou.files.wordpress.com/2011/05/metodo-montecarlo-03.pdf>, 2005
- [4] Palisade, C. *Monte Carlo Simulation*. Obtenido de Palisade: https://www.palisade.com/risk/monte_carlo_simulation.asp, 2019
- [5] Flanagan, R., Norman, G., & Chapman, R. *Risk Management And Construction*. London: John Wiley & Sons Incorporated, 2006
- [6] Morano, C., & Ferreira, M. Aplicação do Método de Monte Carlo em Análise de Riscos em Projetos de Construção. XXXV SBPO - Simpósio Brasileiro de Pesquisa Operacional (pág. 1). Natal: XXXV SBPO - Simpósio Brasileiro de Pesquisa Operacional, 2003
- [7] Montenegro, R. Diseño e implementación de un sistema de inventarios, aplicando simulación Montecarlo, en una empresa de Servicios Petroleros. Obtained from BIBDIGITAL Ecuola Politécnica Nacional: <https://bibdigital.epn.edu.ec/handle/15000/7871>, 2011
- [8] Amat, J. *Introducción a la Regresión Lineal Múltiple*. Obtained from RPubS - Regresión Lineal Múltiple en R: https://rpubs.com/Joaquin_AR/226291, 2020
- [9] Molina, P. Repositorio Institucional de la Universidad de las Fuerzas Armadas ESPE. de Valoración financiera utilizando opciones reales y el método de simulación Montecarlo: una metodología indispensable para escenarios de incertidumbre.: <http://repositorio.espe.edu.ec/bitstream/21000/4717/1/T-ESPE-032842%20-MED.pdf>, 2011
- [10] Medina, S., & Restrepo, J. Estimación de la utilidad en riesgo de una empresa de transmisión de energía eléctrica considerando variables económicas. *Cuadernos de Economía*, 32(59), pp. 103-13, 2013
- [11] Carollo, C. Regresión Lineal Simple. Obtained from Universidad de Santiago de Compostella: http://eio.usc.es/eipc1/BASE/BASEMASTER/FORMULARIOS-PHP-DPTO/MATERIALES/Mat_50140116_Regr_%20simple_2011_12.pdf, 2020
- [12] Salazar E., Alzate, W. Aplicación de la simulación de Monte Carlo en la proyección de estados de resultados. Un estudio de caso. *Revista Espacios*, Vol 39, no. 51, p.11. 2018
- [13] García, F., Sierra, J., & Guzmán, V. *Cálculo de Número Óptimo de Simulaciones*. Obtained from Espacio de Formación Multimodal - Uaem.mx: http://metabase.uaem.mx/bitstream/handle/123456789/564/Calculo_de_numero_optimo_de_simulaciones.pdf?sequence=1&isAllowed=y, 2005
- [14] Barcia, K., Cárdenas-Escobar, N., Gonzalez, V., Optimización del servicio de multiservidores basado en simulación matemática. *Proceedings of the 14th Latin American and Caribbean Conference for Engineering and Technology*, 2016.
- [15] Law, A. Simulation, Modeling and Analysis. En A. Law, *Simulation, Modeling and Analysis*, Tucson: McGraw-Hill International Editions, 2014, pp. 722-723.
- [16] López, M. *Pruebas de Normalidad: Prueba de Anderson-Darling*. Obtained from es.slideshare.net: <https://es.slideshare.net/leugimxw/pruebas-de-normalidad-prueba-de-andersondarling>, 2020
- [17] Santianello, S., & Filippo, F. Saber UCV: Aplicación del método de Monte Carlo en exploración y explotación de petróleo y gas natural. Obtained from Saber UCV: <http://saber.ucv.ve/handle/123456789/556>, 2011