Universidad de Guayaquil

Facultad de Ingeniería Química

INQUIDE

Ingeniería Química y Desarrollo

https://revistas.ug.edu.ec/index.php/iqd

ISSN – p: 1390 –9428 / ISSN – e: 3028-8533 / INQUIDE / Vol. XX / Nº XX

Prediction of moisture content in the cocoa drying process by simple linear regression.

Predicción del contenido de humedad en el proceso de secado del cacao mediante regresión lineal simple.

Francisco Javier Duque-Aldaz [1] * ; Edwin Ronny Haymacaña Moreno [2] ; Leonor Alejandrina Zapata Aspiazu [3] ; & Freddy Carrasco Choque [4]

Recibido: 29/04/2024 – Aceptado: 12/06/2024 – Publicado: 01/07/2024

*Autor para la correspondencia.

Abstract.
The research addressed the development of a predictive model for moisture control in the cocoa production process. Cocoa is an important crop for Ecuador, being the fourth largest exporter in the world in the last ten years. Moisture control during drying is critical to guarantee the quality and safety of the final product. The general objective was to establish a forecasting model for moisture control in the cocoa drying process using simple linear regression. First, the factors that affect the drying process were identified. Then, the variability of each factor was analyzed using historical data. Next, a mathematical model was developed using simple linear regression. Finally, the model was validated with real production data. The results showed that the model had a high predictive capacity of 90.16%, meaning that the variation in moisture could be explained by the independent variable. Validation with real data confirmed the goodness of fit. Initial moisture was the most influential factor, explaining this variation. It was concluded that the simple linear regression model was an effective tool for forecasting final moisture based on initial moisture. The model will allow companies to improve control of this critical parameter through informed measurements. The research was able to successfully validate the methodology proposed for this production problem.

Keywords.
Forecasting, Cocoa moisture, Production process, Simple linear regression, Quality control.

Resumen.
La investigación abordó el desarrollo de un modelo predictivo para el control de humedad en el proceso de producción de cacao. El cacao es un cultivo importante para Ecuador, siendo el cuarto exportador mundial en los últimos diez años. El control de humedad durante el secado es crítico para garantizar la calidad y seguridad del producto final. El objetivo general fue establecer un modelo de pronóstico para el control de humedad en el proceso de secado de cacao utilizando regresión lineal simple. En primera instancia, se identificaron los factores que inciden en el proceso de secado. Luego, se analizó la variabilidad de cada factor mediante datos históricos. Seguidamente, se desarrolló un modelo matemático utilizando regresión lineal simple. Finalmente, se validó el modelo con datos de producción reales. Los resultados mostraron que el modelo tuvo una alta capacidad predictiva de 90.16%, es decir que la variación de la humedad podía ser explicada por la variable independiente. La validación con datos reales confirmó la bondad del ajuste. La humedad inicial fue el factor más influyente, explicando esta variación. Se concluyó que el modelo de regresión lineal simple fue una herramienta eficaz para pronosticar la humedad final en base a la humedad inicial. El modelo permitirá a las empresas mejorar el control de este parámetro crítico mediante medidas informadas. La investigación pudo validar satisfactoriamente la metodología planteada para este problema productivo.

Palabras clave.
Pronóstico, Humedad del cacao, Proceso de producción, Regresión lineal simple, Control de calidad.

1. Introduction

Ecuador has established itself as the fourth largest cocoa exporter worldwide in the last 10 years. The provinces that stand out for their cocoa production are Guayas, Los Ríos, Esmeraldas, Manabí, El Oro, and Santa Elena. Cocoa represents an important item within the Ecuadorian economy, generating significant sources of income[1].

In food safety control environments, the parameters, characteristics, and specifications requested by external customers must be met, thus achieving customer satisfaction and a positive impact on production. Determining a forecast for cocoa moisture control in the production process is fundamental, as it is the starting point of said process. The company has large quantities of raw material entering the production line, but it presents critical points that must be evaluated, such as laboratory analyses to ensure the suitability of the cocoa bean [2].

When carrying out a forecast on cocoa moisture control, it is found that at strategic points of grain reception, the ranges are outside the specifications of the quality parameters, which generates anomalies during the production process.

Through a brainstorming session, some problems within the semi-finished products company have been identified. Firstly, complaints have been received from customers

[1] Universidad de Guayaquil; franscico.duquea@ug.edu.ec; https://orcid.org/0000-0001-9533-1635; Guayaquil; Ecuador.
[2] Instituto Superior Tecnológico ARGOS; e_haymacana@tecnologicoargos.edu.ec; https://orcid.org/0000-0002-8708-3894; Guayaquil; Ecuador.
[3] Universidad Técnica de Babahoyo; lzapata@utb.edu.ec ; https://orcid.org/0009-0003-1497-2273 ; Babahoyo; Ecuador.
[4] Universidad Nacional de Frontera; fcarrasco@unf.edu.pe ; https://orcid.org/0000-0002-4493-5567 ; Sullana; Perú.

because the final product is outside the specifications required in the technical sheet. This may be due to various factors within the process. Additionally, the raw material often has high ranges of microbiological loads when it enters the plant, as the cocoa bean is exposed to different factors from harvest to drying, which affects both the raw material and the final producto [3].

On the other hand, during the cocoa roasting process, there is no record of the bean's moisture, which harms the process and generates bottlenecks due to potential reprocessing. Having this subprocess controlled would be of great importance [4].

The objective of this research is to establish a forecast model for cocoa moisture control in the roasting production process, using multivariate regression.

To achieve the objective, the first step is to identify the key factors that influence the cocoa roasting production process. Next, we will proceed to analyze the variability of each of the factors that affect the cocoa roasting production process. Finally, we will present a mathematical model that ensures an accurate forecast for moisture control in the cocoa roasting process, using multivariate regression.

1.1.- Cocoa Roasting.

Cocoa roasting is an exothermic process that involves subjecting the beans to heating. It is a crucial stage that determines the final flavor and aroma of the product. The roasting temperature varies according to the type of bean, being higher for the "forastero bean" and medium to low for the "criollo bean" or "trinitario bean" [5].

This process pursues several fundamental objectives; firstly, it facilitates the separation of the shell from the bean, cracking it and allowing subsequent dehulling. Additionally, it sterilizes the beans by eliminating pathogens such as Salmonella or E. Coli, as well as other undesirable microorganisms. It is necessary to carefully control the temperature to avoid excessive roasting that could negatively affect the flavor [6].

Another key objective of roasting is to reduce the moisture content of the cocoa bean. Initially, the beans may have up to 8% moisture, but after roasting, this percentage decreases to approximately 2%. This moisture reduction is crucial for the subsequent stages of cocoa processing [7].

1.2.- Food Safety and Wholesomeness

In the food industry, specifically in the production of cocoa and its derivatives, safety and quality are fundamental aspects. It is essential to ensure that the final products are safe and suitable for human consumption. This implies that the raw material, that is, cocoa beans, must be free of impurities, contaminants, or any element that could be harmful to the health of the end consumer [8].

To achieve this objective, it is necessary to implement strict quality controls at all stages of the supply chain, from bean collection to finished product packaging. Work teams must assume responsibility for complying with applicable national and international requirements and regulations, both in production processes and final products. This includes following good manufacturing practices, implementing quality management systems, and conducting periodic analytical tests [9].

In addition to safety, sensory quality is also a key factor in the cocoa industry. Producers must ensure that the final products meet the flavor, aroma, and texture standards expected by consumers. This is achieved through rigorous control of processing conditions, the use of high-quality raw materials, and continuous training of personnel involved in production. Only through a comprehensive approach to quality and safety can consumer satisfaction and long-term success of the cocoa industry be guaranteed [10].

1.3.- Moisture Content and Quality Criteria in Cocoa Beans

For cocoa bean manufacturers, controlling the moisture content of the raw material is crucial. Cocoa beans are required to have approximately 7% moisture content. If this percentage exceeds 8%, it can lead to several negative consequences. Firstly, it would imply a loss of edible material, as excess moisture can promote the growth of molds and bacteria, representing a potential risk to food safety. Furthermore, moisture content above 8% can affect the yield of the production process [11].

On the other hand, if the moisture content of cocoa beans is below 6.5%, the shell becomes too fragile and the beans tend to disintegrate during processing. This would result in a high proportion of broken beans, which would also negatively impact the yield and quality of the final product. Therefore, maintaining an optimal moisture level between 6.5% and 8% is crucial to ensure quality and efficiency in cocoa production [12].

The excellence of cocoa encompasses various essential elements, such as its flavor, authenticity, and physical attributes, which directly influence production yield. Additionally, aspects such as traceability, geographical indications, and certification must be considered, reflecting the sustainability of production methods and product traceability. These factors are fundamental to ensuring quality and consumer confidence [13].

Cocoa quality specifications include: flavor, food safety and wholesomeness, physical characteristics (consistency and edible material yield), cocoa butter characteristics, color potential ("colorability"), and traceability, geographical indicators, and certification. Each of these aspects must be carefully evaluated and controlled to guarantee a final

ISSN – p: 1390 –9428 / ISSN – e: 3028-8533 / INQUIDE / Vol. XX / Nº XX

product of excellent quality that meets the highest standards of the cocoa industry [14].

1.4.- Production Forecast.

Forecasts play a critical role in the business world, as they provide an anticipated vision of the future and allow for informed and strategic decision-making. Based on projections and estimates of future events and trends, forecasts provide a solid foundation for financial planning, supply chain management, product development, market expansion, and human resource management. Thanks to these predictions, companies can anticipate potential changes, challenges, and opportunities, minimizing risks and maximizing competitive advantages [15].

There are different methods for forecasting production demand, and the choice of these methods depends on factors such as the time period of available data, the presence of patterns or trends, the seasonality of the product, and mainly the behavior or trend observed in product demand. Understanding the underlying causes that generate such demand is fundamental to selecting the appropriate forecasting method [16].

Some of the most commonly used methods are time series, simple and multiple linear regressions, and qualitative methods. Time series and regression methods are statistical or quantitative approaches that require the use of historical demand data to predict future demand by analyzing past patterns and trends. On the other hand, qualitative methods are based on incorporating value judgments from experts, focusing on their experience and subjective knowledge to evaluate non-quantifiable factors [17].

Forecasts are essential for business decision-making, allowing anticipation of changes and leveraging opportunities. The choice of forecasting method depends on various factors, such as available data, observed trends and patterns, and demand behavior. Both quantitative and qualitative approaches play an important role in developing accurate and reliable predictions [18].

1.5.- Simple Linear Regression

Simple linear regression is a statistical method used to model the relationship between two variables: a dependent variable (Y) and an independent variable (X). This model assumes that there is a linear relationship between both variables, represented by an equation of the form $Y = \beta 0 + \beta 1X + \varepsilon$, where $\beta 0$ is the y-intercept, $\beta 1$ is the slope of the line, and ε is the random error term. The objective of simple linear regression is to find the values of $\beta 0$ and $\beta 1$ that best fit the straight line to the observed data, minimizing the sum of squared residuals [19].

For simple linear regression to be valid and its results reliable, certain fundamental assumptions must be met. First, the relationship between the variables must be truly linear. Additionally, the residuals or errors must be normally distributed with a mean of zero and constant variance (homoscedasticity). It is also assumed that the errors are independent of each other and that there is no multicollinearity between independent variables (in the case of simple linear regression, there is only one independent variable) [20].

Simple linear regression finds application in various fields, including forecast models. In the context of forecasting, this technique can be used to predict the future value of a dependent variable (for example, product demand) based on a known independent variable (such as price or advertising). By fitting a straight line to historical data, the linear relationship can be extrapolated to make predictions about future values of the dependent variable [21].

However, it is important to note that simple linear regression is only an appropriate forecasting technique when the linearity assumption is met and when a relevant independent variable that significantly influences the dependent variable has been identified. Otherwise, it may be necessary to explore other forecasting methods, such as time series or non-linear models, to obtain more accurate predictions [22].

In addition to its use in forecasting, simple linear regression is also used in other areas, such as experimental data analysis, investigation of cause-effect relationships, and evaluation of the strength of association between two variables. Its simplicity and ease of interpretation make it a valuable tool in various areas of study and application [23].

2. Materials and Methods

2.1.- Materials

The materials used in this research are as follows:

- Cocoa beans: Cocoa beans were sourced from a plantation located in the province of Manabí, Ecuador.
- Drying equipment: A convection oven was used for drying the cocoa samples.
- Analytical balance: An analytical balance was used to determine the weight of the cocoa samples before and after drying.
- Statistical software: The statistical software R (version 4.0.2) was used to perform statistical analyses and develop the predictive model.

2.2.- Methods

The methodology used in this research is described below:

2.2.1 Sample Preparation

Initial cocoa bean samples were randomly selected from different lots and weighed in approximate quantities of 10 grams.
Random samples of cocoa were taken at the end of the drying process and weighed in approximate quantities of 10 grams.

INQUIDE
Ingeniería Química y Desarrollo
https://revistas.ug.edu.ec/index.php/iqd
ISSN – p: 1390 –9428 / ISSN – e: 3028-8533 / INQUIDE / Vol. XX / Nº XX

2.2.2. Determination of Moisture Content
The initial moisture content (IM) of the cocoa samples was recorded.
The final moisture content (FM) of the cocoa samples was recorded at the end of the drying process.

2.2.3. Development of the Predictive Model
To develop the predictive model, the simple linear regression method was used. The dependent variable was the final moisture content (FM) of the cocoa (in %), and the independent variable was the initial moisture content (IM) (in %). The statistical software R was used to perform the regression analysis and determine the coefficients of the model.

2.2.4. Hypothesis Testing and Assumptions of Simple Linear Regression
To verify the goodness of fit of the predictive model, hypothesis tests were conducted, and the assumptions of simple linear regression were evaluated. The following tests were used:

- Breusch-Pagan test, scatter plots, and residual plots: To verify linearity and the absence of patterns in the model's residuals.
- Kolmogorov-Smirnov (KS) normality test: To verify the normality of the model's residuals.
- Breusch-Pagan homoscedasticity test: To verify the homoscedasticity of the model's residuals.
- Durbin-Watson independence test: To verify the independence of the model's residuals.

The results of these tests are presented and discussed in the results and discussion section of this document.

3. Results.
3.1.- Data Visualization

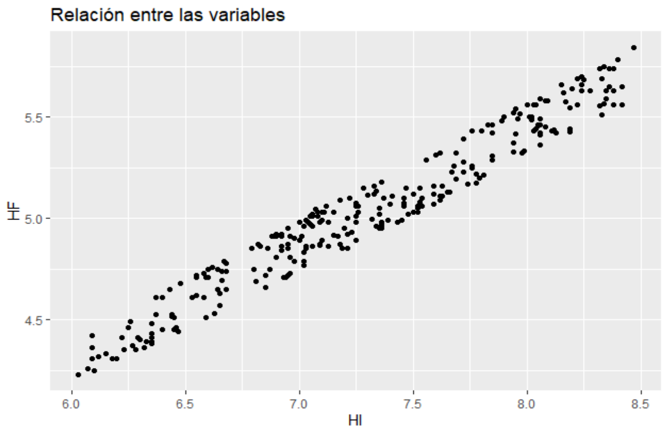

Figure 1: Relationship between the Variables

Figure 1 shows the relationship between the input variable IM and the output variable FM. As can be observed, there is a directly proportional linear relationship. Therefore, it can be visually inferred that a simple linear regression model can be applied.

3.2.- Model Summary:
Table 1: Residuals.

Min	1Q	Median	3Q	Max
-0.13875	-0.07564	-0.01100	0.08972	0.14476

Table 1, shows the descriptive statistics of the model's residuals, which are the differences between the observed values and the values predicted by the regression model.

Analysis of Each Statistic:

1. Minimum (Min): -0.13875
   - This value indicates that the minimum residual (or the lowest prediction error) is -0.13875.
   - A negative value implies that the model underestimated the observed value in that observation.
2. First Quartile (1Q): -0.07564
   - This value represents the residual at the 25th percentile of the lowest observations.
   - This suggests that 25% of the residuals are less than or equal to -0.07564.
3. Median: -0.01100
   - This is the value of the residual at the 50th percentile of the observations, i.e., the midpoint of the residual distribution.
   - A median value close to zero indicates that the model is predicting adequately, on average.
4. Third Quartile (3Q): 0.08972
   - This value represents the residual at the 75th percentile of the lowest observations.
   - This means that 75% of the residuals are less than or equal to 0.08972.
5. Maximum (Max): 0.14476
   - This value indicates that the maximum residual (or the highest prediction error) is 0.14476.
   - A positive value implies that the model overestimated the observed value in that observation.

Table 1 provides information on the distribution of the prediction errors from the linear regression model. Some key points to consider:

- The median close to zero indicates that, on average, the model is predicting adequately.
- The minimum and maximum values indicate the maximum magnitude of the prediction errors, both below and above the observed values.
- The quartiles give an idea of the dispersion of the residuals, which can be useful for evaluating the goodness of fit of the model.

ISSN – p: 1390 –9428 / ISSN – e: 3028-8533 / INQUIDE / Vol. XX / Nº XX

Table 2.- Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.805147	0.058285	13.81	<2e-16 ***
HI	0.579562	0.007946	72.94	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Interpretation of Table 2: Coefficients

1. Intercept:
   - The intercept value is 0.805147.
   - This value represents the expected value of the dependent variable (cocoa moisture) when the independent variable (IM) is equal to zero.
   - The standard error of the intercept is 0.058285.
   - The t-value of the intercept is 13.81, and the associated p-value is less than 2e-16 ($p < 0.001$), indicating that the intercept is statistically significant.

2. Coefficient of IM (HI):
   - The coefficient of the variable IM (HI) is 0.579562.
   - This value represents the expected change in cocoa moisture for each unit change in the IM variable.
   - The standard error of the coefficient of IM (HI) is 0.007946.
   - The t-value of the coefficient of IM (HI) is 72.94, and the associated p-value is less than 2e-16 ($p < 0.001$), indicating that the coefficient of IM (HI) is statistically significant.

The results from Table 2: Coefficients indicate that:

- The intercept of 0.805147 is statistically significant, suggesting a baseline value of cocoa moisture when IM is zero.
- The coefficient of IM (HI), 0.579562, is statistically significant, indicating that a change in IM is associated with a change in cocoa moisture.
- Since both terms are statistically significant, it can be concluded that the linear regression model is suitable for predicting cocoa moisture based on the IM variable.

Table 3: Summary of the Linear Regression Model

Residual Standard Error: 0.08599 with 287 degrees of freedom.
Multiple R-squared: 0.9488
Adjusted R-squared: 0.9486
F-statistic: 5320 on 1 and 287 DF, p-value: < 2.2e-16

Analysis and Interpretation of Table 3:

1. **Residual Standard Error:** 0.08599 with 287 degrees of freedom
   - The residual standard error is a measure of the precision of the regression model.
   - A value of 0.08599 indicates that, on average, the predicted values from the model deviate from the observed values by approximately 0.08599 units.
   - The 287 degrees of freedom represent the number of observations in the dataset minus the number of estimated parameters in the model.

2. **Multiple R-squared:** 0.9488, **Adjusted R-squared:** 0.9486
   - The Multiple R-squared is a measure of the goodness of fit of the model, indicating the proportion of variance in the dependent variable explained by the model.
   - A Multiple R-squared value of 0.9488 means that the model explains approximately 94.88% of the variance in the dependent variable.
   - The Adjusted R-squared is a version of the Multiple R-squared that adjusts for the number of predictors in the model, with a value of 0.9486.

3. **F-statistic:** 5320 on 1 and 287 DF, p-value: < 2.2e-16
   - The F-statistic is a hypothesis test that evaluates whether at least one of the regression coefficients is different from zero.
   - An F-value of 5320 with 1 and 287 degrees of freedom, and a p-value less than 2.2e-16 ($p < 0.001$), indicates that the regression model as a whole is statistically significant.
   - This suggests that at least one of the independent variables (in this case, IM) is useful for predicting the dependent variable (cocoa moisture).

The results shown in Table 3 indicate that the linear regression model is a good fit for the data, as it explains a high proportion of the variance in cocoa moisture (94.88%), and the model as a whole is statistically significant. This suggests that the IM variable is a good predictor of final cocoa moisture.

3.3.- Visualization of the regression line

ISSN – p: 1390 –9428 / ISSN – e: 3028-8533 / INQUIDE / Vol. XX / Nº XX

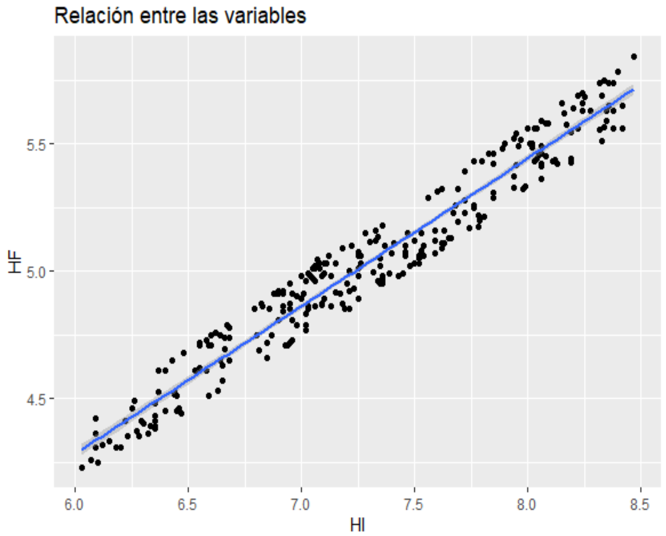

Figure 2: Regression Line

In Figure 2, the trend line between IM and FM can be observed.
The blue line represents the linear trend line, showing that the data increase in a straight line at a constant rate.

Assumptions of Simple Linear Regression
3.4 Assumption 1: Linearity
Breusch-Pagan Test
Data: Model
BP = 1.8583, df = 1, p-value = 0.1728
Pearson Correlation Coefficient: 0.9740694

Breusch-Pagan Test:

- The p-value of the Breusch-Pagan test is 0.1728, which is greater than the significance level (0.05).
- This indicates that there is not enough evidence to reject the null hypothesis that the relationship between the dependent variable and the independent variable is linear.
- Therefore, the results of the Breusch-Pagan test suggest that the linearity assumption is met.

Pearson Correlation Coefficient:

- The Pearson correlation coefficient is 0.9740694, indicating a strong positive linear relationship between the dependent variable and the independent variable.

The results of the Breusch-Pagan test and the high Pearson correlation coefficient provide evidence that the linearity assumption is met for the proposed simple linear regression model.

Residuals vs. Fitted Values Plot

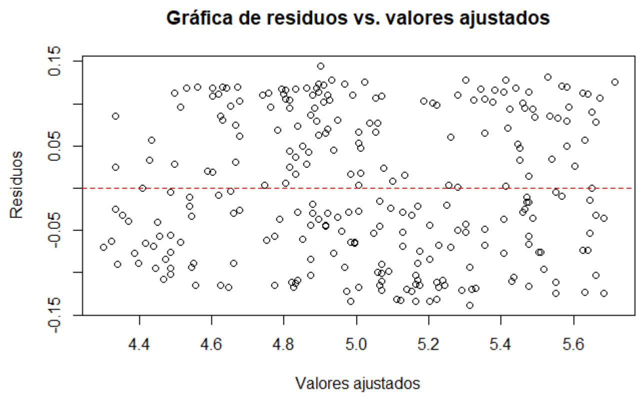

Figure 3: Residuals vs. Fitted Values Plot

Figure 3 visually shows that the residuals are evenly distributed across the range of fitted values.
These residuals do not exhibit a pattern, indicating that the model is acceptable in the sense that the residuals are independent of the fitted values.

Residuals vs. Independent Variable Plot

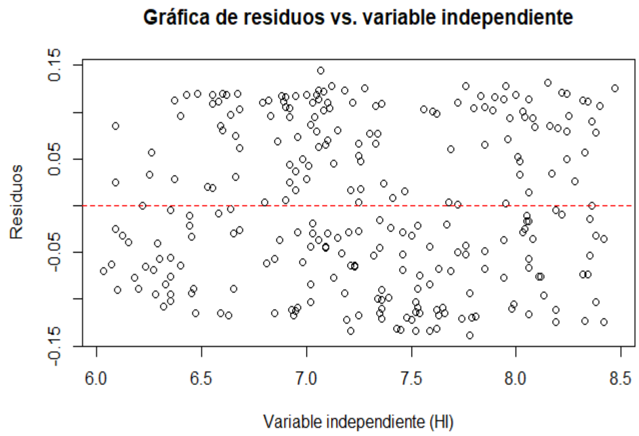

Figure 4: Residuals vs. Independent Variable Plot

Figure 4 visually shows that the residuals are evenly distributed across the values of the independent variable IM.

These residuals do not exhibit a pattern, indicating that the model is acceptable in the sense that the residuals are independent of the independent variable IM.

3.5 Assumption 2: Normality of Residuals
Kolmogorov-Smirnov Test
Kolmogorov-Smirnov Test (KS)
Test Statistic: 1.4448233
P-value: 0.0628

Interpretation:

ISSN – p: 1390 –9428 / ISSN – e: 3028-8533 / INQUIDE / Vol. XX / Nº XX

- The p-value of the Kolmogorov-Smirnov test is 0.0628, which is greater than the significance level (0.05).
- This indicates that there is not enough evidence to reject the null hypothesis that the residuals follow a normal distribution.
- Therefore, the results of the Kolmogorov-Smirnov test suggest that the assumption of normality of residuals is met.

Implications:

- With the assumption of normality of residuals met, statistical inferences such as confidence intervals and hypothesis tests will be valid.

The results of the Kolmogorov-Smirnov test indicate that the assumption of normality of residuals is met for the proposed simple linear regression model.

3.6 Assumption 3: Homoscedasticity of Residuals

Breusch-Pagan Test
Data: Model
BP = 1.8583, df = 1, p-value = 0.1728

Breusch-Pagan Test:

- Breusch-Pagan statistic: 1.8583
- Degrees of freedom (df): 1
- P-value (p-value): 0.1728

Interpretation:

- The p-value of the Breusch-Pagan test is 0.1728, which is greater than the significance level (0.05).
- This indicates that there is not enough evidence to reject the null hypothesis of homoscedasticity.
- Therefore, the results of the Breusch-Pagan test suggest that the assumption of homoscedasticity of residuals is met. Implications:
- When the assumption of homoscedasticity is met, it means that the variance of the residuals is constant across predicted values.

The results of the Breusch-Pagan test indicate that the proposed simple linear regression model meets the assumption of homoscedasticity of residuals.

3.7 Assumption 4: Independence of Residuals.

Durbin-Watson Test
Data: Model
DW = 1.5302

Independence of residuals can be tested with the Durbin-Watson statistic; this statistic takes values between 0 and 4.

If the Durbin-Watson statistic is between 1.5 and 2.5, it is assumed that the residuals are independent.

For our case, the Durbin-Watson statistic is 1.5302, indicating that the residuals are independent.

The results of the Durbin-Watson test indicate that the proposed simple linear regression model meets the assumption of independence of residuals.

3.8.- Proposed Linear Regression Equation Model

Based on the results from Table 2 Coefficients, the model is developed as follows:

FM = 0.805147 + 0.579562 IM

4. Discussion

The present study aimed to establish a forecasting model for controlling cocoa moisture during the roasting production process using multivariate regression. The results obtained demonstrate that a simple linear regression equation model was successfully developed to forecast the final cocoa moisture (FM) based on initial moisture (IM) with high precision.

The proposed model fits the data well, explaining a high proportion of the variance in cocoa moisture (94.86%). Furthermore, the model as a whole is statistically significant, indicating that IM is a strong predictor of FM. This finding is consistent with prior research demonstrating that initial cocoa moisture is a critical factor in the roasting process and significantly influences final product quality [24].

Moreover, the assumptions of simple linear regression were verified, including linearity, normality, homoscedasticity, and independence of residuals. The results of the Breusch-Pagan test and the high Pearson correlation coefficient provide evidence that the linearity assumption is met. Additionally, the Kolmogorov-Smirnov test results indicate that the normality assumption of residuals is satisfied. The Breusch-Pagan test also yielded results suggesting that the assumption of homoscedasticity of residuals holds true, and the Durbin-Watson test indicated that residuals are independent. These findings align with the assumptions of simple linear regression, affirming the validity and reliability of the model.

It is crucial to highlight that controlling cocoa moisture in the production process is essential for ensuring the quality and food safety of the final product. Previous studies have shown that appropriate cocoa moisture content allows for achieving desired sensory characteristics in the final product and prevents the proliferation of pathogenic microorganisms, mold, and bacteria [25]. Therefore, the model proposed in this study can serve as a valuable tool for cocoa producers and processing industries, enabling accurate forecasting of final cocoa moisture and informed decision-making in the production process.

The findings of this study demonstrate the successful development of a simple linear regression equation model to predict final cocoa moisture based on initial moisture with high precision. Additionally, the assumptions of simple linear regression were validated, suggesting that the model is valid and reliable. These findings are consistent

ISSN – p: 1390 –9428 / ISSN – e: 3028-8533 / INQUIDE / Vol. XX / Nº XX

with previous research and hold significant implications for cocoa quality control and food safety in the production process. Further research in this area is recommended to enhance the model's accuracy and applicability in the industry.

5.- Conclusions

This research aimed to develop a predictive model for controlling moisture in the cocoa production process using simple linear regression. The results from analyzing 289 observations reveal that the proposed model exhibits high predictive capability, with an R-squared value of 0.9016. This indicates that 90.16% of the variation in moisture can be explained by the independent variable, which in this case is the moisture content of the input material (HI).

The findings of this study hold significant implications for cocoa exporting companies, as moisture control is a critical factor in ensuring the quality and safety of the final product. The proposed model can be utilized to predict the final moisture content of cocoa based on the input material's moisture level, enabling more efficient and effective moisture control measures. This can lead to cost savings, improved product quality, and increased customer satisfaction.
One of the key contributions of this study is the application of simple linear regression to the moisture control problem in cocoa production. While previous studies have employed more complex statistical models, this study demonstrates that a simple linear regression model can be highly effective in predicting moisture levels. This has practical implications, as it means companies can implement moisture control measures without the need for expensive and complex statistical programs.

Another significant contribution of this study is the identification of key factors influencing moisture levels in the cocoa production process. The results indicate that the moisture content of the input material is the most critical factor, accounting for 90.16% of the moisture variation. This suggests that efforts to control moisture should focus on managing the moisture content of the input material, rather than attempting to manipulate other factors that have a lesser impact on moisture levels.

The conclusions drawn from this study also have important implications for future research. Further studies could explore the use of more complex statistical models to enhance the predictive capability of the proposed model. Additionally, future research could investigate the impact of other factors on moisture levels, such as temperature, airflow, and processing time. This could lead to the development of more sophisticated moisture control measures that take into account multiple factors.

In conclusion, this study has demonstrated the effectiveness of simple linear regression in predicting final moisture levels in the cocoa production process. The proposed model has practical implications for cocoa exporting companies, as it can be used to enhance moisture control measures and ensure the quality and safety of the final product. The findings also underscore the need for continued research into factors influencing moisture levels and the development of more sophisticated moisture control measures.

6.- Referencias.

[1] «Ministerio de Agricultura y Ganadería,» 2023. [En línea]. Available: https://www.agricultura.gob.ec/ecuador-es-el-primer-exportador-de-cacao-en-grano-de-america/.

[2] F. Duque-Aldaz, E. Pazán Gómez, W. Villamagua Castillo y A. López Vargas, «Sistema de gestión de seguridad y salud ocupacional según ISO:45001 en laboratorio cosmético y natural,» *Revista Científica Ciencia Y Tecnología*, vol. 24, nº 41, 2024.

[3] J. Aldas-Morejon, O.-T. Víctor, K. Revilla-Escobar, M. Carrillo-Pisco y D. Sánchez-Aguilera, «Incidencia del tostado sobre las características fisicoquímicas y alcaloides de la cascarilla de cacao (Theobroma cacao) y su efecto en las propiedades organolépticas de una infusión,» *Agroindustrial Science*, vol. 13, nº 1, pp. 15-21, 2023.

[4] V. E. García Casas, F. J. Duque-Aldaz y M. Cárdenas Calle, «Diseño de un plan de buenas prácticas de manufactura para las cabañas restaurantes en el cantón General Villamil Playas,» *Magazine De Las Ciencias: Revista De Investigación E Innovación*, vol. 8, nº 4, p. 58–76, 2023.

[5] V. Rejas Heredia, «Cambios fisicoquímicos y organolépticos en el tostado del cacao,» *Revista Ingeniería,*, vol. 5, nº 11, p. 39–58, 2021.

[6] V. E. García Casas y F. J. Duque-Aldaz, «Mejora de capacidades en el manejo de protocolos de manipulación, higiene y bioseguridad para las cabañas-restaurantes del cantón Playas en tiempos de Covid-19,» *Journal of Science and Research,*, vol. 8, nº 3, p. 192–209, 2022.

[7] M. C. J. Ruiz Lau y S. Vegas Chiyón, «Evaluación paramétrica en tostado de cacao piurano con diseño factorial 3k, y determinación del perfil sensorial,» Universidad de Piura, Piura, 2020.

[8] L. F. Pastorino, «Seguridad alimentaria: un concepto exagerado,» *Przegląd Prawa Rolnego*, vol. 2, nº 27, p. 183–206, 2020.

[9] J. M. M. Barandiarán Falla, E. S. Cuyo Gonzales, D. Medina Aguilar, M. Medina Simpertigues y R. J. Tuesta Tello, «SEGURIDAD ALIMENTARIA EN EL ESTADO DE SALUD DE LA POBLACIÓN DEL DEPARTAMENTO LAMBAYEQUE-PERÚ,» *REVISTA CURAE*, vol. 4, nº 4, p. 1–11, 2022.

[10] G. R. Pérez y Q. Y. Silva, «Enfoques y factores asociados a la inseguridad alimentaria,» *Revista Salud Pública y Nutrición*, vol. 18, nº 1, 2019.

[11] J. N. Saza Coaji y J. A. Jiménez Forero, «DETERMINACIÓN DE CONDICIONES AMBIENTALES PARA LA CONSERVACIÓN DE GRANOS DE CACAO (THEOBROMA CACAO L) DESHIDRATADO DURANTE EL ALMACENAMIENTO.,» *Sistemas de Producción Agroecológicos*, vol. 11, nº 1, pp. 2-32, 2020.

[12] E. Garcia Gonzalez, A. M. Serna Murillo, D. A. Córdoba Pantoja, J. G. Marín Aricapa, C. Montalvo Rodríguez y G. A. Ordoñez Narváez, «Estudio de la fermentación espontánea de cacao (Theobroma cacao L.) y evaluación de la calidad de los granos en una unidad productiva a pequeña escala,» *AGRICULTURAL BIOTECHNOLOGY*, vol. 6, nº 1, p. 29–40, 2019.

[13] R. Valverde - Zurita, R. Castillo - Bermeo, N. Jumbo - Benites y P. Fernández - Guarnizo, «El cacao fino de aroma (Theobroma cacao

L.) del cantón El Pangui- Ecuador, posible alternativa para elaborar chocolate gourmet,» *Revista Investigación Agraria*, vol. 5, nº 3, p. 14–27, 2023.

[14] J. Nogales y D. Ruíz, «La calidad del Cacao ¿Dónde comienza y dónde termina?,» *INIA Divulga*, vol. 42, nº 42, pp. 35-43, 2019.

[15] J. C. Jiménez Novillo, H. Carvajal Romero y H. Vite Cevallos, «Análisis del pronóstico de las exportaciones del camarón en el Ecuador a partir del año 2019,» *REMCA*, vol. 4, nº 1, 2021.

[16] J. M. Pastorino y M. Cornejo, «Pronóstico de Demanda como herramienta para la producción de vinos,» Universidad de Torcuato Di Tella, Buenos Aires,, 2023.

[17] R. Perdigón Llanes y N. González Benítez, «Una revisión bibliográfica sobre modelos para predecir las producciones de leche,» *Revista Ingeniería Agrícola*, vol. 10, nº 4, 2020.

[18] D. Bermúdez y M. González, «Producción de petróleo y gas en Venezuela: análisis mediante la función de Cobb-Douglas,» *Revista UIS Ingenierías*, vol. 18, nº 3, pp. 183-191, 2019.

[19] R. Vilá Baños, M. Torrado-Fonseca y M. Reguante Alvarez, «Análisis de regresión lineal múltiple con SPSS: un ejemplo práctico,» *REIRE Revista de Innovación E Investigación En Educación*, vol. 12, nº 2, pp. 1-10, 2019.

[20] J. Hernández-Lalinde, J.-F. Espinosa-Castro, D. García Álvarez y V. Bermúdez-Pirela, «Sobre el uso adecuado de la regresión lineal: conceptualización básica mediante un ejemplo aplicado a las ciencias de la salud,» *AVFT – Archivos Venezolanos De Farmacología Y Terapéutica*, vol. 38, nº 5, 2020.

[21] A. Cárdenas-Pérez y I. Benavides Echeverría, «Explicación del crecimiento económico en la Economía Popular y Solidaria mediante la aplicación del modelo econométrico de Regresión Lineal y Múltiple,» *Revista Publicando*, vol. 8, nº 28, 2021.

[22] C. M. Bermejo Salmon, «Tratamiento del nivel de competencias laborales desde la regresión lineal simple,» *Retos de la Dirección*, vol. 14, nº 1, 2020.

[23] A. P. Garcia Barreda y M. E. Velázquez Tejeda, «Propuesta metodológica para el análisis de regresión lineal simple en los estudiantes de la carrera de marketing de un instituto superior privado de Lima,» Universidad San Ignacio de Loyola, Lima, 2022.

[24] B. S. Rosales-Valdívia, García-Curiel, Laura, J. G. Pérez-Flores, E. Contreras-López, E. Pérez-Escalante y C. García-Mora, «Influencia de la fermentación del cacao y del uso de cultivos iniciadores sobre las características organolépticas del chocolate: un análisis integral,» *Pädi Boletín Científico De Ciencias Básicas E Ingenierías Del ICBI*, vol. 12, nº 23, 2024.

[25] J. E. Pujota Quimbiamba, «Evaluación de los parámetros tiempo y temperatura en el proceso de tostado de dos variedades de cacao sobre la actividad antioxidante y atributos sensoriales en pasta,» Universidad Técnica del Norte, 2023.

7.- Anexos (En caso de que existan)

Código en R utilizado para el desarrollo de la investigación.

```
# 1. Carga de las librerías:
#install.packages("tidyverse")
#install.packages("car")
#install.packages("lmtest")
library(tidyverse) # Librería que contiene varias funciones útiles para el análisis de datos
library(ggplot2)  # Librería para la creación de gráficos
library(openxlsx)
library(readxl)
library(lmtest)
library(stats)

#-------------------
setwd("D:/Lenovo/Desktop/ELABORACIÓN DE ARTÍCULO CIENTÍFICO")
getwd()
dir()
# 2 Cargar el archivo Excel
##salaries <- read.xlsx("salario.xlsx")
install.packages("openxlsx")
library(openxlsx)
datos <- read.xlsx("DT4.xlsx")
View(datos)
# Ver el contenido del data frame
head(datos)

#=========================
# INSTALO PAQUETES
# ========================
install.packages("dplyr")
install.packages("ggplot2")
install.packages("readxl")
install.packages("cowplot")
install.packages("gmodels")
install.packages("Hmisc")
install.packages("ggthemes")
#=========================
# ACTIVO PAQUETES
# ========================
library("dplyr")
library("ggplot2")
library("readxl")
library("gmodels")
library("Hmisc")
library("ggthemes")
library("cowplot")

# 3. Visualización de los datos:
ggplot(datos, aes(x = HI, y = HF)) +
  geom_point() +
  labs(title = "Relación entre las variables")

#----------------------------
# 4. Estimación del modelo de regresión lineal:

modelo <- lm(HF ~ HI, data = datos)

#----------------------------
# 5. Resumen del modelo:
summary(modelo)

# install.packages("knitr")
# library(knitr)
# knitr::kable(summary(modelo)$coefficients)
# knitr::kable(summary(modelo))
```

ISSN – p: 1390 –9428 / ISSN – e: 3028-8533 / INQUIDE / Vol. XX / Nº XX

```
#------------------------------
# 6. Visualización de la línea de regresión:
ggplot(datos, aes(x = HI, y = HF)) +
  geom_point() +
  labs(title = "Relación entre las variables")

ggplot(datos, aes(x = HI, y = HF)) +
  geom_point() +
  labs(title = "Relación entre las variables") +
  geom_smooth(method = "lm")

# Ecuación de regresión lineal
ecuacion <- paste("HF ~", format(coef(modelo), digits =
2))

# Mostrar la ecuación
ecuacion

# ==================================
# 7. Gráfica de dispersión
ggplot(datos, aes(x = HI, y = HF)) +
  geom_point() +
  labs(title = "Relación entre las variables")

# ==================================
# Supuestos de La Regresión Lineal Simple.

# ===============================
# 8. Prueba de linealidad

# Crear la gráfica QQ
ggplot(modelo, aes(x = ".resid", y = ".fitted")) +
  geom_abline(lty = 2) +
  labs(title = "Gráfica QQ de los residuos") +
  annotate("point", x = ".resid", y = ".fitted", size = 1.5)

# Realizar la prueba de Breusch-Pagan
library(lmtest)
bptest(modelo)

# Calcular el coeficiente de correlación de Pearson
cor_pearson <- cor(datos$HF, datos$HI)

# Imprimir el resultado
print(cor_pearson)

# Gráfica de residuos vs. valores ajustados:
# Obtener los residuos y valores ajustados del modelo
residuos <- residuals(modelo)
valores_ajustados <- fitted(modelo)

# Crear la gráfica de residuos vs. valores ajustados
plot(valores_ajustados, residuos, xlab = "Valores
ajustados", ylab = "Residuos",
     main = "Gráfica de residuos vs. valores ajustados")
abline(h = 0, lty = 2, col = "red")  # Agregar una línea
horizontal en y = 0

# Gráfica de residuos vs. variable independiente:

# Obtener los residuos del modelo
residuos <- residuals(modelo)

# Crear la gráfica de residuos vs. variable independiente
plot(datos$HI, residuos, xlab = "Variable independiente
(HI)",
     ylab = "Residuos", main = "Gráfica de residuos vs.
variable independiente")
abline(h = 0, lty = 2, col = "red")

# ================================
# 9. Prueba de Homocedasticidad:

library(lmtest)
# Creamos un modelo de regresión lineal simple
fit <- lm(HF ~ HI, data = datos)

# Aplicamos la prueba de Breusch-Pagan
bptest(fit)

# ================================
# 10. Prueba de Normalidad:

# Utilizamos la función ks.test() para realizar la prueba de
normalidad
# Prueba de Kolmogorov-Smirnov (KS)
install.packages("nortest")
librar("nortest")
residuos <- unique(residuos)
resultados <- ks.test(residuos, "pnorm")
# Imprimimos los resultados
cat("Statística de la prueba:", resultados$statistic, "\n")
cat("Valor crítico:", resultados$critical, "\n")
cat("P-valor:", format(resultados$p.value, digits = 10),
"\n")
cat("P-valor:", resultados$p.value, "\n")

# Obtener los residuos del modelo
residuos <- residuals(modelo)

# Gráfico de histograma de los residuos
ggplot(data.frame(residuos = residuos), aes(x = residuos))
+
  geom_histogram(aes(y = ..density..), color = "black", fill
= "white") +
  geom_density(alpha = 0.2, fill = "#FF6666") +
  labs(title = "Histograma de los residuos",
       x = "Residuos", y = "Densidad")

# Gráfico de probabilidad normal (Q-Q plot)
ggplot(data.frame(residuos = residuos), aes(sample =
residuos)) +
```

ISSN – p: 1390 –9428 / ISSN – e: 3028-8533 / INQUIDE / Vol. XX / Nº XX

```
  stat_qq() +
  stat_qq_line() +
  labs(title = "Gráfico de probabilidad normal (Q-Q plot)",
       x = "Valores teóricos", y = "Valores observados")

# ================================
# 11. Prueba de Independencia:

# Prueba de Durbin-Watson
# install.packages("lmtest")
library(lmtest)
dwtest(modelo)

# ===================================
# 12. Prueba de No hay colinealidad:
#En el caso de la regresión lineal simple, este supuesto se
cumple automáticamente, ya que solo hay una variable
independiente.
```