



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

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

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.

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.

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.

1. Introducción

Ecuador se ha establecido como el cuarto mayor exportador de cacao a nivel mundial en los últimos 10 años. Las provincias que se destacan por su producción de cacao son Guayas, Los Ríos, Esmeraldas, Manabí, El Oro y Santa Elena. El cacao representa un rubro importante dentro de la economía ecuatoriana, generando grandes fuentes de ingresos [1].

En los entornos de control de inocuidad alimentaria, se deben cumplir con los parámetros, características y especificaciones solicitadas por los clientes externos, logrando así la satisfacción del cliente y un impacto positivo en la producción. La determinación de un pronóstico para el control de la humedad del cacao en el proceso productivo es

fundamental, ya que es el punto de partida de dicho proceso. La empresa cuenta con grandes cantidades de materia prima que ingresan a la línea de producción, pero presenta puntos críticos que deben ser evaluados, como los análisis de laboratorio para asegurar la aptitud del grano de cacao [2].

Al llevar a cabo un pronóstico sobre el control de la humedad del cacao, se presenta que en puntos estratégicos de la recepción del grano, los rangos se encuentran fuera de las especificaciones de los parámetros de calidad, lo que genera anomalías durante el proceso de producción.

A través de una lluvia de ideas, se han identificado algunas problemáticas dentro de la empresa de semielaborados. En primer lugar, se han recibido reclamos por parte de los clientes debido a que el producto final se encuentra fuera de

¹ Universidad de Guayaquil; francisco.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.



las especificaciones requeridas en la ficha técnica. Esto puede deberse a diversos factores dentro del proceso. Además, la materia prima suele tener altos rangos de cargas microbiológicas cuando ingresa a la planta, ya que el grano de cacao se expone a diferentes factores desde la cosecha hasta el secado, lo que afecta tanto a la materia prima como al producto final [3].

Por otro lado, durante el proceso de tostado de cacao no se lleva un registro de la humedad del grano, lo cual perjudica el proceso y genera cuellos de botella por los reprocesos que puedan surgir. Tener este subproceso controlado sería de gran importancia [4].

La presente investigación tiene por objetivo establecer un modelo de pronóstico para el control de humedad del cacao en el proceso de producción de tostado, utilizando regresión multivariada.

Para el cumplimiento del objetivo se plantea como primer paso el identificar los factores claves que inciden en el proceso de producción de tostado del cacao, a continuación, se procederá a analizar la variabilidad de cada uno de los factores que inciden en el proceso de producción de tostado del cacao; y finalmente se mostrará un modelo matemático que asegure un correcto pronóstico para el control de la humedad en el proceso de tostado del cacao, utilizando regresión multivariada.

1.1.- Tostado del cacao.

El tostado del cacao es un proceso exotérmico que implica someter los granos a calentamiento. Es una etapa crucial que determina el sabor y aroma final del producto. La temperatura de tostado varía según el tipo de grano, siendo más alta para el "grano forastero" y media a baja para el "grano criollo" o "grano trinitario" [5].

Este proceso persigue varios objetivos fundamentales; en primer lugar, facilita la separación de la cáscara del grano, agrietándola y permitiendo el descascarillado posterior. Además, esteriliza los granos eliminando patógenos como Salmonella o E. Coli, así como otros microorganismos indeseables. Es necesario controlar cuidadosamente la temperatura para evitar un tostado excesivo que pueda afectar negativamente al sabor [6].

Otro objetivo clave del tostado es reducir la humedad del grano de cacao. Inicialmente, los granos pueden tener hasta un 8% de humedad, pero tras el tostado, este porcentaje disminuye hasta aproximadamente un 2%. Esta reducción de humedad es crucial para las etapas posteriores del procesamiento del cacao [7].

1.2.- Seguridad Alimentaria y Salubridad.

En la industria alimentaria, específicamente en la producción de cacao y sus derivados, la inocuidad y la calidad son aspectos fundamentales. Es esencial garantizar

que los productos finales sean seguros y aptos para el consumo humano. Esto implica que la materia prima, es decir, los granos de cacao, debe estar libre de impurezas, contaminantes o cualquier elemento que pueda resultar dañino para la salud del consumidor final [8].

Para lograr este objetivo, es necesario implementar estrictos controles de calidad en todas las etapas de la cadena de suministro, desde la recolección de los granos hasta el envasado del producto terminado. Los equipos de trabajo deben asumir la responsabilidad de cumplir con los requisitos y normativas nacionales e internacionales aplicables, tanto en los procesos de producción como en los productos finales. Esto incluye el seguimiento de buenas prácticas de manufactura, la implementación de sistemas de gestión de calidad y la realización de pruebas analíticas periódicas [9].

Además de la inocuidad, la calidad sensorial también es un factor clave en la industria del cacao. Los productores deben asegurarse de que los productos finales cumplan con los estándares de sabor, aroma y textura esperados por los consumidores. Esto se logra mediante un control riguroso de las condiciones de procesamiento, el uso de materias primas de alta calidad y la capacitación continua del personal involucrado en la producción. Sólo a través de un enfoque integral en la calidad e inocuidad se puede garantizar la satisfacción del consumidor y el éxito a largo plazo de la industria del cacao [10].

1.3.- Contenido de humedad y Criterios de calidad en el grano del cacao

Para los fabricantes de cacao en grano, es crucial controlar el contenido de humedad de la materia prima. Se requiere que los granos de cacao tengan aproximadamente un 7% de humedad. Si este porcentaje supera el 8%, puede provocar varias consecuencias negativas. En primer lugar, implicaría una pérdida de material comestible, ya que el exceso de humedad puede favorecer el crecimiento de mohos y bacterias, representando un riesgo potencial para la seguridad alimentaria. Además, una humedad superior al 8% puede afectar el rendimiento del proceso productivo [11].

Por otro lado, si el contenido de humedad de los granos de cacao es inferior al 6.5%, la cáscara se vuelve demasiado frágil y los granos tienden a desintegrarse durante el procesamiento. Esto daría lugar a una alta proporción de granos rotos, lo que también impactaría negativamente en el rendimiento y la calidad del producto final. Por lo tanto, mantener un nivel óptimo de humedad entre el 6.5% y el 8% es crucial para garantizar la calidad y la eficiencia en la producción de cacao [12].

La excelencia del cacao abarca diversos elementos esenciales, como su sabor, autenticidad y atributos físicos, los cuales influyen directamente en el rendimiento de la



producción. Además, se deben considerar aspectos como el rastreo, los indicios geográficos y la certificación, que reflejan la sostenibilidad de los métodos de elaboración y la trazabilidad del producto. Estos factores son fundamentales para asegurar la calidad y la confianza de los consumidores [13].

Las especificaciones de calidad del cacao incluyen: el sabor, la inocuidad y salubridad de los alimentos, las características físicas (consistencia y rendimiento de material comestible), las características de la manteca de cacao, el potencial de color ("colorabilidad"), y la trazabilidad, indicadores geográficos y certificación. Cada uno de estos aspectos debe ser cuidadosamente evaluado y controlado para garantizar un producto final de excelente calidad, que cumpla con los estándares más altos de la industria del cacao [14].

1.4.- Pronóstico de producción.

Los pronósticos desempeñan un papel crítico en el ámbito empresarial, ya que brindan una visión anticipada del futuro y permiten tomar decisiones informadas y estratégicas. Al basarse en proyecciones y estimaciones de eventos y tendencias futuras, los pronósticos proporcionan una base sólida para la planificación financiera, la gestión de la cadena de suministro, el desarrollo de productos, la expansión del mercado y la gestión de recursos humanos. Gracias a estas predicciones, las empresas pueden anticiparse a posibles cambios, desafíos y oportunidades, minimizando los riesgos y maximizando las ventajas competitivas [15].

Existen diferentes métodos para pronosticar la demanda de producción, y la elección de estos métodos depende de factores como el periodo de tiempo de los datos disponibles, la presencia de patrones o tendencias, la estacionalidad del producto y, principalmente, el comportamiento o tendencia observada en la demanda del producto. Comprender las causas subyacentes que generan dicha demanda es fundamental para seleccionar el método de pronóstico adecuado [16].

Algunos de los métodos más utilizados son las series de tiempo, las regresiones lineales simples y múltiples, y los métodos cualitativos. Las series de tiempo y los métodos de regresión son enfoques estadísticos o cuantitativos que requieren el uso de datos históricos de la demanda para predecir la demanda futura mediante el análisis de patrones y tendencias pasadas. Por otro lado, los métodos cualitativos se basan en la incorporación de juicios de valor por parte de expertos, centrándose en su experiencia y conocimiento subjetivo para evaluar factores no cuantificables [17].

En resumen, los pronósticos son esenciales para la toma de decisiones empresariales, permitiendo anticiparse a cambios y aprovechar oportunidades. La elección del método de pronóstico depende de diversos factores, como

los datos disponibles, las tendencias y patrones observados, y el comportamiento de la demanda. Tanto los enfoques cuantitativos como los cualitativos juegan un papel importante en la elaboración de predicciones precisas y confiables [18].

1.5.- Regresión lineal simple.

La regresión lineal simple es un método estadístico utilizado para modelar la relación entre dos variables: una variable dependiente (Y) y una variable independiente (X). Este modelo supone que existe una relación lineal entre ambas variables, representada por una ecuación de la forma $Y = \beta_0 + \beta_1 X + \varepsilon$, donde β_0 es la ordenada al origen, β_1 es la pendiente de la recta y ε es el término de error aleatorio. El objetivo de la regresión lineal simple es encontrar los valores de β_0 y β_1 que mejor ajusten la línea recta a los datos observados, minimizando la suma de los residuos al cuadrado [19].

Para que la regresión lineal simple sea válida y sus resultados sean confiables, se deben cumplir ciertos supuestos fundamentales. En primer lugar, la relación entre las variables debe ser verdaderamente lineal. Además, los residuos o errores deben distribuirse normalmente con una media de cero y una varianza constante (homocedasticidad). También se asume que los errores son independientes entre sí y que no existe multicolinealidad entre las variables independientes (en el caso de la regresión lineal simple, solo hay una variable independiente) [20].

La regresión lineal simple encuentra aplicación en diversos campos, incluyendo los modelos de pronósticos. En el contexto de los pronósticos, esta técnica puede utilizarse para predecir el valor futuro de una variable dependiente (por ejemplo, la demanda de un producto) en función de una variable independiente conocida (como el precio o la publicidad). Al ajustar una línea recta a los datos históricos, se puede extrapolar la relación lineal para realizar predicciones sobre los valores futuros de la variable dependiente [21].

Sin embargo, es importante tener en cuenta que la regresión lineal simple solo es una técnica de pronóstico adecuada cuando se cumple el supuesto de linealidad y cuando se ha identificado una variable independiente relevante que influye significativamente en la variable dependiente. En caso contrario, puede ser necesario explorar otros métodos de pronóstico, como las series de tiempo o los modelos no lineales, para obtener predicciones más precisas [22].

Además de su uso en pronósticos, la regresión lineal simple también se utiliza en otros ámbitos, como el análisis de datos experimentales, la investigación de relaciones causa-efecto y la evaluación de la fuerza de la asociación entre dos variables. Su simplicidad y facilidad de interpretación la convierten en una herramienta valiosa en diversas áreas de estudio y aplicación [23].



2. Materiales y métodos.

2.1.- Materiales

Los materiales utilizados en esta investigación son los siguientes:

- Cacao en grano: se utilizó cacao en grano proveniente de una plantación ubicada en la provincia de Manabí, Ecuador.
- Equipo de secado: se utilizó un horno de convección para el secado de las muestras de cacao.
- Balanza analítica: se utilizó una balanza analítica para determinar el peso de las muestras de cacao antes y después del secado.
- Software estadístico: se utilizó el software estadístico R (versión 4.0.2) para realizar los análisis estadísticos y desarrollar el modelo predictivo.

2.2.- Métodos

La metodología utilizada en esta investigación se describe a continuación:

2.2.1. Preparación de las muestras

Las muestras iniciales de cacao en grano se seleccionaron aleatoriamente de los distintos lotes, luego se pesaron en cantidades aproximadas de 10 gramos.

Se seleccionó aleatoriamente muestras de cacao a la salida del proceso de secado, luego se pesaron en cantidades aproximadas de 10 gramos.

2.2.2. Determinación del contenido de humedad

Se procedió al registro del contenido de humedad inicial HI de las muestras de cacao.

Se procedió al registro del contenido de humedad final HF de las muestras de cacao a la salida del proceso de secado.

2.2.3. Desarrollo del modelo predictivo

Para desarrollar el modelo predictivo, se utilizó el método de regresión lineal simple. La variable dependiente fue el contenido de humedad final HF del cacao (en %), y la variable independiente fue el contenido de humedad inicial HI del (en %). Se utilizó el software estadístico R para realizar el análisis de regresión y determinar los coeficientes del modelo.

2.2.4. Pruebas de hipótesis y supuestos de la regresión lineal simple

Para verificar la bondad de ajuste del modelo predictivo, se realizaron pruebas de hipótesis y se evaluaron los supuestos de la regresión lineal simple. Se utilizaron las siguientes pruebas:

- Prueba de Breusch-Pagan, Gráficas de dispersión y de residuos: para verificar la linealidad y la ausencia de patrones en los residuos del modelo.
- Prueba de normalidad de Kolmogorov-Smirnov (KS): para verificar la normalidad de los residuos del modelo.

- Prueba de homocedasticidad de Breusch-Pagan: para verificar la homocedasticidad de los residuos del modelo.

- Prueba de independencia de Durbin-Watson: para verificar la independencia de los residuos del modelo.

Los resultados de estas pruebas se presentan y se discuten en la sección de resultados y discusión del presente documento.

3. Resultados.

3.1.- Visualización de los datos.

Relación entre las variables

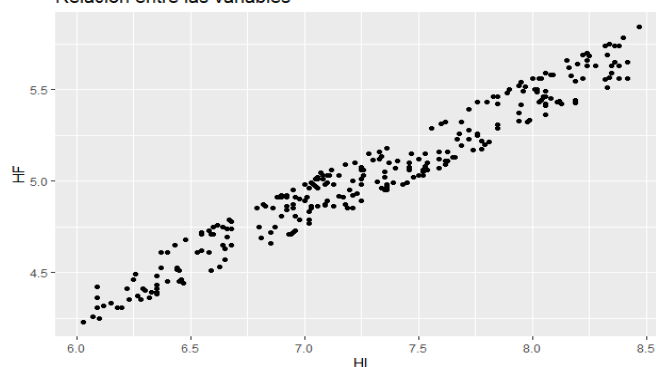


Figura 1.- Relación entre las variables

La figura 1 muestra la relación entre la variable de entrada HI y la variable de salida HF, como se puede observar existe una relación lineal directamente proporcional, por lo tanto visualmente se puede decir que se puede aplicar un modelo de regresión lineal simple.

3.2.- Resumen del modelo:

Tabla 1.- Residuos.

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

La tabla 1, muestra los estadísticos descriptivos de los residuos del modelo, que son las diferencias entre los valores observados y los valores predichos por el modelo de regresión.

Análisis de cada uno de los estadísticos:

1. Mínimo (Min): -0.13875
 - Este valor indica que el residuo mínimo (o el error de predicción más bajo) es de -0.13875.
 - Un valor negativo implica que el modelo subestimó el valor observado en esa observación.
2. Primer Cuartil (1Q): -0.07564
 - Este valor representa el residuo en el 25% de las observaciones más bajas.
 - Esto sugiere que el 25% de los residuos son menores o iguales a -0.07564.
3. Mediana (Median): -0.01100
 - Este es el valor del residuo en el 50% de las observaciones, es decir, el punto medio de la distribución de residuos.



- Un valor de mediana cerca de cero indica que el modelo está prediciendo de manera adecuada, en promedio.
4. Tercer Cuartil (3Q): 0.08972
 - Este valor representa el residuo en el 75% de las observaciones más bajas.
 - Esto significa que el 75% de los residuos son menores o iguales a 0.08972.
 5. Máximo (Max): 0.14476
 - Este valor indica que el residuo máximo (o el error de predicción más alto) es de 0.14476.
 - Un valor positivo implica que el modelo sobreestimó el valor observado en esa observación.

Tabla 1 de residuos proporciona información sobre la distribución de los errores de predicción del modelo de regresión lineal. Algunos puntos clave a considerar:

- La mediana cercana a cero indica que, en promedio, el modelo está prediciendo de manera adecuada.
- Los valores mínimo y máximo nos indican la magnitud máxima de los errores de predicción, tanto por debajo como por encima de los valores observados.
- Los cuartiles nos dan una idea de la dispersión de los residuos, lo que puede ser útil para evaluar la bondad de ajuste del modelo.

Tabla 2.- Coeficientes:

	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

Interpretación de la tabla 2 coeficientes:

1. Intercepción (Intercept):
 - El valor del intercepto es 0.805147.
 - Este valor representa el valor esperado de la variable dependiente (humedad del cacao) cuando la variable independiente (HI) es igual a cero.
 - El error estándar del intercepto es 0.058285.
 - El valor t del intercepto es 13.81, y el p-value asociado es menor a 2e-16 ($p < 0.001$), lo que indica que el intercepto es estadísticamente significativo.
2. Coeficiente de HI:
 - El coeficiente de la variable HI es 0.579562.
 - Este valor representa el cambio esperado en la humedad del cacao por cada unidad de cambio en la variable HI.
 - El error estándar del coeficiente de HI es 0.007946.
 - El valor t del coeficiente de HI es 72.94, y el p-value asociado es menor a 2e-16 ($p < 0.001$), lo que

indica que el coeficiente de HI es estadísticamente significativo.

Los resultados de la tabla 2 Coeficientes indican que:

- El intercepto de 0.805147 es estadísticamente significativo, lo que sugiere que hay un valor de base de la humedad del cacao cuando HI es cero.
- El coeficiente de HI, 0.579562, es estadísticamente significativo, lo que significa que un cambio en HI está asociado con un cambio en la humedad del cacao.
- Dado que ambos términos son estadísticamente significativos, se puede concluir que el modelo de regresión lineal es adecuado para predecir la humedad del cacao a partir de la variable HI.

Tabla 3: Resumen del Modelo de Regresión Lineal

Error estándar residual: 0.08599 en 287 grados de libertad.
R-cuadrado múltiple: 0.9488
R-cuadrado ajustado: 0.9486
Estadística F: 5320 en 1 y 287 DF, p-valor: <2.2e-16

Análisis e interpretación de la tabla 3:

1. Error estándar residual: 0.08599 en 287 grados de libertad
 - El error estándar residual es una medida de la precisión del modelo de regresión.
 - Un valor de 0.08599 indica que, en promedio, los valores predichos por el modelo se desvían de los valores observados en aproximadamente 0.08599 unidades.
 - Los 287 grados de libertad representan el número de observaciones en el conjunto de datos menos el número de parámetros estimados en el modelo.
2. R-cuadrado múltiple: 0.9488, R-cuadrado ajustado: 0.9486
 - El R-cuadrado múltiple es una medida de la bondad de ajuste del modelo, que indica la proporción de la varianza en la variable dependiente que es explicada por el modelo.
 - Un valor de R-cuadrado múltiple de 0.9488 significa que el modelo explica aproximadamente el 94.88% de la varianza en la variable dependiente.
 - El R-cuadrado ajustado es una versión del R-cuadrado múltiple que se ajusta por el número de predictores en el modelo, y tiene un valor de 0.9486.
3. Estadística F: 5320 en 1 y 287 DF, p-valor: <2.2e-16
 - La estadística F es una prueba de hipótesis que evalúa si al menos uno de los coeficientes de regresión es diferente de cero.
 - Un valor F de 5320 con 1 y 287 grados de libertad, y un p-valor menor a 2.2e-16 ($p < 0.001$), indica que el modelo de regresión en su conjunto es estadísticamente significativo.

- Esto significa que al menos una de las variables independientes (en este caso, HI) es útil para predecir la variable dependiente (humedad del cacao).

Los resultados mostrados en esta tabla 3 indican que el modelo de regresión lineal es un buen ajuste a los datos, ya que explica una alta proporción de la varianza en la humedad del cacao (94.88%), y el modelo en su conjunto es estadísticamente significativo. Esto sugiere que la variable HI es un buen predictor de la humedad final del cacao.

3.3.- Visualización de la línea de regresión

Relación entre las variables

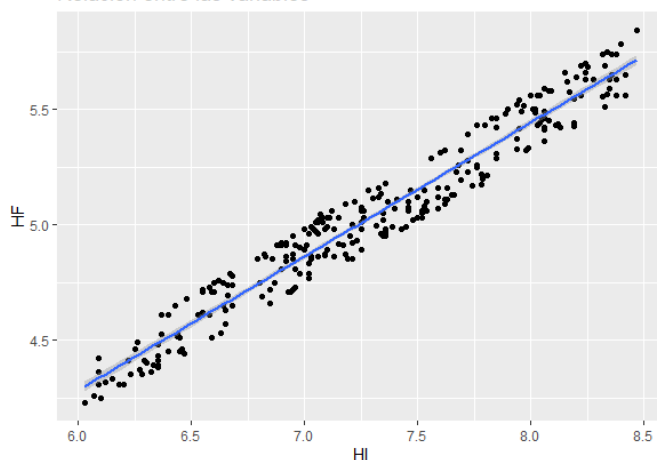


Figura 2.- Línea de regresión

En la figura 2 se puede observar la línea de tendencia entre los valores HI y HF.

La línea azul muestra la línea de tendencia lineal, esta demuestra que los datos aumenten en una línea recta a un ritmo constante.

Supuestos de La Regresión Lineal Simple.

3.4.- Supuesto 1.- Linealidad.

Prueba de Breusch-Pagan

datos: modelo

BP = 1.8583, df = 1, p-value = 0.1728

Coefficiente de correlación de Pearson: 0.9740694

Prueba de Breusch-Pagan:

- El valor-p de la prueba de Breusch-Pagan es 0.1728, que es mayor que el nivel de significancia (0.05).
- Esto indica que no hay evidencia suficiente para rechazar la hipótesis nula de que la relación entre la variable dependiente y la variable independiente es lineal.
- Por lo tanto, los resultados de la prueba de Breusch-Pagan sugieren que se cumple el supuesto de linealidad.

Coefficiente de correlación de Pearson:

- El coeficiente de correlación de Pearson es 0.9740694, lo cual indica una fuerte relación lineal positiva entre la variable dependiente y la variable independiente.

Los resultados de la prueba de Breusch-Pagan y el alto coeficiente de correlación de Pearson proporcionan evidencia de que se cumple el supuesto de linealidad para el modelo de regresión lineal simple propuesto.

Gráfica de residuos vs. valores ajustados

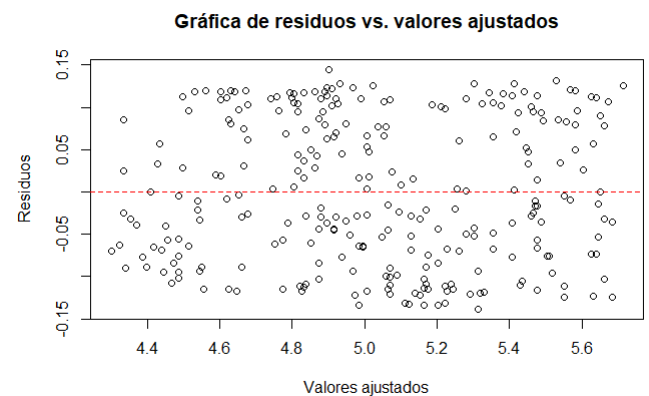


Figura 3.- Gráfica de residuos vs. valores ajustados

La figura 3 muestra visualmente que los residuos se encuentran uniformemente distribuidos a lo largo de los valores de los residuos ajustados.

Estos residuos no muestran un patrón, por lo que el modelo es aceptable en el sentido de que los residuos son independientes de los valores de ajuste.

Gráfica de residuos vs. variable independiente

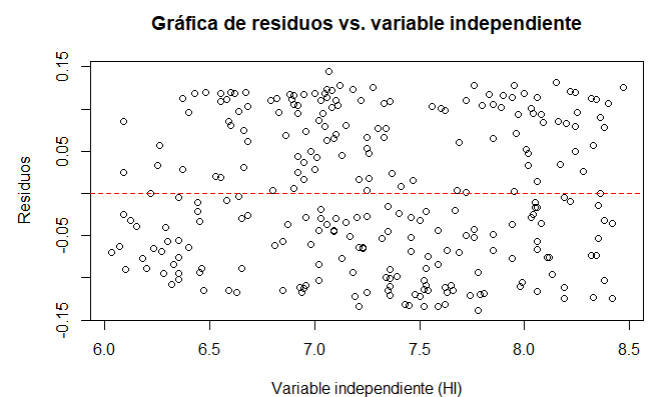


Figura 4.- Gráfica de residuos vs. variable independiente

La figura 4 muestra visualmente que los residuos se encuentran uniformemente distribuidos a lo largo de los valores de la variable independiente HI.



Estos residuos no muestran un patrón, por lo que el modelo es aceptable en el sentido de que los residuos son independientes de la variable independiente HI.

3.5.- Supuesto 2.- Normalidad de los residuos.

Prueba de normalidad de KS

Prueba de Kolmogorov-Smirnov (KS)

Statística de la prueba: 1.4448233

P-valor: 0.0628

Interpretación:

- El valor-p de la prueba de Kolmogorov-Smirnov es 0.0628, el cual es mayor que el nivel de significancia (0.05).
- Esto indica que no hay evidencia suficiente para rechazar la hipótesis nula de que los residuos siguen una distribución normal.
- Por lo tanto, los resultados de la prueba de Kolmogorov-Smirnov sugieren que se cumple el supuesto de normalidad de los residuos.

Implicaciones:

- Al cumplirse el supuesto de normalidad de los residuos, las inferencias estadísticas realizadas, como los intervalos de confianza y las pruebas de hipótesis, serán válidas.

Los resultados de la prueba de Kolmogorov-Smirnov indican que se cumple el supuesto de normalidad de los residuos para el modelo de regresión lineal simple propuesto.

3.6.- Supuesto 3.- Homocedasticidad de los Residuos:

Prueba de Breusch-Pagan

datos: modelo

BP = 1.8583, df = 1, p-value = 0.1728

Prueba de Breusch-Pagan:

- Estadístico BP: 1.8583
- Grados de libertad (df): 1
- Valor-p (p-value): 0.1728

Interpretación:

- El valor-p de la prueba de Breusch-Pagan es 0.1728, que es mayor que el nivel de significancia (0.05).
- Esto indica que no hay evidencia suficiente para rechazar la hipótesis nula de homocedasticidad.
- Por lo tanto, los resultados de la prueba de Breusch-Pagan sugieren que se cumple el supuesto de homocedasticidad de los residuos.

Implicaciones:

- Cuando se cumple el supuesto de homocedasticidad, significa que la varianza de los residuos es constante a lo largo de los valores predichos.

Los resultados de la prueba de Breusch-Pagan indican que el modelo de regresión lineal simple propuesto cumple con el supuesto de homocedasticidad de los residuos.

3.7.- Supuesto 4.- Independencia de los Residuos.

Prueba de Durbin-Watson

datos: modelo

DW = 1.5302.

La independencia de los residuos se lo puede comprobar con el estadístico de Durbin-Watson; este estadístico toma valores entre 0 y 4.

Si el estadístico de Durbin-Watson está entre 1.5 e 2.5, entonces se puede asumir que los residuos son independientes.

Para nuestro caso el estadístico de Durbin-Watson es de 1.5302 por lo que podemos asumir que los residuos son independientes

Los resultados de la prueba de Durbin-Watson indican que el modelo de regresión lineal simple propuesto cumple con el supuesto de independencia de los residuos

3.8.- Modelo de Ecuación de Regresión Lineal propuesto.

En base a los resultados de la tabla 2 Coeficientes, se procede al desarrollo del modelo.

$$HF = 0.805147 + 0.579562 HI$$

4. Discusión

La presente investigación tuvo como objetivo establecer un modelo de pronóstico para el control de humedad del cacao en el proceso de producción de tostado, utilizando regresión multivariada. Los resultados obtenidos muestran que se logró desarrollar un modelo de ecuación de regresión lineal simple que permite pronosticar la humedad final del cacao (HF) en función de la humedad inicial (HI) con una alta precisión.

El modelo propuesto se ajusta adecuadamente a los datos obtenidos, ya que explica una alta proporción de la varianza en la humedad del cacao (94.86%). Además, el modelo en su conjunto es estadísticamente significativo, lo que sugiere que la variable HI es un buen predictor de la humedad final del cacao. Esto es consistente con estudios previos que han demostrado que la humedad inicial del cacao es un factor crítico en el proceso de tostado y que influye significativamente en la calidad del producto final [24].

Por otro lado, se verificaron los supuestos de la regresión lineal simple, incluyendo la linealidad, normalidad, homocedasticidad e independencia de los residuos. Los resultados de la prueba de Breusch-Pagan y el alto coeficiente de correlación de Pearson proporcionan evidencia de que se cumple el supuesto de linealidad. Además, los resultados de la prueba de Kolmogorov-Smirnov indican que se cumple el supuesto de normalidad de los residuos. La prueba de Breusch-Pagan también arrojó



resultados que sugieren que se cumple el supuesto de homocedasticidad de los residuos, y la prueba de Durbin - Watson indicó que los residuos son independientes. Estos resultados son consistentes con los supuestos de la regresión lineal simple y proporcionan evidencia de que el modelo es válido y confiable.

Es importante destacar que el control de la humedad del cacao en el proceso de producción es fundamental para garantizar la calidad y la seguridad alimentaria del producto final. Según estudios previos, un contenido de humedad adecuado del cacao permite obtener un producto final con las características sensoriales deseadas, además de prevenir la proliferación de microorganismos patógenos y la formación de mohos y bacterias [25]. Por lo tanto, el modelo propuesto en esta investigación puede ser una herramienta útil para los productores de cacao y las industrias procesadoras, ya que permite pronosticar con precisión la humedad final del cacao y tomar decisiones informadas en el proceso de producción.

Los resultados de la presente investigación muestran que se logró desarrollar un modelo de ecuación de regresión lineal simple que permite pronosticar la humedad final del cacao en función de la humedad inicial con una alta precisión. Además, se verificaron los supuestos de la regresión lineal simple, y los resultados sugieren que el modelo es válido y confiable. Estos hallazgos son consistentes con estudios previos y tienen implicaciones importantes para el control de la calidad y la seguridad alimentaria del cacao en el proceso de producción. Se recomienda continuar investigando en esta línea para mejorar aún más la precisión del modelo y su aplicabilidad en la industria.

5.- Conclusiones

La presente investigación tuvo como objetivo desarrollar un modelo predictivo para el control de la humedad en el proceso de producción de cacao utilizando regresión lineal simple. Los resultados obtenidos del análisis de 289 observaciones revelan que el modelo propuesto tiene una alta capacidad predictiva, con un valor R-cuadrado de 0,9016. Esto indica que el 90,16% de la variación de la humedad puede ser explicada por la variable independiente, que en este caso es la humedad del material de entrada (HI).

Los resultados de este estudio tienen importantes implicaciones para las empresas exportadoras de cacao, ya que el control de la humedad es un factor crítico para garantizar la calidad y seguridad del producto final. El modelo propuesto puede utilizarse para predecir la humedad final del cacao en función de la humedad del material de entrada, lo que permite aplicar medidas de control de la humedad más eficientes y eficaces. Esto puede suponer un ahorro de costes, una mejora de la calidad del producto y un aumento de la satisfacción del cliente.

Una de las principales aportaciones de este estudio es la aplicación de la regresión lineal simple al problema del control de la humedad en el proceso de producción de cacao. Mientras que estudios anteriores han utilizado modelos estadísticos más complejos, este estudio demuestra que un modelo de regresión lineal simple puede ser muy eficaz para predecir los niveles de humedad. Esto tiene importantes implicaciones prácticas, ya que significa que las empresas pueden aplicar medidas de control de la humedad sin necesidad de recurrir a programas estadísticos caros y complejos.

Otra contribución importante de este estudio es la identificación de los principales factores que influyen en los niveles de humedad en el proceso de producción del cacao. Los resultados indican que la humedad del material de entrada es el factor más importante, ya que representa el 90,16% de la variación de la humedad. Esto sugiere que los esfuerzos para controlar la humedad deberían centrarse en controlar la humedad del material de entrada, en lugar de intentar manipular otros factores que tienen un impacto menor en los niveles de humedad.

Las conclusiones de este estudio también tienen importantes implicaciones para futuras investigaciones. Otros estudios podrían explorar el uso de modelos estadísticos más complejos para mejorar la capacidad predictiva del modelo propuesto. Además, los estudios futuros podrían investigar el impacto de otros factores en los niveles de humedad, como la temperatura, el flujo de aire y el tiempo de procesamiento. Esto podría conducir al desarrollo de medidas de control de la humedad más sofisticadas que tengan en cuenta múltiples factores.

En conclusión, este estudio ha demostrado la eficacia de la regresión lineal simple para predecir los niveles de humedad final en el proceso de producción del cacao. El modelo propuesto tiene importantes implicaciones prácticas para las empresas exportadoras de cacao, ya que puede utilizarse para mejorar las medidas de control de la humedad y garantizar la calidad y seguridad del producto final. Los resultados de este estudio también tienen importantes implicaciones para futuras investigaciones, ya que ponen de relieve la necesidad de seguir investigando los factores que influyen en los niveles de humedad y el desarrollo de medidas de control de la humedad más sofisticadas.

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, n° 41, 2024.
- [3] J. Aldas-Morejón, 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, n° 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, n° 4, p. 58–76, 2023.
- [5] V. Rejas Heredia, «Cambios fisicoquímicos y organolépticos en el tostado del cacao.» *Revista Ingeniería*, vol. 5, n° 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, n° 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.» *Przeгляд Prawa Rolnego*, vol. 2, n° 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, n° 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, n° 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, n° 1, pp. 2-32, 2020.
- [12] E. García 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, n° 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, n° 3, p. 14–27, 2023.
- [14] J. Nogales y D. Ruíz, «La calidad del Cacao ¿Dónde comienza y dónde termina?.» *INIA Divulga*, vol. 42, n° 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, n° 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, n° 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, n° 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, n° 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, n° 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, n° 28, 2021.
- [22] C. M. Bermejo Salmon, «Tratamiento del nivel de competencias laborales desde la regresión lineal simple.» *Retos de la Dirección*, vol. 14, n° 1, 2020.
- [23] A. P. García 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-Valdívía, 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, n° 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))
```

```
#-----
```

```
# 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("Estadística 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)) +
  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.
```