



## Determinación de un modelo de regresión lineal múltiple para la predicción de Pol en la caña de azúcar (*Saccharum officinarum*)


*Determination of a multiple linear regression model for the prediction of Pol in sugarcane (*Saccharum officinarum*)*

*Determinação de modelo de regressão linear múltipla para predição de Pol em cana-de-açúcar (*Saccharum officinarum*)*

**Flavio Reyes<sup>1</sup>**

Universidad Hipócrates, Acapulco - Estado de Guerrero, México

Instituto Universitario de Innovación Ciencia y Tecnología Inudi-Perú, Puno - Puno Perú

 <https://orcid.org/0000-0002-9694-705X>

freyes@inudi.edu.pe (correspondencia)

**Estuardo Monroy**

Universidad Hipócrates, Acapulco - Estado de Guerrero, México

Instituto Universitario de Innovación Ciencia y Tecnología Inudi-Perú, Puno - Puno Perú

 <https://orcid.org/0009-0002-5493-0767>

emonroyb@inudi.edu.pe

DOI: <https://doi.org/10.35622/j.rca.2024.02.003>

Recibido: 11/09/2024 Aceptado: 17/12/2024 Publicado: 20/12/2024

### PALABRAS CLAVE

agroindustria, calidad, eficiencia, optimización, producción.

**RESUMEN.** En el proceso de fabricación de azúcar, la Pol en caña es un parámetro de calidad. Sin embargo, existe una problemática en su determinación, ya que está condicionada por la complejidad de cálculo, lo que obligó a realizar esta investigación, para brindar una solución factible, dinámica y económica. El objetivo consistió en elaborar un modelo predictivo que permita predecir la Pol en caña de manera más rápida y de forma sencilla a partir de seis variables independientes: brix en jugo, Pol en jugo, no Pol en jugo, pureza de jugo, fibra en caña y jugo en caña. La investigación fue cuantitativa con diseño explicativo y transversal, donde se disponía de datos recolectados durante la zafra 2023-2024 en un ingenio de la región sur de Guatemala. Se analizaron 23,470 registros para cada variable. Los resultados del análisis de regresión lineal múltiple demostraron que las variables Pol en jugo y fibra en caña inciden directamente en la predicción de Pol en caña al evidenciar coeficientes estandarizados con significancia estadística. Se generó un modelo matemático sólido que atribuye a las variables Pol en jugo y fibra en caña una alta capacidad explicativa. La ecuación de la fórmula es: Pol en caña = 3.642 + (0.80 x Pol en jugo) - (0.242 x fibra en caña). Se concluye que la nueva propuesta para el cálculo de Pol en caña, establece equilibrio entre su simplicidad y precisión, facilitará su aplicación e impactará positivamente en la toma de decisiones en la producción de azúcar.

<sup>1</sup> Doctor en Investigación Social por la Universidad Panamericana, Guatemala.



#### KEYWORDS

agroindustry, efficiency, optimization, production, quality.

**ABSTRACT.** In the sugar manufacturing process, Pol in cane is a quality parameter. However, there is a problem in its determination, since it is conditioned by the complexity of the calculation, which forced this research to be carried out to provide a feasible, dynamic and economical solution. The objective was to develop a predictive model that allows predicting Pol in cane more quickly and easily from six independent variables: brix in juice, Pol in juice, non-Pol in juice, purity of juice, fiber in cane and cane juice. The research was quantitative with an explanatory and transversal design, where data collected during the 2023-2024 harvest at a sugar mill in the southern region of Guatemala was available. 23,470 records were analyzed for each variable. The results of the multiple linear regression analysis demonstrated that the variables Pol in juice and fiber in cane directly affect the prediction of Pol in cane by evidencing standardized coefficients with statistical significance. A solid mathematical model was generated that attributes high explanatory capacity to the variables Pol in juice and fiber in cane. The formula equation is: Pol in cane = 3.642 + (0.80 x Pol in juice) – (0.242 x fiber in cane). It is concluded that the new proposal for calculating Pol in cane establishes a balance between its simplicity and precision, will facilitate its application and will positively impact decision-making in sugar production.

#### PALAVRAS-CHAVE

agroindústria, eficiência, otimização, produção, qualidade.

**RESUMO.** No processo de fabricação do açúcar, o Pol na cana é um parâmetro de qualidade. Contudo, existe um problema na sua determinação, uma vez que está condicionada pela complexidade do cálculo, o que obrigou à realização desta investigação para fornecer uma solução viável, dinâmica e econômica. O objetivo foi desenvolver um modelo preditivo que permitisse prever Pol em cana de forma mais rápida e fácil a partir de seis variáveis independentes: brix em caldo, Pol em caldo, não Pol em caldo, pureza de caldo, fibra em cana e caldo de cana. A pesquisa foi quantitativa com desenho explicativo e transversal, onde foram disponibilizados dados coletados durante a safra 2023-2024 em uma usina de açúcar da região sul da Guatemala. Foram analisados 23.470 registros para cada variável. Os resultados da análise de regressão linear múltipla demonstraram que as variáveis Pol no caldo e fibra na cana afetam diretamente a predição da Pol na cana, evidenciando coeficientes padronizados com significância estatística. Foi gerado um modelo matemático sólido que atribui alta capacidade explicativa às variáveis Pol no caldo e fibra na cana. A equação da fórmula é: Pol na cana = 3,642 + (0,80 x Pol no caldo) – (0,242 x fibra na cana). Conclui-se que a nova proposta de cálculo do Pol na cana estabelece um equilíbrio entre sua simplicidade e precisão, facilitará sua aplicação e impactará positivamente na tomada de decisões na produção de açúcar.

## 1. INTRODUCCIÓN

La caña de azúcar es un cultivo que se practica a nivel mundial (De Almeida Silva et al., 2022). Es un cultivo bioenergético con producciones más eficientes de las regiones tropicales y subtropicales (Waclawovsky et al., 2010). Actualmente, existe una gran demanda de azúcar en el mercado mundial, por lo tanto, la industria azucarera es la industria líder y depende principalmente de la caña de azúcar (Chauhan et al., 2011). El propósito de todo ingenio azucarero es producir azúcar con altos estándares de calidad y rendimientos óptimos, pero lograr una excelente calidad del producto con mínimas pérdidas, es muy importante comprender la calidad de las materias primas, que afectan el proceso de producción de azúcar y la calidad del producto final (Cavalcante & de Albuquerque, 2015). La calidad de la caña se determina a través de su contenido de Pol (sacarosa), pero su cálculo contempla el empleo de fórmulas que son difíciles de entender por ser complejas y sofisticadas (Lejars et al., 2010). La sacarosa es un disacárido que se encuentra en la caña de azúcar (Plaza-Díaz & Gil, 2015). La Pol en caña es la sacarosa presente en la caña, permite estimar el rendimiento de azúcar, por tanto, es un dato crucial para la optimización del proceso de extracción de azúcar.

Para la determinación del contenido de sacarosa en la caña (Pol en caña) se debe invertir mucho tiempo, es una actividad muy laboriosa y representa una amenaza para el medio ambiente (Chiatrakul et al., 2022). El



problema detectado en este estudio fue la mala calidad de la producción de azúcar y es causado por la dificultad de analizar y calcular la Pol en caña que es un indicador importante en la materia prima y como efectos negativos, genera la inversión considerable de tiempo, equipo analítico de medición y altos costos operacionales. De no resolverse, puede repercutir negativamente en la toma de decisiones y no generar acciones preventivas y correctivas justo a tiempo en el proceso productivo lo que conduce a altas pérdidas de azúcar que se traducen en mermas económicas. Esta problemática se puede resolver a través de un modelo de regresión lineal práctico y sencillo para calcular la Pol en caña que sea utilizado de forma oportuna y justo a tiempo.

El brix en jugo se refiere a la cantidad de sólidos solubles presentes en la caña de azúcar y consiste en no azúcares y azúcares totales, entre ellos, sacarosa, glucosa y fructosa (Da Costa et al., 2021). La Pol en jugo es la sacarosa aparente y es un parámetro principal que se analiza en la caña para evaluar su calidad (Sorol et al., 2021). La no Pol en el jugo de la caña se encuentra al realizar la resta de brix en jugo menos Pol en jugo. La pureza del jugo es la relación que se calcula al dividir la Pol en jugo sobre el brix del jugo, se calcula así,  $Pureza = \left( \frac{sacarosa}{brix} \right) * 100$ . La fibra en caña es un componente seco e insoluble en el agua del tallo muy importante en la valoración de la calidad de la caña porque es inversamente proporcional con la extracción de jugo y la eficiencia en la operación de la molienda (Islam et al., 2021). El jugo en caña, está constituido por agua, sacarosa y no azúcares, su extracción en la etapa de molienda es muy importante para la eficiencia de la fábrica (Serrano Febles et al., 2022). Por último, la Pol en caña es el parámetro más importante de calidad y representa la sacarosa aparente contenida en 100 partes de caña (Bastidas et al., 2009).

En el ingenio azucarero donde se realizó el estudio, se ubica en la región sur de Guatemala y para la zafra 2023-2024 enfrentó el problema que se ha descrito anteriormente, lo que conlleva a la necesidad de resolverlo y presentar para las zafras venideras una solución factible, aplicable y de uso amigable que incentive la productividad con calidad en el proceso azucarero. Por lo tanto, el objetivo de este estudio fue determinar un modelo de regresión lineal que permita calcular de forma predictiva la Pol en caña de manera más rápida y de forma sencilla con el empleo de variables de fácil medición que se correlacionen significativamente y proporcione un resultado preciso y confiable en la toma de decisiones operativas en el proceso para garantizar una producción eficiente y económicamente sostenible.

## 2. MÉTODO

El estudio se realizó en un sector de la agroindustria azucarera porque actualmente enfrenta muchos problemas y oportunidades para ser sustentables (Eggleston & Lima, 2015). Se empleó un enfoque metodológico explicativo porque se encontró relación entre variables (Hernández Sampieri et al., 2014). Además, fue transversal, ya que los datos fueron recolectados en un momento determinado (Díaz Narváez, 2009).

Para este estudio se utilizaron 23,420 registros de los datos analíticos de laboratorio de un ingenio azucarero ubicado en la región sur de Guatemala, los datos de cada una de las variables fueron obtenidos a partir de ensayos de laboratorio que incluyeron mediciones de brix en jugo, Pol en jugo, no Pol en jugo, fibra en caña, jugo en caña como variables independientes, y Pol en caña como variable dependiente. El período del estudio fue comprendido desde el mes de noviembre del año 2024 al mes de marzo del año 2025. Se consideraron todos los datos generados durante la zafra (cosecha de la caña), lo que eliminó la necesidad de calcular el tamaño de la muestra, dado que se trabajó con la totalidad de la información disponible.

Se utilizó la prueba de normalidad de Kolmogorov-Smirnov para evaluar la distribución de los datos y determinar cuáles serían las pruebas estadísticas idóneas a realizar (Drezner & Turel, 2011). Además, se calcularon estadísticas descriptivas como la media aritmética, la desviación estándar, el grado de asimetría y la curtosis para comprender todas las diferencias entre las distribuciones de las variables. Por otra parte, fue vital realizar un análisis de evaluación no paramétrica mediante los coeficientes Tau-b de Kendall y Rho de Spearman para identificar las relaciones entre variables independientes y dependientes sin necesidad de asumir que las distribuciones siguen una normal (Li et al., 2021). Dado el gran tamaño de la muestra, se asumió que el promedio sigue una distribución normal y con este supuesto se procedió a realizar un modelo de regresión lineal múltiple con el método de los mínimos cuadrados (Sawada, 2021).

### 3. RESULTADOS

#### a) Caracterización de las variables independientes y variable dependiente

El análisis de las variables brix en jugo, Pol en jugo, no Pol en jugo, pureza de jugo, fibra en caña y jugo en caña es esencial para calcular la Pol en caña, ya que cada una de ellas proporciona información clave sobre la calidad del jugo y la caña de azúcar. Los estadísticos descriptivos de 23,470 datos de jugo que agrupan análisis de brix en jugo, Pol en jugo, no Pol en jugo, pureza de jugo, fibra en caña, jugo en caña y Pol en caña pueden verse en la siguiente tabla.

Tabla 1

*Estadísticos descriptivos de las pruebas de laboratorio de caña de azúcar*

Descripción	Media	Desviación estándar	Asimetría	Curtosis
Brix en jugo	19.21	2.1	-0.07	0.13
Pol en jugo	16.30	2.04	-0.14	0.13
No Pol en jugo	2.91	0.84	0.54	2.40
Pureza de jugo	0.85	0.04	-0.46	1.40
Fibra en caña	14.00	1.38	1.22	10.65
Jugo en caña	71.50	2.58	-1.50	20.29
Pol en caña	13.23	1.65	-0.18	0.31

*Nota.* Total, de datos por cada variable: 23,470.

Los análisis realizados sobre las variables del jugo y de la caña ofrecen perspectivas relevantes sobre sus distribuciones. Tanto el brix en jugo como la Pol en jugo presentan características que sugieren que siguen distribuciones normales, esto es porque muestran asimetrías y curtosis cercanas a cero. Contrariamente, la variable no Pol en jugo muestra una asimetría positiva y una alta curtosis, lo que indica que no se ajusta a una distribución normal. Lo mismo se observa en el caso de la pureza del jugo; presenta asimetría negativa y curtosis que sugiere un comportamiento no normal. Por otro lado, la fibra en caña presenta alta asimetría positiva y una curtosis extremadamente elevada, lo que claramente indica que no sigue una distribución normal. Asimismo, el jugo en caña muestra características de asimetría negativa y una curtosis considerablemente alta, por lo que tampoco se ajusta a una distribución normal. Finalmente, la Pol de caña se aproxima a una distribución normal, presenta asimetría y curtosis que indican una forma equilibrada. En resumen, las variables brix en jugo, Pol en jugo y Pol en caña son normales, mientras que no Pol en jugo, pureza del jugo, fibra en caña y jugo en caña son distribuciones no normales.

**Tabla 2**

*Pruebas de normalidad*

Descripción	Estadístico (Kolmogorov-Smirnov)	Significancia (p Valor)
Brix en jugo	0.013	0.000
Pol en jugo	0.014	0.000
No Pol en jugo	0.043	0.000
Pureza de jugo	0.070	0.000
Fibra en caña	0.067	0.000
Jugo en caña	0.070	0.000
Pol en caña	0.015	0.000

La prueba de Kolmogorov-Smirnov para verificar la normalidad muestra que no cumple para todas las variables examinadas. Todos los valores de significancia son menores al nivel crítico de 0.05. En consecuencia, brix en jugo, Pol en jugo, no Pol en jugo, pureza de jugo, fibra en caña, jugo en caña y Pol de caña difieren significativamente de la normalidad. En cuanto al análisis estadístico posterior, los hallazgos tienen una influencia significativa: la falta de normalidad significa que no deben utilizarse métodos paramétricos de correlación lineal. En este caso, los métodos no paramétricos son más adecuados para estos datos. Por lo tanto, se optó por métodos no paramétricos, dado que no son distribuciones normales.

**b) Análisis de correlación**

Dado que las distribuciones no son normales, se optó por emplear métodos estadísticos no paramétricos para analizar la correlación entre estas variables. En lugar de utilizar coeficientes de correlación como el de Pearson, porque asume normalidad en los datos, como alternativa, se utilizó el coeficiente de correlación de Spearman y el coeficiente de correlación de Kendall. Ambos métodos poseen supuestos más relajantes y son más aptos para evaluar las relaciones en los datos ordinales o no distribuidos normalmente.

La interpretación de la tabla de correlaciones no paramétricas indica que Pol en jugo presenta la correlación más fuerte con Pol en caña, tiene el mejor valor frente a las demás variables analizadas. Aunque brix en jugo también muestra una correlación notable, es Pol en jugo la que demuestra un vínculo más robusto, lo que sugiere que puede servir como un predictor de aumento de Pol en caña.

**Tabla 3**

*Coefficientes de correlación no paramétricas*

Descripción	Pol en caña			
	Tau b de Kendall	Significancia p Valor	Rho de Spearman	Significancia p Valor
Brix en jugo	0.67**	0.00	0.85**	0.00
Pol en jugo	0.85**	0.00	0.96**	0.00
No Pol en jugo	-0.13**	0.00	-0.19**	0.00
Pureza jugo	0.38**	0.00	0.53**	0.00
Fibra en caña	-0.05**	0.00	-0.08**	0.00
Jugo en caña	0.06**	0.00	0.08**	0.00

*Nota.* \*\* La correlación es significativa en el nivel 0.01 (bilateral)



Este hallazgo subraya la relevancia de Pol en jugo en los análisis de calidad de la caña de azúcar y sugiere que, para optimizar el rendimiento en la industria azucarera, es crucial prestar especial atención a esta variable. En comparación, aunque brix en jugo aporta información valiosa, su valoración, aunque positiva, es menos fuerte que la observada con Pol en jugo.

Las correlaciones analizadas son significativas, ya que los valores de  $p$  obtenidos son iguales a 0.00. Esto significa que hay suficiente evidencia para afirmar que las relaciones entre las variables independientes y la variable dependiente Pol en caña no son producto del azar, sino patrones consistentes en los datos. Aunque todas son estadísticamente relevantes, la intensidad de cada variación varía, pero destacan las más altas como las de mayor importancia para el estudio.

En el contexto de la tabla analizada, aunque ambas correlaciones revelan relaciones fuertes entre Pol en jugo y Pol en caña, la clasificación de Spearman proporciona valores más altos en términos absolutos. Esto sugiere que, aunque ambas correlaciones indican una relación significativa, la evaluación de Spearman podría interpretarse como más optimista, pues refleja una mayor asociación en el contexto de los datos analizados. Sin embargo, si se considera la robustez en presencia de potenciales valores atípicos y la estabilidad en muestras más pequeñas, la clasificación de Kendall podría considerarse más confiable.

Si se prioriza la robustez y la resistencia a influencias externas, la clasificación de Kendall es la opción preferida. Sin embargo, si se busca interpretar la fuerza de la relación entre los dos en términos absolutos, la magnitud de Spearman parecería más fuerte. La elección entre los dos debe basarse en los datos y los objetivos del análisis ya que resalta hacia la importancia de interpretar estas correlaciones no solo en función de los números, sino también de las características de la muestra y la naturaleza de los datos.

### c) Análisis de regresión lineal múltiple

Se utilizó el método de mínimos cuadrados para el análisis de regresión lineal múltiple. Aunque la distribución de las variables no es normal, pero el tamaño de la muestra es lo suficientemente grande con 23,470 datos para cada distribución, se puede confiar en el teorema del límite central porque a medida que el tamaño de la muestra aumenta, la distribución de las medias muestrales tiende a ser normal, lo que permite el uso de técnicas de inferencia estadística como la regresión lineal y las pruebas  $t$  sin respetar el supuesto de normalidad (Lumley et al., 2002). En el contexto de muestras grandes, el supuesto de normalidad de los errores de regresión lineal se vuelve menos importante debido al teorema del límite central (Schmidt & Finan, 2018).

**Tabla 4**

*Resumen del modelo*

Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación	Durbin-Watson
1	0.985 <sup>a</sup>	0.970	0.970	0.28561	0.506

*Nota.* <sup>a</sup>. Predictores: (Constante), Jugo caña, Pol jugo, No Pol jugo, Fibra caña, Pureza jugo.

<sup>b</sup>. Variable dependiente: Pol caña.

La tabla 4 revela información muy importante. La correlación múltiple entre las variables independientes y la variable dependiente, Pol en caña es de 0.985. Es una correlación muy fuerte entre las variables predictoras y la variable dependiente, por lo tanto, el modelo explica bien la variación en la variable dependiente. El coeficiente de determinación es de 0.97 esto significa que el porcentaje de la variación en Pol en caña que es explicado

por las variables independiente, es decir, el 97% de la variación en Pol en caña es explicado por el conjunto de variables independientes, por lo que es un ajuste excelente. Se puede utilizar el modelo muy confiadamente para predecir la Pol en caña.

Es importante mencionar que el R cuadrado ajustado es una versión ajustada del coeficiente de determinación que tiene en cuenta el número de variables en el modelo y el tamaño de la muestra, esto ayuda a evitar una sobreestimación de la varianza explicada, el valor de 0.97 es igual al coeficiente de determinación, por lo tanto, el modelo es sólido y la utilización de las variables independientes está bien justificada. El error estándar de la estimación con un valor de 0.28561 indica la desviación estándar de los residuos del modelo (significa, la diferencia entre los valores observados y los valores predichos de la Pol en caña). Este error estándar sugiere que, en promedio, las predicciones del modelo se desvían en esa cantidad de los valores reales de la Pol en caña. Como el error estándar es pequeño, las predicciones son bastante que se realicen son el modelo son muy precisas.

**Tabla 5**

*Modelo de regresión lineal múltiple*

Modelo	Coeficientes no estandarizados		Coeficientes estandarizados			95.0% intervalo de confianza para B		Correlaciones		
	B	Desv. Error	Beta	t	Sig	Límite inferior	Límite superior	Orden cero	Parcial	Parte
(Constante)	5.386	0.333		16.188	.000	4.734	6.038			
Pol en jugo	0.800	0.003	0.987	294.995	.000	0.794	0.805	0.964	0.887	0.333
No Pol en jugo	-0.078	0.014	-0.040	-5.632	.000	-.105	-.051	-0.196	-0.037	-
Pureza de jugo	-0.772	0.309	-0.020	-2.498	.012	-1.377	-.166	0.541	-0.016	0.003
Fibra en caña	-0.254	0.004	-0.213	-66.919	.000	-.261	-.247	-0.113	-0.400	-
Jugo en caña	-0.011	0.002	-.0017	-5.292	.000	-.015	-.007	0.123	-0.035	0.006

*Nota.* <sup>a</sup> Variable dependiente: Pol en caña.

En la tabla anterior se puede apreciar los resultados del modelo de regresión múltiple, la relación entre variables independientes y variable dependiente y presenta la significancia estadística de los efectos. Se encuentran las variables explicativas que son Pol en jugo, no Pol en jugo, pureza de jugo, fibra en caña y jugo en caña. La Pol en jugo tiene un coeficiente de 0.80 esto significa que tiene un efecto positivo significativo sobre la variable dependiente Pol en caña, el coeficiente estandarizado Beta de 0.987 sugiere un impacto fuerte y positivo y el valor t es de 294.995, con una significancia de 0.000 debe interpretarse que este coeficiente es altamente significativo. La no Pol en jugo tiene un coeficiente de -0.078, por lo que tiene un efecto negativo sobre la variable dependiente, el coeficiente estandarizado Beta de -0.040 es pequeño, con un impacto es débil, sin embargo, el valor t de -5.632 y la significancia de 0.000 muestran significancia estadística

La pureza de jugo tiene un coeficiente de -0.772, esto indica un efecto negativo sobre la variable dependiente, el coeficiente estandarizado Beta de -0.020 brinda un impacto relativamente pequeño, con un valor t de -2.498 y una significancia de 0.012. La fibra en caña presenta un coeficiente de -0.254 esta variable impacta

negativamente y muy fuerte sobre la variable dependiente, con un coeficiente estandarizado Beta de -0.213 y un valor t muy alto de -66.919, con una significancia de 0.000, esto significa que la fibra de caña tiene un efecto importante y negativo, es decir a medida que aumenta la fibra en caña, disminuye la Pol en caña. el jugo en caña con un coeficiente de -0.011 presenta un efecto negativo muy pequeño, el coeficiente estandarizado Beta de -0.017 dice que la relación es débil, pero el valor t de -5.292 y la significancia de 0.000 indican que este efecto tiene significancia estadística. El valor p indica que todas las variables son estadísticamente significativas, ya que sus valores están por debajo de 0.05, lo que sugiere que todas las variables contribuyen de manera significativa al modelo.

Los intervalos de confianza para los coeficientes B no son iguales a cero, lo que refuerza la idea de que todos los efectos de las variables independientes son significativos. La correlación de orden cero, muestra la relación directa entre cada variable independiente y la variable dependiente sin considerar el efecto de otras variables. En este caso, Pol en jugo tiene la elevación más alta con un valor de 0.964, e indica una fuerte relación positiva inicial con la Pol en caña. Las demás variables muestran correlaciones menores o negativas. La correlación parcial refleja la relación entre cada variable independiente y la variable dependiente después de controlar las otras variables. Aquí, Pol en jugo sigue siendo la variable con una correlación positiva más fuerte 0.887, incluso al considerar el impacto de otras variables. La fibra en caña también muestra una relación significativa, aunque negativa con un valor de -0.400, significa que su efecto sigue siendo relevante y debe ser tomado en cuenta. Y la correlación de parte representa la relación única de cada variable independiente con la variable dependiente, que aísla el efecto que cada una aporta de forma exclusiva. Nuevamente, la Pol en jugo muestra el valor más alto 0.333, seguido por fibra en caña -0.076, lo que confirma que estas dos variables contribuyen más de forma individual al modelo.

**Tabla 6**

*Análisis de varianza*

Modelo	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Regresión	61936.701	5	12387.340	151851.793	.000 <sup>b</sup>
Residuo	1914.080	23464	.082		
Total	63850.782	23469			

*Nota.* <sup>a</sup>. Variable dependiente: Pol caña.

<sup>b</sup>. Predictores: (Constante), jugo en caña, Pol en jugo, no Pol en jugo, fibra en caña, pureza en jugo.

El valor de F en el análisis de varianza sugiere que el modelo de regresión múltiple explica significativamente más la variabilidad de la variable dependiente en comparación con lo que sería esperado por azar. Este valor es una medida de la efectividad global del modelo para predecir la variable dependiente. Es decir, el modelo es muy adecuado para explicar las relaciones entre las variables independientes y la variable dependiente. La significancia con valor de 0.000 indica que el valor de p es inferior a 0.05, y esto significa que el modelo es estadísticamente significativo, quiere decir, al menos una de las variables independientes tiene una relación significativa con la variable dependiente. Entonces se debe rechazar la hipótesis nula, que establece que todas las variables predictoras no tienen efecto en la variable dependiente.

Al analizar el modelo de regresión lineal múltiple, respecto a la teoría azucarera, solo las variables independientes de Pol en jugo y fibra en caña tienen un comportamiento lógico con fundamento científico, pues a medida que



aumenta la Pol en jugo debe aumentar la Pol en caña y a medida que aumenta la Pol en caña, disminuye la fibra en caña (Eggleston, 2002). Bajo ese concepto, entonces se corrió nuevamente el modelo con las variables antes mencionadas y se obtuvieron los siguientes resultados.

**Tabla 7**

*Modelo de regresión lineal múltiple, Pol en jugo y fibra en caña para Pol en caña*

Modelo	Coeficientes no estandarizados		Coeficientes estandarizados		t	Sig	95.0% intervalo de confianza para B		Correlaciones		
	B	Desv. Error	Beta				Límite inferior	Límite superior	Orden cero	Parcial	Parte
(Constante)	3.642	0.023			155.933	.000	3.597	3.688			
Pol en jugo	0.800	0.001	0.982		857.678	.000	0.794	0.798	0.964	0.948	0.992
Fibra en caña	-0.242	0.001	-0.203		-177.11	.000	-0.245	-0.239	-0.113	-0.756	0.992

*Nota.* <sup>a</sup> Variable dependiente: Pol en caña.

Comparamos los dos modelos de regresión lineal múltiple (tablas 5 y 7) para identificar la mejor estructura predictiva para la variable dependiente Pol en caña. El primer modelo considera cinco variables predictoras: Pol en jugo, no Pol en jugo, pureza de jugo, fibra en caña y jugo en caña, de las cuales Pol en jugo y fibra en caña son las variables más influyentes por su significancia estadística y altos valores de Beta estandarizado 0.987 y -0.213, respectivamente, presentan una considerable relevancia en la explicación de Pol en caña. Sin embargo, las otras tres variables muestran coeficientes y correlaciones de menor magnitud, que sugieren un impacto menor en el modelo, por lo que se omitieron para calcular un segundo modelo.

En la tabla 7 observamos el segundo modelo más sencillo pues solo incluye las variables Pol en jugo y fibra en caña, que conservan sus coeficientes y significancia estadística observados en el primer modelo con valores Beta de 0.982 para Pol en jugo y -0.203 para fibra en caña, lo que confirma su influencia significativa en la variable dependiente. Además, las correlaciones parciales de estas variables son superiores en el segundo modelo con valores de 0.948 para Pol en jugo y -0.756 para fibra en caña y esto significa una mayor capacidad explicativa sin la necesidad de incorporar variables adicionales.

En función de estos análisis, el segundo modelo se presenta como la opción más robusta y confiable, ya que permite una interpretación más directa y eficaz al capturar la variabilidad esencial de la Pol en caña mediante un menor conjunto variables predictoras. El segundo modelo es simple y facilita su utilización sin poner en riesgo la precisión estadística, lo que lo convierte en la mejor alternativa para la predicción.

**Tabla 8**

*Análisis de varianza para el nuevo modelo*

Modelo	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Regresión	61902.168	2	372741.519	372741.519	.000 <sup>b</sup>
Residuo	1948.613	23467	.083		
Total	63850.782	23469			

*Nota.* <sup>a</sup> Variable dependiente: Pol caña. <sup>b</sup> Predictores: (Constante), Pol jugo y Fibra caña.

El análisis de la varianza del segundo modelo de regresión lineal múltiple confirma la efectividad en la predicción de la variable dependiente Pol en caña, al considerar el valor F y con significancia estadística

#### d) Modelo predictivo

El segundo modelo, con su elevado valor F en análisis de varianza y su simplicidad en el número de predictores, se presenta como el más eficiente y confiable para la predicción de Pol en caña, pues proporciona una alta capacidad explicativa con menos complejidad estructural. El modelo de regresión lineal múltiple para la valorización de la Pol en caña queda de acuerdo a la siguiente ecuación.

$$\text{Pol en caña} = 3.642 + (0.80 \times \text{Pol en jugo}) - (0.242 \times \text{Fibra en caña})$$

El modelo de regresión lineal múltiple establece una relación predictiva entre Pol en caña como variable dependiente y dos variables independientes: Pol en jugo y fibra en caña. La constante de 3.642 indica el valor esperado de Pol en caña cuando las variables predictoras son cero. A medida que el valor de Pol en jugo aumenta una unidad, también la Pol en caña aumenta 0.80 unidades esto es una relación directamente proporcional. Por otro lado, a medida que el valor de fibra en caña aumenta una unidad, disminuye en 0.242 unidades la Pol en caña y presenta una relación inversamente proporcional.

#### 4. DISCUSIÓN

Las variables independientes que estimulan la Pol en caña son la Pol en jugo y la fibra en caña. Esto se debe en primer lugar porque la variación de la Pol en el jugo impacta positivamente en la variación de la Pol en la caña. Lo anterior concuerda con Jackson (2005) porque un mayor contenido de sacarosa en el jugo tiende a reflejarse directamente en la caña, lo que promueve un mejor rendimiento de azúcar en la producción. También (Sajid et al., 2023) determinaron que el porcentaje de Pol tiene una correlación positiva con el porcentaje de azúcar (Pol) en la caña comercial.

En segundo plano, la fibra en caña influye en la Pol caña, pero de forma negativa, es decir, a medida que la fibra aumenta la Pol disminuye, eso porque fibra es un material vegetativo que no contiene sacarosa, como lo que encontraron Robertson y Donaldson (1998) en un estudio que demostró un aumento en el rendimiento de sacarosa cuando hubo disminución en la masa seca (fibra) del tallo. Lo anterior también concuerda con (Reyes-Hernández et al., 2022) quienes encontraron que el aumento de la fibra en la caña, implica una disminución en los valores teóricos de sacarosa (Pol) en la caña.

El estudio realizado es importante porque con el modelo de regresión múltiple encontrado se evitará la inversión de mucho tiempo y recursos para obtener el resultado de Pol en caña como ocurre actualmente (Phuphaphud et al., 2020). Por el contrario, se optimizarán los tiempos de respuesta, la confiabilidad de los resultados y relación costo beneficio, tal como lo indican Lanza et al., (2016) pues son requerimientos más frecuentes de los laboratorios de análisis de alimentos y de sus usuarios. También, Guimarães et al. (2016) en su estudio, concuerdan que es necesario desarrollar métodos rápidos basados en mínimos cuadrados para la determinación de parámetros de calidad en la materia prima.

La limitación de este estudio radica en que la determinación de Pol en caña aún se realiza de forma destructiva, es decir, para analizar su contenido es necesario cortar los tallos de la planta y extraer el jugo para su posterior análisis. En ese sentido se propone ampliar esta investigación para la determinación de Pol en caña mediante

un método de análisis no destructivo de los tallos de la caña con la utilización de espectrometría de infrarrojo cercano (NIRS) como lo propone Corrêdo et al. (2021) pues se perfila como un método rápido, confiable y preciso para predecir la calidad de la caña de azúcar como materia prima. También se debe considerar la posibilidad de validar el modelo de predicción de Pol en caña con datos de la zafra 2024-2025.

## 5. CONCLUSIONES

Las variables que influyen en la predicción de la Pol en caña son la Pol en jugo y la fibra en la caña, esto se debe a que la relación entre la Pol en jugo y Pol en caña son directamente proporcionales, es decir un contenido alto de sacarosa en el jugo va a repercutir positivamente en la sacarosa presente en la caña, caso contrario ocurre con la fibra en caña, puesto que esta es materia vegetativa que no contiene sacarosa a medida que aumenta esta variable habrá disminución de Pol en caña. Estas dos variables son importantes en el proceso de producción de azúcar, desde el cultivo de la caña hasta la recuperación de la sacarosa en la fábrica, por tal razón es imperante mantener los controles adecuados en todas las operaciones unitarias para su correcto análisis y medición lo que va a contribuir enormemente en la mitigación de pérdidas de azúcar y consecuentemente en una mejor sostenibilidad económica agroindustrial.

La facilidad de predecir la Pol en caña permite tener tiempos de respuesta de resultados justo a tiempo, lo que permite tomar decisiones en el control del proceso. Estas decisiones sirven prever la producción de azúcar a recuperarse, así como el azúcar a perderse dentro del proceso físico-químico industrial, que puede ser azúcar cruda, blanca, refinada o cualquier calidad comercial, puesto que la Pol en caña bien calculada también genera indicadores de calidad que se asocian con el tipo de azúcar a producir.

Los ganadores finales de la información de este estudio, son los productores de caña, los fabricantes o sea la parte industrial que extrae el azúcar, todos los empleados y/o trabajadores de dichas empresas, los inversionistas con sus exportaciones e ingreso de divisas y por supuesto la población guatemalteca, al existir empleo, salarios y salud.

### **Conflicto de intereses / Competing interests:**

Los autores declaran que no incurrieron en conflictos de intereses.

### **Rol de los autores / Authors Roles:**

Flavio Reyes: conceptualización, curación, análisis formal de los datos, investigación, metodología, redacción, administración del proyecto y redacción-revisión y edición.

Estuardo Monroy: recursos materiales, software, supervisión, validación, redacción-borrador original, revisión y edición.

### **Fuentes de financiamiento / Funding:**

Los autores declaran que no recibieron financiamiento para la realización de la investigación.

### **Aspectos éticos / legales; Ethics / legals:**

Los autores declaran no haber incurrido en aspectos antiéticos, ni haber omitido aspectos legales en la realización de la investigación.

## REFERENCIAS

Bastidas, L., De Sousa, O., Briceño, R., & Hernández, E. (2009). Potencial azucarero y panelero de cinco cultivares de caña de azúcar en el Valle de Santa Cruz de Bucaral, Estado Falcón, Venezuela. *Agronomía Tropical*, 59(2), 137-148.

- Cavalcante, C. S., & de Albuquerque, F. M. (2015). The Sugar Production Process. In F. Santos, A. Borém & C. Caldas (Eds.), *Sugarcane: Agricultural Production, Bioenergy and Ethanol* (pp. 285–310). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-802239-9.00014-1>
- Chauhan, M. K., Chaudhary, S., & Kumar, S. (2011). Life cycle assessment of sugar industry: A review. *Renewable and Sustainable Energy Reviews*, *15*(7), 3445-3453. <https://doi.org/10.1016/j.rser.2011.04.033>
- Chiatrakul, J., Terdwongworakul, A., Phuangsombut, K., & Phuangsombut, A. (2022). Improved evaluation of commercial cane sugar content in sugarcane stalk using near infrared hyperspectral imaging and stalk axis rotation technique. *Biosystems Engineering*, *223*, 161-173. <https://doi.org/10.1016/J.BIOSYSTEMSENG.2022.08.019>
- Corrêdo, L. P., Wei, M. C., Ferraz, M. N., & Molin, J. P. (2021). Near-infrared spectroscopy as a tool for monitoring the spatial variability of sugarcane quality in the fields. *Biosystems Engineering*, *206*, 150-161. <https://doi.org/10.1016/j.biosystemseng.2021.04.001>
- Da Costa, M. V. A., Fontes, C. H., Carvalho, G., & Júnior, E. C. de M. (2021). Ultrabrix: A device for measuring the soluble solids content in sugarcane. *Sustainability*, *13*(3), 1227. <https://doi.org/10.3390/su13031227>
- De Almeida Silva, M., Véliz, J. G. E., Sartori, M. M. P., & Santos, H. L. (2022). Glyphosate applied at a hormetic dose improves ripening without impairing sugarcane productivity and ratoon sprouting. *Science of the Total Environment*, *806*, 150503. <https://doi.org/10.1016/j.scitotenv.2021.150503>
- Díaz Narváez, V. P. (2009). *Metodología de la investigación científica y bioestadística para profesionales y estudiantes de las ciencias de la salud* (1ª ed.). Editorial Universidad Evangélica de El Salvador.
- Drezner, Z., & Turel, O. (2011). Normalizing variables with too-frequent values using a Kolmogorov–Smirnov test: A practical approach. *Computers & Industrial Engineering*, *61*(4), 1240-1244. <https://doi.org/10.1016/J.CIE.2011.07.015>
- Eggleston, G. (2002). Deterioration of cane juice—sources and indicators. *Food chemistry*, *78*(1), 95-103. [https://doi.org/10.1016/S0308-8146\(01\)00390-9](https://doi.org/10.1016/S0308-8146(01)00390-9)
- Eggleston, G., & Lima, I. (2015). Sustainability issues and opportunities in the sugar and sugar-bioproduct industries. *Sustainability*, *7*(9), 12209-12235. <https://doi.org/10.3390/su70912209>
- Guimarães, C. C., Assis, C., Simeone, M. L. F., & Sena, M. M. (2016). Use of near-infrared spectroscopy, partial least-squares, and ordered predictors selection to predict four quality parameters of sweet sorghum juice used to produce bioethanol. *Energy & fuels*, *30*(5), 4137-4144. <https://doi.org/10.1021/acs.energyfuels.6b00408>
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la Investigación* (6ª ed.). Mc Graw Hill Education.
- Islam, M. S., Pan, Y. B., Lomax, L., & Grisham, M. P. (2021). Identification of quantitative trait loci (QTL) controlling fibre content of sugarcane (*Saccharum hybrids* spp.). *Plant Breeding*, *140*(2), 360-366. <https://doi.org/10.1111/pbr.12912>



- Jackson, P. A. (2005). Breeding for improved sugar content in sugarcane. *Field Crops Research*, 92(2-3), 277-290. <https://doi.org/10.1016/J.FCR.2005.01.024>
- Lanza, J. G., Churión, P. C., & Gómez, N. (2016). Comparison between Kjeldahl traditional method and automated Dumas (N cube) method for determination of proteins in several kinds of food. *Saber*, 28(2), 245-249. <https://www.redalyc.org/articulo.oa?id=427749623006>
- Lejars, C., Auzoux, S., Siegmund, B., & Letourmy, P. (2010). Implementing sugarcane quality-based payment systems using a decision support system. *Computers and Electronics in Agriculture*, 70(1), 225-233. <https://doi.org/10.1016/J.COMPAG.2009.10.010>
- Li, Z., Gao, X., & Lu, D. (2021). Correlation analysis and statistical assessment of early hydration characteristics and compressive strength for multi-composite cement paste. *Construction and Building Materials*, 310, 125260. <https://doi.org/10.1016/j.conbuildmat.2021.125260>
- Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual review of public health*, 23(1), 151-169. <https://doi.org/10.1146/annurev.publhealth.23.100901.140546>
- Phuphaphud, A., Saengprachatanarug, K., Posom, J., Maraphum, K., & Taira, E. (2020). Non-destructive and rapid measurement of sugar content in growing cane stalks for breeding programmes using visible-near infrared spectroscopy. *Biosystems Engineering*, 197, 76-90. <https://doi.org/10.1016/j.biosystemseng.2020.06.012>
- Plaza-Díaz, J., & Gil, A. (2015). Sucrose: Dietary Importance. In B. Caballero, P. M. Finglas & F. Toldrá (Eds.), *Encyclopedia of Food and Health* (pp. 199-204). Elsevier Ltd. <https://doi.org/10.1016/B978-0-12-384947-2.00668-1>
- Reyes-Hernández, J., Torres-de los Santos, R., Hernández-Torres, H., Hernández-Robledo, V., Alvarado-Ramírez, E., & Joaquín-Cancino, S. (2022). Rendimiento y calidad de siete variedades de caña de azúcar en El Mante, Tamaulipas. *Revista Mexicana De Ciencias Agrícolas*, 13(5), 883-892. <https://doi.org/10.29312/remexca.v13i5.3232>
- Robertson, M. J., & Donaldson, R. A. (1998). Changes in the components of cane and sucrose yield in response to drying-off of sugarcane before harvest. *Field Crops Research*, 55(3), 201-208. [https://doi.org/10.1016/S0378-4290\(97\)00065-8](https://doi.org/10.1016/S0378-4290(97)00065-8)
- Sajid, M., Amjid, M., Munir, H., Valipour, M., Rasul, F., Khil, A., Alqahtani, M. D., Ahmad, M., Zulfiqar, U., Iqbal, R., Ali, M. F., & Ibtahaj, I. (2023). Enhancing sugarcane yield and sugar quality through optimal application of polymer-coated single super phosphate and irrigation management. *Plants*, 12(19), 3432. <https://doi.org/10.3390/plants12193432>
- Salgado, S., Núñez, R., J Peña, J., Etchevers, J. D., Palma, D. J., & Soto, R. M. (2003). Manejo de la fertilización en el rendimiento, calidad del jugo y actividad de invertasas en caña de azúcar. *Interciencia*, 28(10), 576-580. <https://www.redalyc.org/articulo.oa?id=33908503>
- Sawada, T. (2021). Conditions of the central-limit theorem are rarely satisfied in empirical psychological studies. *Frontiers in Psychology*, 12, 762418. <https://doi.org/10.3389/fpsyg.2021.762418>



- Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of clinical epidemiology*, 98, 146-151. <https://doi.org/10.1016/j.jclinepi.2017.12.006>
- Serrano Febles, J., Luis León, M., & Luis Orozco, J. (2022). Análisis de la situación operacional de la etapa de extracción de un Central azucarero. *Ingeniería y Desarrollo*, 40(2), 114-130. <https://doi.org/10.14482/inde.40.02.624.749>
- Sorol, N., Zossi, S., Juarez, B., Diez, P., Medina, S., & Ruiz, M. (2021). Espectroscopía NIRS en el estudio de calidad de caña de azúcar Parte I: Correlaciones para Brix, Pol, Sacarosa, Glucosa y Fructosa. *Revista industrial y agrícola de Tucumán*, 98(1), 61-66.
- Waclawovsky, A. J., Sato, P. M., Lembke, C. G., Moore, P. H., & Souza, G. M. (2010). Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. *Plant biotechnology journal*, 8(3), 263-276. <https://doi.org/10.1111/j.1467-7652.2009.00491.x>

