

Análisis de la incidencia de variables morfológicas en el peso del tallo de caña de azúcar según tres variedades diferentes, en una finca ubicada en Alajuela, Costa Rica

Karen Acuña Poveda¹, Elizabeth Araya Murillo¹, Andrés Padilla Loria¹

karen.acunapoveda@ucr.ac.cr, elizabeth.arayamurillo@ucr.ac.cr,
andres.padillaloria@ucr.ac.cr

RESUMEN

Este estudio se enfoca en la caña de azúcar (*Saccharum officinarum*), destacada por su importancia en la economía mundial debido a su versatilidad y adaptabilidad a diferentes condiciones ambientales, así como los constantes desafíos que enfrentan los agricultores de mejorar la producción, buscando incrementar tanto el rendimiento como la calidad de tallos, además de las remuneraciones. La investigación tiene como objetivo analizar la incidencia de variables morfológicas en el peso de los tallos de caña en tres variedades específicas: RB 86-7515, LAICA 07-20 y MEX 79-431. Se emplea un enfoque analítico con modelos de regresión lineal múltiple, considerando altura, grosor, cantidad y longitud de entrenudos, así como la variedad, como variables predictoras y el peso como variable respuesta. La metodología incluye un diseño observacional con 60 tallos seleccionados aleatoriamente y distribuidos equitativamente entre las tres variedades. Se realiza la verificación de supuestos, y para abordar la multicolinealidad, se aplica el método de Análisis de Componentes Principales. Los resultados indican que estas variables afectan diferencialmente el peso en cada variedad, destaca que la variedad RB 86-7515 muestra mayor influencia de las variables, lo cual generaría que produzca tallos más pesados en comparación con las otras variedades. Sin embargo, se observa una discrepancia con otros estudios que posicionan a LAICA 07-20 como una de la variedad más sobresaliente, por lo cual se sugiere que factores como condiciones climáticas y gestión del cultivo pueden influir en el potencial genético de las variedades. En la conclusión se resalta la importancia de realizar estudios complementarios en diversos entornos agrícolas para orientar decisiones en la selección de variedades y lograr rendimientos óptimos en la producción de caña de azúcar.

PALABRAS CLAVES: Modelo de regresión, supuestos, muestra, componentes principales

INTRODUCCIÓN

La caña de azúcar (*Saccharum officinarum*) se ha destacado como un cultivo de gran importancia en la economía de muchas regiones del mundo, gracias a su versatilidad y adaptabilidad a diferentes condiciones ambientales, según resaltan Lagos-Burbano y Castro-Rincón(2019). La optimización de su producción se ha convertido en un desafío constante para los agricultores, quienes buscan mejorar por un lado la calidad de la caña que depende de la condición del jugo influido por la morfología de los tallos, el contenido de sacarosa

¹Estudiantes de Bachillerato de Estadística de la Universidad de Costa Rica

(determinada por el peso como porcentaje del aparente contenido de este en relación con el peso de la caña); finalmente, por los compuestos no sacarosos, según Larrahondo (1995, p.338-343). Por otro lado, el peso de la caña influye significativamente en la remuneración de los productores, ya que la fórmula utilizada por los ingenios² para establecer el pago por la producción contempla las toneladas de caña por hectárea recolectada (TCH) (Vargas, E., comunicación personal, 3 de octubre del 2023). Por ende, esta investigación se centra en un aspecto fundamental: la incidencia de diversas variables en el peso de los tallos de caña de azúcar, con un enfoque específico en tres variedades cultivadas en la finca del señor Marlon Cruz Espinoza, ubicada en Tacaes, Grecia, en la provincia de Alajuela, Costa Rica.

En este contexto, la aplicación de modelos de regresión lineal múltiple se presenta como una herramienta analítica poderosa, permitiendo no solo entender las relaciones existentes entre las variables predictoras y la variable respuesta, sino también realizar descripciones valiosas para la producción agrícola. Con este enfoque, el objetivo es analizar la incidencia de las variables morfológicas en el peso de los tallos de caña de azúcar en función de la variedad a la que pertenecen. Esto con el fin de comprender cómo características como la altura, grosor, cantidad de entrenudos, la longitud del entrenudo influye en el peso de los tallos, según la variedad a la que pertenezca. Todo ello con el deseo de proporcionar información valiosa para la toma de decisiones en la optimización de la producción agrícola de caña de azúcar.

A nivel específico, se pretende analizar las relaciones entre las variables morfológicas y la variedad de caña de azúcar para comprender su incidencia en el peso de los tallos. Además, se busca identificar si alguna de las variedades tiende a producir tallos significativamente más pesados o más ligeros en comparación con las demás. Los resultados de este análisis ayudarán a proporcionar recomendaciones basadas en datos concretos, orientadas a optimizar las prácticas de cultivo y la calidad de los tallos de caña de azúcar.

METODOLOGÍA

El presente trabajo se realizó mediante el diseño de un estudio observacional, que implica el registro de datos sobre las variables de interés de la unidad de estudio sin ejercer influencia en dichas mediciones (Kelmansky, 2009). La unidad de estudio seleccionada para el análisis estadístico fue el tallo de la caña de azúcar, situado en una plantación propiedad del señor Marlon Cruz Espinoza en Tacaes, Grecia. La muestra consistió en 60 tallos seleccionados aleatoriamente, distribuidos equitativamente en grupos de 20 tallos para cada una de las tres variedades de caña de azúcar presentes: RB 86-7515, LAICA 07-20 y MEX 79-431.

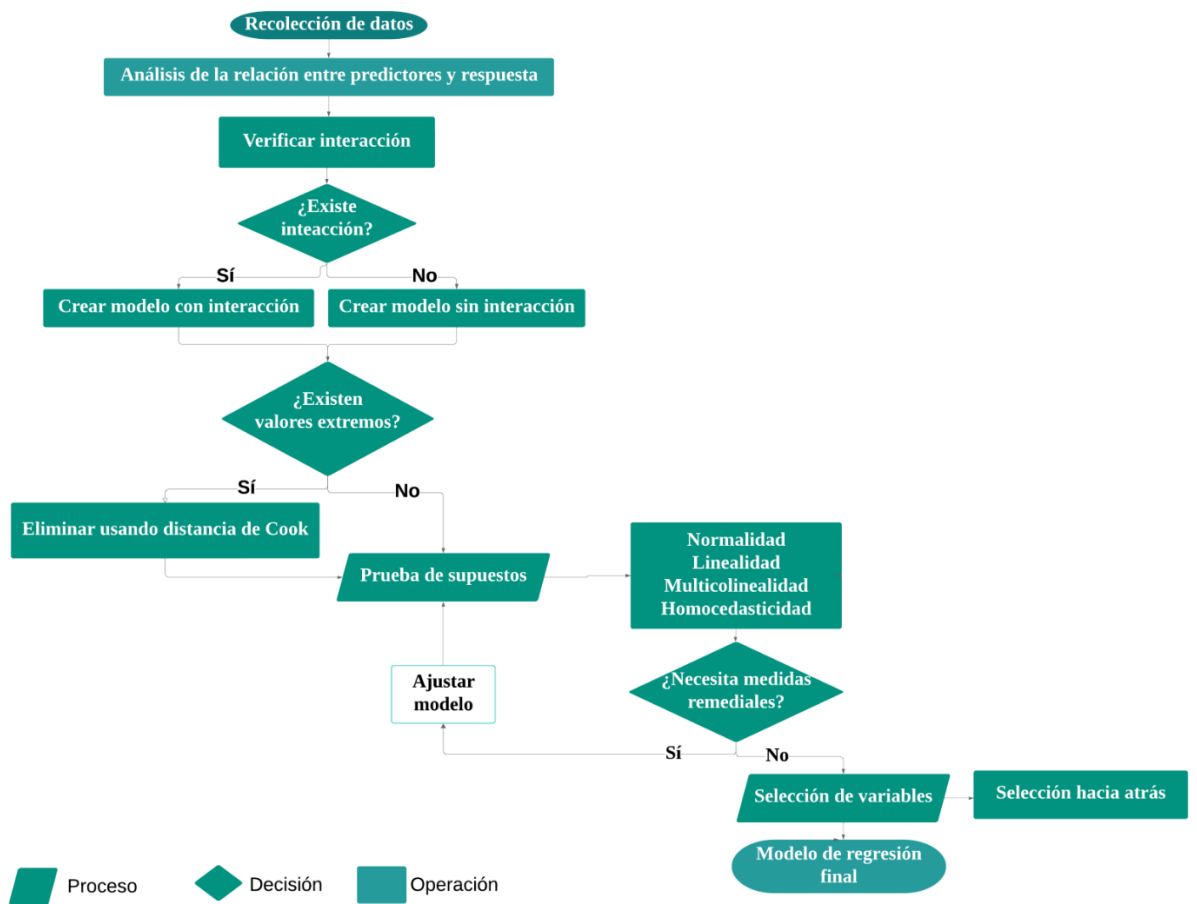
La línea de trabajo seguida para realizar el análisis de los datos se puede observar en la Figura 2. Sin embargo, a continuación, se presenta una descripción detallada de cada paso. La selección de los tallos se llevó a cabo utilizando una muestra aleatoria con distribución normal,

² Ingenios: “Finca o hacienda donde se encuentran las instalaciones utilizadas para procesar caña de azúcar con el objeto de obtener azúcar, ron, alcohol y otros productos” (Diloné, 2019).

que permitió la elección de un surco³ y una cepa⁴, donde posteriormente se eligió un tallo para ser cuidadosamente cortado. Después, se hicieron las mediciones de las variables de interés. Este proceso de recolección de datos, llevado a cabo durante tres días, contó con la asistencia del experto en Ingeniería agrónoma, el Lic. Eduardo Vargas Miranda, asegurando así la calidad y rigurosidad del procedimiento.

Figura 1

Diagrama de flujo



Para llevar a cabo el análisis de datos, se definieron cinco variables predictoras, cuatro de naturaleza numérica y una de carácter categórico. Las variables numéricas comprenden la altura, medida en centímetros desde la base del tallo hasta la primera lígula⁵ visible; el grosor, medido en centímetros mediante la circunferencia en la mitad del tallo según su altura; la

³ Surco: “Hendidura que se hace en la tierra con el arado.” (Real Academia Española, s.f., definición 1).

⁴ Cepa: “Conjunto de varias plantas que tienen una raíz común” (Real Academia Española, s.f., definición 8).

⁵ Lígula: “Es una estructura con forma de lengüeta que sobresale y forma parte del cuello de la hoja” (Mendez, 2018).

cantidad de entrenudos, cuantificando los presentes en cada tallo; y la longitud del entrenudo, medida en centímetros como la distancia entre los entrenudos ubicados en la mitad del tallo.

La elección de la variedad de caña de azúcar como variable categórica se basó en el hecho de que las tres variedades presentes compartían condiciones climáticas y de mantenimiento idénticos. Según el Lic. Eduardo Vargas Miranda, mantener condiciones consistentes proporciona una base sólida para realizar comparaciones significativas y extraer conclusiones más confiables sobre las diferencias entre las variedades de caña de azúcar. Es crucial resaltar que la variable respuesta se definió como el peso de cada tallo, medido en kilogramos.

Después de la definición de las variables, se visualizaron los datos para comprender su comportamiento y su relación con la variable respuesta. Acto seguido, se desarrolló un modelo que englobó todas las posibles interacciones entre las variables numéricas y la categórica. Es importante aclarar que se tomó como referencia la variedad RB 86-7515, la cual desempeña el papel de variedad testigo destacada por ser la de mayor uso en la zona, de acuerdo con la Oficina Nacional de Semillas (s. f).

Este modelo se utilizó con el fin de identificar la presencia de valores extremos a través de la matriz sombrero o leverages, debido a que este método posee una mayor precisión para distinguir valores atípicos en los valores de X, tomando en cuenta el número de variables utilizados, en comparación a gráficos de dispersión, como menciona Hoaglin y Welsch. Seguidamente, se llevó a cabo un análisis de la influencia en dos aspectos: sobre un valor ajustado (DFFITS) y con la distancia de Cook. El modelo resultante es:

$$\begin{aligned} \mu_{Y|A,CE,G,LE,V} = & \beta_0 + \beta_1A + \beta_2CE + \beta_3G + \beta_4LE + \beta_5LAI + \beta_6MEX + \beta_7A \times LAI \\ & + \beta_8CE \times LAI \\ & + \beta_9G \times LAI + \beta_{10}LE \times LAI + \beta_{11}A \times MEX + \beta_{12}CE \times MEX + \beta_{13}G \times MEX \\ & + \beta_{14}LE \times MEX \end{aligned}$$

Donde:

- Y= Peso del tallo
- A= Altura del tallo
- CE= Cantidad de entrenudos
- G= Grosor del tallo
- LE= Longitud de entrenudo
- V= Variedad
- LAI= LAICA 07-20
- MEX= MEX 79-431

Después de examinar la presencia de valores extremos, se procedió a verificar el cumplimiento de los supuestos que, según Pardo y Ruiz (2005), constituyen un conjunto de condiciones necesarias para asegurar la validez del modelo. En este contexto, se evaluó la normalidad mediante un gráfico de probabilidad normal (QQ Plot); la linealidad, a través de

gráficos de los residuales contra los predictores; la homocedasticidad con gráficos de los residuales contra los valores ajustados, además de la Prueba formal del Breusch-Pagan; y la no-colinealidad mediante un gráfico de la matriz de correlación, así como el factor de inflación de la varianza generalizado (GVIF).

Sin embargo, este último supuesto no se cumplió debido a la presencia de multicolinealidad entre las variables predictoras. Para abordar este problema, se optó por centrar los predictores con la finalidad de reducir la presencia de multicolinealidad. No obstante, esta medida no fue suficiente para resolver el problema por lo cual se decidió aplicar adicionalmente el método remedial de Análisis de Componentes Principales (PCA, siglas en inglés), que consiste en transformar las variables predictoras altamente correlacionadas en un nuevo conjunto de variables no correlacionadas utilizadas como nuevas variables predictoras, conocidas como componentes principales (Kutner et al, 2004, p.403).

Posteriormente, se ajustó el modelo utilizando los dos componentes obtenidos del PCA y su interacción con la variable categórica. Se evaluaron nuevamente los supuestos, los cuales se cumplieron satisfactoriamente. Luego, se llevó a cabo una reducción en el modelo mediante el método de selección por pasos hacia atrás, que consiste en eliminar el predictor cuya probabilidad asociada es mayor a α de 0,15. El modelo ajustado es el siguiente:

$$\mu_{Y|Z_1, Z_2, V} = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 \text{LAI} + \beta_4 \text{MEX} + \beta_5 Z_1 \times \text{LAI} + \beta_6 Z_1 \times \text{MEX} + \beta_7 Z_2 \times \text{LAI} + \beta_8 Z_2 \times \text{MEX}$$

El predictor eliminado después de aplicar dicho método fue el de la interacción entre el Z_1 y la variedad, ya que su contribución no resultó significativa, presentando una probabilidad asociada superior al α establecido. El modelo reducido, que cumplió con todos los supuestos y en el cual todas sus variables eran significativas, se presenta a continuación:

$$\mu_{Y|Z_1, Z_2, V} = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 \text{LAI} + \beta_4 \text{MEX} + \beta_5 Z_2 \times \text{LAI} + \beta_6 Z_2 \times \text{MEX}$$

Finalmente, es relevante señalar que todos los análisis fueron realizados utilizando R, un lenguaje de programación especializado en estadística (Mendoza Vega, 2018), en su versión 4.3.1 (R Core Team, 2023), a través de la interfaz RStudio (2020). Asimismo, se emplearon los paquetes lattice (Sarkar D, 2008), car (Fox, J., y Weisberg, S., 2019) y readxl (Wickham y Bryan, 2022).

RESULTADOS

En el análisis del comportamiento de las productoras contra la variable respuesta, se distingue que ninguna de ellas exhibe una tendencia distinta a la lineal. Asimismo, se sugiere la posible existencia de interacciones entre las variables numéricas con la categórica. Además, no se identifican valores extremos ni puntos de influencia que puedan afectar los resultados obtenidos. Al evaluar los supuestos del modelo completo, que incorpora las interacciones, se evidencia que el factor de inflación de la varianza generalizado (GVIF) presenta valores que superan el límite recomendado (ver Cuadro 1 en Anexos). Conforme a Nahhas (2023), este

límite generalmente cuando hay interacciones se establece como la $\sqrt{10}$; por lo tanto, se indica la presencia de multicolinealidad.

Para abordar esta situación, se centran las variables; pero, dado que los valores del GVIF son altos, se opta por aplicar el método de Análisis de Componentes Principales. Es importante mencionar que fue necesario utilizar solamente dos componentes, ya que estos explicaban un 82% de la variabilidad en los datos, a continuación, se presentan dichos componentes:

$$Z_1 = -0.48A + 0.52CE + 0.19G - 0.68LE$$

$$Z_2 = -0.56A - 0.39CE - 0.72G - 0.09LE$$

Posteriormente, se ajusta el modelo y se vuelve a comprobar los supuestos, los cuales se cumplen todos. Seguidamente se emplea el método de selección por pasos hacia atrás para reducir el modelo, conservando únicamente aquellas variables significativas. La interacción entre el primer componente y la variedad, al poseer una probabilidad asociada de 0.99, se elimina.

Con el modelo reducido, se realiza una nueva verificación de los supuestos. En esta fase, se observa que todos los valores del GVIF son inferiores al límite establecido (ver Cuadro 2 en Anexos), indicando que se ha solventado el problema de multicolinealidad. Respecto a la homocedasticidad, según la Figura 3 (ver en Anexos), se puede inferir que se está cumpliendo. Para validar esta suposición, se examina la probabilidad asociada a la prueba de Breusch-Pagan, que es igual a 0.32, indicando que no hay suficiente evidencia estadística para rechazar la hipótesis nula de varianzas iguales.

De acuerdo a Kutner et al(2005, p.104-110), se puede suponer normalidad al analizar los residuales parciales contra cuantiles normales (ver Figura 4 en Anexos), gracias a como los residuales se distribuyen de forma casi lineal. Con respecto a la linealidad, para decidir si la función de regresión lineal es adecuada, residuales parciales contra los predictores (ver Figura 5 en Anexos) se identifica una relación linealidad negativa.

Por otro lado, se observa que el modelo final se expresa en términos de los componentes principales, los cuales eran combinaciones lineales de las variables numéricas originales. Estos componentes presentan la desventaja de limitar la capacidad para realizar interpretaciones, ya que carecían de un significado claro en relación con las variables originales (González, 2023). Por consiguiente, se procedió a desarrollar los componentes principales para que el modelo quedará en términos de las variables originales lo cual posibilitó la interpretación de los resultados mediante las ecuaciones estimadas para cada variedad. Por lo tanto, el modelo final se presenta a continuación:

$$\begin{aligned} \mu_{Y|A,CE,G,LE,V} = & Y_0 + Y_1A + Y_2CE + Y_3G + Y_4LE + Y_5LAI + Y_6MEX - 0.56Y_7A \times LAI \\ & - 0.39Y_8CE \times LAI - 0.72Y_9G \times LAI - 0.09Y_{10}LE \times LAI - 0.56Y_{11}A \times MEX \end{aligned}$$

$$-0.39Y_{12}CE \times MEX - 0.72Y_{13}G \times MEX - 0.09Y_{14}LE \times MEX$$

Donde:

- $Y_0 = \beta_0$
- $Y_1 = (-0.48 \beta_1 - 0.56 \beta_2)$
- $Y_2 = (0.52 \beta_1 - 0.39 \beta_2)$
- $Y_3 = (0.19 \beta_1 - 0.72 \beta_2)$
- $Y_4 = (-0.68 \beta_1 - 0.09 \beta_2)$

Al confirmar el cumplimiento de todos los supuestos, se procede a estimar las ecuaciones de las tres variedades que conforman la variable categórica. Dichas ecuaciones se presentan a continuación:

- Ecuación estimada para la variedad RB 86-7515

$$\hat{y} = 0.91 + 0.23A + 0.11CE + 0.25G + 0.07LE$$
- Ecuación estimada para la variedad LAICA 07-20

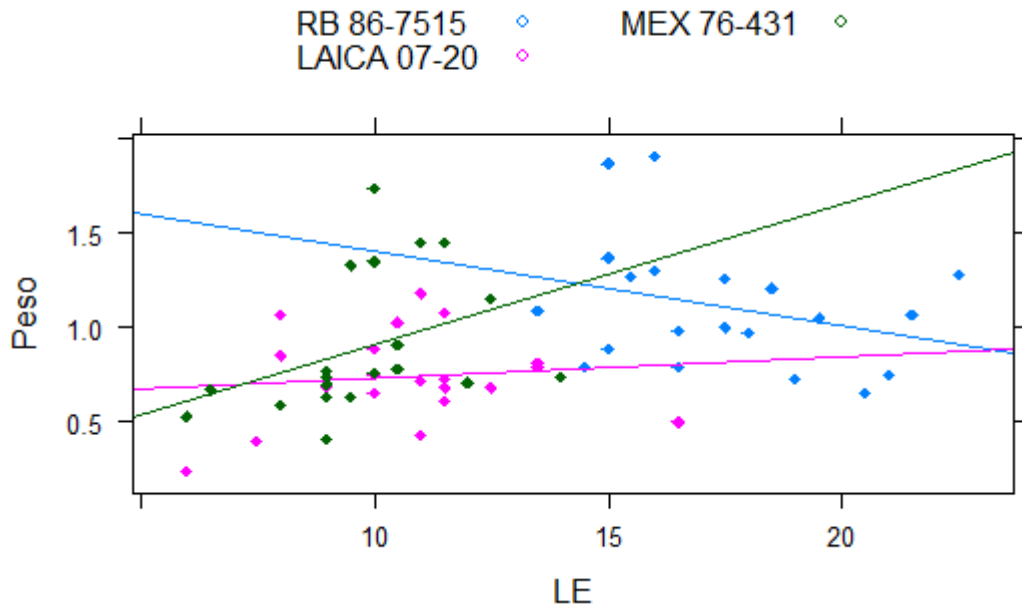
$$\hat{y} = 0.88 + 0.15A + 0.05CE + 0.14G + 0.06LE$$
- Ecuación estimada para la variedad MEX 79-431

$$\hat{y} = 0.90 + 0.19A + 0.08CE + 0.20G + 0.06LE$$

A partir de estas ecuaciones, se puede destacar que existe una relación parcial positiva entre la altura, el grosor, la cantidad de entrenudos y la longitud del entrenudo con el peso del tallo en las tres variedades. Asimismo, es relevante señalar el contraste evidente que se identificó entre las pendientes en los gráficos donde se evaluó la interacción con el valor de obtenido en las ecuaciones. Por ejemplo, en la Figura 2 se puede visualizar que la pendiente de MEX 76-431 es la más pronunciada; sin embargo, en la ecuación, tiene el mismo valor que LAICA 07-20, siendo a su vez ligeramente inferior al valor RB 86-7515.

Figura 2

Relación entre Longitud de entrenado y la variedad.



Nota: Verificación de interacción

Al profundizar en el análisis, destaca que la altura, el grosor y la cantidad de entrenudos influyen más en el peso del tallo en la variedad RB 86-7515 en comparación con las otras dos variedades, especialmente en comparación con LAICA 07-20. Este hallazgo respalda la consideración de RB 86-7515 como una variedad altamente productiva, lo que se refleja en su extenso cultivo en la región y su posición como una variedad testigo. Por otro lado, se observa que la diferencia en el efecto de la longitud del entrenudo sobre el peso del tallo es insignificante.

Entre las variedades MEX 79-431 y LAICA 07-20, se identifica que la altura, el grosor y la cantidad de entrenudos influyen más en el peso de los tallos en la variedad MEX 79-431, aunque la longitud del entrenudo muestra el mismo efecto en ambas variedades. En este contexto, se destaca una diferencia más marcada en el caso del grosor, este último factor exhibe una variación más significativa en su impacto en el peso del tallo entre ambas variedades.

En resumen, a través del análisis detallado de las ecuaciones, se revela que las variables de altura, grosor, cantidad de entrenudos y longitud del entrenudo impactan de manera distinta en el peso del tallo en cada una de las tres variedades estudiadas. Estas diferencias en la magnitud de los efectos proporcionan información valiosa sobre las características específicas de cada variedad, cuando comparten las mismas condiciones, y pueden ser fundamentales para la toma de decisiones en el cultivo y selección de las mismas.

CONCLUSIONES

El presente estudio se ha enfocado en analizar la incidencia de las variables morfológicas en el peso de los tallos de caña de azúcar, según la variedad a la que pertenecen. Con la finalidad de comprender cómo estas características podían afectar el peso de los tallos. Los resultados obtenidos en los análisis pertinentes permitieron identificar que la variedad en la que las variables ejercen mayor influencia es la RB 86-7515. Esto podría ser un factor de aumento en el peso de los tallos, hecho que le permitiese seguir destacándose como una de las variedades con mayor rendimiento en la producción de toneladas de caña por hectárea (Durán Alfaro & Oviedo Alfaro, 2002). No obstante, cabe aclarar que esta afirmación queda fuera del alcance de la investigación.

Sin embargo, estos han presentado una discrepancia al compararlos con afirmaciones realizadas por otros autores, quienes destacan en sus investigaciones a diversas variedades LAICA, entre ellas LAICA 07-20. Como una de las más sobresalientes puesto que han mostrado un mayor rendimiento tanto en toneladas de caña como de azúcar por hectárea, superando a las variedades RB 86-7515 y Mex 79-431 (Chávez Solera, 2016, p.27).

Esta discrepancia podría ser atribuida a diferentes razones, entre ellas las condiciones climáticas, fertilización, tipo de suelo u otras características referentes a la gestión del cultivo lo cual podría estar afectando su potencial genético, esto se fundamenta con lo afirmado por Chávez Solera (2016): “Cabe por ello la posibilidad de que un excelente material promisorio no alcance a expresar sus atributos por inadaptación del lugar donde se cultiven; o también, que un manejo inconveniente y deficiente minimice su potencial real.”

En conclusión, la divergencia entre nuestros resultados y los informados por otros autores destaca la importancia de la validación y la replicación de estudios en distintos entornos agrícolas donde se posea mayor control de las condiciones de siembra para cada variedad. La realización de estudios complementarios permitirá una evaluación más integral de las variedades de caña de azúcar y proporcionará información valiosa para orientar las decisiones en la selección de variedades para obtener un rendimiento óptimo en la producción de caña de azúcar.

BIBLIOGRAFÍA

- Chávez Solera, M. (2016). Programa de variedades. Informe de resultados 2015. LAICA. <https://servicios.laica.co.cr/laica-cv-biblioteca/index.php/Library/download/PJkiMiseensDuzndbKwGNUpKlxMETCoW>
- Diloné, C. M. (2019, October 9). *Sobre los ingenios azucareros*. Familia Bateyera. <https://familiabateyera.com/sobre-los-ingenios-azucareros/>
- Durán Alfaro, J. R., & Oviedo Alfaro, M. (2002). Catálogo de variedades de caña de azúcar cultivadas en algunas regiones cañeras de Costa Rica. LAICA. <https://servicios.laica.co.cr/laica-cv-biblioteca/index.php/Library/download/gmEwSbPflLsZtgpWfzdVfQMdzOjHYmzw>
- Lagos-Burbano, E., & Castro-Rincón, E. (2019). Caña de azúcar y subproductos de la agroindustria azucarera en la alimentación de rumiantes. *Agronomía Mesoamericana*, 917–934. <https://doi.org/10.15517/am.v30i3.34668>
- Larrahondo, J. E. (1995). Calidad de la caña de azúcar. Repositorio Institucional Agrosavia. Corporación colombiana de investigación agropecuaria. <http://hdl.handle.net/20.500.12324/1362>

- Fox J., & Weisberg S. (2019). *An R Companion to Applied Regression*, Third Edition. Thousand Oaks CA: Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- González, F. (2023, 5 mayo). *Análisis de Componentes principales (PCA)*. <https://www.linkedin.com/pulse/an%C3%A1lisis-de-componentes-principales-pca-francisco-gonz%C3%A1lez/?originalSubdomain=es>
- Hoaglin, D. y Welsch, R. (1978). The Hat Matrix in Regression and ANOVA. *The American Statistician*, 32:1, 17-22. <https://doi.org/10.1080/00031305.1978.10479237>
- Kelmansky, D. M. (2009). *Estadística para todos: Estrategias de pensamiento y herramientas para la solución de problemas* (1.ª ed.). Ministerio de Educación de la Nación.
- Kutner, M. H., Nachtsheim, C., Neter, J., Li, W. (2005). *Applied Linear Statistical Models*. India: McGraw-Hill Irwin.
- Mendez, L. O. Z. (2018, 20 marzo). *Técnica para la identificación de variedades y especies de caña de azúcar*. Engormix. https://www.engormix.com/agricultura/cana-azucar/tecnica-identificacion-variedades-especies_a42077/
- Mendoza Vega, J. B. (2018). *R Para principiantes*. <https://bookdown.org/jboscomendoza/r-principiantes4/introduccion-que-es-r-y-para-que-es-usado.html>
- Nahhas, R. W. (2023, 21 noviembre). *5.19 Collinearity | Introduction to Regression Methods for Public Health using R*. <https://www.bookdown.org/rwnahhas/RMPH/mlr-collinearity.html#generalized-vifs-when-at-least-one-predictor-is-categorical>
- Oficina Nacional de Semillas. (s. f.). *PROTOCOLO PARA LA REALIZACION DE ENSAYOS DE VALORACIÓN AGRONÓMICA DE ARROZ, A SER INSCRITOS EN EL REGISTRO DE VARIEDADES COMERCIALES (R.V.C.)*. https://ofinase.go.cr/wp-content/uploads/2017/09/rvc_protocoloregistroarroz.pdf
- Pardo, A., & Ruiz, M. Á. (2005). Análisis de regresión lineal: El procedimiento Regresión lineal. En C, Sánchez González (Ed.), *Análisis de datos con SPSS 13 Base* (1.ª ed., p. 494). McGraw-Hill.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Versión: 4.3.1. <https://www.R-project.org/>
- Real Academia Española. (s.f.). *Surco | Diccionario de la Lengua Española*. «Diccionario de la lengua española» - Edición del Tricentenario. <https://dle.rae.es/surco>
- Real Academia Española. (s.f.). *CEPA | Diccionario de la Lengua Española*. «Diccionario de la lengua española» - Edición del Tricentenario. <https://dle.rae.es/cepa>
- RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA URL. Versión: 4.3.1. <http://www.rstudio.com/>
- Sarkar D (2008). *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5. Versión: 0.21.8. <http://lmdvr.r-forge.r-project.org>
- Wickham, H y Bryan, j. (2022). *readxl: Read Excel Files*. R package version 1.4.1 <https://cran.rproject.org/web/packages/readxl/index.htm>

ANEXOS

Cuadro 1

Factor de Inflación de la varianza (VIF) del modelo de regresión inicial.

Variable	GVIF
Altura	3,74
Grosor	4.74
Cantidad de entrenudos	3.90
Longitud de entrenudos	3.71
Interacción de Altura con Variedad	2.69
Interacción de Grosor con Variedad	2.76
Interacción de Cantidad de entrenudos con Variedad	2.33
Interacción de Longitud de entrenudo con Variedad	2.67

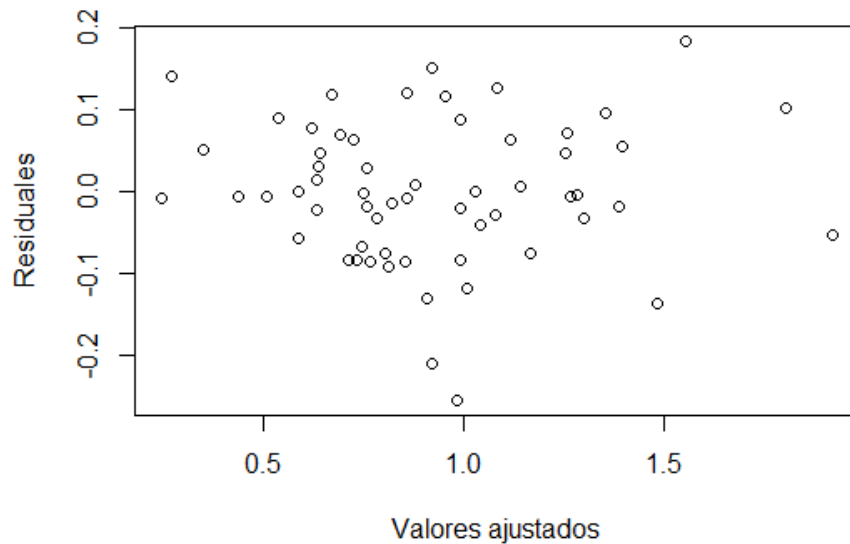
Cuadro 2

Factor de Inflación de la varianza (VIF) del modelo de regresión reducido.

Variable	GVIF
Z_1	2.01
Z_2	2.18
Variedad	1.49
Interacción de Z_2 con Variedad	1.49

Figura 1

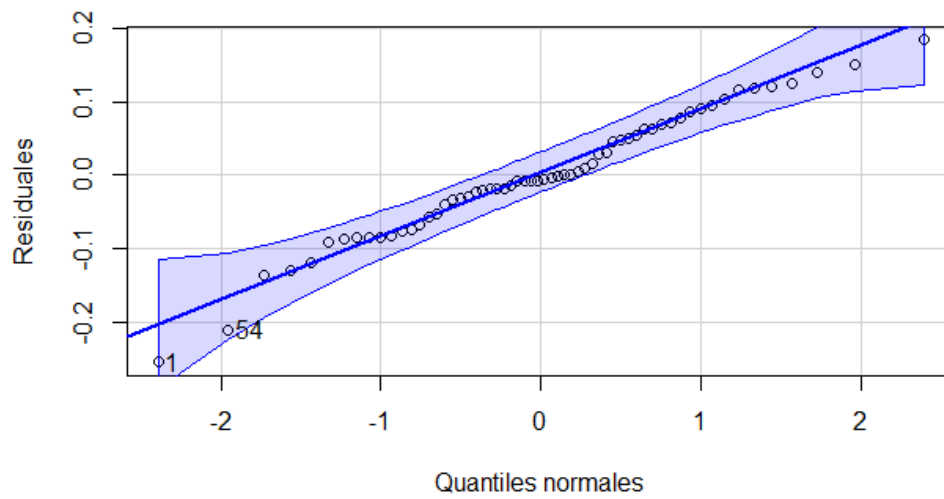
Residuales parciales del modelo de regresión final.



Nota: Supuesto de homocedasticidad

Figura 2

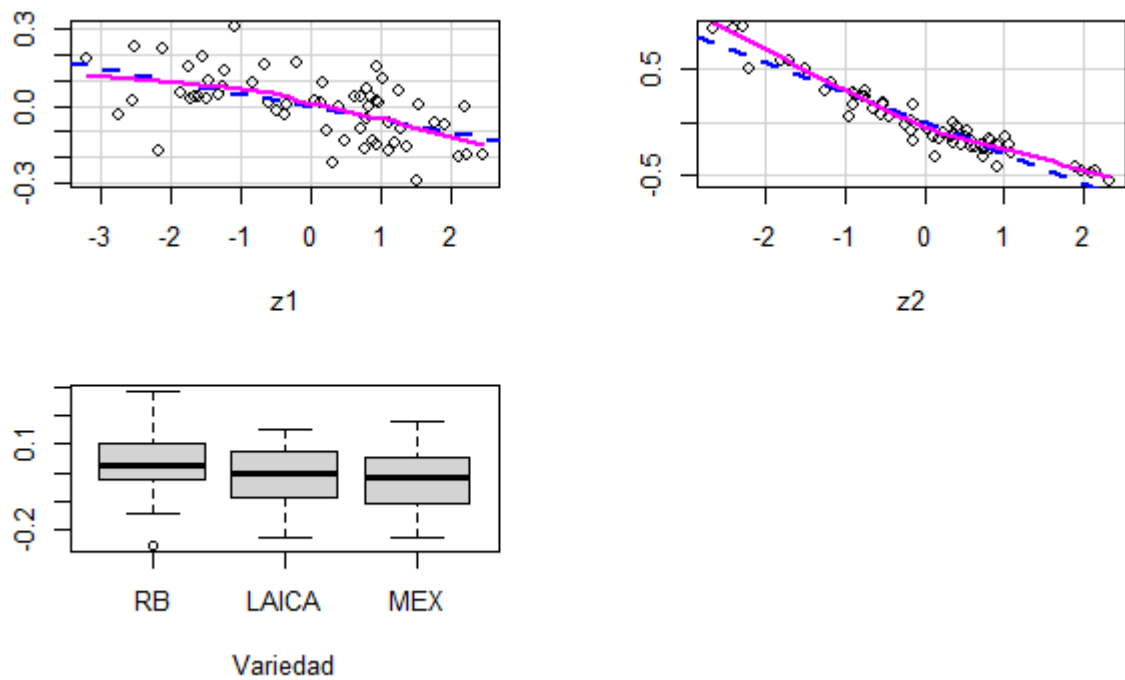
Residuales parciales del modelo inicial contra cuantiles normales



Nota: Supuesto de normalidad

Figura 3

Residuales parciales contra predictores del modelo final.



Nota: Supuesto de linealidad