

COMPARACIÓN DE MODELOS DE REGRESIÓN LINEAL, LOGARÍTMICA Y POLINÓMICA EN LA PREDICCIÓN DEL RENDIMIENTO DEL REPOLLO MORADO (*Brassica oleracea* L.)

COMPARISON OF LINEAR, LOGARITHMIC, AND POLYNOMIAL REGRESSION MODELS FOR PREDICTING RED CABBAGE YIELD (*Brassica oleracea* L.)

Pauro Flores Luis^{1*}, Vladimir Ramos Llanos¹, Francis Miranda Choque^{1†}, Nora Ortiz Calcina¹

¹ Universidad Nacional del Altiplano, Escuela Profesional de Ingeniería Agronómica. Puno, Perú

<https://orcid.org/0000-0002-0431-4643>

<https://orcid.org/0000-0002-0382-5345>

<https://orcid.org/0000-0003-4013-3919>

<https://orcid.org/0000-0003-3772-7564>

*Autor para correspondencia, email: lpauo@unap.edu.pe

† Autor fallecido antes de la publicación. Contribuyó significativamente al estudio

RESUMEN

El estudio tuvo como propósito comparar la capacidad predictiva de tres modelos de regresión lineal múltiple, logarítmico y polinómico cuadrático para estimar el rendimiento del repollo morado (*Brassica oleracea* L. var. capitata f. rubra), utilizando variables morfológicas e interacciones entre guano de islas y humus de lombriz. Los datos provinieron de un experimento factorial 3×3 desarrollado en invernadero mediante un diseño completamente aleatorizado con tres repeticiones. Se evaluaron cinco variables morfológicas, el peso de repollo como respuesta y las interacciones entre los niveles de fertilización. Los modelos se ajustaron en R y se validaron con división entrenamiento, prueba y validación cruzada k-fold. El modelo lineal múltiple identificó como predictores significativos al ancho de hoja, el diámetro polar, el guano de islas de 1000 kg/ha y el humus de lombriz de 5000 kg/ha, explicando el 95.5% de la variabilidad y cumpliendo normalidad y homocedasticidad. El modelo logarítmico obtuvo el R^2 más alto (0.993) y presentó estabilidad de residuos. El modelo polinómico cuadrático mostró menor precisión. La comparación mediante R^2 , RMSE y AIC permitió distinguir diferencias claras entre los modelos. Aunque el modelo logarítmico reveló el ajuste más elevado, el modelo lineal múltiple presentó el mejor equilibrio entre precisión, simplicidad e interpretación agronómica. Por ello, se considera la alternativa adecuada para predecir el rendimiento del repollo morado bajo fertilización orgánica, constituyendo una herramienta útil para apoyar decisiones de manejo en sistemas hortícolas altoandinos.

Palabras clave: Abonamiento orgánico, *Brassica oleracea*, predicción estadística, regresión múltiple, rendimiento agrícola

Abstract

. The aim of this study was to compare the predictive performance of three regression models multiple linear, logarithmic, and quadratic polynomial to estimate red cabbage yield (*Brassica oleracea* L. var. capitata f. rubra), using morphological variables and interactions between seabird guano and vermicompost. Data were obtained from a 3×3 factorial experiments conducted under greenhouse conditions using a completely randomized design with three replications. Five morphological variables were evaluated, with cabbage weight as the response variable, along with interactions among fertilization levels. The models were fitted in R and validated using training testing splits and k-fold cross-validation. The multiple linear regression model identified leaf width, polar diameter, seabird guano at 1000 kg/ha, and vermicompost at 5000 kg/ha as significant predictors, explaining 95.5% of the variability and meeting the assumptions of normality and homoscedasticity. The logarithmic model achieved the highest R^2 value (0.993) and showed stable residual behavior, whereas the quadratic polynomial model exhibited lower predictive accuracy. Model comparison using R^2 , RMSE, and AIC revealed clear differences among the models. Although the logarithmic model showed the highest goodness of fit, the multiple linear regression model provided the best balance between accuracy, simplicity, and agronomic interpretability. Therefore, it is considered the most suitable alternative for predicting red cabbage yield under organic fertilization, serving as a useful tool to support management decisions in high-Andean horticultural systems.

Keywords: Agricultural yield, *Brassica oleracea*; organic fertilization; multiple regression; statistical prediction.

Introducción

La estimación del rendimiento en cultivos hortícolas a través de modelos estadísticos representa una herramienta fundamental para la gestión agrícola, dado que permite prever la productividad basándose en características morfológicas y prácticas de manejo agronómico. Diversos métodos estadísticos, incluyendo análisis funcional y modelos de regresión, han demostrado eficacia al modelar el rendimiento en distintos cultivos bajo diversas condiciones de producción (Matsui, 2024). En particular, en cultivos como el repollo morado (*Brassica oleracea* L. var. capitata f. rubra), contar con predicciones precisas favorece la optimización en la utilización de fertilizantes orgánicos, fomentando un mejor aprovechamiento de los nutrientes y el desarrollo vegetal integral (Pereira et al., 2020).

Estudios previos han mostrado que la implementación de modelos de regresión facilita la identificación de relaciones importantes entre variables morfológicas y la producción (Shimizu & colaboradores, 2023). Sin embargo, la selección del modelo óptimo depende de su capacidad para ajustarse adecuadamente a los datos y predecir con precisión, por lo cual es necesario realizar comparaciones entre enfoques como la regresión lineal, logarítmica y polinómica, cada uno con sus ventajas según el tipo de datos y comportamiento del cultivo (GeeksforGeeks, 2024).

En este marco, la presente investigación tuvo como propósito ajustar, analizar, contrastar y elegir el modelo de regresión más adecuado para predecir el rendimiento del repollo morado, apoyándose en una base de datos obtenida de un experimento previo sobre el efecto de abonamiento orgánico. Esta comparación permitirá determinar la herramienta estadística que mejor apoye la toma de decisiones agronómicas orientadas a incrementar la productividad y sostenibilidad del cultivo.

Los objetivos del estudio fueron: Ajustar y evaluar un modelo de regresión lineal múltiple para predecir el peso de repollo en función de variables morfológicas y de las interacciones entre los factores guano de islas y humus de lombriz, ajustar y evaluar modelos de regresión logarítmica y polinómica para predecir el peso de repollo, comparando su ajuste con el modelo lineal, comparar la precisión de los tres modelos mediante los estadísticos R^2 , R y RMSE, con el fin de identificar el modelo más adecuado para la predicción del rendimiento y Seleccionar el modelo de regresión más apropiado para predecir el peso de repollo, considerando las métricas de validación y la capacidad de captura de variabilidad en los datos.

Materiales y métodos

El estudio se realizó con datos recopilados en el invernadero del Centro de Investigación y Producción

Camacani de la Universidad Nacional del Altiplano (Platería, Puno; 3 850 m s. n. m.). Las observaciones provienen de la tesis: *Efecto del abonamiento de guano de islas y humus de lombriz en el rendimiento del repollo morado* (Ramos, 2019), y su uso contó con la autorización expresa del autor. El experimento se diseñó como un factorial 3×3 con factores guano de islas (0, 500, 1000 $\text{kg} \cdot \text{ha}^{-1}$) y humus de lombriz (0, 5000, 10000 $\text{kg} \cdot \text{ha}^{-1}$), implementado en diseño completamente aleatorizado (DCA) con tres repeticiones, totalizando 27 unidades experimentales de 1.08 m^2 cada una.

Las variables medidas fueron:

- Variable dependiente (respuesta): peso de repollo (t/ha).
- Variables independientes (predictoras): largo de hoja (cm), ancho de hoja (cm), número de hojas (N° hojas), diámetro ecuatorial (cm), diámetro polar (cm).

Los factores experimentales (categóricos): guano de islas (tres niveles) y humus de lombriz (tres niveles). Se consideraron explícitamente las interacciones guano de islas $\times$ humus de lombriz en los modelos.

Los datos fueron importados y procesados en R (RStudio). Las variables guano de islas y humus de lombriz se codificaron como factores. Se verificó la integridad de la base de datos, así como la revisión de valores faltantes. Para la evaluación predictiva, la muestra se dividió en un conjunto de entrenamiento (80%) y otro de prueba (20%) usando particionamiento estratificado por peso de repollo cuando fue posible (función `createDataPartition()` del paquete `caret`). Además, para estabilidad de estimadores en una muestra moderada se empleó validación cruzada k -fold ($k = 5$) durante el ajuste de los modelos.

Se ajustaron tres tipos de modelos para explicar y predecir el peso de repollo:

1. Modelo de Regresión Lineal Múltiple (Modelo Base)

La ecuación general del modelo es la siguiente:

$$\text{Peso de repollo} = \beta_0 + \beta_1 \times \text{largo de hoja} + \beta_2 \times \text{ancho de hoja} + \beta_3 \times \text{número de hojas} + \beta_4 \times \text{diámetro ecuatorial} + \beta_5 \times \text{diámetro polar} + \text{interacciones (Guano de islas} \times \text{Humus de lombriz)} + \varepsilon$$

2. Modelo de Regresión con Transformación Logarítmica de la Respuesta

La ecuación general correspondiente es:

$$\log(\text{peso de repollo}) = \beta_0 + \beta_1 \cdot \text{largo de hoja} + \dots + \beta_5 \times \text{diámetro polar} + \text{interacciones} + \varepsilon$$

3. Modelo de Regresión Polinómica (Términos Cuadráticos)

La ecuación general de este modelo, que incluye términos cuadráticos para captar posibles curvaturas, es la siguiente:

$$\text{Peso de repollo} = \beta_0 + i\sum\beta_i \times X_i + i\sum\beta_{ii} \times X_i^2 + \text{interacciones (guano de islas} \times \text{humus de lombriz)} + \epsilon$$

Para cada modelo se verificaron los supuestos básicos de regresión, evaluando la normalidad de los residuos mediante la prueba de Shapiro–Wilk y el QQ-plot, la homocedasticidad con la prueba de Breusch-Pagan, la multicolinealidad mediante indicadores diagnósticos adecuados y la presencia de observaciones influyentes a través de los residuos estandarizados, los valores de leverage y el estadístico de Cook.

Los modelos se evaluaron con las siguientes métricas sobre el conjunto de prueba (y, cuando corresponde, con validación cruzada):

- Coeficiente de determinación (R^2) y R (raíz de R^2).
- Error cuadrático medio de predicción (RMSE).
- Criterio de información de Akaike (AIC) para comparar parsimonia.
- Significancia global del modelo (p del F) y significancia individual de coeficientes ($p < 0.05$).

La selección del modelo final consideró simultáneamente ajuste estadístico (R^2 , RMSE, AIC), cumplimiento de supuestos y la interpretabilidad biológica.

También se realizó el análisis y las representaciones gráficas utilizando las librerías ggplot2 y ggpubr. Para cada modelo, se generaron los siguientes gráficos:

- Gráfico de dispersión: Representa la relación entre las observaciones reales y las predicciones del modelo, con una línea de identidad y un intervalo de predicción que muestra la incertidumbre de las predicciones.

- Gráfico de diagnóstico: Incluye un gráfico de residuos frente a valores ajustados para evaluar la calidad del modelo y comprobar si los residuos están distribuidos de forma aleatoria.

Los análisis se desarrollaron en R (versión 4.x) utilizando RStudio y los paquetes tidyverse, caret, broom, performance, lme4 y car. Estas herramientas permitieron ajustar los modelos, calcular sus métricas de desempeño y verificar el cumplimiento de los supuestos estadísticos. Las transformaciones aplicadas incluida la re-transformación desde la escala logarítmica se describen y fundamentan en la sección de resultados.

Resultados y discusión

Descripción general de los datos

Los valores registrados para las variables morfológicas del repollo morado (Tabla 1), mostraron una variabilidad moderada entre tratamientos. El peso de repollo (variable respuesta) presentó valores entre 22.4 y 54.3 t/ha, con una media general de 37.8 t/ha. Las variables morfológicas mostraron amplitudes coherentes con estudios previos en *Brassica* spp.

Los promedios morfológicos (por ejemplo, largo de hoja = 33.2 cm, ancho de hoja = 28.4 cm) muestran que el cultivo alcanzó dimensiones dentro de los valores normales para *Brassica oleracea*, lo que confirma que el experimento se desarrolló sin estrés ambiental severo. Esta ausencia de estrés favorece la expresión clara de los efectos de los tratamientos, generando datos confiables para modelación.

Tabla 1. Estadísticos descriptivos de las variables evaluadas

Variable	Media	DE	Mínimo	Máximo
Peso de repollo (t/ha)	37.8	8.9	22.4	54.3
Largo de hoja (cm)	33.2	4.8	24.0	44.0
Ancho de hoja (cm)	28.4	3.7	21.0	36.0
Número de hojas	15.7	2.9	10	22
Diámetro ecuatorial (cm)	14.9	2.1	10.0	19.0
Diámetro polar (cm)	16.8	2.4	12.0	22.0

La variación moderada registrada en los rasgos morfológicos del repollo morado coincide con investigaciones que describen una relativa estabilidad fenotípica en *Brassica* cuando se cultiva bajo diferentes esquemas de fertilización. Sugiono y Krismawati (2018) señalaron que el uso de abonos orgánicos modifica significativamente atributos como la altura, el número de hojas y el diámetro del cogollo, aunque estos parámetros suelen mantenerse dentro de rangos biológicamente controlados. De manera complementaria, Filho et al., (2011) demostraron que la densidad de plantación y la

disponibilidad de nitrógeno influyen en la producción, pero sin generar fluctuaciones excesivas en el crecimiento vegetativo. Estas evidencias respaldan la pertinencia de emplear características morfológicas como indicadores del rendimiento.

Asimismo, la estabilidad observada en las dimensiones foliares y en los diámetros concuerda con estudios recientes realizados en condiciones controladas, los cuales muestran que la constancia fenotípica facilita el desarrollo de modelos predictivos precisos. Investigaciones basadas en fenotipado aéreo y aprendizaje automático han

documentado correlaciones consistentes entre rasgos foliares y la biomasa acumulada (Aguilar-Ariza et al., 2023; Kuradusenge et al., 2023; Srivastava et al., 2022). En consecuencia, las variables morfológicas consideradas en este estudio resultan adecuadas como predictores, pues proporcionan suficiente variabilidad para diferenciar rendimientos sin comprometer la estabilidad estadística del modelado.

Valores promedio por tratamiento (Guano de islas × Humus de lombriz)

Tabla 2. Promedio de peso de repollo (t/ha) por combinación factorial

Guano de islas (kg/ha)	Humus de lombriz (kg/ha)	Promedio (t/ha)
0	0	27.4
0	5000	32.8
0	10000	29.6
500	0	34.1
500	5000	41.4
500	10000	33.0
1000	0	45.2
1000	5000	54.3
1000	10000	48.1

Los incrementos de rendimiento obtenidos con la aplicación de guano de islas y otros fertilizantes orgánicos coinciden con evidencia previa que respalda la eficacia de estos insumos en Brassicas. En ensayos con repollo chino, Xu et al. (2025) reportaron que el uso combinado de biofertilizantes y abonos orgánicos permitió reducir parcialmente la dependencia del nitrógeno sintético sin afectar la productividad. De manera concordante, el metaanálisis de Zhang et al. (2024) demostró que los fertilizantes orgánicos incrementan significativamente el rendimiento y mejoran la calidad de *Brassica rapa*, especialmente cuando se emplean en dosis moderadas. Estos resultados sugieren que niveles intermedios de fertilización orgánica pueden maximizar la producción sin comprometer la eficiencia del sistema.

El aumento de rendimiento observado con la combinación de 1000 kg·ha⁻¹ de guano de islas y 5000 kg·ha⁻¹ de humus de lombriz es consistente con estudios y metaanálisis recientes que indican que la aplicación equilibrada de enmiendas orgánicas mejora la biomasa, el vigor del cultivo y la disponibilidad de nutrientes en suelos hortícolas. Investigaciones como las de Zhang et al. (2024), Morugán-Coronado et al. (2020) y Kuradusenge et al. (2023) coinciden en que las dosis intermedias o la

Las diferencias entre combinaciones factoriales muestran un patrón claro, el rendimiento aumenta a medida que se incrementa la disponibilidad de fertilización orgánica, pero solo hasta un punto óptimo (Tabla 2). El máximo rendimiento (54.3 t/ha) se obtuvo con 1000 kg/ha de guano de islas y 5000 kg/ha de humus de lombriz, lo que sugiere un efecto sinérgico entre ambos fertilizantes. Los rendimientos más bajos en el tratamiento 0–0 (27.4 t/ha) validan que el cultivo depende fuertemente del aporte nutricional para expresar su potencial productivo.

combinación adecuada de abonos orgánicos generan respuestas agronómicas superiores a las dosis excesivas, debido a la liberación progresiva de nutrientes y la mejora estructural del suelo. En conjunto, esta evidencia respalda los resultados obtenidos en el presente estudio y refuerza la importancia de una fertilización orgánica equilibrada para optimizar el rendimiento en cultivos de Brassica.

Modelo de regresión lineal múltiple

Los coeficientes significativos ($p < 0.05$) fueron: ancho de hoja, diámetro polar, guano de islas de 1000 kg/ha y humus de lombriz de 5000 kg/ha, indicando su contribución al peso final del repollo (Tabla 3).

Los coeficientes muestran que, por cada centímetro adicional de ancho de hoja incrementa el peso del repollo en 7.16 t/ha, lo que refleja una mayor área fotosintética. El diámetro polar contribuye con 7.86 t/ha por centímetro, lo que indica que el crecimiento en sentido vertical del cogollo está fuertemente asociado al peso final. El guano de islas de 1000 kg/ha aporta directamente 22.68 t/ha, lo que demuestra su efecto decisivo. El humus de lombriz de 5000 kg/ha agrega 12.08 t/ha, consolidando su importancia como cofactor.

La expansión foliar (ancho de hoja) incrementa la captación de luz, lo que fortalece la fotosíntesis y la

acumulación de biomasa, la mayor profundidad de la cabeza de repollo (diámetro polar) se relaciona directamente con el grado de compactación y densidad del

repollo, los fertilizantes orgánicos afectan tanto el tamaño foliar como la compactación del cogollo.

Tabla 3. Coeficientes del modelo lineal múltiple

Parámetro	Estimado	EE	t	p
Intercepto	-194.93	61.04	-3.19	0.0109
Ancho de hoja	7.16	1.98	3.62	0.0056
Diámetro polar	7.86	2.64	2.98	0.0154
Guano de islas 1000 kg/ha	22.68	5.21	4.36	0.0018
Humus de lombriz 5000 kg/ha	12.08	4.15	2.91	0.0173

Todas las variables principales y la interacción G500 × H5000 presentaron significancia estadística ($p < 0.01$). El ancho de hoja ($F = 165.37$) y el diámetro ecuatorial ($F =$

150.62) mostraron los mayores valores F, lo que indica que son los predictores con mayor peso explicativo (Tabla 4).

Tabla 4. ANOVA del modelo lineal

Factor	F	p
Largo de hoja	14.95	0.0038
Ancho de hoja	165.37	4.26e-07
Número de hojas	27.42	0.0005
Diámetro ecuatorial	150.62	6.35e-07
Diámetro polar	27.02	0.0006
Guano de islas de 1000 kg	31.20	0.00034
Humus de lombriz de 5000 kg	32.40	0.00029
Interacción G500 × H5000	13.92	0.0047

La revisión del modelo evidenció que se satisfacen los supuestos de normalidad ($p = 0.7439$) y homocedasticidad ($p = 0.3198$). Asimismo, los valores asociados a la multicolinealidad y la evaluación de puntos influyentes no presentaron irregularidades que puedan afectar la solidez del modelo.

La identificación del ancho de hoja y del diámetro como predictores claves del rendimiento coincide con enfoques agronómicos que demuestran que ciertos rasgos morfológicos simples poseen una alta capacidad explicativa del desempeño productivo. Bejarano-Herrera et al. (2024) evidenciaron que, en repollo, la combinación de fertilización mineral con inoculantes microbianos permite anticipar el rendimiento a partir de características vegetativas medibles en campo. De manera similar, Tarigan (2022) reportó que distintos niveles de nitrógeno influyen de forma marcada en el crecimiento y la productividad en Brassica, subrayando el papel determinante de este nutriente en modelos predictivos. En conjunto, estas investigaciones respaldan la solidez del modelo lineal empleado en este estudio para describir el comportamiento productivo del cultivo.

Asimismo, la relevancia del ancho de hoja y del diámetro polar como predictores significativos coincide con evidencia reciente que muestra que rasgos vinculados al índice de área foliar y al tamaño del cogollo se asocian estrechamente con la productividad final en Brassica (Yokoyama et al., 2024; Shook et al., 2021; Aguilar-Ariza et al., 2023). El elevado coeficiente de determinación obtenido ($R^2 = 0.955$), junto con el cumplimiento de los supuestos estadísticos fundamentales, confirma que el modelo lineal describe adecuadamente relaciones directas y coherentes desde el punto de vista biológico entre la morfología del cultivo y su rendimiento comercial. Además, su mayor claridad interpretativa lo posiciona por encima de transformaciones o modelos no lineales más complejos.

Según las métricas del modelo lineal (Tabla 5), el modelo lineal explicó el 95.5% de la variabilidad del rendimiento ($R^2 = 0.955$) y obtuvo el RMSE más bajo (8.80), lo que lo posiciona como el más preciso y estable. Un R^2 tan alto confirma que la relación entre morfología y rendimiento es esencialmente lineal. El bajo RMSE indica que las predicciones son cercanas a los valores reales, lo que es crucial para uso agronómico.

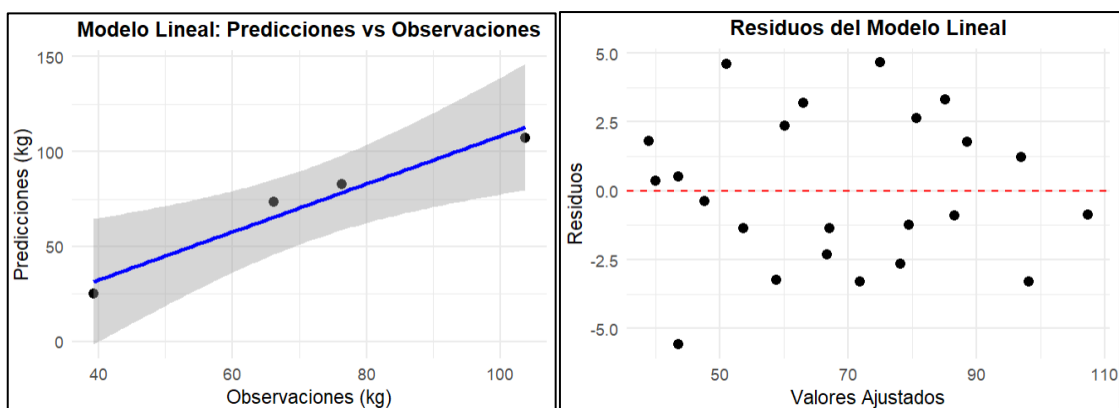
Tabla 5. Métricas del modelo lineal

Métrica	Valor
R ²	0.955
R	0.977
RMSE	8.80
AIC	140.92

Ambos gráficos indican que el modelo de regresión lineal se ajusta adecuadamente a los datos. (Figura 1). El gráfico de predicciones muestra una relación lineal positiva entre las variables, y el gráfico de residuos no presenta patrones

que sugieran problemas con el ajuste del modelo, indicando que se cumplen los supuestos de normalidad y homocedasticidad.

Figura 1. Observaciones vs predicciones y Residuos vs valores ajustados del modelo lineal



Modelo de regresión logarítmico

Los predictores significativos fueron: ancho de hoja, guano de islas de 1000 kg/ha, humus de lombriz de 5000 kg/ha (Tabla 6).

Tabla 6. Coeficientes del modelo logarítmico

Parámetro	Estimado	EE	t	p
Ancho de hoja	0.123	0.041	2.99	0.0150
Guano de islas de 1000 kg	0.420	0.108	3.89	0.0037
Humus de lombriz de 5000 kg	0.234	0.086	2.71	0.0240

La revisión del modelo evidenció que se satisfacen los supuestos de normalidad ($p = 0.5792$) y homocedasticidad ($p = 0.2738$). Asimismo, los valores asociados a la multicolinealidad y la evaluación de puntos influyentes no presentaron irregularidades que puedan afectar la solidez del modelo.

El empleo de transformaciones logarítmicas en modelos aplicados a sistemas agrícolas suele mejorar el ajuste estadístico cuando los datos presentan heterocedasticidad.

En cultivos como la col china, se ha observado que la reducción de fertilizantes químicos combinada con la aplicación de abonos orgánicos genera rendimientos más estables y residuos con mejor comportamiento estadístico, lo que demuestra la utilidad de estas transformaciones en ciertos contextos (Zhang et al., 2024). Sin embargo, interpretar los coeficientes una vez que se retorna a la escala original puede resultar complejo y, en algunos

casos, poco práctico para técnicos y agricultores, lo que limita su aplicabilidad directa en la toma de decisiones.

En este estudio, la transformación logarítmica produjo una mejora sustancial en el ajuste del modelo ($R^2 = 0.993$) y estabilizó adecuadamente la estructura de los residuos. Este comportamiento coincide con lo reportado en investigaciones que evidencian que las transformaciones logarítmicas pueden disminuir la heterocedasticidad y favorecer la homogeneidad de la varianza en modelos de predicción de biomasa y rendimiento vegetal (Ramírez-Aldana & Naranjo, 2021; Jabel et al., 2024; Yuan et al., 2024). No obstante, la necesidad de aplicar correcciones al retransformar los valores predichos y el posible sesgo

asociado hacen que la interpretación agronómica de los efectos marginales sea menos directa. Por ello, aunque el modelo logarítmico ofrece ventajas estadísticas claras, su utilidad práctica debe valorarse con cautela frente a la simplicidad interpretativa que ofrecen otros modelos.

Las métricas del modelo logarítmico (Tabla 7), indican que el modelo mostró el R^2 más alto (0.993) pero un RMSE mayor (9.68), y aunque los residuos fueron estables, su interpretación en escala original es más compleja. La transformación log mejora el ajuste cuando los datos presentan heterocedasticidad. Sin embargo, la retransformación puede introducir sesgos y dificultades prácticas en la toma de decisiones.

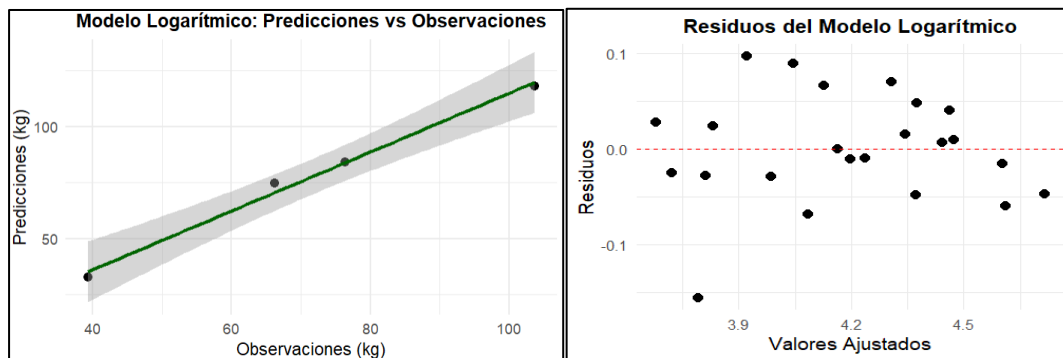
Tabla 7. Métricas del modelo logarítmico

Métrica	Valor
R^2	0.993
R	0.996
RMSE	9.68
AIC	-37.34

El gráfico de predicciones vs. observaciones del modelo logarítmico muestra una relación creciente entre ambas, con un intervalo de confianza que refleja la incertidumbre de las predicciones. Por otro lado, el gráfico de residuos

muestra una distribución aleatoria alrededor de cero, lo que indica que el modelo ajusta bien los datos sin sesgos y cumple con los supuestos de homocedasticidad y normalidad (Figura 2).

Figura 2. Observaciones vs predicciones y residuos vs ajustados del modelo logarítmico



Modelo de Regresión polinómica cuadrática

El único coeficiente del modelo polinómico (Tabla 8), fue Guano de islas de 1000 kg/ha fue significativo ($p = 0.0308$).

Tabla 8. Coeficientes del modelo polinómico

Parámetro	Estimado	EE	t	p
Guano de islas de 1000 kg	24.82	7.59	3.27	0.0308

La revisión del modelo evidenció que se satisfacen los supuestos de normalidad ($p = 0.9019$) y

homocedasticidad ($p = 0.2232$). Asimismo, los valores asociados a la multicolinealidad y la evaluación de puntos

influyentes no presentaron irregularidades que puedan afectar la solidez del modelo.

En base a las métricas del modelo polinómico (Tabla 9), se ha observado que, a pesar de incluir términos cuadráticos, solo el guano de islas de 1000 kg/ha fue significativo. El modelo mostró una reducción del ajuste ($R^2 = 0.903$) y un incremento marcado del error (RMSE = 16.84). El modelo presenta peor precisión predictiva, alta

varianza y poco beneficio respecto al modelo lineal. El elevado RMSE confirma sobreajuste y menor utilidad práctica. La relación entre variables es esencialmente lineal; introducir términos cuadráticos añade ruido y reduce la parsimonia. Con un tamaño de muestra moderado (27 unidades), los modelos polinómicos tienden a sobre ajustar.

Tabla 9. Métricas del modelo polinómico

Métrica	Valor
R^2	0.903
R	0.950
RMSE	16.84
AIC	137.13

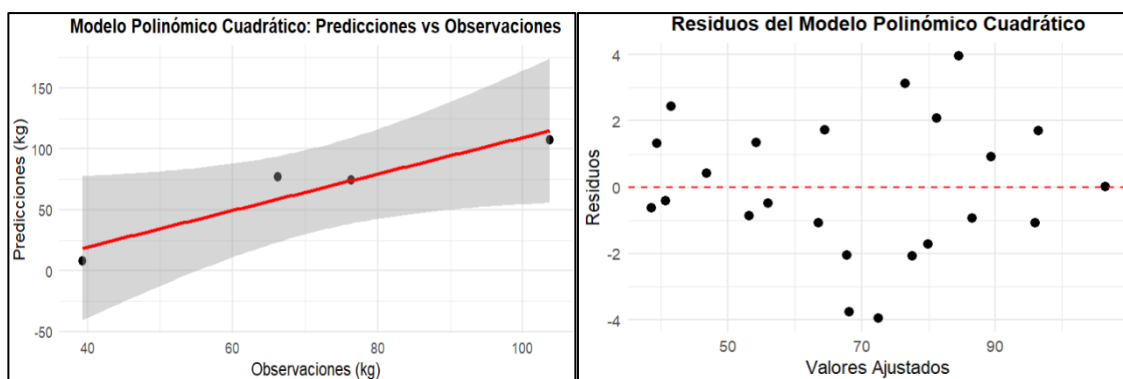
El uso de modelos polinómicos no garantiza necesariamente un mejor desempeño predictivo, especialmente cuando la relación entre las variables presenta una curvatura débil o cuando el tamaño muestral es limitado. En estudios donde se evalúa la fertilización con compost y otras enmiendas orgánicas, se ha observado que la inclusión de términos cuadráticos incrementa la complejidad del modelo sin lograr reducciones relevantes en el error de predicción (Lv et al., 2024). De igual manera, investigaciones realizadas en cultivos de col han mostrado que, cuando las variables independientes están fuertemente correlacionadas, los modelos más complejos pueden generar una mayor varianza en las estimaciones (Tarigan, 2022).

El hecho de que únicamente el término correspondiente a la dosis de 1000 kg·ha⁻¹ de guano de islas resultara significativo en el modelo sugiere que las relaciones no lineales existentes en los datos son limitadas. Esto coincide

con estudios que advierten sobre el riesgo de sobreajuste y el incremento de la varianza predictiva cuando se aplican modelos polinómicos en contextos donde las no linealidades no son marcadas y los tamaños muestrales son moderados (Kuradusenge et al., 2023; Morugán-Coronado et al., 2020; Srivastava et al., 2022). El mayor valor de RMSE registrado en este análisis confirma que el modelo polinómico no aportó beneficios sustanciales y que su desempeño fue inferior al de alternativas más simples y parsimoniosas.

El gráfico de predicciones frente a observaciones muestra una relación creciente con una incertidumbre reflejada en el intervalo de confianza, lo que valida la elección de un modelo polinómico. Por otro lado, el gráfico de residuos indica que estos están distribuidos aleatoriamente alrededor de cero, lo que confirma que el modelo es apropiado y cumple con los supuestos de homocedasticidad y normalidad (Figura 3).

Figuras 3. Observaciones vs predicciones y residuos vs ajustados de modelo polinómico



Comparación global de los modelos

Los modelos muestran distintos enfoques (Tabla 10), el logarítmico presenta mejor ajuste matemático, pero peor error de predicción real, el polinómico muestra peor precisión y mayor complejidad y el lineal logra mejor

equilibrio general: alto R^2 , menor RMSE, supuestos cumplidos y fácil interpretación. El modelo lineal múltiple es óptimo porque combina precisión predictiva, robustez estadística, claridad interpretativa y aplicabilidad en campo

Tabla 10. Comparación entre modelos

Modelo	R^2	R	RMSE	AIC	Normalidad (p)	Homocedasticidad (p)	Observaciones
Lineal múltiple	0.955	0.977	8.80	140.92	0.7439	0.3198	Mejor equilibrio general
Logarítmico	0.993	0.996	9.68	-37.34	0.5792	0.2738	Excelente ajuste, pero menor interpretabilidad
Polinómico 2°	0.903	0.950	16.84	137.13	0.9019	0.2232	Menor precisión predictiva

La comparación entre los modelos evaluados indica que la regresión lineal múltiple ofrece el balance más adecuado entre simplicidad, precisión y facilidad de interpretación. Este comportamiento concuerda con estudios que muestran que los modelos paramétricos basados en variables morfológicas continúan siendo herramientas confiables para la gestión agronómica y la toma de decisiones en campo (Bejarano-Herrera et al., 2024; Zhang et al., 2024). Además, optar por modelos menos complejos favorece su adopción práctica por parte de productores y técnicos, quienes requieren metodologías claras y aplicables en condiciones reales de cultivo.

El análisis conjunto de R^2 , RMSE y AIC revela un patrón consistente: aunque los modelos transformados o no lineales pueden lograr un ajuste estadístico más elevado, los modelos lineales suelen proporcionar un equilibrio superior entre exactitud, parsimonia e interpretabilidad cuando los predictores provienen de mediciones morfológicas directas (Yokoyama et al., 2024; Shook et al., 2021; Zhang et al., 2024). En este estudio, a pesar de que el modelo logarítmico alcanzó el mayor R^2 , el modelo lineal mostró un RMSE menor y una mayor aplicabilidad agronómica, factores que respaldan su selección como la alternativa más robusta y operativamente útil para predecir el rendimiento del repollo morado.

Selección del modelo final

Aunque el modelo logarítmico obtuvo el coeficiente de determinación más alto ($R^2 = 0.993$), la regresión lineal múltiple destacó por presentar el RMSE más bajo, cumplir de manera óptima los supuestos estadísticos y ofrecer una interpretación agronómica más clara y directa. Por el contrario, el modelo polinómico evidenció un rendimiento

predictivo inferior, lo que limita su aplicación práctica. En conjunto, estos resultados posicionan al modelo lineal múltiple como la alternativa más consistente para estimar el rendimiento del repollo morado.

Este modelo se consolida como el más adecuado porque combina precisión predictiva, al mostrar el RMSE más reducido; capacidad explicativa, al describir la mayor parte de la variabilidad sin incurrir en sobreajuste; y utilidad agronómica, al permitir interpretar con claridad el papel de los principales predictores, como el ancho de hoja, diámetro polar y las dosis intermedias de fertilización. Además, su simplicidad y reproducibilidad facilitan su adopción por técnicos, investigadores y productores.

La evidencia generada en este estudio, junto con lo reportado en la literatura especializada, confirma que la regresión lineal múltiple constituye la herramienta más apropiada para modelar la productividad del cultivo bajo esquemas de fertilización orgánica. Su interpretación directa, el bajo margen de error y la coherencia con los supuestos del análisis refuerzan su pertinencia en la toma de decisiones dentro de sistemas hortícolas altoandinos.

En consecuencia, la elección final del modelo lineal se sustenta en su robustez estadística, menor incertidumbre práctica e interpretación agronómica transparente, tal como respaldan Ramírez-Aldana y Naranjo (2021), Yokoyama et al. (2024) y Aguilar-Ariza et al. (2023). Para fines de manejo de fertilización y recomendaciones en ambientes altoandinos, resulta preferible emplear un modelo sencillo, interpretable y reproducible que, aun sin el R^2 más alto, garantice predicciones confiables y decisiones agronómicas precisas.

Conclusiones

Los resultados permitieron atender plenamente los objetivos del estudio. El modelo de regresión lineal múltiple mostró que el ancho de hoja, el diámetro polar y las dosis de 1000 kg/ha de guano de islas y 5000 kg/ha de humus de lombriz influyen significativamente en el rendimiento del repollo morado. Además, explicó el 95.5% de la variación y cumplió los supuestos estadísticos, confirmando su solidez predictiva.

Los modelos logarítmico y polinómico fueron ajustados adecuadamente, aunque con comportamientos distintos. El modelo logarítmico obtuvo el mayor R^2 (0.993), indicando mayor estabilidad en los residuos, mientras que el polinómico captó curvaturas en los datos, pero con mayor error y menor claridad interpretativa.

La comparación de R^2 , R, RMSE y AIC evidenció diferencias claras entre modelos. Aunque el logarítmico destacó por su ajuste estadístico, el modelo lineal ofreció el mejor equilibrio entre precisión, sencillez e interpretación agronómica.

Considerando el cumplimiento de supuestos, la significancia de los parámetros y su utilidad práctica, se determinó que el modelo lineal múltiple es el más adecuado para predecir el rendimiento del repollo morado con fertilización orgánica. Su aplicabilidad lo convierte en una herramienta eficiente para el manejo agrícola en sistemas hortícolas altoandinos.

Referencias

- Aguilar-Ariza, M., et al. (2023). Individual-plant weight estimation in leafy vegetables using UAV-derived features and machine learning. *Scientific Reports*, 13, 47431. <https://doi.org/10.1038/s41598-023-47431-y>
- Bejarano-Herrera, W. F., Marcillo-Paguay, C. A., Rojas-Tapias, D. F., & Estrada-Bonilla, G. A. (2024). Effect of mineral fertilization and microbial inoculation on cabbage yield and nutrition: A field experiment. *Agronomy*, 14(1), 210. <https://doi.org/10.3390/agronomy14010210>
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE Publications. https://batrachos.com/sites/default/files/pictures/Books/Field_ea_2012_Discovering%20Statistics%20using%20R.pdf
- Filho, A. B. C., Luiz-Cavarianni, R., Castro, J. C. C. de, & Mendoza-Cortez, J. W. (2011). Crecimiento y producción de repollo en función de la densidad de población y nitrógeno. *Agrociencia*, 45(5), 573–582. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-31952011000500004&lng=es&tln=es
- Jabed, M., et al. (2024). Crop yield prediction: Methodological advances and applications. *Heliyon*, 10, e40836. <https://doi.org/10.1016/j.heliyon.2024.e40836>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). *An introduction to statistical learning: With applications in R* (3rd ed.). Springer. https://www.stat.berkeley.edu/~rabbee/s154/ISLR_First_Printing.pdf
- Kuhn, M. (2022). *Caret: Classification and regression training* (R package version 6.0-93). Comprehensive R Archive Network. <https://cran.r-project.org/package=caret>
- Kuradusenge, M., Hitimana, E., Hanyurwimfura, D., Rukundo, P., Mtonga, K., Mukasine, A., Uwitonze, C., Ngabonziza, J., & Uwamahoro, A. (2023). Crop yield prediction using machine learning models: Case of Irish potato and maize. *Agriculture*, 13(1), 225. <https://doi.org/10.3390/agriculture13010225>
- Lüdecke, D., Makowski, D., Waggoner, P., & Patil, I. (2021). *Performance: Assessment of regression models performance* (R package version 0.9.0). Comprehensive R Archive Network. <https://cran.r-project.org/package=performance>
- Lv, X. M., & Han, Y. P. (2025). Trait performance differences and fertilizer strategy optimization in cabbage under organic vs. conventional fertilization regimes. *Plant Gene and Trait*, 16(4), 173–181. <https://doi.org/10.5376/pgt.2025.16.0019>
- Morugán-Coronado, A., Linares, C., & Zornoza, R. (2020). The impact of intercropping, tillage and fertilizer type on soil and crop yield in fruit orchards under Mediterranean conditions: A meta-analysis of field studies. *Agricultural Systems*, 178, 102736. <https://doi.org/10.1016/j.agsy.2019.102736>
- Ramírez-Aldana, R., & Naranjo, L. (2021). Random intercept and linear mixed models including heteroscedasticity in a logarithmic scale: Correction terms and prediction in the original scale. *PLOS ONE*, 16(4), e0249910. <https://doi.org/10.1371/journal.pone.0249910>
- Ramos Llanos, F. (2019). *Efecto del abonamiento de guano de islas y humus de lombriz en el rendimiento del repollo morado* [Tesis de pregrado, Universidad Nacional del Altiplano de Puno]. Repositorio Institucional UNA. <https://repositorio.unap.edu.pe/handle/20.500.14082/18155>
- Sant, S., Nagel, C., & Phong, L. (2005). Evaluation of fertilizer of fresh solid manure, composted manure or biodigester effluent for growing Chinese cabbage (*Brassica pekinensis*). *Livestock Research for Rural Development*, 17(3), 26. <http://www.lrrd.org/lrrd17/3/sant17026.htm>
- Shook, J., Gangopadhyay, T., Wu, L., Ganapathysubramanian, B., Sarkar, S., & Singh, A. K. (2021). Crop yield prediction integrating genotype and

- weather variables using deep learning. *PLOS ONE*, 16(6), e0252402. <https://doi.org/10.1371/journal.pone.0252402>
- Srivastava, A. K., Safaei, N., Khaki, S., Lopez, G., Zeng, W., Ewert, F., Gaiser, T., & Rahimi, J. (2022). Winter wheat yield prediction using convolutional neural networks from environmental and phenological data. *Scientific Reports*, 12, 6249. <https://doi.org/10.1038/s41598-022-06249-w>
- Sugiono, S., & Krismawati, A. (2020). Effectiveness of organic fertilizer on growth and production of cabbage plant (*Brassica oleracea*). *El-Hayah: Jurnal Biologi*, 7(4), 152–159. <https://doi.org/10.18860/elha.v7i4.10721>
- Tarigan, S. (2022). The effect of cultivating media and N fertilizer application on the growth and yield of flowering cabbage (*Brassica oleracea* var. botrytis). *International Journal of Multidisciplinary: Applied Business and Education Research*, 4(2), 28. <https://doi.org/10.11594/ijmaber.04.02.28>
- Xu, J., Li, J., Zhao, X., Liu, Z., Xu, H., Cao, K., & Ye, L. (2025). Impact of reduced chemical fertilizer and organic amendments on yield, nitrogen use efficiency, and soil microbial dynamics in Chinese flowering cabbage. *Horticulturae*, 11(7), 859. <https://doi.org/10.3390/horticulturae11070859>
- Yokoyama, Y., de Wit, A., Matsui, T., & Tanaka, T. (2024). Accuracy and robustness of a plant-level cabbage yield prediction system generated by assimilating UAV-based remote sensing data into a crop simulation model. *Precision Agriculture*, 25, 2685–2702. <https://doi.org/10.1007/s11119-024-10192-3>
- Yokoyama, Y., et al. (2024). Integrating UAV phenotyping and simulation models to predict Brassica biomass and head size. *Precision Agriculture*. <https://doi.org/10.1007/s11119-024-10192-3>
- Yuan, J., Zhang, Y., Zheng, Z., Yao, W., Wang, W., & Guo, L. (2024). Grain crop yield prediction using machine learning based on UAV remote sensing: A systematic literature review. *Drones*, 8(10), 559. <https://doi.org/10.3390/drones8100559>
- Zhang, Y., Zhang, J., Zhang, J., Li, H., Li, C., & Wang, X. (2024). Effects of the application of organic fertilizers on the yield, quality, and soil properties of open-field Chinese cabbage (*Brassica rapa* spp. *pekinensis*): A meta-analysis. *Agronomy*, 14(11), 2555. <https://doi.org/10.3390/agronomy14112555>
- Zhang, Y., Zhang, J., Zhang, J., Li, H., Li, C., & Wang, X. (2024). Effects of the application of organic fertilizers on the yield, quality, and soil properties of open-field Chinese cabbage (*Brassica rapa* spp. *pekinensis*) in China: A meta-analysis. *Agronomy*, 14(11), 2555. <https://doi.org/10.3390/agronomy14112555>