

INFORME DEL ANÁLISIS DE LOS DETERMINANTES DE LA PRODUCTIVIDAD DE MAÍZ AMARILLO DURO 2023

PIDARA

Proyecto Integral de Diversificación
Agroproductiva y Reconversión Agrícola



Ministerio de
Agricultura y Ganadería

Informe del análisis de los determinantes de la productividad de maíz amarillo duro 2023

Proyecto Integral de Diversificación Agroproductiva y Reconversión Agrícola-PIDARA

Quito – Ecuador

2024

RESUMEN

Este informe analiza los rendimientos objetivos del cultivo de maíz amarillo duro en Ecuador, abordando factores agronómicos y socioeconómicos. El estudio se basa en los datos recolectados en el ciclo de invierno del año 2023, en seis provincias; utilizando metodologías como la regresión lineal múltiple en valores absolutos, así como estandarizados, para identificar los principales determinantes de la productividad de maíz. Los resultados destacan la importancia de la densidad existente entre las plantas, de las prácticas de manejo fitosanitario y la aplicación adecuada de fertilizantes, especialmente el nitrógeno, como factores críticos para incrementar el rendimiento, así como la impureza, como principal variable que influye de manera negativa en la productividad. Se recomienda una mayor investigación en combinaciones óptimas de prácticas agrícolas para mejorar la productividad y sostenibilidad del cultivo.

Palabras clave: Maíz amarillo duro, rendimientos objetivos, regresión lineal múltiple

ABSTRACT

This report analyzes the objective yields of maize (corn) crops in Ecuador, addressing agronomic and socioeconomic factors. The study is based on data collected in the winter cycle of 2023, in six provinces; using methodologies such as multiple linear regression in absolute values, as well as standardized values, to identify the main determinants of maize productivity. The results highlight the importance of the existing density between plants, phytosanitary management practices and the adequate application of fertilizers, especially nitrogen, as critical factors to increase yield, as well as impurity, as the main variable that influences negatively on productivity. Further research into optimal combinations of agricultural practices is recommended to improve crop productivity and sustainability.

Keywords: Maize, crop yields, multiple linear regression

Contenido

| | |
|---|-----------|
| 1. ANTECEDENTES | 5 |
| 1.1. Operativos de rendimientos objetivos | 5 |
| 1.1. Justificación | 5 |
| 2. METODOLOGÍA | 6 |
| 2.1. Selección de características | 6 |
| 2.2. Tratamiento de datos atípicos y faltantes | 6 |
| 2.3. Transformación de variables | 7 |
| 2.4. Análisis estadístico y modelado | 7 |
| 3. RESULTADOS | 9 |
| 3.1. Regresión lineal múltiple | 10 |
| 3.2. Regresión lineal múltiple con variables estandarizadas | 14 |
| 4. CONCLUSIONES | 15 |
| 5. BIBLIOGRAFÍA | 15 |

1. ANTECEDENTES

1.1. Operativos de rendimientos objetivos

La metodología de rendimientos objetivos permite obtener información consistente de los principales factores que inciden en la producción, basada en un marco metodológico que permite obtener parámetros robustos para determinar la productividad a nivel nacional y provincial. En el caso del rubro de maíz amarillo, se realizan dos operativos en el año (para cada ciclo de siembra) en las principales provincias productoras: Los Ríos, Manabí, Guayas, Loja, Santa Elena y El Oro.

A continuación, se detalla el proceso metodológico del levantamiento de la información:

Figura 1. Proceso metodológico de Rendimientos Objetivos



FUENTE: Coordinación General de Información Nacional Agropecuaria – MAG

El proceso inicia con el monitoreo permanente de las siembras y cosechas del rubro de maíz amarillo, a través de la supervisión del personal técnico de territorio. Para determinar el tamaño de la muestra se implementa el método de Muestreo Aleatorio Simple Estratificado, en el que se incluyen parámetros de nivel de confianza y términos de error, con la finalidad de obtener información representativa de la productividad, a nivel nacional y desagregada a nivel provincial.

Para el levantamiento de la información en territorio se utiliza una encuesta que contiene información de variables socioeconómicas, productivas y de manejo agronómico del cultivo. La metodología se considera objetiva debido a que se recolectan muestras de cada predio que se visita; posteriormente, se envían al Instituto Nacional de Investigaciones Agropecuarias (INIAP) para medir variables propias del rubro, como la humedad e impurezas.

1.1. Justificación

La producción de maíz es una de las actividades agrícolas más importantes en el Ecuador, por ser la materia prima para la elaboración de productos balanceados destinados a la industria animal, es esta precisamente la razón que justifica la expansión, tanto en superficie cultivada como en producción y rendimiento.

La dinámica que genera este cultivo, principalmente basado en su cadena productiva, hace que el gobierno lo haya considerado como uno de los puntos de principal atención en el sector agrícola, interviniendo en su producción y comercialización, con el fin de cubrir la demanda local de maíz amarillo.

Debido a esta importancia, el Ministerio de Agricultura y Ganadería (MAG) decide poner en marcha el Operativo de Rendimientos Objetivos para el cultivo de maíz amarillo duro, en el año 2014; con el objetivo de levantar información precisa, oportuna y en tiempo real, acerca de la producción nacional, rendimiento del cultivo, sus características, los insumos utilizados, los problemas que enfrenta; de esta manera, establecer y definir las mejores políticas en beneficio de las personas productoras.

Para el año 2023, esta actividad se realizó en las provincias de Los Ríos, Manabí, Guayas, Santa Elena, Loja y El Oro, con una superficie total cosechada de 308,373 hectáreas; de las cuales, el 80 % se registra durante el ciclo de invierno y el 20 % en verano.

El presente documento pretende establecer una metodología para analizar la relación que existe entre el rendimiento y los principales factores que lo afectan.

Así, al identificar cuáles son los factores que tienen el mayor impacto en la productividad, los agricultores pueden optimizar el uso de recursos como agua, fertilizantes y otros insumos. Esto no solo mejora la eficiencia económica, sino que también promueve prácticas de agricultura sostenible.

Los resultados de este estudio pueden generar políticas agrícolas efectivas que apoyen la innovación, la sostenibilidad y la rentabilidad, en el sector agrícola. Esto incluye políticas sobre subsidios, apoyo a la investigación agrícola y desarrollo de infraestructura.

En base a lo mencionado, para el presente estudio se seleccionó a los datos obtenidos durante el ciclo de invierno del año 2023, al ser el de mayor aporte en cuanto a superficie y producción de maíz amarillo duro.

2. METODOLOGÍA

La metodología utilizada para evaluar los determinantes de la productividad se centra en varios pasos clave, para analizar y modelar los factores que afectan la productividad agrícola:

2.1. Selección de características

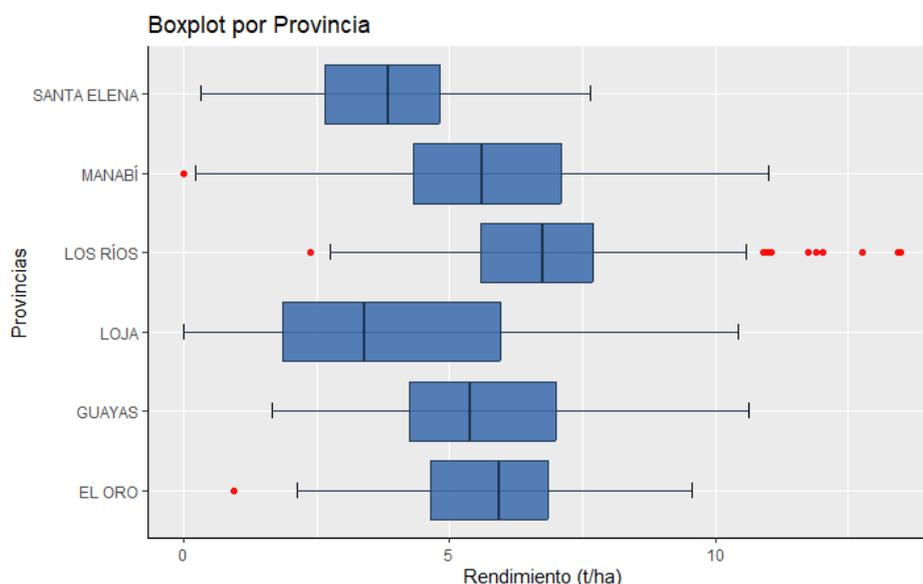
Se seleccionan las posibles variables predictoras relevantes para el estudio, incluyendo variables socioeconómicas y variables agrícolas específicas.

Para la variable de respuesta se debe elegir uno de los dos posibles enfoques: seleccionar el rendimiento neto extrapolado por parcela o seleccionar el rendimiento de las plantas específicas muestreadas.

2.2. Tratamiento de datos atípicos y faltantes

Se determinan valores atípicos utilizando métodos basados en el rango intercuartílico y pruebas estadísticas especializadas (test de Grubs y test de Rosner), para evitar distorsiones en los análisis.

Figura 2. Gráfico de cajas y bigotes del ciclo de invierno de maíz amarillo duro



FUENTE: Coordinación General de Información Nacional Agropecuaria – MAG

Además, se imputan valores faltantes o se codifican de manera adecuada con el fin de preparar los datos para el análisis econométrico.

2.3. Transformación de variables

Otro paso es la codificación de variables categóricas, la cual es crucial en el preprocesamiento de datos para análisis estadísticos y econométricos, especialmente porque la mayoría de los algoritmos de modelado solo pueden manejar datos numéricos.

2.4. Análisis estadístico y modelado

2.4.1. Regresión lineal múltiple

El modelo de regresión lineal múltiple es un método estadístico que modela la relación entre una variable dependiente y dos o más variables independientes, ajustando una ecuación lineal a los datos observados (Faraway, 2021). Cada valor de la variable independiente x está asociado con un valor de la variable dependiente y . La ecuación para el modelo de regresión lineal múltiple se expresa como:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i$$

Donde:

- y_i es la variable dependiente,
- $x_{i1}, x_{i2}, \dots, x_{ip}$ son las variables independientes,

- $\beta_0, \beta_1, \dots, \beta_p$ son los coeficientes del modelo, y
- ϵ_i es el término de error aleatorio, que es una variable no observada que añade ruido al modelo lineal.

Análisis de los Coeficientes

Valor: Indica la magnitud y dirección del efecto de cada variable independiente sobre la variable dependiente. Un coeficiente positivo sugiere que, a medida que la variable independiente aumenta, la variable dependiente también tiende a aumentar; mientras que, un coeficiente negativo indica una relación inversa.

Magnitud: La cantidad promedio en la que la variable dependiente aumenta, cuando la variable independiente aumenta en una unidad; mientras las otras variables independientes se mantienen constantes.

Magnitud estandarizada: Un coeficiente más grande (en valor absoluto) sugiere un impacto más significativo en la variable dependiente, todo lo demás constante.

Significancia Estadística ($P > |t|$)

Este valor indica, si la influencia de la variable independiente en la variable dependiente es estadísticamente significativa. Por lo regular, un *p-valor* menor que 0.05 se considera estadísticamente significativo, lo que implica que hay menos de un 5 % de probabilidad de que la relación observada sea debido al azar.

Identificación de Factores Influyentes

Para determinar cuáles son los factores más influyentes:

- Estandarizar las variables seleccionadas
- Ordenar los coeficientes por magnitud y observar cuáles variables tienen los coeficientes más grandes en valor absoluto; ya que estas son las que tienen el mayor impacto por cada unidad de cambio.
- Revisar la significancia estadística, para confirmar que estos coeficientes sean también estadísticamente significativos (*p-value* < 0.05).

Considerar la relación práctica, a veces, un coeficiente puede ser estadísticamente significativo, pero no necesariamente importante en un contexto práctico. Considerar la relevancia práctica de cada variable en el contexto del estudio.

2.4.2. Estandarización de variables

Se realizan transformaciones de variables para estandarizar los datos, como la estandarización Z-score, que ajusta los datos para que tengan media cero y desviación estándar uno. Esto es crucial para algunos modelos estadísticos y econométricos, que asumen que todas las variables están en la misma escala.

La fórmula utilizada es:

$$X_{\text{estandarizado}} = \frac{X - \mu}{\sigma}$$

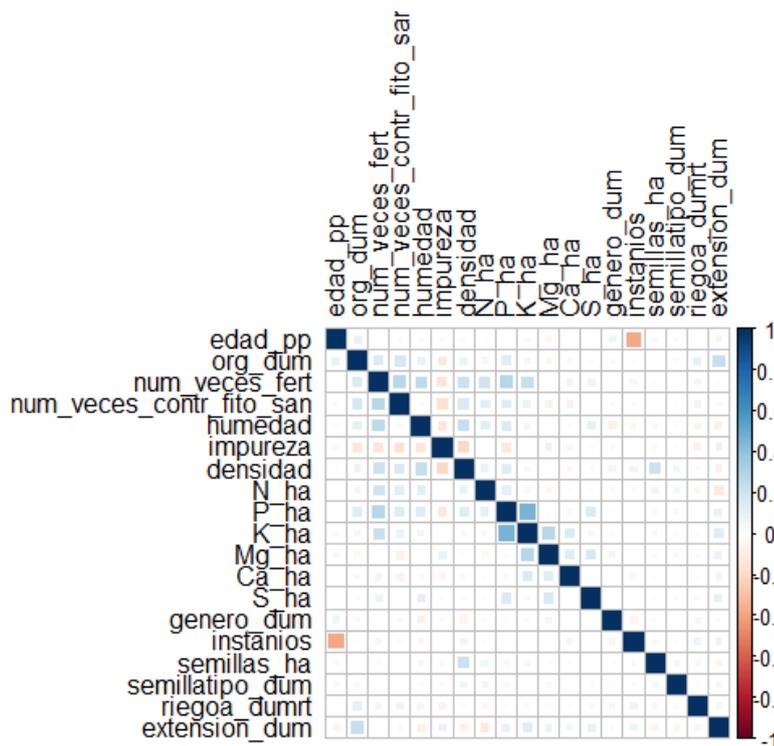
Donde:

- X es el valor original de la característica.
- μ es la media de la característica.
- σ es la desviación estándar de la característica.

3. RESULTADOS

El siguiente mapa de calor muestra la matriz de correlación entre las variables predictoras utilizadas en el análisis. Las correlaciones se presentan en una escala de -1 a 1, donde 1 indica una correlación positiva perfecta, -1 una correlación negativa perfecta y 0 ninguna correlación. Este gráfico es fundamental para identificar variables altamente correlacionadas, que podrían influir en la multicolinealidad de los modelos de regresión.

Figura 3. Gráfico de correlaciones de variables



FUENTE: Coordinación General de Información Nacional Agropecuaria – MAG

La multicolinealidad, en el Modelo de regresión lineal múltiple, se presenta cuando las variables independientes muestran un alto nivel de correlación. Como se puede observar en el gráfico, no existen correlaciones significativas entre las variables explicativas del modelo.

3.1. Regresión lineal múltiple

Al estimar el primer modelo de regresión lineal múltiple con las variables seleccionadas previamente, se obtienen los siguientes resultados:

Tabla 1. Resultados del modelo 1 de regresión lineal múltiple

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|--------------------------|------------|------------|---------|----------|-----|
| (Intercept) | -1.625e+00 | 6.310e-01 | -2.576 | 0.010121 | * |
| edad_pp | 4.134e-03 | 3.208e-03 | 1.289 | 0.197791 | |
| num_veces_fert | 6.280e-02 | 7.153e-02 | 0.878 | 0.380169 | |
| num_veces_contr_fito_san | 2.394e-01 | 4.644e-02 | 5.154 | 2.98e-07 | *** |
| humedad | -1.591e-03 | 9.242e-03 | -0.172 | 0.863334 | |
| impureza | -7.654e-02 | 2.982e-03 | -25.663 | < 2e-16 | *** |
| densidad | 7.663e-05 | 4.882e-06 | 15.696 | < 2e-16 | *** |
| N_ha | 4.622e-03 | 7.675e-04 | 6.023 | 2.28e-09 | *** |
| P_ha | -1.836e-03 | 1.876e-03 | -0.979 | 0.327871 | |
| K_ha | 2.013e-03 | 1.929e-03 | 1.044 | 0.296849 | |
| Mg_ha | 5.980e-02 | 1.795e-02 | 3.332 | 0.000888 | *** |
| Ca_ha | -5.296e-04 | 4.386e-02 | -0.012 | 0.990368 | |
| S_ha | -2.531e-03 | 4.652e-03 | -0.544 | 0.586562 | |
| genero_dum | 3.471e-02 | 1.074e-01 | 0.323 | 0.746547 | |
| instancios | 3.776e-02 | 1.232e-02 | 3.064 | 0.002232 | ** |
| semillas_ha | 1.615e-05 | 6.864e-06 | 2.353 | 0.018776 | * |
| semillatipo_dum | 9.629e-01 | 3.350e-01 | 2.874 | 0.004119 | ** |
| riegoa_dumrt | -4.066e-01 | 2.098e-01 | -1.938 | 0.052842 | . |
| extension_dum | 8.178e-02 | 9.137e-02 | 0.895 | 0.370931 | |
| org_dum | 1.566e-01 | 9.573e-02 | 1.636 | 0.102143 | |

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

Residual standard error: 1.472 on 1198 degrees of freedom
Multiple R-squared: 0.5546, Adjusted R-squared: 0.5475
F-statistic: 78.52 on 19 and 1198 DF, p-value: < 2.2e-16

FUENTE: Coordinación General de Información Nacional Agropecuaria – MAG

Al analizar la significancia de cada variable, se observa que existen variables con un valor p superior a 0.05, lo que indica que la influencia de la variable independiente en la variable explicada no es estadísticamente significativa, por lo que se realiza una segunda estimación de modelo, con los siguientes resultados:

Tabla 2. Resultados del modelo 2 de regresión lineal múltiple

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|--------------------------|------------|------------|---------|----------|-----|
| (Intercept) | -1.192e+00 | 5.468e-01 | -2.179 | 0.02953 | * |
| num_veces_contr_fito_san | 2.606e-01 | 4.444e-02 | 5.865 | 5.81e-09 | *** |
| impureza | -7.626e-02 | 2.912e-03 | -26.193 | < 2e-16 | *** |
| densidad | 7.709e-05 | 4.696e-06 | 16.418 | < 2e-16 | *** |
| N_ha | 4.731e-03 | 7.441e-04 | 6.359 | 2.88e-10 | *** |
| Mg_ha | 6.537e-02 | 1.652e-02 | 3.957 | 8.03e-05 | *** |
| instancios | 2.987e-02 | 1.123e-02 | 2.660 | 0.00791 | ** |
| semillas_ha | 1.635e-05 | 6.832e-06 | 2.394 | 0.01683 | * |
| semillatipo_dum | 8.920e-01 | 3.336e-01 | 2.674 | 0.00759 | ** |

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

Residual standard error: 1.473 on 1209 degrees of freedom
 Multiple R-squared: 0.5501, Adjusted R-squared: 0.5472
 F-statistic: 184.8 on 8 and 1209 DF, p-value: < 2.2e-16

FUENTE: Coordinación General de Información Nacional Agropecuaria – MAG

Como se puede observar en el segundo modelo de regresión, todas las variables son significativas dentro del modelo.

Variables Significativas:

- **Número de veces que se realiza control fitosanitario (num_veces_contr_fito_san):** Con un coeficiente de $2.606e^{-01}$ y un *p-valor* significativo (*p* cercano a 0). Esta variable indica que, un incremento en una unidad en el número de veces que se realizan controles fitosanitarios, incrementa el rendimiento en la magnitud que expresa el coeficiente beta, sugiriendo que las prácticas de manejo de la salud de las plantas son cruciales para la productividad.
- **Impureza (impureza):** Presenta un coeficiente de $-7.626e^{-02}$ con un *p-valor* significativo (*p* cercano a 0). Al tener un coeficiente negativo, denota que esta variable tiene una relación inversa con el rendimiento del maíz amarillo duro. Esto revela que, el correcto manejo de la post cosecha influye en la productividad.
- **Densidad de las plantas (densidad):** Registra un coeficiente de $7.907e^{-05}$ con un *p-valor* cercano a cero, mostrando un impacto positivo significativo en el rendimiento. Demuestra que, el correcto manejo del cultivo en la etapa de siembra, influye en la productividad.
- **Nitrógeno por hectárea (N_ha):** Tiene un coeficiente de $4.731e^{-03}$ con un *p-valor* cercano a cero, mostrando un impacto positivo significativo en el rendimiento. Lo que puede indicar la importancia de una nutrición adecuada, específicamente del nitrógeno en el rendimiento del maíz duro.
- **Magnesio por hectárea (Mg_ha):** Tiene un coeficiente de $6.537e^{-02}$ con un *p-valor* cercano a cero, mostrando un impacto positivo significativo en el rendimiento. Puede

indicar la importancia de una nutrición adecuada, específicamente del magnesio en el rendimiento del maíz duro.

- **Años aprobados de instrucción formal (instanios):** Con un coeficiente de $2.987e^{-02}$ y un *p-valor* significativo (*p* cercano a 0.001). Esta variable indica que un incremento en una unidad, en el número de años aprobados de la instrucción formal, por parte de la persona productora, incrementa el rendimiento del maíz duro en la magnitud que expresa el coeficiente beta; lo que sugiere que la preparación académica influye de manera positiva en la productividad.
- **Cantidad de semillas por hectárea (semillas_ha):** Registra un coeficiente de $1.635e^{-05}$ con un *p-valor* cercano a 0.01, mostrando un impacto positivo significativo en el rendimiento. Lo que indica que, el correcto manejo del cultivo en la etapa de siembra, influye en la productividad.
- **Tipo de semilla (semillatipo_dum):** Un coeficiente de $8.920e^{-01}$ y *p-valor* de cercano a 0.001, sugiere que los datos de las semillas certificadas asociados con la producción de maíz duro, tienen un rendimiento mayor en el cultivo, comparado con las semillas recicladas. Lo que refleja diferencias en la productividad al utilizar distintos tipos de semillas.

Variables no Significativas

Muchas variables como la edad de los productores, humedad, fósforo, potasio, asociatividad, género, entre otras; mostraron *p-valores* altos, sugiriendo que no tienen un efecto estadísticamente significativo sobre el rendimiento en este modelo. Lo que puede deberse a la falta de variabilidad en estos factores entre las muestras o a que, otros factores no medidos puedan estar distorsionando estas relaciones.

Interpretación del Modelo

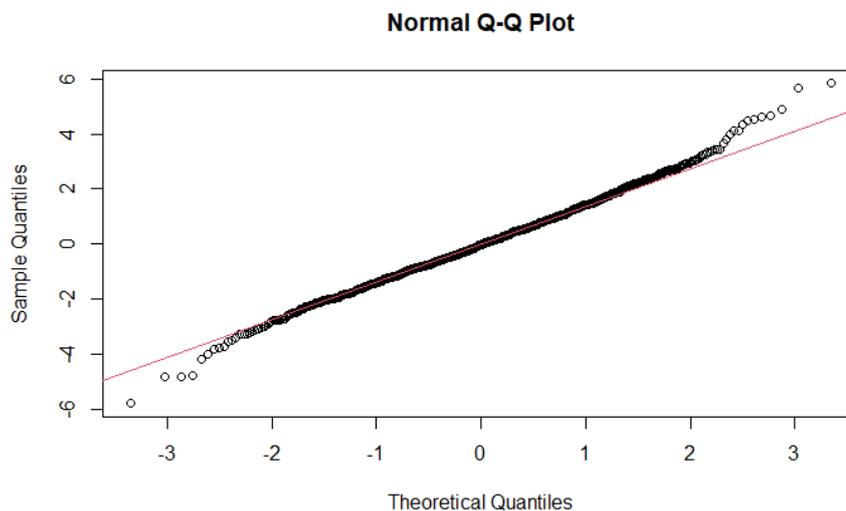
- **R-cuadrado:** El valor indica que, el modelo explica aproximadamente el 55 % de la variabilidad en el rendimiento del maíz amarillo duro. Aunque es una cantidad considerable, también sugiere que hay otros factores no incluidos en el modelo, que están afectando la productividad. En este sentido, el R-cuadrado ajustado (0.5472), el cual toma en cuenta el número de variables incluido en el modelo, es cercano al valor del coeficiente de determinación mencionado, lo que indica que, el número de variables utilizadas en el modelo es correcto.
- **F-estadístico:** El valor de la prueba F y su probabilidad asociada ($2.2e^{-16}$) denotan que el modelo es globalmente significativo, lo que significa que, al menos algunas de las variables independientes están relacionadas significativamente con la variable dependiente.

Diagnósticos del Modelo

Para determinar la validez de los resultados del modelo, este debe cumplir con supuestos mínimos de calidad, basados principalmente en el análisis de residuales, los cuales se describen a continuación:

- **Normalidad:** El supuesto de normalidad establece que, los errores de la regresión deben seguir una distribución normal. Como se puede observar, los puntos están sobre la recta, por lo que infiere que los residuos del modelo tienen una distribución normal.

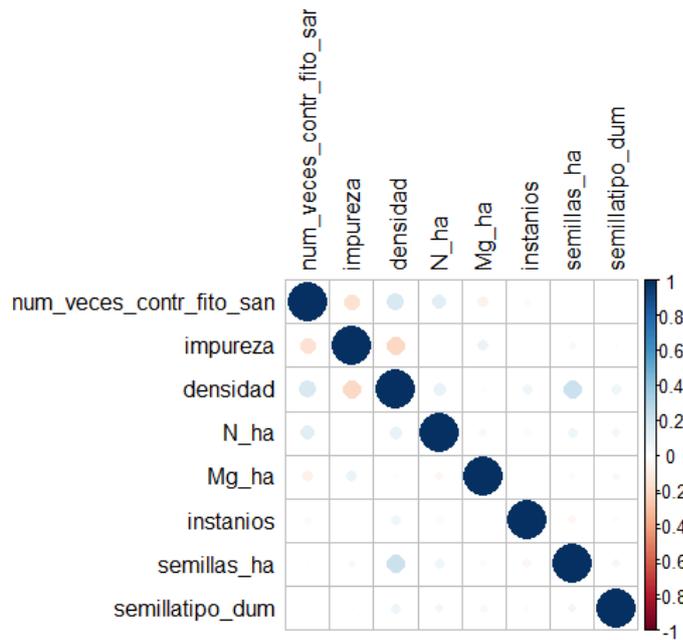
Figura 4. Gráfico de probabilidad normal



FUENTE: Coordinación General de Información Nacional Agropecuaria – MAG

- **Autocorrelación:** Este supuesto establece que los errores del modelo no están correlacionados. En este sentido, el test de Durbin-Watson somete a prueba la autocorrelación de Primer orden, obteniendo un valor de 1.44, siendo cercano al valor deseado de 2 ($DW \approx 2$).
- **Multicolinealidad:** La multicolinealidad en el modelo de regresión, se da cuando las variables independientes presentan alto nivel de correlación. Como se puede observar en el gráfico, no existen correlaciones significativas entre las variables explicativas del modelo.

Figura 5. Gráfico de correlaciones de variables del modelo

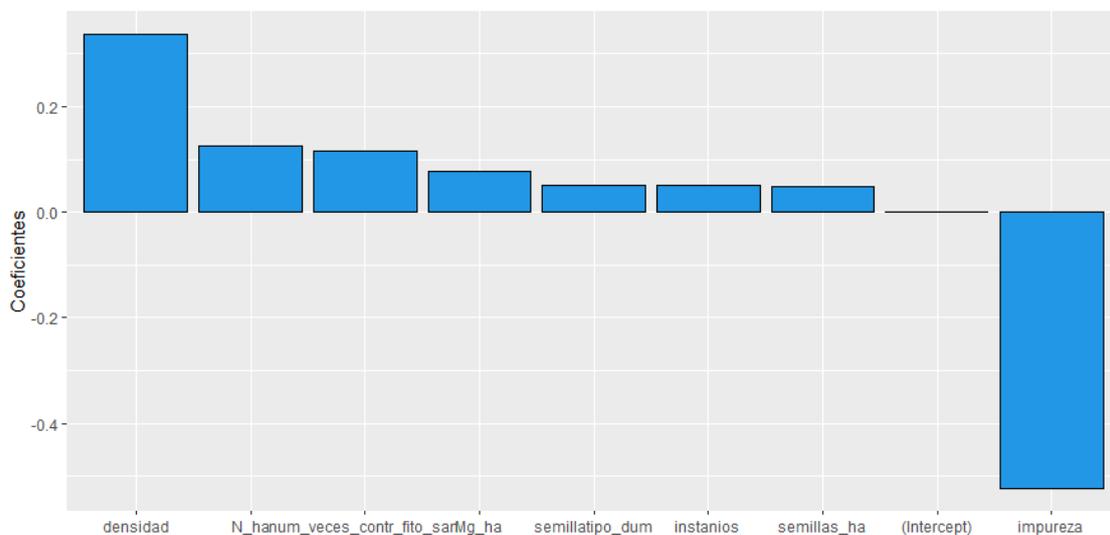


FUENTE: Coordinación General de Información Nacional Agropecuaria – MAG

3.2. Regresión lineal múltiple con variables estandarizadas

Como se mencionó en los anteriores apartados, al estandarizar las variables se ajusta los datos para que tengan media cero y desviación estándar uno, lo que elimina las unidades de medida de cada variable, con el fin de realizar una mejor interpretación de la magnitud del impacto de cada variable independiente en la variable dependiente.

Figura 6. Coeficientes del modelo con variables estandarizadas



FUENTE: Coordinación General de Información Nacional Agropecuaria – MAG

Como se puede observar, las variables que ejercen una mayor influencia positiva para el rendimiento del maíz amarillo duro, es la densidad de las plantas con las que fueron sembradas, seguido por la cantidad de nitrógeno aplicado por hectárea y el número de veces que se realizó un control fitosanitario.

También se observa el aporte del resto de variable en menor magnitud, como la cantidad de magnesio aplicada, el tipo de semilla utilizada, la instrucción formal de la persona productora y, la cantidad de semillas por hectárea.

Por otro lado, la variable que ejerce una mayor influencia de forma negativa en la productividad del cultivo, es el nivel de impureza de los granos, medido en laboratorio.

4. CONCLUSIONES

El modelo de regresión lineal múltiple seleccionado, cumple con los parámetros de calidad establecidos, lo que determina la validez de los resultados obtenidos, tanto con las variables, como con las unidades de medida originales; así como los resultados, una vez que fueron estandarizadas.

Este análisis destaca la importancia de ciertas prácticas agrícolas y características del cultivo y la persona productora en la productividad del maíz amarillo duro. Además, identifica áreas donde la investigación adicional podría ser necesaria para entender mejor otros factores que afectan la productividad y cómo mejorar las estrategias de gestión agrícola.

Considerando las limitaciones del modelo, como la explicación parcial de la variabilidad y una posible autocorrelación positiva, es necesario realizar análisis adicionales y considerar la inclusión de más variables o datos de seguimiento para futuros modelos.

5. BIBLIOGRAFÍA

Gujarati, D. (2003). *Econometría*. México. McGraw-Hill.

Novales, A. (1993). *Econometría*. España. McGraw-Hill.

Peñaherrera, D. (2011). *Manejo Integrado del Cultivo de Maíz de Altura. Módulos de Capacitación para Capacitadores*. Ecuador. INIAP.