

INFORME DEL ANÁLISIS DE LOS DETERMINANTES DE LA PRODUCTIVIDAD DE CAFÉ ARÁBIGO 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 café arábigo 2023

Proyecto Integral de Diversificación Agroproductiva y Reconversión Agrícola-PIDARA

Quito – Ecuador

2024

RESUMEN

Este informe analiza los rendimientos objetivos del cultivo de café arábigo en Ecuador, abordando factores agronómicos y socioeconómicos. El estudio se basa en los datos recolectados durante el año 2023, en doce provincias, utilizando metodologías como la regresión lineal múltiple en valores absolutos, así como estandarizados, para identificar los principales determinantes de la productividad del café arábigo. Los resultados destacan la importancia de las horas de trabajo en el cultivo, la superficie del cultivo; es decir el número de hectáreas disponibles para la producción, el número de veces que se realiza el control fitosanitario, la densidad de plantas existente y la aplicación adecuada de fertilizantes, especialmente el potasio, como factores críticos para incrementar el rendimiento; sin embargo, se lo debe manejar de manera óptima ya que el uso excesivo del azufre, 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: café, café arábigo, rendimientos objetivos, regresión lineal múltiple

ABSTRACT

This report analyzes the objective yields of arabica coffee in Ecuador, addressing agronomic and socioeconomic factors. The study is based on data collected during the year 2023, in twelve 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 working hours in the crop, crop area, that is the number of hectares available for production, the number of times phytosanitary control is performed, the existing density and the adequate application of fertilizers, especially potassium, as critical factors to increase yield; however, it must be optimally managed since the excessive use of sulfur has a negative influence on productivity. Further research into optimal combinations of agricultural practices is recommended to improve crop productivity and sustainability.

Keywords: Coffee, arabica coffee, multiple linear regression

Contenido

1. ANTECEDENTES	5
1.1. Operativos de rendimientos objetivos	5
1.2. Justificación	6
2. METODOLOGÍA	6
2.1. Selección de características	6
2.2. Transformación de variables	6
2.3. Elección de edad de la plantación	7
2.4. Análisis estadístico y modelado	7
3. RESULTADOS	8
3.1. Regresión lineal múltiple	9
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 café arábigo, se realiza un operativo al año en las principales provincias productoras: Azuay, Carchi, Chimborazo, El Oro, Galápagos, Imbabura, Loja, Manabí, Morona Santiago, Pastaza, Pichincha y Zamora Chinchipe.

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 café arábigo, a través de la supervisión del personal técnico del 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 del manejo agronómico del cultivo. La metodología se considera objetiva debido a que se recolecta información de variables como distanciamiento entre plantas e hileras, así como identificación de ejes y ramas productivas; además, se realiza el conteo de frutos sanos en tres ramas productivas seleccionadas, este procedimiento se realiza en cuatro plantas del terreno monitoreado.

Para la estimación de los rendimientos se calcula, a partir de la estimación de frutos total por planta, posteriormente se realiza la transformación de frutos a grano oro usando los índices transformación de grano cereza a grano oro.

1.2. Justificación

La producción de café es una de las actividades agrícolas más importantes en el Ecuador; a nivel económico porque es una fuente de generación de divisas debido a las exportaciones realizadas, en el aspecto social, ya que en el cultivo de este producto se encuentran involucradas alrededor de 42,528 personas.

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.

Debido a esta importancia, el Ministerio de Agricultura y Ganadería (MAG) decide poner en marcha el Operativo de Rendimientos Objetivos para el cultivo de café, en el año 2016, con la finalidad de levantar información precisa, oportuna y en tiempo real acerca de la producción nacional, su rendimiento, características, los insumos utilizados, los problemas que enfrenta; de esta manera, definir las mejores políticas en beneficio de las personas productoras.

Para el año 2023, esta actividad se realizó en las provincias de Azuay, Carchi, Chimborazo, El Oro, Galápagos, Imbabura, Loja, Manabí, Morona Santiago, Pastaza, Pichincha y Zamora Chinchipe, con una superficie total cosechada de 30,314 hectáreas.

El presente documento pretende establecer una metodología para analizar la relación que existe entre el rendimiento del cultivo 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 del estudio pueden generar políticas agrícolas efectivas que apoyen la innovación, la sostenibilidad y la rentabilidad, en el sector agrícola. Lo cual, incluye políticas de incentivos a las personas productoras, apoyo a la investigación agrícola y desarrollo de infraestructura.

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 seleccionar uno de los dos posibles enfoques: seleccionar el rendimiento neto extrapolado por parcela o seleccionar el rendimiento de las plantas específicas muestreadas.

2.2. 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, para lo cual se crean variables dummy.

2.3. Elección de edad de la plantación

Tomando en cuenta que el café es un cultivo permanente, y por lo tanto se puede obtener el producto por más de 20 años, es necesario realizar el análisis de rendimiento en el periodo en el cual, el cultivo se encuentra en el pico máximo de producción; entre los 4 y 7 años de edad.

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 ($p\text{-value} < 0.05$).

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:

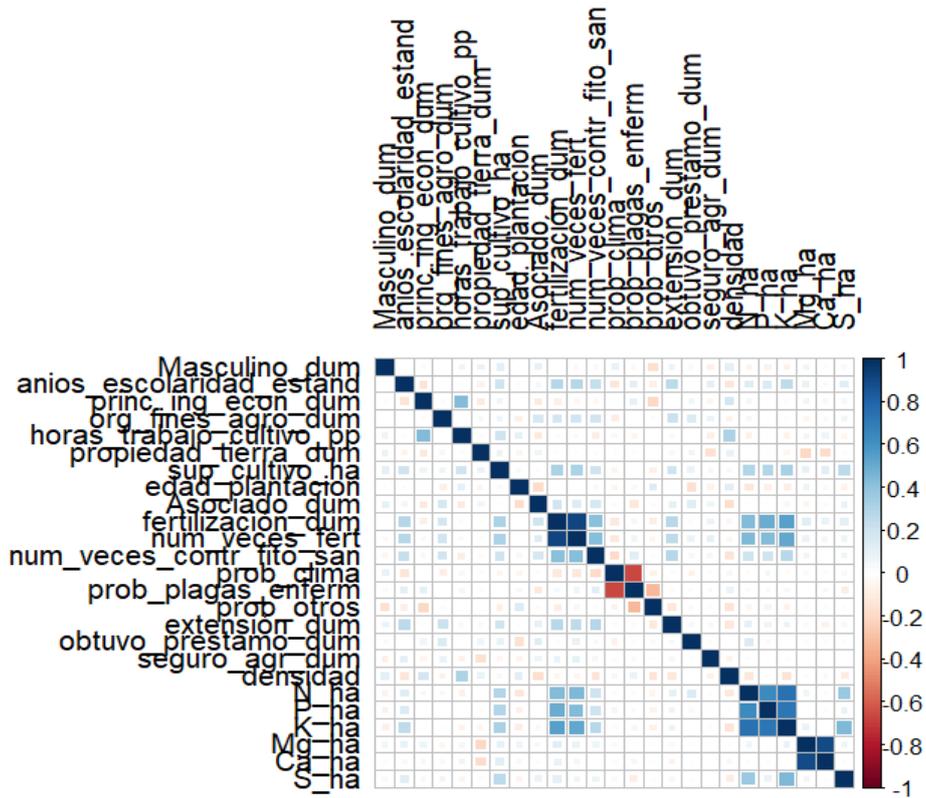
- X es el valor original de la característica.
- μ es la media de la característica.
- σ es la desviación estándar de la característica.

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 2. 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 manifiestan 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)	-3.228e-01	3.222e-01	-1.002	0.317221	
Masculino_dum	-7.236e-02	5.407e-02	-1.338	0.181850	
anios_escolaridad_estand	3.387e-04	5.739e-03	0.059	0.952975	
princ_ing_econ_dum	9.246e-02	5.762e-02	1.605	0.109693	
org_fines_agro_dum	-4.021e-02	5.576e-02	-0.721	0.471440	
horas_trabajo_cultivo_pp	7.538e-03	2.078e-03	3.628	0.000339	***
propiedad_tierra_dum	2.179e-02	8.311e-02	0.262	0.793349	
sup_cultivo_ha	8.145e-02	2.466e-02	3.303	0.001081	**
edad_plantacion	3.530e-02	5.099e-02	0.692	0.489385	
Asociado_dum	2.797e-02	5.675e-02	0.493	0.622487	
fertilización_dum	-2.998e-02	1.386e-01	-0.216	0.828983	
num_veces_fert	-7.495e-02	7.830e-02	-0.957	0.339227	
num_veces_contr_fito_san	8.315e-02	2.181e-02	3.812	0.000169	***
prob_clima	-5.886e-03	1.119e-01	-0.053	0.958077	
prob_plagas_enferm	-3.169e-02	9.282e-02	-0.341	0.733084	
prob_otros	-1.140e-01	1.604e-01	-0.711	0.477843	
extensión_dum	-3.071e-02	5.497e-02	-0.559	0.576832	
obtuvo_prestamo_dum	-8.506e-02	1.447e-01	-0.588	0.556991	
seguro_agr_dum	2.860e-01	4.307e-01	0.664	0.507220	
densidad	1.382e-04	2.076e-05	6.659	1.44e-10	***
N_ha	7.075e-04	1.725e-03	0.410	0.681963	
P_ha	-2.390e-03	1.929e-03	-1.239	0.216428	
K_ha	9.860e-03	4.084e-03	2.414	0.016414	*
Mg_ha	-2.456e-02	9.514e-02	-0.258	0.796487	
Ca_ha	-1.016e-03	2.450e-02	-0.041	0.966964	
S_ha	-1.193e-01	4.253e-02	-2.806	0.005365	**

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

Residual standard error: 0.4122 on 282 degrees of freedom
Multiple R-squared: 0.3501, Adjusted R-squared: 0.2925
F-statistic: 6.076 on 25 and 282 DF, p-value: 1.724e-15

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)	-0.1693367	0.0698049	-2.426	0.01586	*
horas_trabajo_cultivo_pp	0.0085153	0.0018253	4.665	4.65e-06	***
sup_cultivo_ha	0.0597283	0.0226067	2.642	0.00867	**
num_veces_contr_fito_san	0.0630967	0.0194493	3.244	0.00131	**
densidad	0.0001364	0.0000193	7.070	1.08e-11	***
K_ha	0.0049498	0.0022425	2.207	0.02805	*
S_ha	-0.0766593	0.0360291	-2.128	0.03417	*

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

Residual standard error: 0.4106 on 301 degrees of freedom
Multiple R-squared: 0.3118, Adjusted R-squared: 0.298
F-statistic: 22.72 on 6 and 301 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:

- **Horas de trabajo en el cultivo por parte de la persona productora (horas_trabajo_cultivo_pp):** Con un coeficiente de $8.52e^{-03}$ y un *p-valor* significativo (*p* cercano a 0.01). Esta variable indica que un incremento en una unidad en el número de horas trabajadas en el cultivo, aumenta el rendimiento en la magnitud que expresa el coeficiente beta, sugiriendo que es necesario dedicar un mayor tiempo de trabajo al cultivo ya que sería crucial para la productividad.
- **Superficie del cultivo en hectáreas (sup_cultivo_ha):** Este presenta un coeficiente de $5.97e^{-02}$ con un *p-valor* significativo (*p* cercano a 0). Revela que el incrementar en una unidad el número de hectáreas destinadas a este cultivo, incrementará el rendimiento esperado del mismo.
- **Número de veces control fitosanitario (num_veces_contr_fito_san):** Registra un coeficiente de $6.31e^{-02}$ con un *p-valor* cercano a 0, mostrando un impacto positivo significativo en el rendimiento. Esto revela que incrementar el número de veces que se realiza el control fitosanitario, influye en la productividad.
- **Densidad de las plantas (densidad):** Expone un coeficiente de $1.36e^{-04}$ con un *p-valor* cercano a cero, mostrando un impacto positivo significativo en el rendimiento. Esto revela que el correcto manejo de las distancias de siembra entre plantas e hileras influye en la productividad; es decir, que una falta o un exceso de plantas por hectáreas afecta de manera negativa al rendimiento.

- **Potasio por hectárea (K_ha):** Tiene un coeficiente de $4.95e^{-03}$ con un *p-valor* cercano a cero, mostrando un impacto positivo significativo en el rendimiento. Esto puede indicar la importancia de una nutrición adecuada, específicamente del potasio en el rendimiento del café arábigo, basado en que este elemento se ha asociado como el nutrimento de calidad para la producción de cultivos. Debido a su papel fundamental en la fotosíntesis, la respiración y la activación de enzimas, el potasio tiene una influencia significativa tanto en el crecimiento como en la calidad de la fruta.
- **Azufre por hectárea (S_ha):** Tiene un coeficiente de $-7.67e^{-02}$ con un *p-valor* cercano a cero. Al tener un coeficiente negativo, denota que esta variable tiene una relación inversa con el rendimiento del café arábigo. Es decir, el exceso en el uso de este fertilizante, puede ser tóxico afectando a la asimilación de otros nutrientes que necesita la planta; por lo tanto, podría afectar negativamente al rendimiento. Cabe resaltar que el azufre es un nutriente secundario; por lo cual, la necesidad que tiene la planta es en menor cantidad con respecto a los macronutrientes como el potasio; por lo tanto, la cantidad adecuada es la ya existente en el suelo o la que es absorbida de fuentes naturales; por lo que, para el caso de estudio, no es necesario fertilizar esta plantación con azufre sintético.

Variables no Significativas

Muchas variables como el género de los productores, la propiedad de la tierra, asociatividad, el calcio, 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

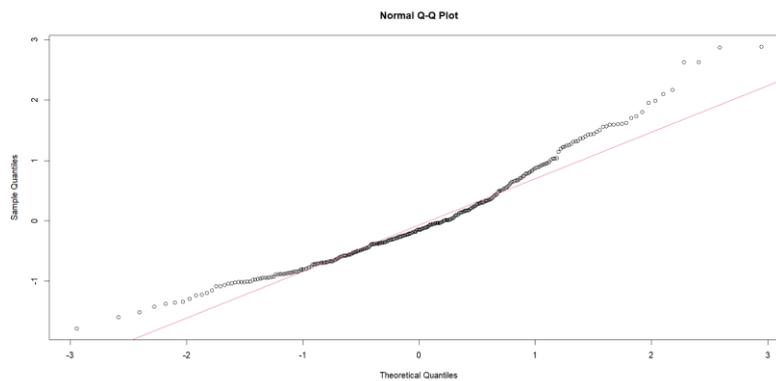
- **R-cuadrado:** El valor indica que, el modelo explica aproximadamente el 31 % de la variabilidad en el rendimiento del café arábigo. Aunque es una cantidad considerable, también sugiere que hay otros factores no incluidos en el modelo, que están afectando la productividad. En este sentido, el R-cuadrado ajustado (0.298), 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, se 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.

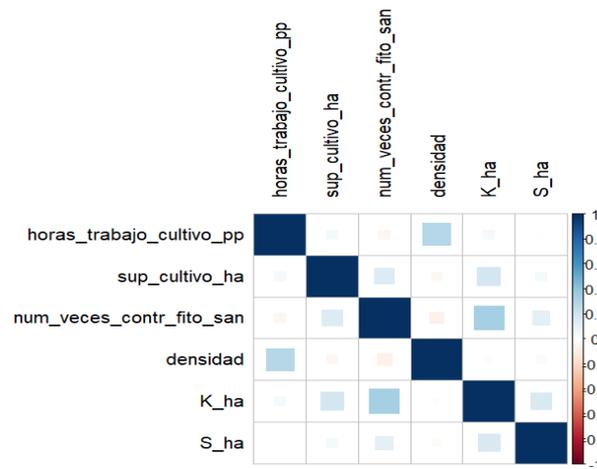
Figura 3. Gráfico de probabilidad normal



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 1.88, siendo cercano al valor deseado de 2 ($DW \approx 2$).
- **Multicolinealidad:** La multicolinealidad en el modelo de regresión, se presenta 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 4. Gráfico de correlaciones de variables del modelo

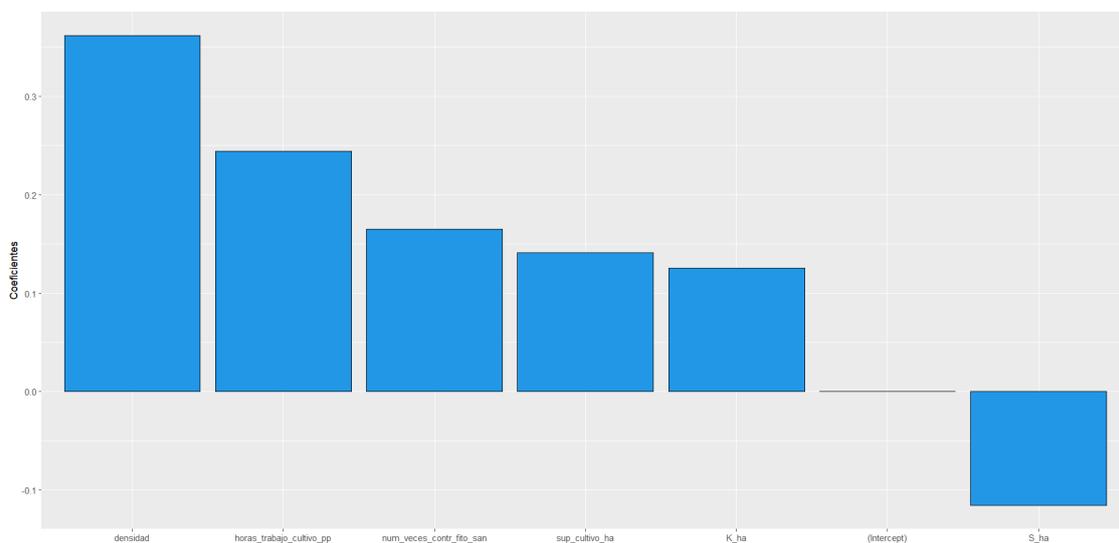


FUENTE: Coordinación General de Información Nacional Agropecuaria – MAG

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 5. 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 café arábigo, es la densidad de las plantas con las que fueron sembradas, seguido por las horas de trabajo dedicadas al cultivo, por parte de la persona productora y el número de

veces que se realizó un control fitosanitario; también se observa el aporte de la superficie del cultivo y la cantidad de potasio aplicado por hectárea.

Por otro lado, la variable que ejerce una mayor influencia de forma negativa a la productividad del cultivo, es el azufre; debido a que, como se dijo anteriormente, existe azufre en el suelo y la planta lo absorbe de varias fuentes naturales; por lo que, no es necesario el uso de este nutriente de origen sintético.

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 con las unidades de medida originales, así como los resultados una vez que fueron estandarizadas.

Este análisis destaca la importancia de ciertas prácticas agrícolas, como tener una buena densidad de siembra y características del cultivo y la persona productora, en el rendimiento del café arábigo. 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.

Sotomayor, I. (1993). *Manual del Cultivo del Café*. Quevedo, Ecuador. INIAP.