

Modelo de la dinamica del Ca

Agrocomanche s.r.l.

Richard Buitrago

2025-04-09

Índice

1. TÍTULO, PLANTEAMIENTO DEL PROBLEMA Y OBJETIVO GENERAL	2
Título	2
Planteamiento del problema	2
Objetivo general	2
2. ANÁLISIS EXPLORATORIO DE LOS DATO	3
Metodología de la investigación	3
Los datos y la metodología	3
Area de estudio	3
Fuente de los datos	5
Unidades de medidas de los datos	5
Resumen estdistico	6
Analisis de una variable cualitativa nominal	8
Graficas relevantes de variable usando ggplot2	10
3. MODELACION Y ANÁLISIS ECONOMÉTRICO	12
Modelacion	12
Estrategia de Identificación	20
3. CONCLUSIONES	21
Bibliografia	21

1. TÍTULO, PLANTEAMIENTO DEL PROBLEMA Y OBJETIVO GENERAL

Título

Efecto de la Materia Organica (M.O.) del suelo en la disponibilidad del Calcio (Ca).

Planteamiento del problema

El departamento de Santa Cruz a nivel de Bolivia, es conocida por su actividad agrícola, enfrenta desafíos importantes para llevar adelante un sistema sostenido de producción agropecuaria necesaria para atender la creciente demanda de alimentos a nivel local y regional.

En los últimos 5 años la region de San Ignacio de Velasco se ha constituido en el nuevo foco de desarrollo de la agricultura extensiva como es la soya (*Glycine max*). La formación de los suelos en la región partir de la meteorización y mineralización de las rocas, han dado lugar a la formacion de suelos con presencia de diversos minerales, los cuales conjuntante con la condiciones climáticas y presencia de materia organica, caracterizan a los suelos como fertiles desde el punto de vista agronómico. La materia orgánica es una parte intrínseca del suelo se forma a partir de restos de plantas que se incorporan al suelo por el hecho de procesos naturales, o por la intervención del ser humano. El proceso de descomposición se debe a la actividad microbiana(Julca-Otiniano et al., 2006) del suelo, este proceso está acompañado por la liberación de dióxido de carbono contenido en la materia orgánica. El efecto directo de la materia orgánica en el suelo analizado en el presente estudio es con la capacidad de intercambio cationico (CIC) de las bases como ser Calcio, Potasio, Magnesio y Sodio.

El análisis de la relacion de la materia orgánica y las bases intercambiable en el suelo permite una comprensión de la dinámica de asimilación de nutrientes por las plantas, de esta manera permite implementar estrategias de gestión sostenible del recurso suelo.

Objetivo general

Determinar la disponibilidad del calcio (Ca) en presencia de potasio (K), magnesio (Mg) y Materia orgánica (M.O.).

2. ANÁLISIS EXPLORATORIO DE LOS DATO

Metodología de la investigación

El enfoque de la investigación es cuantitativo, quasi-experimental, descriptivo y explicativo, que busca caracterizar y analizar la dinámica del calcio (Ca), en la zona de San Ignacio de Velasco, departamento de Santa Cruz.

Los datos y la metodología

El procedimiento para la obtención de los datos, en el presente trabajo de investigación esta basado en el análisis de laboratorio de las muestras de suelo obtenida en la zona de San Ignacio, en la propiedad agrícola Agrocomanche s.r.l., dedicada a la producción de soya en la campaña 2023. Las etapas y procesos que comprende, se observa en la Figura 1.

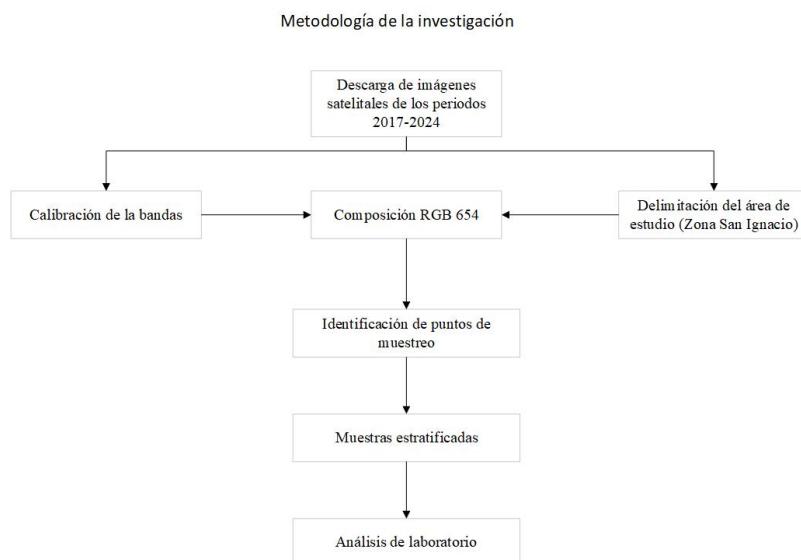


Figura 1: Metodología

Área de estudio

La zona de estudio se encuentra a 500 kilómetros al este de la ciudad de Santa Cruz de la Sierra, comprende el Municipio de San Ignacio de Velasco. Esta región es conocida por su actividad ganadera, la cual ha dado paso a la actividad agrícola de forma intensiva (producción de soya). La zona tambien es conocida por ser parte de las misiones jesuiticas Figura 1.

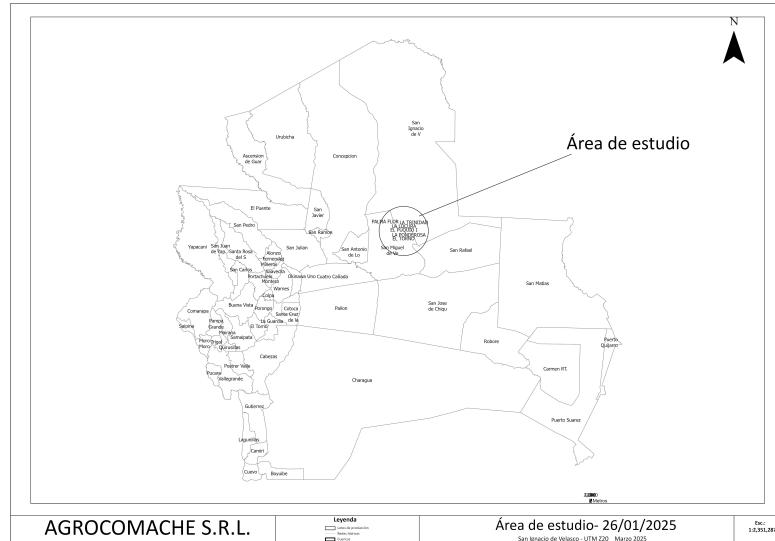


Figura 2: Área de estudio

Con el objetivo de realizar el análisis fisico-químico en la zona de estudio, se procede a identificar los puntos de muestreo a partir de imágenes satelitales (Figura 1), dando como resultado 186 muestras estratificadas (Figura 2 y 3).

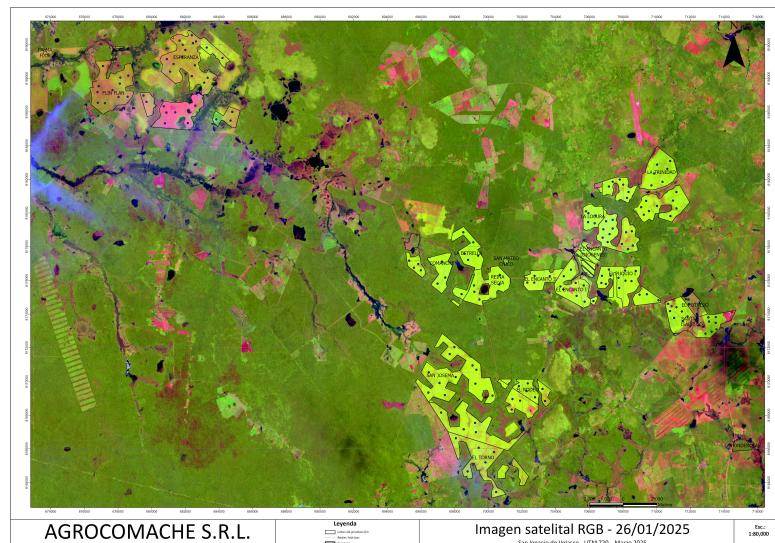


Figura 3: Imagen_muestras

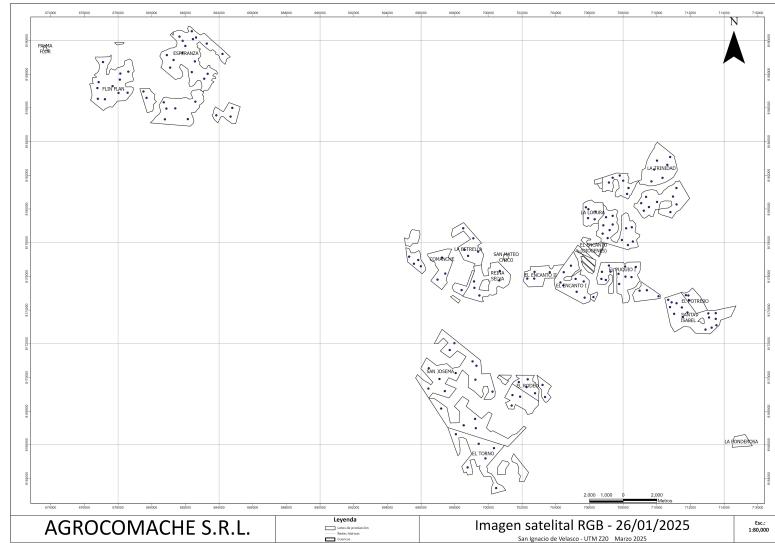


Figura 4: Puntos de muestreo

Fuente de los datos

El análisis de laboratorio realizado en el periodo 2023 por la empresa brasileña **Plante Certo**, en la zona de San Ignacio, a partir de un muestreo estratificado comprende 186 muestras (**Base de datos**), las cuales se constituyen en datos de línea base, para muestras futuras en el área de estudio.

Unidades de medidas de los datos

Las unidades de medida para las variables:

Ca, K, Mg corresponde a $cmolc/dm^3$, es decir $1\ cmolc/dm^3$ de Ca, equivale a 801.6 kg CaO/ha; $1\ cmolc/dm^3$ de K, equivale a 37,6 kg de K/ha y finalmente $1\ cmolc/dm^3$ de Mg, equivale a 484 kg MgO/ha.

La Materia Organica es medida en porcentaje(%).

Resumen estadistico

Cargar librerias

```
library(haven) #para leer datos .dta
library(fixest) #para regresiones con efectos fijos y regresiones lineales
library(ggplot2) #gráficas
library(tidyr) #manipulación de datos
library(dplyr) #manipulación de datos
library(lmtest) #tiene múltiples test
library(MASS)
library(car)
library(readxl)
```

Abrir base de datos

```
library(readxl)
Base_datos <- read_excel("Base_datos.xlsx")
View(Base_datos)
head(Base_datos)
```

```
# A tibble: 6 x 9
  K     Ca     Mg   M.O.   M01  MOC      MOD  logK  MOD1
  <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <dbl> <dbl> <dbl>
1 0.280 3.70 1.31  30.5  3.05 alta    1 -1.27  0
2 0.340 4.98 0.990 23.7  2.37 alta    1 -1.08  0
3 0.230 3.75 0.960 24.0  2.40 alta    1 -1.47  0
4 0.220 3.76 0.960 22.7  2.27 alta    1 -1.51  0
5 0.300 4.04 0.670 14.6  1.46 Baja   0 -1.20  1
6 0.210 4.94 0.710 15.1  1.51 Baja   0 -1.56  1
```

```
str(Base_datos)
```

```
tibble [186 x 9] (S3: tbl_df/tbl/data.frame)
$ K     : num [1:186] 0.28 0.34 0.23 0.22 0.3 ...
$ Ca    : num [1:186] 3.7 4.98 3.75 3.76 4.04 ...
$ Mg    : num [1:186] 1.31 0.99 0.96 0.96 0.67 ...
$ M.O.  : num [1:186] 30.5 23.7 24 22.7 14.6 ...
$ M01   : num [1:186] 3.05 2.37 2.4 2.27 1.46 ...
$ MOC   : chr [1:186] "alta" "alta" "alta" "alta" ...
$ MOD   : num [1:186] 1 1 1 1 0 0 1 1 1 0 ...
$ logK  : num [1:186] -1.27 -1.08 -1.47 -1.51 -1.2 ...
$ MOD1  : num [1:186] 0 0 0 0 1 1 0 0 0 1 ...
```

Resumen estadistico de las variables

```
library(summarytools)
summary(Base_datos)
```

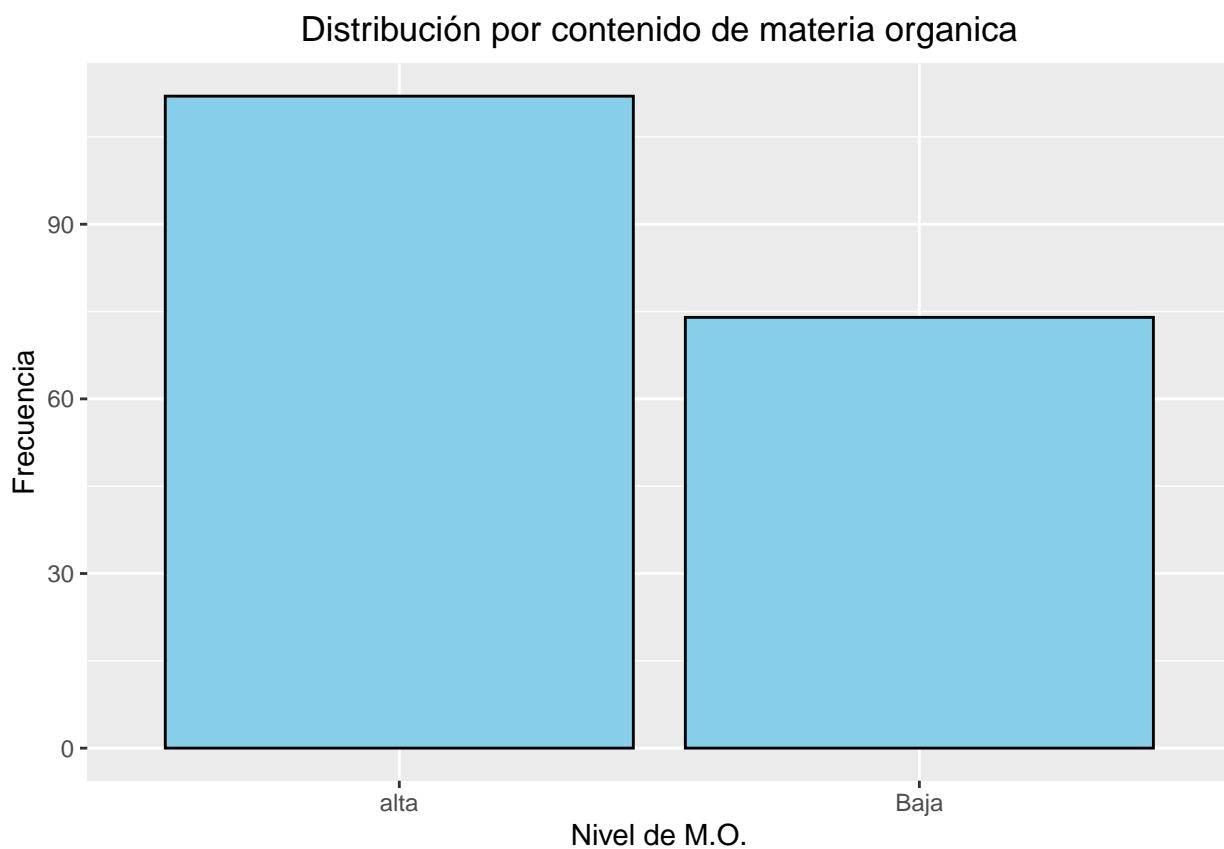
	K	Ca	Mg	M.O.
Min.	:0.0400	Min. :0.750	Min. :0.4100	Min. :10.54
1st Qu.	:0.2300	1st Qu.:3.420	1st Qu.:0.7100	1st Qu.:18.23
Median	:0.3000	Median :4.080	Median :0.8600	Median :20.93
Mean	:0.3162	Mean :4.291	Mean :0.8937	Mean :21.34
3rd Qu.	:0.3600	3rd Qu.:5.115	3rd Qu.:0.9900	3rd Qu.:24.13
Max.	:0.8500	Max. :9.100	Max. :2.6300	Max. :41.24
	M01	MOC	MOD	logK
Min.	:1.054	Length:186	Min. :0.0000	Min. :-3.2189

```
1st Qu.:1.823  Class :character  1st Qu.:0.0000  1st Qu.:-1.4697
Median :2.093  Mode  :character  Median :1.0000  Median :-1.2040
Mean   :2.134                           Mean   :0.6022  Mean   :-1.2299
3rd Qu.:2.413                           3rd Qu.:1.0000  3rd Qu.:-1.0217
Max.   :4.124                           Max.   :1.0000  Max.   :-0.1625
MOD1
Min.   :0.0000
1st Qu.:0.0000
Median :0.0000
Mean   :0.3978
3rd Qu.:1.0000
Max.   :1.0000
```

Analisis de una variable cualitativa nominal

Distribucion de frecuencia para la variable categorica

```
library(ggplot2)
ggplot(Base_datos, aes(x = MOC)) +
  geom_bar(fill = "skyblue", color = "black") +
  ggtitle("Distribución por contenido de materia organica") +
  labs( x = "Nivel de M.O.", y = "Frecuencia") +
  theme(plot.title = element_text(hjust = 0.5))
```

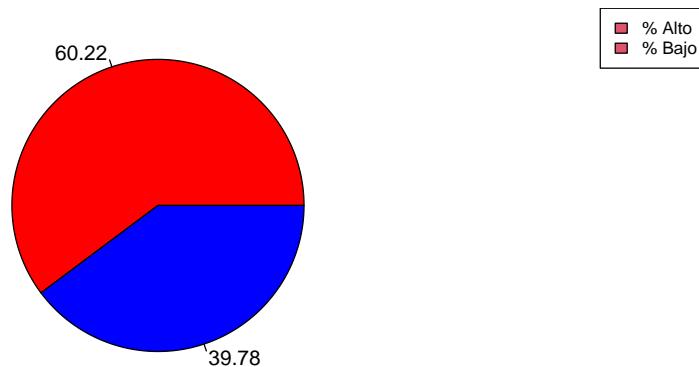


```

piepercent <- round(prop.table(table(Base_datos$MOC))*100,2)
pie(table(Base_datos$MOC),
main="Distribución por contenido de materia organica", cex.main=0.9, # título
col=c("red","blue"), # damos color a los sectores
labels=piepercent)
legend("topright", c("% Alto", "% Bajo"), cex=0.8, fill=c(2,2)) # añadimos la leyenda al gráfico

```

Distribución por contenido de materia organica

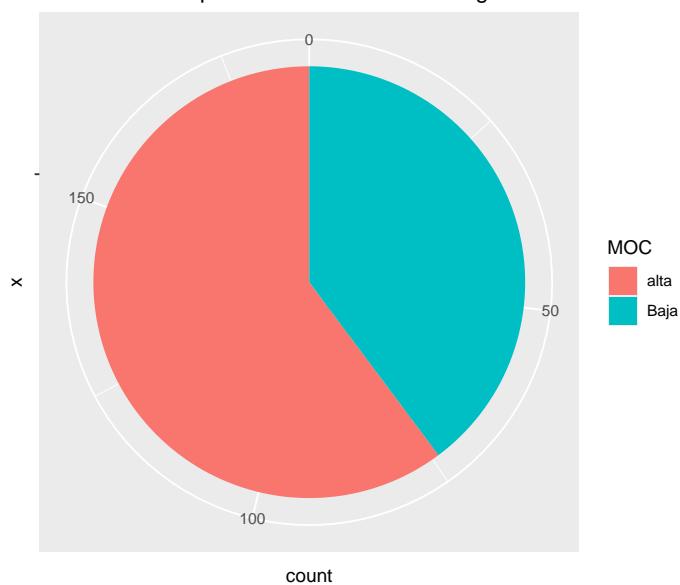


```

library(ggplot2)
ggplot(Base_datos, aes(x = "", fill = MOC)) +
  geom_bar(width = 1, stat = "count") +
  coord_polar(theta = "y") +
  labs(title = "Distribución por contenido de materia organica") +
  theme(plot.title = element_text(hjust = 0.5))

```

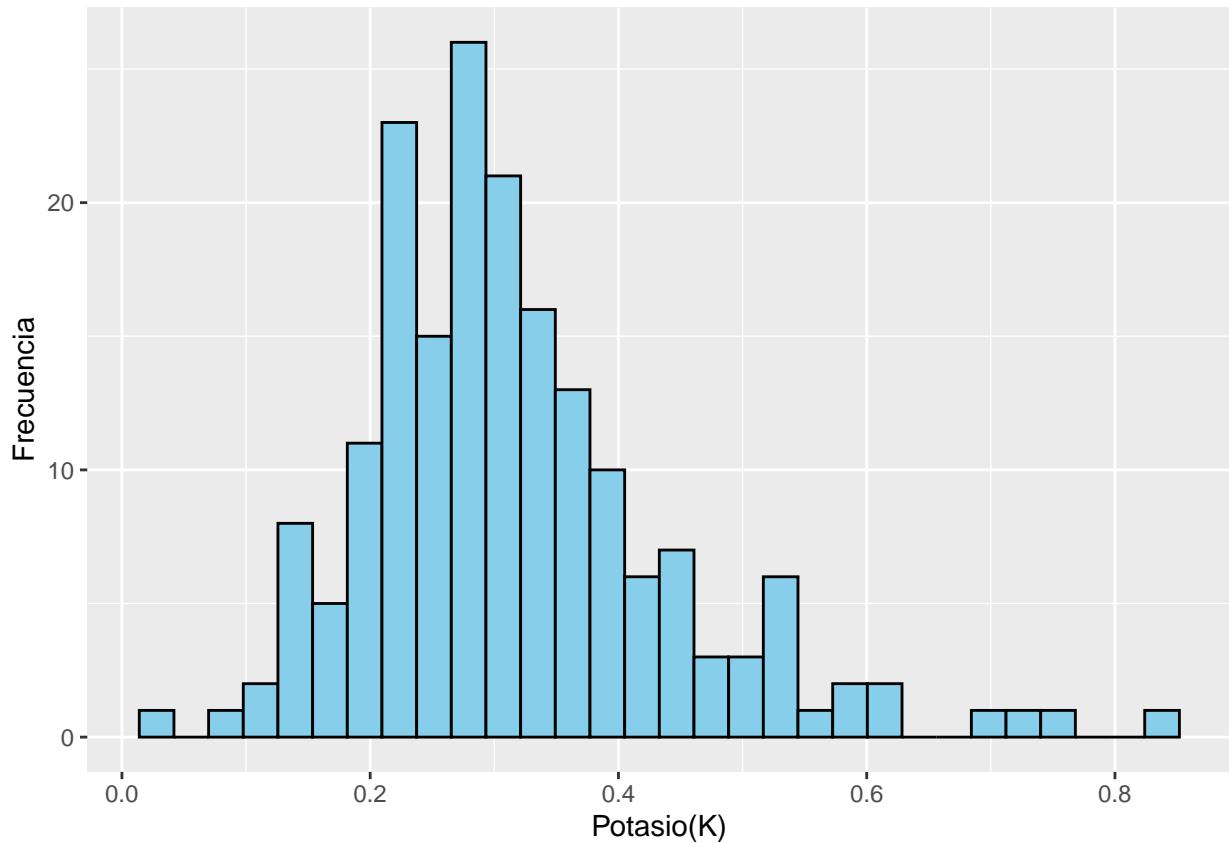
Distribución por contenido de materia organica



Graficas relevantes de variable usando ggplot2

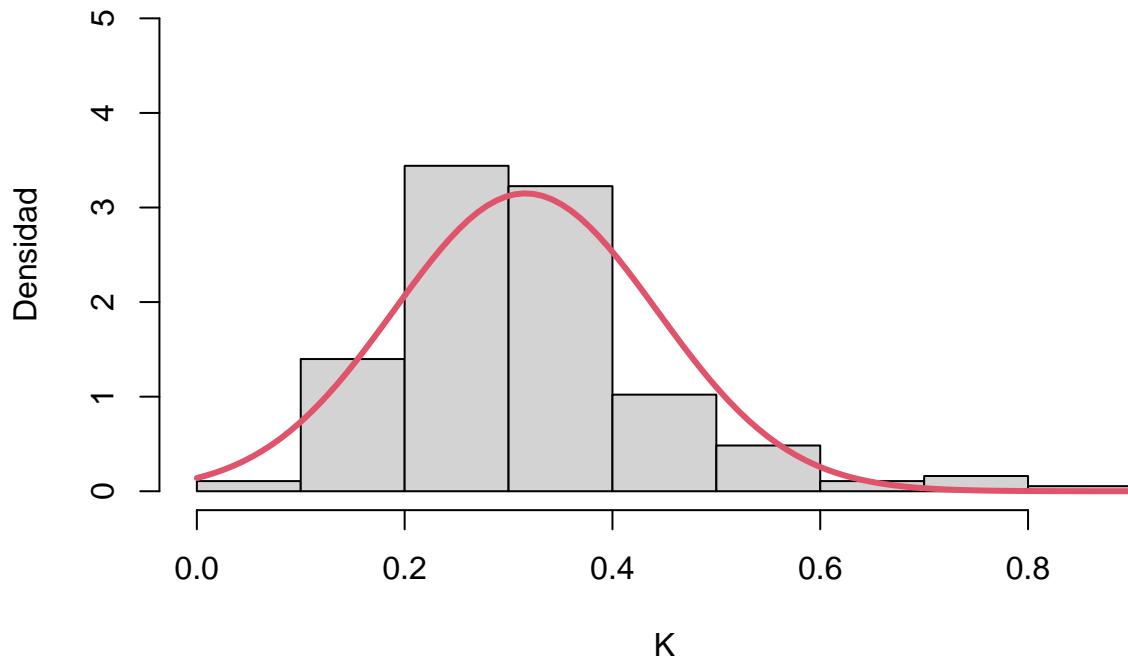
Potasio (K)

```
library(ggplot2)
ggplot(Base_datos, aes(x = K)) +
  geom_histogram(fill = "skyblue", color = "black", bins = 30) +
  labs(title = "Histograma de Potasio K" + theme(plot.title = element_text(hjust = 0.5))),
  x = "Potasio(K)", y = "Frecuencia")
```



```
hist(Base_datos$K,freq = F,main = "Histograma-Densidad Normal",ylab = "Densidad",xlab = "K",
      ylim = c(0,5))
curve(dnorm(x,mean=Base_datos$K),
      sd(Base_datos$K)),add = T, col=2, lwd=3)
```

Histograma–Densidad Normal



3. MODELACION Y ANÁLISIS ECONOMÉTRICO

Modelacion

Proposito del analisis

El proposito de análisis corresponde a determinar el impacto de los cationes intercambiables (Mg, K) en el suelo cuando la materia orgánica se encuentra en niveles altos y bajos.

Especificacion del modelo

Se propone un modelo donde las variables explicativas corresponden a las concentraciones de nutrientes en el suelo, segúin la ecuación siguiente:

$$Ca_i = \beta_0 + \beta_1 Mg_i + \beta_2 K_i + \beta_3 MOD_i + \mu_i;$$

donde:

Ca_i : Calcio disponible, cmol/dm³;

Mg_i : Manganese disponible, cmol/dm³;

MOD_i : Materia organica disponible (Variable dicotomica); MO>=2% = 1; MO<2% = 0

siendo i, el iésimo muestra observada. Mientras $\beta_i, i=1,2,3,4,5$ representan los coeficientes de impacto; siendo adicionalmente, μ_i (perturbación)

Varianza constante.

La hipótesis referida al valor de los parámetros del modelo son las siguientes:

$H_0: \beta_i = 0 ; i=1,2,3,4,5;$

$H_1: \beta_i \neq 0 ; i=1,2,3,4,5;$

mientras que las hipótesis estadísticas, asumen la forma siguiente;

$$E(\epsilon) = 0$$

$$E(\epsilon\epsilon') = \theta^2 I$$

Estimacion en R (1er Modelo)

```
ols<-feols(Ca~Mg + K + MOD, data =Base_datos, vcov="hc1")
summary(ols)

OLS estimation, Dep. Var.: Ca
Observations: 186
Standard-errors: Heteroskedasticity-robust
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.602189   0.260877 6.14156 5.0099e-09 ***
Mg          0.276728   0.209518 1.32078 1.8823e-01
K           6.275271   0.773303 8.11489 7.0451e-14 ***
MOD         0.759205   0.133516 5.68626 5.0924e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.87695  Adj. R2: 0.543045
```

Respuesta

El modelo obtenido indica lo siguiente:

- 1.- El Magesio (Mg) no es significativo en el modelo al 5%.
- 2.- El potasio (K) es altamente significativo menor al 1%, implica que por cada unidad adicional de K el Ca aumenta en 6.27 unidades de cmol/dm³.
- 3.- La presencia de materia organica mayor al 2% en el suelo es significativa a un nivel de significancia

menor al 1% (variable dicotomica), implica que suelos con una unidad adicional de MO superior al 2%, el Ca aumenta en 0.76 cmol/dm³. 4.- El R² ajustado es de 0.5430, es decir las variables regresoras explican el modelo en un 54.30 %.

Estimacion en R (2do Modelo)

```
ols<-feols(Ca~ K + MOD, data =Base_datos, vcov="hc1")
summary(ols)
```

OLS estimation, Dep. Var.: Ca
Observations: 186
Standard-errors: Heteroskedasticity-robust
Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.787557 0.231836 7.71045 7.7551e-13 ***
K 6.343017 0.766064 8.28001 2.5243e-14 ***
MOD 0.826502 0.126217 6.54824 5.7015e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.879807 Adj. R2: 0.542575

Estimacion en R (3er Modelo)

```
ols<-feols(Ca~Mg + MOD, data =Base_datos, vcov="hc1")
summary(ols)
```

OLS estimation, Dep. Var.: Ca
Observations: 186
Standard-errors: Heteroskedasticity-robust
Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.128109 0.242851 12.88076 < 2.2e-16 ***
Mg 0.630549 0.295065 2.13699 3.3927e-02 *
MOD 0.995586 0.195410 5.09485 8.6449e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 1.16773 Adj. R2: 0.19419

Estimacion en R (4to Modelo)

```
ols<-feols(Ca~Mg , data =Base_datos, vcov="hc1")
summary(ols)
```

OLS estimation, Dep. Var.: Ca
Observations: 186
Standard-errors: Heteroskedasticity-robust
Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.06364 0.275156 11.13419 < 2.2e-16 ***
Mg 1.37348 0.297463 4.61731 7.2808e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 1.24728 Adj. R2: 0.085665

Estimacion en R (5to Modelo)

```
ols<-feols(Ca~Mg + K + MOD1, data =Base_datos, vcov="hc1")
summary(ols)

OLS estimation, Dep. Var.: Ca
Observations: 186
Standard-errors: Heteroskedasticity-robust
    Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.361394  0.296210 7.97204 1.6654e-13 ***
Mg          0.276728  0.209518 1.32078 1.8823e-01
K           6.275271  0.773303 8.11489 7.0451e-14 ***
MOD1       -0.759205  0.133516 -5.68626 5.0924e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.87695  Adj. R2: 0.543045
```

Cumplimiento de supuestos del modelo

```
modelo1<-lm(Ca~Mg + K + MOD, data= Base_datos)
summary (modelo1)
```

Call:
lm(formula = Ca ~ Mg + K + MOD, data = Base_datos)

Residuals:

Min	1Q	Median	3Q	Max
-2.5832	-0.6296	0.0038	0.5178	3.3408

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.6022	0.2489	6.438	1.05e-09 ***
Mg	0.2767	0.2539	1.090	0.277
K	6.2753	0.5290	11.862	< 2e-16 ***
MOD	0.7592	0.1490	5.096	8.66e-07 ***

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

Residual standard error: 0.8865 on 182 degrees of freedom
Multiple R-squared: 0.5505, Adjusted R-squared: 0.543
F-statistic: 74.28 on 3 and 182 DF, p-value: < 2.2e-16

Respuesta

De acuerdo a los resultados el p_value es menor al 5%, por tanto el modelo es significativo.

Gráfico de los residuos

```
res<-residuals(modelo1)
plot(res,type="l")
```

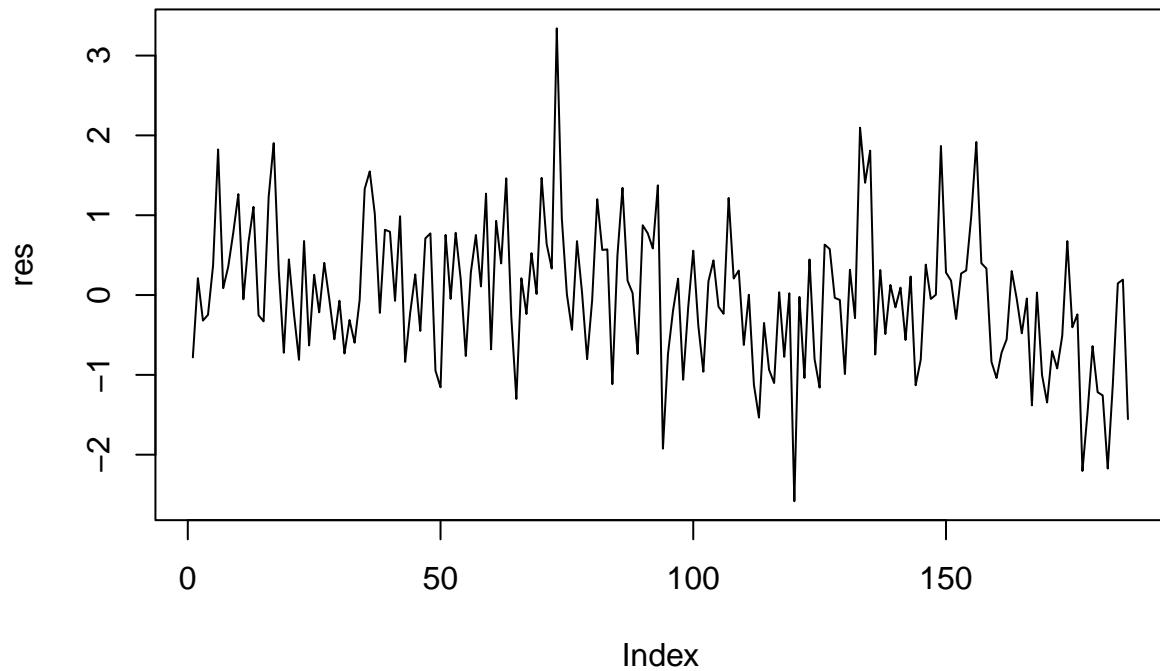


Gráfico de residuos vs valores ajustados con línea horizontal en cero

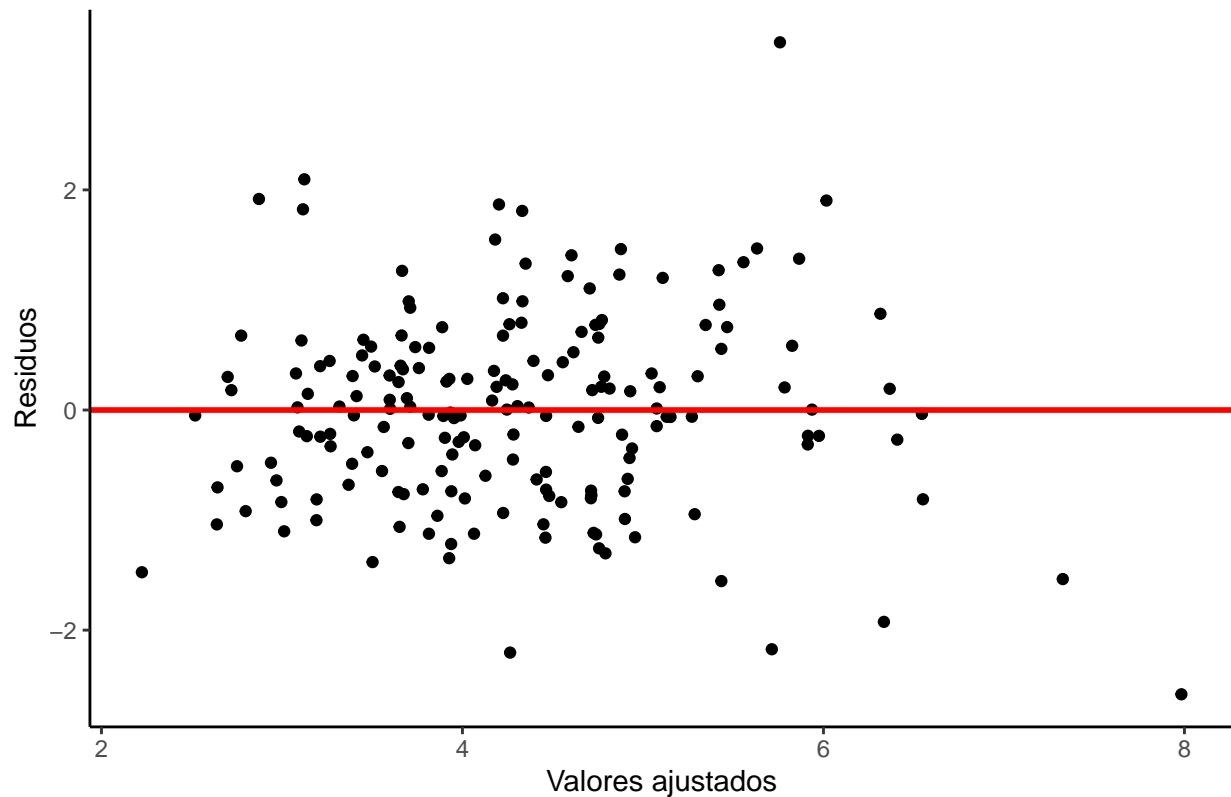
```
ggplot(data.frame(ajustados= fitted(modelo1),
residuos=resid(modelo1)),
aes(x=ajustados,y=residuos))+  
geom_point() +  
geom_hline(yintercept=0,color = "red",size=1) +  
labs(title="Residuos vs valores ajustados",x="Valores ajustados",y="Residuos") +  
theme_classic()
```

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.

This warning is displayed once every 8 hours.

Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.

Residuos vs valores ajustados



Respuesta

Tomando en cuenta los residuos de los errores se aprecia que no hubiera tendencia en la varianza de los errores y por tanto posiblemente existiera homocedasticidad, portanto se procede a realizar la prueba de BP.

Prueba de heterocedasticidad (Breusch-Pagan Test)

```
bptest(modelo1)
```

studentized Breusch-Pagan test

```
data: modelo1  
BP = 12.922, df = 3, p-value = 0.004807
```

Respuesta

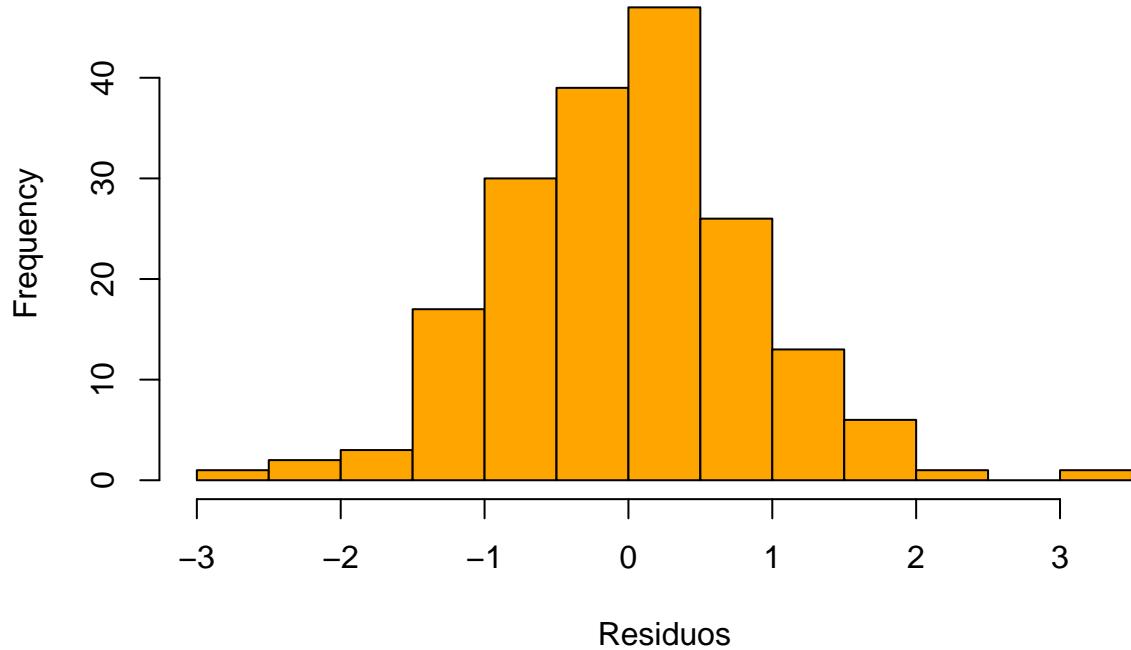
Debido a que `p_value` es menor al nivel de significacia α al 5% se acepta la H_0 y se rechaza la H_1 , por consiguiente los residuos son homocedasticos.

Gráficos de residuos para verificar la normalidad

Histograma de los residuos

```
hist(resid(modelo1), main = "Histograma de residuos",  
      xlab = "Residuos", col = "orange", border = "black")
```

Histograma de residuos

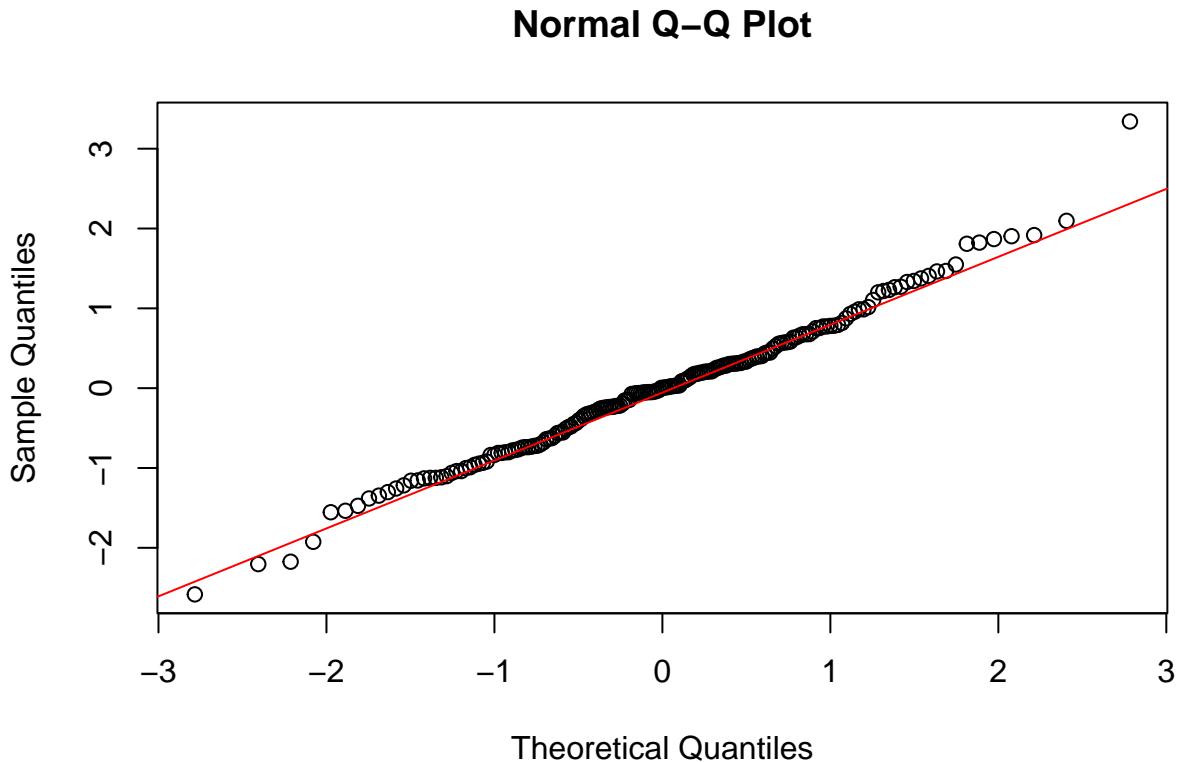


Respuesta

El histograma de los residuos presenta sesgos en la cola derecha.

QQ-plot para verificar la normalidad de los residuos

```
qqnorm(resid(modelo1))
qqline(resid(modelo1), col = "red")
```



Respuesta

El gráfico qq-plot confirma la existencia de sesgos en los residuos en la cola derecha.

Prueba de normalidad de los residuos (Shapiro-Wilk)

```
shapiro.test(resid(modelo1))
```

Shapiro-Wilk normality test

```
data: resid(modelo1)
W = 0.9899, p-value = 0.2132
```

Respuesta

De acuerdo a la prueba Shapiro-Wilk se concluye que el p_value es mayor al nivel de significancia α del 5%, por tanto se acepta la H_0 , es decir la distribución de los residuos es normal.

Especificacion lineal del modelo

Especificacion del modelo

```
resettest(modelo1)
```

RESET test

```
data: modelo1
RESET = 7.4926, df1 = 2, df2 = 180, p-value = 0.0007488
```

Respuesta

De acuerdo al test de especificidad del modelo y tomando en cuenta el p_value y el nivel de significancia α al 5%, se acepta la H_0 es decir el modelo esta correctamente especificado de forma **lineal**. Esta afirmacion se basa en que la distribución de los errores es normal y homocedastico.

Prueba de autocorrelación de los residuos (Durbin-Watson Test)

```
dwtest(modelo1)
```

Durbin-Watson test

```
data: modelo1
DW = 1.4603, p-value = 8.386e-05
alternative hypothesis: true autocorrelation is greater than 0
```

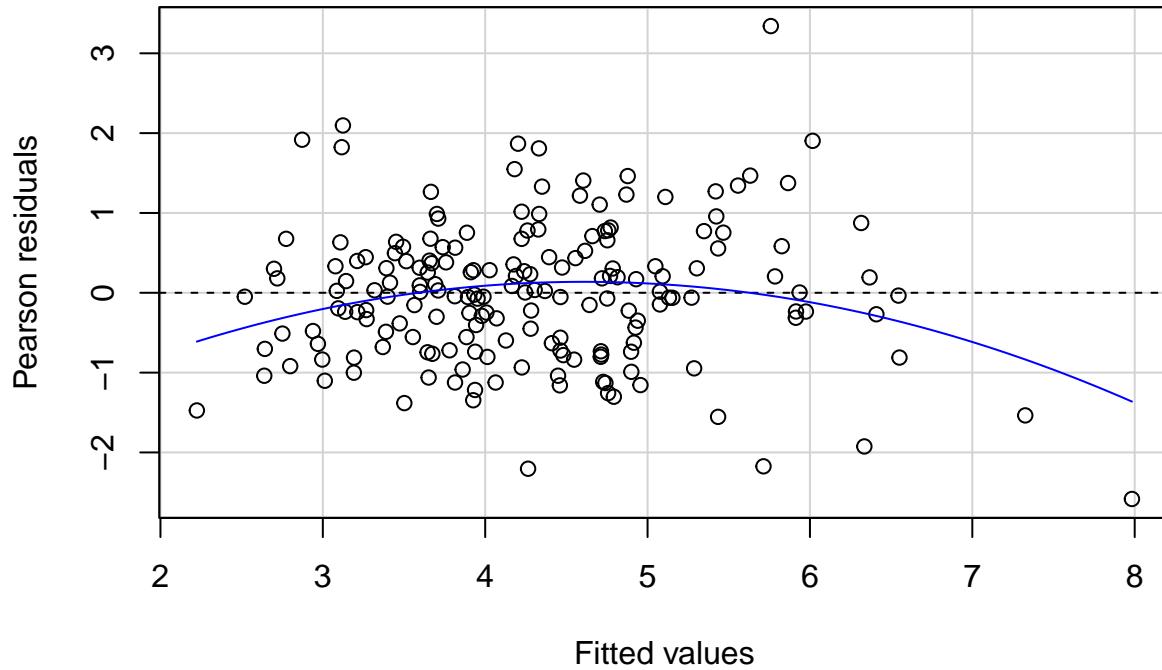
Respuesta

El valor DW = 1.46 y el p_value 0.2533 > a 0.05, se concluye ausencia de autocorrelacion entre las variables.

Es la regresión lineal adecuada

Detectar no linealidad de los residuos

```
residualPlot(modelo1)
```



Respuesta

Si bien los residuos al parecer son lineales, se evidencia la existencia de un patrón que puede mejorar la regresión, por ejemplo una de segundo grado.

Estrategia de Identificación

Respuesta

1.- Como el coeficiente $\beta_1 = 0.2767$; $\beta_2 = 6.2753$ y $\beta_3 = 0.7592$, son positivos se concluye que son significativos en la respuesta del Ca

2.- No se puede afirmar que las únicas causas que afecta la disponibilidad del Ca en suelo se deba exclusivamente de las variables analizadas, pueden haber otras causas omitidas no tomadas en cuenta como ser la presencia de sodio Na, que afecta a la disponibilidad del Ca. Tomando en cuenta el sesgo de la variable omitida (Na), se tiene:

Modelo propuesto

$$Ca_i = \beta_0 + \beta_1 Mg_i + \beta_2 K_i + \beta_3 MO_i + \mu_i;$$

Modelo con una variable omitida

$$Ca_i = \Gamma_0 + \Gamma_1 Mg_i + \Gamma_2 K_i + \Gamma_3 MO_i + \Gamma_4 Na_i + \mu_i;$$

Comparando ambos modelos y tomando en cuenta el efecto marginal del Nivel de ingreso

$$\beta_2 = \Gamma_2 + \Gamma_4 \frac{Cov(K, Na)}{Var(K)}$$

Se concluye el signo del sesgo es positivo ya que el Na infiere de forma negativa en el K y en el Ca, por tanto (-*-=+), implicando el signo positivo, de esta manera el estimador β_2 está sobreestimado.

3. CONCLUSIONES

El modelo es significativo, los residuos cumplen los supuestos de linealidad, distribución normal, aunque pudiera haber cierta autocorrelación según el estadístico Durwin-Watson.

De acuerdo a los resultados obtenidos por el modelo se evidencia que la presencia de la materia orgánica tiene un impacto significativo (Cruz-Macías et al., 2020), en la capacidad de intercambio cationico y principalmente en el calcio (Ca).

Bibliografía

- Cruz-Macías, W. O., Rodríguez-Larramendi, L. A., Salas-Marina, M. Á., Hernández-García, V., Campos-Saldaña, R. A., Chávez-Hernández, M. H., & Gordillo-Curiel, A. (2020). Efecto de la materia orgánica y la capacidad de intercambio catiónico en la acidez de suelos cultivados con maíz en dos regiones de Chiapas, México. *REVISTA TERRA LATINOAMERICANA*, 38(3), 475-480. <https://doi.org/10.28940/terra.v38i3.506>
- Julca-Otiniano, A., Meneses-Florián, L., Blas-Sevillano, R., & Bello-Amez, S. (2006). LA MATERIA ORGÁNICA, IMPORTANCIA Y EXPERIENCIA DE SU USO EN LA AGRICULTURA. *Idesia (Arica)*, 24(1), 49-61. <https://doi.org/10.4067/S0718-34292006000100009>