

Tratamiento de los Modelos de Regresión Multivariantes.

Methodological proposal for the treatment of multivariate regression models.

Yader Alexander Ramírez Bonilla
Máster en Matemática Aplicada
Estudiante del Doctorado en Matemática Aplicada
UNAN – Managua, Nicaragua
<https://orcid.org/0000-0001-5025-2386>
yader.ramirez@unan.edu.ni

Iván Augusto Cisneros Díaz
Doctor en Matemática Aplicada
UNAN – Managua, Nicaragua
<https://orcid.org/0000-0003-2014-1946>
ivan.cisneros@unan.edu.ni

RESUMEN

El análisis estadístico multivariado, es una herramienta poderosa al momento de trabajar con una inmensa cantidad de datos, nos permiten resumir e inferir sobre los parámetros poblacionales de donde procede dicho conjunto de datos, esto hace que sean atractivos para estudiar desde el punto de vista práctico y teórico.

El propósito fundamental de este trabajo consiste en la presentación y abordaje del análisis de regresión múltiple desde un punto de vista más práctico, particularmente enfocándose en su aplicabilidad en diversos fenómenos de la realidad.

Cabe recalcar que este proceso metodológico se caracteriza por analizar de una forma intuitiva, práctica y sencilla el análisis de regresión múltiple, utilizando las operaciones fundamentales matriciales, la teoría de valores y vectores propios al igual que el cálculo de los determinantes. Se abordan los aspectos fundamentales de la teoría matricial y se establece un vínculo de los datos con el análisis estadístico.

El trabajo presenta la deducción de algunas de las fórmulas a utilizar en el análisis de regresión múltiple y se desarrolla un problema práctico, el cual fue ejemplificado con los comandos del lenguaje de programación Visual R.

Palabras claves: Matrices, Regresión múltiple, Tratamiento Metodológico y Valores - Vectores Propios.

ABSTRACT

Multivariate statistical analysis is a powerful tool when working with an immense amount of data, it allows us to summarize and infer about the population parameters from which said data set comes, this makes them attractive to study from a practical point of view. and theoretical.

The fundamental purpose of this work consists in the presentation and approach of multiple regression analysis from a more practical point of view, particularly focusing on its applicability in various phenomena of reality.

It should be noted that this methodological process is characterized by analyzing multiple regression analysis in an intuitive, practical and simple way, using the fundamental matrix operations, the theory of values and eigenvectors as well as the calculation of determinants. Fundamental aspects of matrix theory are addressed and a link between data and statistical analysis is established.

The work presents the deduction of some of the formulas to be used in the multiple regression analysis and a practical problem is developed, which was exemplified with the commands of the Visual R programming language.

Keywords: Matrices, Multiple Regression, Methodological Treatment and Values - Eigenvectors.

INTRODUCCION

El análisis estadístico multivariante son un conjunto de técnicas estadísticas para investigar y modelar la relación existente entre las variables, entre ellas tenemos al análisis de regresión multivariada, esta técnica tiene un sinnúmero de aplicaciones en diversos campos como: ingeniería, econometría, química, ciencias biológicas y administración, quizás el análisis de regresión sea una de las técnicas más usadas y conocidas.

Este análisis también tiene un interés teórico, esto gracias a la matemática que se usan y por una teoría estadística bien fundamentada y desarrollada, para entender y usar bien la regresión se necesita estudiar tanto la teoría como los problemas prácticos que se suelen presentar cuando se emplea esta técnica con datos del mundo real.

Aunque el análisis de regresión lineal es muy usado actualmente, los grandes volúmenes de datos hacen esta tarea tediosa, muchas veces por la complejidad de los cálculos, teniendo que recurrir a software computaciones que nos faciliten esta tarea, uno de los software más utilizados en el análisis de datos es programación en R, este software nos permite una infinidad de paquetes y comando que ayudan a la visualización, modelación y comprensión de los datos para dar a solución a problemas que aquejan las distintas áreas del conocimiento.

MATERIALES Y MÉTODOS

En este acápite se muestra la metodología aplicada, así como una perspectiva general de la deducción de la fórmula de mínimos cuadrados para la estimación de los valores de los coeficientes en el modelo de regresión lineal multivariada

La metodología empleada en este trabajo siguió las siguientes etapas:

Revisión del estado del arte sobre el método de regresión lineal multivariada.

Selección de los datos a utilizar en la construcción del modelo

Estudio de diversos procedimientos matemáticos y estadístico en la verificación de los supuestos del modelo de regresión lineal

Por último, el uso de Programación en R para la verificación de los resultados y la comprobación de los supuestos que debe cumplir el análisis de regresión lineal múltiple.

Obtención de los coeficientes por el método de mínimos cuadrado.

Sea el conjunto $\{X_1, X_2, X_3, \dots, X_k\}$ variables independientes o regresores, sea Y la variable respuesta o dependiente, entonces el modelo de regresión lineal múltiple con k variables independientes es el polinomio de primer grado de la forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon_i$$

Donde los β_i son los parámetros del modelo o mejor conocidos como coeficientes y ε_i es el error aleatorio, con media cero y varianza constante.

El modelo de regresión lineal múltiple se puede escribir en su forma matricial de la siguiente manera:

$$y = X\beta + \varepsilon$$

Donde:

$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$	$X = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{pmatrix}$
Es el vector de Datos observados	Matriz de regresores
$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$	$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$
Es el vector de coeficiente	Error aleatorio

De esta última expresión deducimos que el error es la diferencia entre el valor observado y el valor ajustado, es decir:

$$\varepsilon = y - \hat{y}$$

La sumatoria de los errores es cero, por ende, elévalos al cuadrado, obtenemos la sumatoria de los errores al cuadrado, al cual denotaremos por S .

$$S = \sum \varepsilon_i^2$$

Dado que la sumatoria de los errores al cuadrado es un escalar y ε es una matriz de orden $n \times 1$, se multiplica por su transpuesta ε^t para obtener un escalar, donde:

$$\varepsilon = y - X\beta; \varepsilon^t = (y - X\beta)^t$$

Reescribiendo

$$S = \sum (y - X\beta)^t (y - X\beta)$$

Tenemos.

$$\begin{aligned} S &= (y - X\beta)^t (y - X\beta) \\ S &= (y^t - \beta^t X^t) (y - X\beta) \\ S &= y^t y - 2\beta^t X^t y + \beta^t X^t X \beta \end{aligned}$$

Derivando parcialmente respecto a la variable β y Evaluando en $\beta = \hat{\beta}$

$$S_{\beta}(\hat{\beta}) = -2X^t y + 2X^t X \hat{\beta}$$

Igualando a cero y despejando $\hat{\beta}$.

$$\hat{\beta} = (X^t X)^{-1} X^t y$$

Donde la matriz $(X^t X)$ de orden $n \times n$, se asume que tiene inversa. Por tanto, el modelo ajustado de regresión lineal múltiple es:

$$\hat{y} = X \hat{\beta}$$

Sesgo de los estimadores

Para que $\hat{\beta}_i$ sea un estimador insesgado de β , debemos probar que $E(\hat{\beta}_i) = \beta_i, \forall i$.

$$E(\hat{\beta}_i) = E[(X^t X)^{-1} X^t y]$$

Sustituyendo $y = X\beta + \varepsilon$. Se obtiene

$$E(\hat{\beta}_i) = E[(X^t X)^{-1} X^t (X\beta + \varepsilon)]$$

Haciendo uso de la ley distributiva

$$E(\hat{\beta}_i) = E[(X^t X)^{-1}(X^t X)\beta + (X^t X)^{-1}X^t \varepsilon]$$

Donde la multiplicación $(X^t X)^{-1}(X^t X) = I$, I es la matriz identidad

$$E(\hat{\beta}_i) = E[I\beta + (X^t X)^{-1}X^t \varepsilon]$$

El valor esperado es un operador lineal

$$E(\hat{\beta}_i) = E[I\beta] + E[(X^t X)^{-1}X^t \varepsilon]$$

Valor esperado de los residuales es cero ($E[\varepsilon] = 0$)

$$E(\hat{\beta}_i) = \beta$$

lo que demuestra que $\hat{\beta}_i$ es un estimador insesgado de β_i .

Estimación de la varianza del error

La suma de los residuales es cero, es decir

$$\sum_{i=1}^n \varepsilon_i = 0$$

Donde $\varepsilon_i = y_i - \hat{y}_i$, se eleva al cuadrado para obtener la suma de los errores al cuadrado, esta se denotará por SSE , por tanto

$$SSE = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^t \varepsilon$$

$$SSE = (y - X\hat{\beta})^t (y - X\hat{\beta}) = (y^t - \hat{\beta}^t X^t)(y - X\hat{\beta})$$

$$SSE = y^t y - y^t X\hat{\beta} - \hat{\beta}^t X^t y + \hat{\beta}^t X^t X\hat{\beta}$$

$$SSE = y^t y - \hat{\beta}^t X^t y$$

La SSE tiene $n - k$ grados de libertad, por ende:

$$\hat{\sigma}^2 = \frac{SSE}{n - k} = \frac{y^t y - \hat{\beta}^t X^t y}{n - k}$$

Se concluye que:

$$E[\hat{\sigma}^2] = \sigma^2$$

Supuestos del modelo de regresión lineal

Como todo modelo estadístico, existen supuestos del modelo de regresión lineal, cuyo cumplimiento son necesarios para validar su confiabilidad.

Estos supuestos o condiciones, constituyen la base teórica del análisis, por lo que su incumplimiento, impide su ejecución de forma tácita.

Esto quiere decir que antes de intentar aplicar el modelo de regresión lineal, a nuestros datos, es necesario examinar los supuestos que lo condicionan.

Dichos supuestos son:

Linealidad.

El modelo de regresión lineal, es un polinomio de primer grado, esto implica que el comportamiento de las variables debe ser lineal, para verificar este supuesto se realiza la prueba de significancia sobre la regresión, el cual es una prueba de hipótesis cuyo objetivo es verificar la linealidad de las variables.

Las hipótesis pertinentes son:

$$H_0: \beta_0 = \beta_1 = \beta_2 = \dots = \beta_k$$

$$H_1: \beta_i \neq 0, \text{ para algun } i$$

El rechazo de la hipótesis nula, conlleva que al menos una de las variables regresoras aporta de forma significativa al modelo y existe una relación lineal entre la variable respuesta y al menos una de las variables regresoras.

El estadístico de prueba se obtiene mediante la fórmula $F_e = \frac{CMR}{CME}$

Donde CME es el cuadrado medio del error, que viene dado por $CME = \frac{SSE}{n-k-1}$

Y el CMR es el cuadrado medio debido a la regresión, este se obtiene mediante $CMR = \frac{SSR}{k}$

Tabla ANOVA de la prueba de significancia del modelo de regresión lineal múltiple.

Fuente de variación	de Suma de cuadrados	de Grados de libertad	de Cuadrado medio	de Estadístico de Fisher
Regresión	SSR	K	CMR	$F_0 = \frac{CMR}{CME}$
Residuales	SSE	$n - k - 1$	CME	
Total	SST	$n - 1$	CMT	

Normalidad en los residuales

Para cada valor de la variable independiente o combinación de sus valores, los residuos se distribuyen normalmente con media cero.

Existen diversas pruebas para probar este supuesto, aunque la mayoría del software trabajan con la prueba de Shapiro Wilk para muestras pequeñas.

La hipótesis para probar este supuesto es:

$$H_0: \varepsilon \sim N(0, \sigma^2)$$
$$H_1 \neq N(0, \sigma^2)$$

El estadístico de prueba se calcula con la formula:

$$W = \frac{(\sum_{i=1}^n \alpha_i * x_i)^2}{SSE}$$

Donde

α_i : son los valores ponderados los cuales se obtiene de tabla

x_i : es la diferencia de los residuales ordenados de menor a mayor.

SSE : es la varianza de los residuales.

El valor de $W \leq 1$, se contrasta con un valor de tabla W' , si $W > W'$, no se rechaza la hipótesis nula, dado que entre más cercano sea W a 1, es mas probable que los residuales o cualquier otra variable, se comporte de forma normal.

No colinealidad

No existe relación lineal exacta entre ninguna de las variables independientes, el incumplimiento de este supuesto da origen a la colinealidad o multicolinealidad en dependencia de las variables que afecte.

La multicolinealidad se asocia con aquellas situaciones en las que las variables explicativas están linealmente relacionadas entre sí, es decir una variable es un múltiplo de la otra, esto con lleva problemas serios al momento de calcular los estimadores $\hat{\beta}_i$, dado que estos valores se calculan mediante la expresión:

$$\hat{\beta}_i = (X^t X)^{-1} X^t y$$

Al existir multicolinealidad entre las variables regresores, implica que el determinante de la matriz $X^t X$ es cero o muy cercano a cero.

Si no existe relación lineal entre los regresores, se dice que estos son ortogonales, se puede realizar con facilidad las inferencias y predicciones, aunque en la mayoría de las aplicaciones estas variables no son ortogonales, en ocasiones no es grave ,si la relación no es tan lineal, pero si los regresores tienen una relación casi lineal, las inferencias basadas en el modelo pueden ser erróneas.

Hay cuatro fuentes de multicolinealidad principales:

1. El método de recolección de datos que se empleó.
2. Restricciones en el modelo o en la población (variables correlacionadas de hecho)
3. Especificaciones del modelo (polinomios)
4. Un modelo sobre definido (más variables que observaciones).

Estos cuatro factores conllevan a una serie de problemas asociados a la multicolinealidad, entre ellos tenemos:

1. Grandes varianzas y covarianzas de los estimadores.
2. Estimaciones para los coeficientes demasiado grandes.
3. Pequeños cambios en los datos o en la especificación provocan grandes cambios en las estimaciones de los coeficientes.
4. Las estimaciones de los coeficientes suelen presentar signos distintos a los esperados y magnitudes poco razonables.

Para identificar problemas de multicolinealidad, existen diversos métodos, como:

La matriz de correlación

La correlación indica el grado de asociación que existe entre las variables, es decir la intensidad de relación que hay entre dos variables, es por ello que podemos utilizar la correlación como un indicador de multicolinealidad.

La matriz de correlación es una matriz cuadrada y simétrica que tiene unos en la diagonal y fuera de ella los coeficientes de correlación entre las variables, escribimos

$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \cdots & 1 \end{pmatrix}$$

La matriz de correlación se calcula mediante:

$$R = D^{-1}SD^{-1}$$

Donde S es la matriz de varianza y covarianza, D es una matriz diagonal, donde la diagonal principal son las desviaciones típicas.

Análisis del Eigen sistema de X^tX

Los valores propios de X^tX , definidos por $\lambda_1, \lambda_2, \dots, \lambda_p$ se pueden usar para medir el grado de multicolinealidad en los datos, si hay una dependencia casi lineal, una de las raíces características será pequeña, esto se puede ver al calcular el número de condición de X^tX dado por:

$$k = \frac{\lambda_{max}}{\lambda_{min}}$$

Este valor indica problemas graves de multicolinealidad si excede el valor de 1000, problemas moderados si el valor esta entre 100 y 1000, poco problema de multicolinealidad si es menor a 100.

Los índices de condición de la matriz X^tX viene dado por:

$$k_j = \frac{\lambda_{max}}{\lambda_j}$$

Varianza constante o homocedasticidad

Detectar la heterocedasticidad en un modelo de regresión es importante para verificar si la varianza de los errores del modelo es constante a lo largo de todo el rango de valores de las variables independientes.

La heterocedasticidad puede afectar la precisión y validez de los resultados obtenidos en el análisis de regresión. Aquí hay algunos métodos para detectar la heterocedasticidad:

Gráficos de residuos vs. valores ajustados:

Graficar los residuos del modelo (diferencia entre los valores observados y los valores predichos) en el eje y, y los valores ajustados en el eje x. Si la dispersión de los residuos parece aumentar o disminuir a medida que aumenta los valores ajustados, podría indicar heterocedasticidad.

Prueba de Breusch-Pagan:

Esta prueba evalúa si la varianza de los errores del modelo es constante. La hipótesis nula es que los errores son homocedásticos (constantes), y la alternativa es que existe heterocedasticidad. Se puede calcular la prueba y obtener un valor p para decidir si rechazar o no la hipótesis nula.

Ejemplo

En la Universidad Nacional Agraria (UNA), se analizaron tres variables cuantitativas en su plantación de Maíz, ellos están interesados en la altura de la planta, para tal fin se extrajo una muestra de 20 datos, los cuales se muestran en la tabla de abajo. EL ingeniero Agrónomo encargado de este estudio cree que existen dos variables que podrían incidir en la altura (cm) de la planta de maíz, estas variables son el diámetro del tallo en mm y el número de hojas.

Se busca ajustar los datos a un modelo de regresión lineal múltiple, como variable dependiente la altura de la planta y las variables regresores diámetro del tallo medido en milímetro y el número de hojas.

Los cálculos y gráficos se harán el lenguaje de programación R Studio versión 4.3.1

La tabla con los datos se cargará en R con el nombre de Ejemplo.

N	Altura	Diámetro del tallo	Numero de hojas
1	52.0	15.0	9.0
2	28.3	6.3	6.0
3	20.0	4.0	6.0
4	15.3	3.3	6.0
5	19.3	5.3	6.0
6	83.0	21.0	11.0
7	30.0	12.0	6.0
8	16.0	4.0	6.0
9	25.3	6.3	7.0
10	20.0	4.0	7.0
11	16.0	4.0	6.0
12	20.7	4.0	7.0
13	32.0	9.0	7.0
14	28.0	6.0	7.0
15	31.0	6.0	8.0
16	20.0	5.0	5.0
17	30.0	6.0	6.0
18	20.0	4.0	6.0
19	23.7	5.7	7.0
20	17.0	5.0	7.0

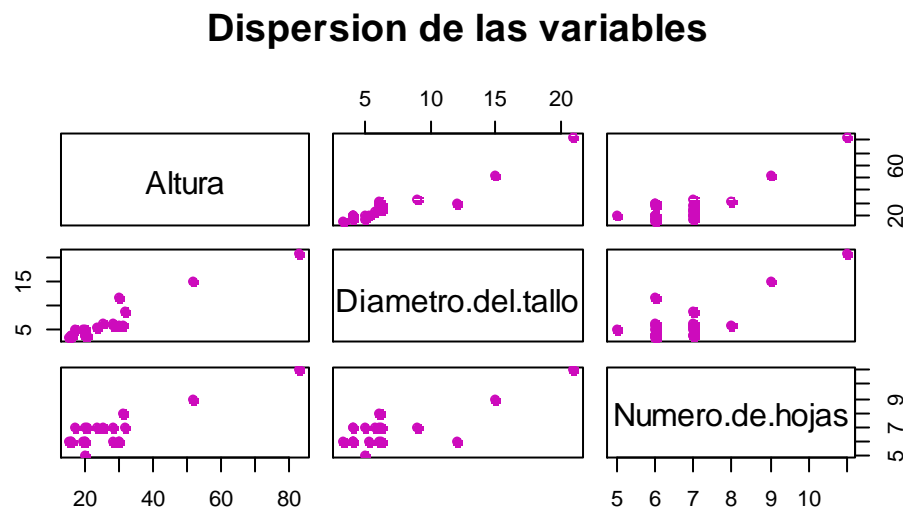
Gráfico de dispersión de las variables.

El primer análisis que se efectúa antes de la obtención de los coeficientes del modelo de regresión lineal, es un gráfico de dispersión, para realizarlo se usa la función **Pairs** de R, cuya sintaxis es la siguiente.

Pairs(data(base de datos),pch(forma de los puntos), col(color de los puntos), main(nombre del gráfico), labels(nombre de los ejes), cex.labels(espacio del texto))

Sintaxis para realizar el grafico con la base de datos Ejemplo.

pairs(Ejemplo, pch=19, col=6, main="Dispersion de las variables", labels = colnames(Ejemplo), cex.labels = 1.5)



El grafico se muestra en forma de matriz, donde en la diagonal principal se observa la variable consigo misma, fuera de la diagonal principal se observa el grafico de dispersión de cada variable con cada una de las demás variables, en el cual se observa una relación lineal fuerte entre la altura con diámetro del tallo y altura con numero de hojas.

Modelo de regresión lineal

El grafico de dispersión planteando anteriormente nos muestra que existe relación lineal entre las variables, por ellos se busca un modelo de regresión lineal múltiple que tiene la forma:

$$\text{Altura} = \text{Diámetro del tallo} + \text{Número de hojas}$$

Para encontrar los coeficientes del modelo de regresión, en R se usa el comando **lm** (linear models), cuya sintaxis es:

lm(data(base datos), variable dependiente~variables independiente 1+variable independiente 2)

el modelo se guardará con el nombre de modelo, por tanto, el código es:

```
modelo<-lm(data=Ejemplo,Altura~Diametro.del.tallo+Numero.de.hojas)
```

para mostrar los coeficientes se llaman con la siguiente sintaxis.

```
modelo$coefficients
```

```
- modelo$coefficients
      (Intercept) Ejemplo$Diametro.del.tallo  Ejemplo$Numero.de.hojas
      -14.960123          2.448090          3.780199
- |
```

El modelo de regresión lineal para estos datos es:

$$Altura = -14.960123 + 2.44809 * \text{Diámetro del tallo} + 3.780199 * \text{Número de hojas}$$

Prueba de Linealidad.

Esta prueba contempla si las variables represoras se comportan de manera lineal con la variable respuesta, para verificar este supuesto se realiza un ANOVA.

El Anova se muestra al emplear el comando `summary()`, sobre el modelo. Cuya hipótesis correspondiente son:

H_0 : No existe linealidad entre las variables

H_1 : Existe linealidad entre las variables

`Summary(modelo)`, obteniendo los siguientes resultados:

```
Residual standard error: 3.964 on 17 degrees of freedom
Multiple R-squared:  0.942,    Adjusted R-squared:  0.9352
F-statistic: 138.1 on 2 and 17 DF,  p-value: 3.08e-11
```

Se observa un P valor de $3.08 * 10^{-11}$, con lo cual se rechaza la hipótesis nula y se concluye que existe linealidad entre las variables regresoras y la variable respuesta.

Normalidad en los residuales.

Otro supuesto del modelo es la normalidad en los residuales, con el fin de verificar este supuesto se usa la prueba de Shapiro Wilks, donde la hipótesis nula es la normalidad de los residuales y la alternativa es la no normalidad.

La sintaxis en R es la siguiente.

```
Shapiro.test(variable)
```

En el modelo solo nos interesa trabajar con los residuales es por ello que el signo `$` se usa para dentro de modelo extraer únicamente los residuales.

Las hipótesis pertinentes son las siguientes:

H_0 : existe normalidad en los residuales
 H_1 : No hay normalidad en los residuales

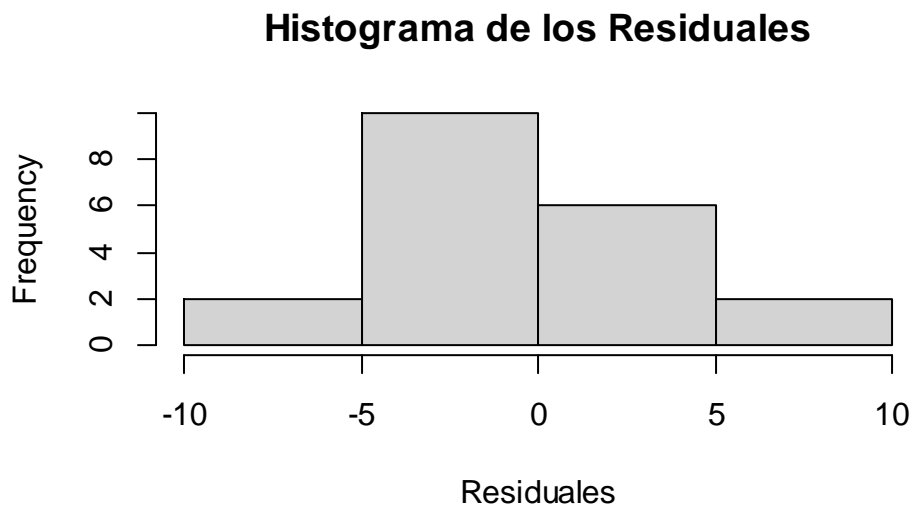
```
shapiro.test(modelo$residuals)
```

la salida

```
shapiro-wilk normality test  
data: modelo$residuals  
w = 0.95527, p-value = 0.4542
```

El P valor es superior a 0.05, No se rechaza la hipótesis nula y se concluye que existe normalidad en los residuales, esto se puede ver en el histograma de los residuales.

```
hist(modelo$residuals, breaks = "fd", main = "Histograma de los Residuales", xlab = "Residuales")
```



Multicolinealidad.

La multicolinealidad está estrechamente relacionada con la matriz (X^tX) , dado que, si el determinante es cero o muy cercano a cero, implica problemas graves de multicolinealidad, es decir existe una combinación lineal entre las variables regresores. Este problema se puede detectar a través de la matriz de correlación o el número de condición.

Matriz de correlación.

Para obtener la matriz de correlación, se usa el comando `cor(variables)`, las variables deben ser cuantitativas, dado que las 3 variables en la base de datos Ejemplo son cuantitativas, la sintaxis es:

cor(Ejemplo)

	Altura	Diámetro.del.tallo	Numero.de.hojas
Altura	1.0000000	0.9507599	0.8734616
Diámetro.del.tallo	0.9507599	1.0000000	0.7939313
Numero.de.hojas	0.8734616	0.7939313	1.0000000

Existe una correlación alta entre las variables Diámetro del tallo y el número de hojas, la cual es de 0.7939313, esto nos hace pensar que existe una relación lineal entre estas variables y por ende hay colinealidad entre las variables.

Análisis del Eigen Sistema.

```
valores<-eigen(cor(Ejemplo[,2:3]))$value
```

```
for(i in 1:length(valores)){
```

```
  k=valores[1]/valores[i]
```

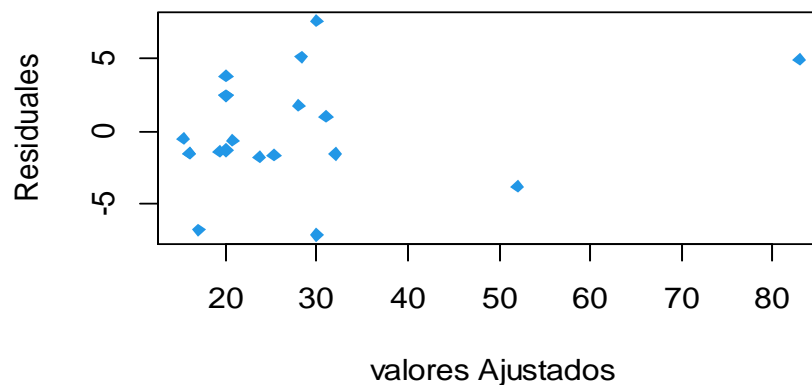
```
  print(k)}
```

```
> valores<-eigen(cor(Ejemplo[,2:3]))$value
> valores
[1] 1.7939313 0.2060687
> for(i in 1:length(valores)){
+   k=valores[1]/valores[i]
+   print(k)
+ }
[1] 1
[1] 8.7055
```

El número de condición 8.7055 muestra que el problema de colinealidad no es tan grave como se esperaba y se puede trabajar con ambas variables.

Varianza Constante.

Para analizar si existe varianza Constante, gráficamente se puede ver mediante un diagrama de dispersión, donde en el eje x irán los valores ajustados y en el eje y los residuales.



La grafica anterior muestra que a medida que los valores ajustados aumentan los residuos muestran mayor dispersión, esto conlleva a sospechar de problemas de heteroscedasticidad.

Prueba de Breusch-Pagan.

En R esta Prueba es muy sencilla se realizar, basta con cargar el paquete “lmtest”, usar el siguiente comando:

```
bptest(modelo)
```

las hipótesis pertinentes son:

$$H_0: \text{Varianza constante}$$
$$H_1: \text{Varianza no constante}$$

salida.

```
data: modelo  
BP = 5.9984, df = 2, p-value = 0.04983
```

Cuyo p valor es inferior a 0.05, por tanto, se rechaza la hipótesis nula y se concluye que hay problemas de heterocedasticidad.

Conclusiones

Los resultados de relevancia alcanzados en este trabajo investigativos se detallan a continuación:

Descripción detallada del proceso de la teoría de los modelos de regresión lineal múltiple.

Análisis de cada una de las salidas de R y su interpretación con relación al contexto del problema planteado a través de los datos obtenidos.

Todos los cálculos y algoritmo fueron realizado en R Studio, esto se realizó con el objetivo de verifica cada supuesto y contrastarlo con los cálculos obtenidos de manera manual.

Referencias Bibliográficas.

Gerber. H. (1992). Algebra Lineal, México D.F: Grupo Editorial Iberoamérica.

Montgomery, P. & Vining G. (2001). Introduction To Linear Regression Analysis (Tercera Edición). John Wiley And Sons: Editorial Continental.

Searly. S. (1982). Matriz Algebra Useful For Statistics, United States of America, John Wiley And Sons.

Walpole M. (1999). Probabilidad y Estadística Para Ingenieros (Sexta Edición), México D.F: Prentice-Hall Hispanoamericana.