UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

División Matemáticas e Ingeniería

Licenciatura en Matemáticas Aplicadas y Computación

Estadística II

Profesor: Vergara Prado Jaime

“Análisis de regresión Lineal: Un enfoque estadístico de la pobreza en México (2014-2016).”

Alumnos:
Flores González Eva Alejandrina
Ramirez Almazan Emilio
Simon Aranda Zahir Hassan

1 Marco teórico

1.1 Definición de Matriz

Las definiciones mostradas, a continuación, tiene como objetivo principal familiarizarnos con los conceptos, ya que después ahondaremos en algunas demostraciones que requieren de operaciones empleadas a partir de las mismas.

Paéz (2013)"Una matriz en $\Re$ es un arreglo rectangular de números reales distribuidos en filas y columnas

En general, una matriz real A que tiene m filas y n columnas es un ordenamiento de números reales de la forma:

donde $a_{ij} \in \Re,\ \forall\ i,\ j \in \aleph\ con\ 1\ \leq\ i\ \leq\ m,\ 1\ \leq\ j\ \leq\ n$." (Matrices y sistemas lineales, pág 4.)

Una de las propiedades empleadas de las matrices es la transpuesta,

1.1.1 Transpuesta

Paéz (2013) "Sea $A \in Mm{\times}n\ (\Re)$, la matriz transpuesta de A, denotada como A^T, es la matriz de tamaño $n{\times}m$, y cumple las siguientes propiedades:

1. $(A\ +\ B)^T = A^T +\ B^T$ 2. $(A^T)^T\ =\ A$

3. $(kA)^T\ =\ kA^T,\ con\ k\ un\ escalar$ 4. $(AB)^T\ =\ B^T A^T$"

(Matrices y sistemas lineales, pág 7)

1.1.2 Multiplicación de matrices

Paéz (2013) "Sea $A\ =\ (a_{ij})\epsilon M_{n\,x\,m}$ y $B\ =\ (b_{ij})\epsilon\ M_{m\,x\,p}$, entonces el producto AB es la matriz $C\ =\ (c_{ij})\epsilon M_{n\,x\,p}$, cuyos elementos son de la forma: $C_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} +...+ a_{im}b_{mk} = \sum_{j=1}^{m} a_{ij}b_{jk}$".

Con las siguientes propiedades:

1. $A(BC)\ =\ (AB)C$ 2. $A\cdot I = I\cdot A\ =\ A$ 3. $A(B\ +\ C)\ =\ A\cdot B\ +\ A\cdot C$" (Matrices y sistemas lineales, pág 16)

A lo largo del documento se estarán mostrando resultados obtenidos por medio del entorno **R**, en consecuencia:

1.2 Entorno de *R*

De acuerdo a Paula Elosua (2011) "R es un entorno de programación, análisis estadístico y software gráfico derivado del lenguaje de programación S (Becker, Chambers y Wilks, 1988; Chambers, 1998; Chambers y Hastie, 1992; Venables y Ripley, 2000)." "R dispone de funciones básicas relacionadas con los análisis descriptivos de datos, y de los modelos más complejos y actuales concernientes con los últimos avances en el campo de la estadística, la psicometría, la econometría o el análisis de datos." (Introducción al entorno R, pág 13)

Respecto al trabajo en cuestión, se define la regresión lineal en primer lugar como:

1.3 Definición de Regresión lineal

Montgomery et al. (2002) aclara que "el análisis de regresión es una técnica estadística para investigar y modelar la relación entre variables. Un aspecto esencial del análisis de regresión es la recolección, recopilación o adquisición de datos. Un buen esquema de recolección de datos puede asegurar un análisis simplificado y un modelo de aplicación más general.

Los modelos de regresión se usan con varios fines, que incluyen los siguientes: descripción de datos, estimación de parámetros, control, predicción y estimación”. (Introducción al análisis de regresión lineal, pp. 1-9).
Al aplicar un modelo de regresión lineal es necesario identificar de manera clara qué papel está jugando cada una de las variables a trabajar, de ahí que,

1.3.1 Variable dependiente

Alfonso Novales (2010) define a una variable dependiente como: “variable cuyo comportamiento se pretende explicar, Y_i. También tome los nombres de endógena o variable a explicar(respuesta)” (Análisis de regresión, pág 14)

1.3.2 Variable independiente

Alfonso Novales (2010) dice “mientras que la variable denotada por X_i recibe el nombre de variable independiente, exógena o explicativa” y permite controlar el modelo. (Análisis de regresión, pág 14).

Se estará definiendo dos tipos de regresión lineal: simple y múltiple, sin embargo, para este caso particular sólo se estará trabajando con la regresión lineal simple o RLS, y añadiendo el tema de mínimos cuadrados para el ajuste de rectas que posteriormente se estará calculando en R.

1.3.3 Tipos de regresión

Regresión lineal simple (RLS)
En este tipo de regresión lineal se toma en cuenta tanto las variables independientes y las dependientes como se refiere a (Bouza,2018)“Cuando el coeficiente de regresión lineal entre dos variables sea “cercano” $a + 1$ o $a - 1$, tiene sentido considerar la ecuación de la recta que “mejor se ajuste” a la nube de puntos. Esta recta permitirá establecer el lugar geométrico que describe la relación y puede ser utilizada para predecir o estimar los valores que tomaría de Y al fijar X. ”(Modelos de regresión y sus aplicaciones,pág 10), donde se puede mostrar mediante un modelo el cual es:

$$Y = \beta_0 + \beta_1 X$$

(Novales,2010) “Los coeficientes β_0 y β_1 se denominan términos constante y pendiente del modelo de regresión lineal siempre, respectivamente.” (Análisis de regresión, pág.14)

Mínimos cuadrados

“Para hacer una estimación del modelo de regresión lineal simple, trataremos de buscar una recta de la forma:

$$Y = \beta_0 + \beta_1 X = b_0 + b_1 X$$

de modo que se ajuste a la nube de puntos.
Para esto utilizaremos el método de mínimos cuadrados.
Este método consiste en minimizar la suma de los cuadrados de los errores: $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$
Es decir, la suma de los cuadrados de las diferencias entre los valores reales observados (Y_i) y los valores estimados ($\widehat{Y}_i$).” (Carrollo, pág 5)

Regresion lineal multiple (RLM)
(Santibañez,2018) “El modelo de regresión lineal múltiple (RLM) es una extensión del modelo (RLS) ya que se modelos una variable aleatoria Y condicional a un conjunto de variables auxiliares $X_1, X_2, ..., X_n$, mismas que se asumen no aleatorias.” (regresion lineal multiple,pp.3)

Tiene la forma : $Y = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \varepsilon$

1.3.4 Diagramas de Dispersión

Juan Pablo de la Guerra (2015) define "Los Diagramas de Dispersión o Gráficos de Correlación permiten estudiar la relación entre 2 variables. Dadas 2 variables X y Y, se dice que existe una correlación entre ambas si cada vez que aumenta el valor de X aumenta proporcionalmente el valor de Y (Correlación positiva) o si cada vez que aumenta el valor de X disminuye en igual proporción el valor de Y (Correlación negativa)." (Las siete herramientas de la calidad, pág. 6)

1.4 Análisis de varianza (ANOVA)

"El *ANOVA* es un conjunto de técnicas estadísticas de gran utilidad y ductilidad. Es útil cuando hay más de dos grupos que necesitan ser comparados, cuando hay mediciones repetidas en más de dos ocasiones, cuando los sujetos pueden variar en una o más características que afectan el resultado y se necesita ajustar su efecto o cuando se desea analizar simultáneamente el efecto de dos o más tratamientos diferentes" (Dagnino, 2014, p. 1).

1.5 Supuestos teóricos del modelo de regresión lineal simple(RLS)

La idea principal de los supuestos teóricos del modelo de regresión lineal simple es el cumplimiento de una serie de supuestos que garanticen su correcta desarrollo y aplicación, partiendo en el análisis de residuales. Estos supuestos se dividen en 4: linealidad, independencia, homocedasticidad y normalidad. Cada uno se estará tratando a partir de test que permite su ejemplificación y verificación de resultados.

1.5.1 Homocedasticidad y Heterocedástico

"El supuesto de homocedasticidad exige que para todo el recorrido de la variable X la varianza del error sea constante. Esto es importante de cara a la predicción de valores en los cuales la desviación tipo de los residuos forma parte del cálculo del intervalo de confianza." (Thompson, 2015, como se citó en Giler, 2021. pág. 35).

Como define Rafael de Arce (2001) "la heterocedasticidad es la existencia de una varianza no constante en las perturbaciones aleatorias de un modelo econométrico." (Conceptos Básicos Sobre la Heterocedasticidad en el Modelo Básico de Regresión Lineal, pág. 3)

Este contraste se hace con el apoyo de gráficos o test, en este caso, empleando el test de Breusch-Pagan para un contraste paramétrico, es decir,

Prueba de Breusch-Pagan

"La idea del contraste es comprobar si se puede encontrar un conjunto de variables Z que sirvan para explicar la evolución de la varianza de las perturbaciones aleatorias, estimada ésta a partir del cuadrado de los errores del modelo inicial sobre el que se pretende comprobar si existe o no heterocedasticidad." (Arce, R. 2001, Conceptos Básicos Sobre la Heterocedasticidad en el Modelo Básico de Regresión Lineal, pág. 8)

1.5.2 No Autocorrelación

Se hace el supuesto de que las perturbaciones aleatorias no están autocorrelacionadas. El origen de la autocorrelación puede ser por diferentes razones y dependerá de la naturaleza del problema. En algunas ocasiones el no incluir algunas variables importantes al modelo o el catalogar un ajuste como incorrecto, originará la autocorrelación.

1.5.2.1 Prueba During-Watson

“Esta prueba presenta probar si los errores no se encuentran autocorrelacionados, de acuerdo al supuesto $Cov(\varepsilon_i, \varepsilon_j) = 0$, con tan objetivo se llevará a efecto la prueba de hipótesis siguiente: $H_0 = No\ existe\ autocorrelación\ en\ los\ \varepsilon_i$

$H_a = Existe\ evidencia\ de\ autocorrelación\ en\ los\ \varepsilon_i$”

(Alonso, M. 1994. Análisis de Residuales en Regresión, pág. 14)

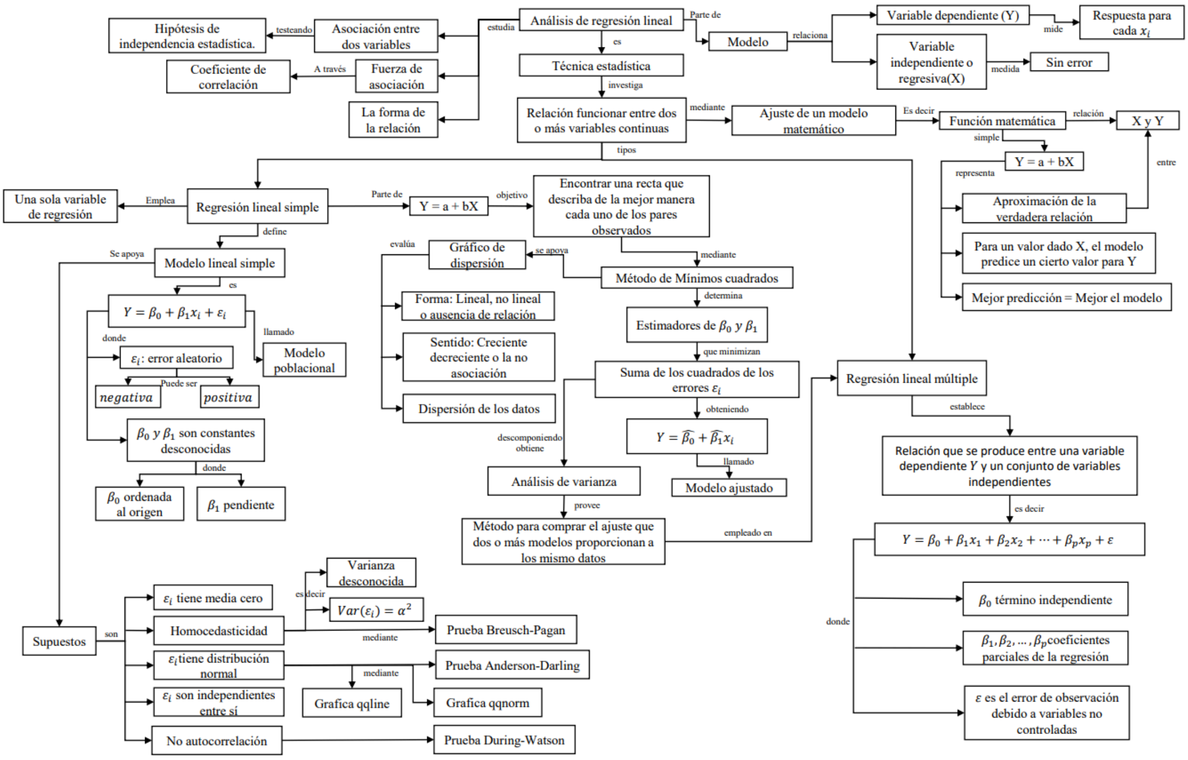
1.5.3 Normalidad

“Para facilitar la estimación por intervalo del modelo de regresión es exigible la normalidad de la distribución de los errores. Por ello, se emplea un análisis gráfico como lo es el gráfico P-P normal” (Thompson, 2015, como se citó en Giler, 2021, pág. 35)

1.5.3.1 Anderson-Darling

El estadístico de bondad de ajuste de Anderson-Darling -AD- mide el área entre la línea ajustada -basada en la distribución normal- y la función de distribución empírica -que se basa en los puntos de los datos-. El estadístico de Anderson-Darling es una distancia elevada al cuadrado que tiene mayor ponderación en las colas de la distribución (Jensen & Alexander, 2016, como se citó en Flores, C. & Flores, K., 2021, pág. 87).

2 Mapa Conceptual



3 Resumen

Tomando lo dicho por el CONEVAL(2010)"La persistente desigualdad económica y social de nuestro país ha ocasionado una creciente preocupación sobre los retos que enfrenta el Estado mexicano para generar alternativas de inclusión y movilidad social entre distintos sectores de la población, en especial aquellos que, a causa de la insuficiencia de recursos económicos o por falta de ejercicio de derechos sociales, carecen de los elementos indispensables para mantener una vida activa y participativa."(p. 4), por tanto, se destaca que el tema de estudio toma gran relevancia permitiendo analizar con rigurosidad la pobreza en México durante el periodo 2014-2016, empleando herramientas estadísticas partiendo de datos concretos. Para poder desarrollar esta práctica se implementó lenguaje de programación R para tener un análisis más detallado, permitiendo documentar los resultados obtenidos paso a paso, mostrando el flujo completo de procesamiento de los datos por medios de *scripts*.

La información derivada de estos análisis no solo contribuirá a una comprensión más profunda de la pobreza en el país, sino que también respaldará la toma de decisiones informadas en relación con los ámbitos políticos, sociales y económicos.

Para hacer el análisis se utilizó la regresión lineal simple con dos variables, partiendo de un análisis de varianza (ANOVA), gráfica de dispersión y residuales, y con apoyo de los supuestos teóricos.

3.1Palabras clave

Regresión lineal Simple, ANOVA, supuestos teóricos.

4 Introducción

La pobreza es uno de los mayores problemas que perjudica a México desde varias décadas así lo describe Jonathan (Ortiz.et,2013)" Es necesario conocer sobre las personas denominadas como pobres: la salud y los servicios médicos a los que pudieran tener acceso y cómo tienen acceso a ellos; conocer su educación, la disponibilidad y la calidad de ésta, debido a que cursar educación básica en regiones marginadas en los estados más pobres del país no implica el mismo aprovechamiento y calidad que en una zona urbana con todos los servicios necesarios "(La pobreza en México, un análisis con enfoque multidimensional, pp.4) así pues, en esta práctica se mostrará cómo podemos analizar e interpretar los datos proporcionados durante el transcurso de los años 2014-2016 en el periodo presidencial de Enrique Peña Nieto en el cual veremos si los datos proporcionados muestran un aumento a diferencia de otros años.

No hay un factor en específico que determine un crecimiento de la pobreza puede ser por varias causas en las cuales se encuentran la desigualdad económica, la falta de empleo formal y bien remunerado, la discriminación, especialmente hacia poblaciones indígenas y afrodescendientes, y la inseguridad alimentaria. Además, factores como la corrupción, la violencia relacionada con el crimen organizado y los desastres naturales también contribuyen a la persistencia de la pobreza en algunas regiones del país.

5 Desarrollo

El desarrollo del siguiente trabajo parte de los datos sobre pobreza en México del sitio del CONEVAL, donde se pretende comprobar la existencia de una relación aproximadamente lineal entre los datos de 2014 y 2016.
Definimos así

$$Y = variable\ dependiente = Medición\ de\ la\ pobreza\ en\ el\ año\ 2016$$
$$X = variable\ independiente = Medición\ de\ la\ pobreza\ en\ el\ año\ 2014$$

Para implementar esta práctica llevamos a cabo la creación de código en R donde obtuvimos lo siguiente:

5.1 Carga de datos en R

Guardamos los datos en un data frame donde se colocan los datos del año 2014 y del 2016

```r
entidades <- c("Aguascalientes", "Baja California", "Baja California Sur", "Campeche", "Coahuila",
               "Colima", "Chiapas", "Chihuahua", "Distrito Federal", "Durango", "Guanajuato",
               "Guerrero", "Hidalgo", "Jalisco", "México", "Michoacán", "Morelos", "Nayarit",
               "Nuevo León", "Oaxaca", "Puebla", "Querétaro", "Quintana Roo", "San Luis Potosí",
               "Sinaloa", "Sonora", "Tabasco", "Tamaulipas", "Tlaxcala", "Veracruz", "Yucatán",
               "Zacatecas")

# Crear un vector para 2014
x2014 <- c(442.9, 984.9, 226.2, 391.0, 885.8, 244.9, 3961.0, 1265.5, 2502.5, 761.2, 2683.3, 2315.4,
           1547.8, 2780.2, 8269.9, 2708.6, 993.7, 488.8, 1022.7, 2662.7, 3958.8, 675.7, 553.0, 1338.1,
           1167.1, 852.1, 1169.8, 1330.7, 745.1, 4634.2, 957.9, 819.8)

# Crear un vector para 2016
x2016 <- c(369.7, 789.1, 175.6, 405.0, 745.9, 248.7, 4114.0, 1150.0, 2434.4, 643.3, 2489.7, 2314.7,
           1478.8, 2560.6, 8230.2, 2565.9, 965.9, 470.1, 737.8, 2847.3, 3728.2, 635.7, 471.0, 1267.7,
           929.7, 831.4, 1228.1, 1156.2, 701.8, 5049.5, 901.9, 780.3)

# Graficar los datos
plot(x2014, x2016, main = "Gráfico de Dispersión", xlab = "X", ylab = "Y", col = "blue", pch = 16)
```

Ilustración 1. Carga de datos en R

5.2 Gráfica de Dispersión

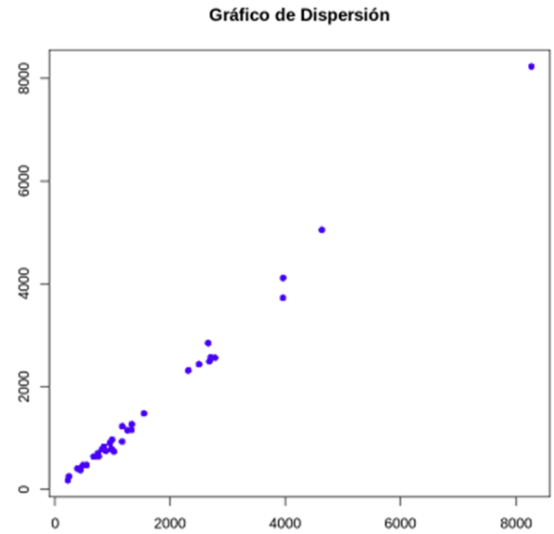


Ilustración 2. Gráfico de Dispersión

Usando la función plot, graficamos los puntos dados y obtuvimos lo siguiente: existe una asociación lineal positiva, significa que a medida que el valor de una variable X (2014), el valor de la otra variable Y también tiende a aumentar (2016).

5.3 Ajuste del modelo de regresión lineal

Obtendremos los datos siguientes para verificar si existe algún modelo

```
RLM<-lm(x2016~x2014)
summary(RLM)
```

```
Call:
lm(formula = x2016 ~ x2014)

Residuals:
    Min      1Q  Median      3Q     Max 
-213.52 -102.57    6.26   50.15  419.35 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -93.46671   34.49160   -2.71    0.011 *  
x2014         1.01930    0.01449   70.32   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 134 on 30 degrees of freedom
Multiple R-squared:  0.994,	Adjusted R-squared:  0.9938 
F-statistic:  4945 on 1 and 30 DF,  p-value: < 2.2e-16
```

Ilustración 3. Modelo

Se puede observar que R^2 muestra el porcentaje de variabilidad de los datos que el modelo está recogiendo, es decir, un 99% que representa una gran cantidad de variabilidad en los datos. Esto podría indicar que se tiene un buen modelo de regresión una vez sabiendo que si hay modelo.
Para la H_0 para $b_1 = 0$, que es el más importante, y para H_a para $b_1 \neq 0$
Como el p-value es de 0.011 para el intercepto no requiere de un α, entonces se rechaza la hipótesis nula, eso quiere decir que el intercepto b_1 es significativo y se incluye en el modelo.

Una vez que se determina la prueba de hipótesis, se analiza que tan buen ajuste tiene a mis datos, que apoyado de la *Ilustración 3*, se observa que es un buen ajuste ya que era prácticamente una línea recta.

La Prueba F y el Test F, solo se toman en cuenta para una regresión múltiple, porque indica en conjunto que tanto los parámetros son significativos.
Para una regresión lineal simple, es el mismo valor de la F, ya que indica la significancia de **todos** los parámetros (la prueba F es equivalente a la prueba T cuando se habla de un modelo de regresión lineal simple).
El valor de p-value para b_1 es menor a 2.2e-16, también se rechaza la hipótesis nula y se incluye en el modelo. Por lo tanto, el modelo queda de la siguiente manera:

$$Y =- 93.46671 + 1.01930X$$

Significado de los coeficientes estimados: En teoría, el valor de la ordenada al origen, es decir, -93.46671 es la medición de pobreza en miles de personas en el año 2016.
La pendiente de la recta es 1.01930, lo cual implica que para el año 2016 la medición de pobreza en miles de personas aumenta el 1.01930 en promedio.

Para hacer una verificación de la existencia del modelo, se emplean los siguientes supuestos teóricos: mediante gráficas y pruebas de hipótesis, partiendo del análisis de los residuos.

5.4 Límites de confianza

Utilizando la función *confit* podemos obtener los límites de confianza del 95% para b_1 , para los miles de personas en el año de 2014 y obteniendo lo siguiente:

```
confianza_cof<-confint(RLM,level = 0.95)
print(confianza_cof)
```

```
                   2.5 %     97.5 %
(Intercept) -163.9079591 -23.025458
x2014          0.9896931   1.048898
```

Ilustración 4. Límites de Confianza

Entonces la función queda de la siguiente manera:

$$Y = - 163.9079591 + 0.9896931X$$

Y observamos que la hipótesis nula será rechazada para todo b fijo que no quede contenido en el intervalo $[0.9896931, 1.048898]$ en base a la muestra observada (esto es lo que se conoce como dualidad entre intervalos de confianza y tests)

5.5 Análisis de Varianza

Utilizando la función en R "anova" podemos construir la tabla de análisis de varianza y para los datos de Y_1 queda de la siguiente manera:

```
resultado_anova <- aov(x2016 ~ x2014)
summary(resultado_anova)
```

```
            Df   Sum Sq  Mean Sq F value Pr(>F)    
x2014        1 88820008 88820008    4945 <2e-16 ***
Residuals   30   538837    17961                   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ilustración 5. ANOVA

EN ESTE CASO EL $p - value$ ES MENOR QUE α POR LO QUE SE RECHAZA LA HIPÓTESIS NULA

Fuente	**df**	**SS**	**MS**
Debido a la regresión	1	88820008	88820008
Sobre la regresión (residuales)	30	538837	17961
Total	31		

5.6 Residuos

Para chequear que los supuestos del modelo lineal son apropiados para un conjunto de datos, suelen hacerse una serie de gráficos.

Verificar que los residuos: Primero se realiza la comprobación de manera gráfica usando el siguiente código en R

```r
# Obtener los residuos
residuos <- residuals(RLM)

# Gráficos
par(mfrow=c(2,2))  # Para mostrar varios gráficos en una cuadrícula

# Residuos vs. Valores Ajustados
plot(RLM$fitted.values, residuos, main="Residuos vs. Valores Ajustados", xlab="Valores Ajustados", ylab="Residuos")
abline(h=0, col="red")  # Agregar línea en y=0

# Histograma de Residuos
hist(residuos, main="Histograma de Residuos", xlab="Residuos", ylab="Frecuencia", col="lightblue")

# Gráfico Q-Q
qqnorm(residuos)
qqline(residuos)

# Gráfico de Autocorrelación de Residuos
acf(residuos, main="Autocorrelación de Residuos")
```

Ilustración 6. Test residuales

5.7 Gráficas

Y obtenemos las siguientes gráficas:

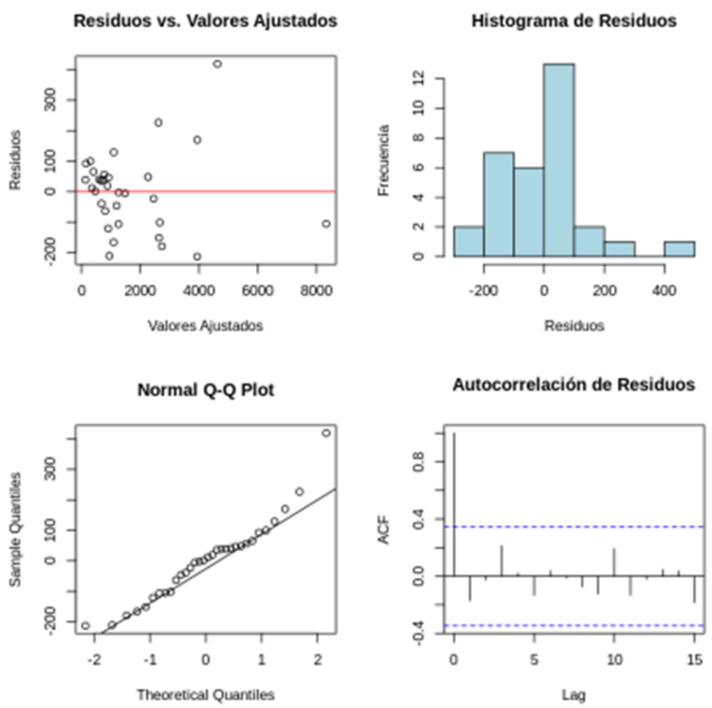


Ilustración 7. Gráficas de los Test

En la gráfica de Residuos vs Valores Ajustados podemos notar que los residuos no están dispersos, por lo tanto tienen heterocedasticidad.

Para el caso del histograma de residuos se muestra que la distribución es sesgada o no normal por lo que podemos decir que no es normal la gráfica.

Por otra parte la tabla Q-Q muestra cómo se distribuyen los datos alrededor de la línea la cual si se muestran casi como una línea recta.

Y finalmente, en la gráfica de autocorrelación podemos notar que los residuos se correlacionan ya que se encuentran dentro de la líneas horizontales

5.8 Supuesto: Media cero

Se requiere verificar que los residuos tengan media cero

Utilizando R obtenemos el siguiente código:

```
> #t-test
> t.test(residuales, mu=0)

        One Sample t-test

data:  residuales
t = 2.9847e-17, df = 31, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -47.53345  47.53345
sample estimates:
   mean of x
6.956241e-16
```

Ilustración 8. T-test

Se obtiene un $p - value = 1$, es decir, no se rechaza la hipótesis nula y la media es cercana a 0. Por tanto existe modelo

5.9 Supuesto: Homocedasticidad

Tengan varianza constante: homocedasticidad (cuando no se cumple hay heterocedasticidad)
Tendiendo el código en R:

```
restultlevel<- bptest(RLM)
print(restultlevel)
```

```
	studentized Breusch-Pagan test

data:  RLM
BP = 6.3333, df = 1, p-value = 0.01185
```

Ilustración 9. Test Breusch-Pagan

En este caso si se rechaza la hipótesis nula, es decir, se tiene un $p - value = 0.01185$; por lo tanto existe heterocedasticidad.

5.10 Supuesto: No correlación

No correlación
En base de este código obtenido en R:

```
# Realizar la prueba de hipótesis de Durbin-Watson
resultado_dw <- dwtest(RLM)

# Mostrar el resultado de la prueba
print(resultado_dw)
```

```
	Durbin-Watson test

data:  RLM
DW = 2.3366, p-value = 0.8253
alternative hypothesis: true autocorrelation is greater than 0
```

Ilustración 10. Test Durbin-Wattson

Se muestra un $p - value = 0.8253$, por lo que no se rechaza la prueba de hipótesis y se concluye que hay autocorrelación en los residuos.

<u>5.11 Supuestos: Distribución normal</u>

Se distribuyen normalmente

Se muestra el código en R:

```r
residuos3 <- residuals(RLM)

# Realizar la prueba de Anderson-Darling
resultado_ad <- ad.test(residuos3)

# Mostrar el resultado de la prueba
print(resultado_ad)
```

```
        Anderson-Darling normality test

data:  residuos3
A = 0.4773, p-value = 0.2214
```

Ilustración 11. Test Anderson-Darling

Se obtiene un $p - value = 0.2214$, por lo tanto se acepta la hipótesis nula, eso quiere decir que se distribuye normalmente.

Los resultados obtenidos durante las pruebas fueron que: la única que rechazó la hipótesis nula fue la *Prueba de Breusch-Pagan*, dando como resultado la existencia de heterocedasticidad, es decir, la existencia de una varianza no constante de un modelo lineal.

7 Referencias

- De la Guerra, J. P. (2015). Las siete herramientas de la calidad. CORE. https://core.ac.uk/download/pdf/356676474.pdf
- Elousa, P. (2011) . Introducción al entorno R. Bilbao: Euskal Herriko Unibertsitateko Argitalpen Zerbitzua/Servicio Editorial de la Universidad del País Vasco.
- Novales. A. (20 de septiembre de 2010). Análisis de Regresión. Universidad Complutense. https://www.ucm.es/data/cont/docs/518-2013-11-13-Analisis%20de%20Regresion.pdf
- Carrillo, M. C. (2011). Regresión Lineal Simple. Departamento de Estadística, Análisis Matemático y Optimización. http://eio.usc.es/eipc1/BASE/BASEMASTER/FORMULARIOS-PHP-DPTO/MATERIALES/Mat_50140116_Regr_%20simple_2011_12.pdf
- Consejo Nacional de Evaluación de la Política de Desarrollo Social.(2010). La pobreza por ingresos en México, CONEVAL. https://www.coneval.org.mx/rw/resource/coneval/info_public/pdf_publicaciones/pobreza_ingresos_mexico_web.pdf
- Ortiz Galindo. (2013).La pobreza en México, un análisis con enfoque multidimensional https://www.redalyc.org/pdf/413/41331033010.pdf
- Montgomery, D., Peck, E., Vining, G. (2002). Introducción al Análisis de Regresión Lineal. Tercera edición. Grupo Patria Cultural S. A. de C. V.
- Bouza, C. (2018). Modelos de Regresión y sus Aplicaciones. University of Havana. at: https://www.researchgate.net/publication/323227561
- Santibáñez, J. (2018). Regresión lineal múltiple. Departamento de Probabilidad y Estadística. iimas.http://sigma.iimas.unam.mx/jsantibanez/Cursos/Regresion/2018_2/notas/3_rlm.pdf
- Dagnino, J. (2014). Análisis de Varianza. Bioestadística y Epidemiología. Revista Chilena de Anestesia. pág.1.
- Giler, Y. (2021). El impuesto al valor agregado y a los consumos especiales sobre la recaudación tributaria en el Ecuador: estimación y evaluación en el periodo 2008 - 2019. Tesis en opción al título de Magister. Universidad Tecnológica Empresarial de Guayaquil.
- Arce, R. (2001). Conceptos Básicos Sobre la Heterocedasticidad en el Modelo Básico de Regresión Lineal Tratamiento con E-Views. Universidad Autónoma de Madrid. https://tabarefernandez.tripod.com/dearce.pdf
- Alonso, M. (1994). Análisis de Residuales en Regresión. Trabajo para obtener un diploma en: Especialidad en Métodos Estadísticos. Universidad Veracruzana.
- Flores, C., Flores, K. (2021). Pruebas Para Comprobar la Normalidad de Datos en Procesos Productivos: Anderson-Darling, Ryan-Joiner, Shapiro-Wilk y Kolmogórov-Smirnov. Societas: Revistas de Ciencias Sociales y Humanísticas. pág.87.