ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Instituto de Ciencias Matemáticas

Ingeniería en Estadística Informática

"SOFTWARE ESTADÍSTICO PARA REGRESIÓN LINEAL ERLA: VERIFICACIÓN DE CALIDAD DE MODELOS EN REGRESIÓN LINEAL"

INFORME DE LA MATERIA DE GRADUACIÓN

Previo a la obtención del Título de:

INGENIERÍA EN ESTADÍSTICA INFORMÁTICA

Presentado por:

JUAN CARLOS BUENAÑO CORDERO

CELIA ANNABELL DE LA CRUZ CEDEÑO

Guayaquil - Ecuador

2011

AGRADECIMIENTO

A Jehová, Dios, por estar siempre a nuestro lado y por habernos permitido llegar a la culminación de nuestra carrera para ser profesionales de bien y al servicio de la humanidad.

A nuestros padres y familia por su apoyo incondicional en todo momento.

DEDICATORIA

A nuestros padres, y demás familiares que de alguna u otra manera siempre estuvieron a nuestro lado brindándonos su apoyo, comprensión y amor en el transcurso de nuestra carrera.

A nuestros profesores, compañeros y a todos quienes hicieron posible el desarrollo de este informe de graduación.

TRIBUNAL GRADUACIÓN

M.Sc. Gaudencio Zurita Herrera

PROPFESOR DE LA MATERIA DE GRADUACIÓN

Ing. Dalton Noboa Macías

DELEGADO DEL ICM

DECLARACIÓN EXPRESA

"La responsabilidad del contenido de este Informe de Graduación, nos corresponde exclusivamente; y el patrimonio intelectual de la misma a la Escuela Superior Politécnica del Litoral".

Juan Carlos Buenaño Cordero

Celia Annabell De La Cruz Cedeño

RESUMEN

En un análisis de Regresión Lineal, existen varios supuestos o premisas que deben ser considerados al momento de determinar la validez de un modelo, puesto que el no cumplimiento de alguno de estos supuestos podría conducirnos a modelos inestables, de ser así, un valor alto del estadístico R^2 o R^2 ajustado no garantiza que el modelo se ajuste bien a los datos.

Entre los principales supuestos que se realizan están: La distribución del error normal con media 0 y varianza σ^2 constante, la no correlación de los errores y la relación no lineal entre las variables de explicación. Por otra parte los estadísticos de resumen como t, F o R^2, los Coeficientes de Regresión y la Media Cuadrática del Error son sensibles a la presencia de valores aberrantes o atípicos.

Este proyecto presenta los métodos más usados, para verificar el cumplimiento de estos supuestos, detectar la presencia de valores aberrantes y puntos de influencia a través de su implementación y correspondiente validación en un software estadístico especializado en la técnica de Regresión Lineal llamado ERLA (Estadística de Regresión Lineal Avanzada), desarrollado por estudiantes del Instituto de Ciencias Matemáticas de la ESPOL.

ÍNDICE GENERAL

ABREVIATURAS

ANOVA	Análisis de Varianza
ERLA	Estadística de Regresión Lineal Avanzada
FIV	Factor de Inflación de la Varianza
MCR	MATLAB Component Runtime
SCE	Suma Cuadrática del Error
SCR	Suma Cuadrática de la Regresión
SCS	Sumas de Cuadrados Secuenciales
SCT	Suma Cuadrática Total

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

INTRODUCCIÓN

La Regresión Lineal es una de las técnicas estadísticas más usadas en áreas como los negocios y la medicina, ya que permite estudiar la relación lineal entre dos o más variables de tipo cuantitativo. El correcto uso de esta técnica exige el cumplimiento de ciertos supuestos que deben ser analizados cuidadosamente, ya que las violaciones a estos supuestos podrían conducirnos a modelos inestables y con ello a conclusiones equivocadas. Para verificar el cumplimiento de los supuestos que se realizan existen varios métodos que serán presentados en capítulos posteriores, los cuales han sido implementados en el software estadístico de regresión ERLA.

ERLA es un software estadístico, especializado en la técnica de Regresión Lineal, desarrollado por estudiantes de Ingeniería en Estadística Informática de la Escuela Superior Politécnica del Litoral. Mediante el uso del MCR (MATLAB Component Runtime) y una interfaz gráfica desarrollada en .NET, el software ERLA permite al usuario la ejecución de funciones programadas en lenguaje MATLAB, dichas funciones son compiladas y referenciadas para que sean accesibles a través de ERLA. Esta aplicación está dirigida a estudiantes y profesionales de distintas áreas que hacen uso de la estadística.

Junto a ERLA se han escrito dos documentos: El Reporte Técnico y el Manual de Usuario. El Manual de Usuario ofrece al lector la información necesaria para empezar a utilizar ERLA; detalla datos técnicos, instalación y cómo utilizar cada una de sus funcionalidades. El Reporte Técnico proporciona los fundamentos teóricos que hicieron posible desarrollar el Software. En el Capítulo 1 se presentan los fundamentos teóricos que sustentan la técnica de Regresión Lineal. El Capítulo 2 trata los métodos y técnicas más utilizados para evaluar la calidad del modelo de regresión; esto es: Verificar supuestos sobre el error, detectar valores aberrantes, puntos de influencia, problemas de enmascaramiento y multicolinealidad. En el Capítulo 3 se presenta toda la información sobre los recursos utilizados para el desarrollo del Software. Finalmente, en el Capítulo 4 se presenta la validación estadística del Software en base a simulaciones. Para la comprensión del presente documento, se supone que el lector tiene conocimientos básicos de Computación, Estadística, Cálculo y Álgebra lineal.

CAPÍTULO 1

1. REGRESIÓN LINEAL

1.1. Introducción

En estadística, estamos constantemente trabajando con un gran número de variables, y es razonable pensar que pueda existir algún tipo de relación funcional, que explique a una de ellas en términos de una o varias de las variables restantes. La técnica estadística que permite resolver el problema planteado, cuando las variables involucradas son cuantitativas es la Regresión Lineal, dicho de otra forma es la técnica estadística que se utiliza para explorar y cuantificar la relación entre una variable llamada variable de respuesta, explicada o pronosticada Y, y una o más variables predictoras o variables de explicación $X_1, X_2, ..., X_{p-1}$ siempre y cuando éstas sean cuantitativas.

En éste capítulo se desarrollan los fundamentos teóricos básicos de la técnica de Regresión Lineal distribuidos en dos partes. La primera parte corresponde al modelo de Regresión Lineal Simple en la cual se explica detenidamente cómo obtener el modelo de regresión, y; la segunda parte corresponde al modelo de Regresión Lineal Múltiple, para la cual se utilizará la notación matricial.

1.2. Regresión Lineal Simple

Supongamos que queremos analizar la relación funcional entre dos variables cuantitativas denominadas X y Y, pero que experimentalmente se pueden fijar n valores de X y leer luego los n valores que corresponden a Y. Lo primero que podemos hacer, es realizar un gráfico de dispersión X vs. Y, y; si el gráfico se asemeja a una línea recta estaríamos tentados a proponer la siguiente relación entre X y Y:

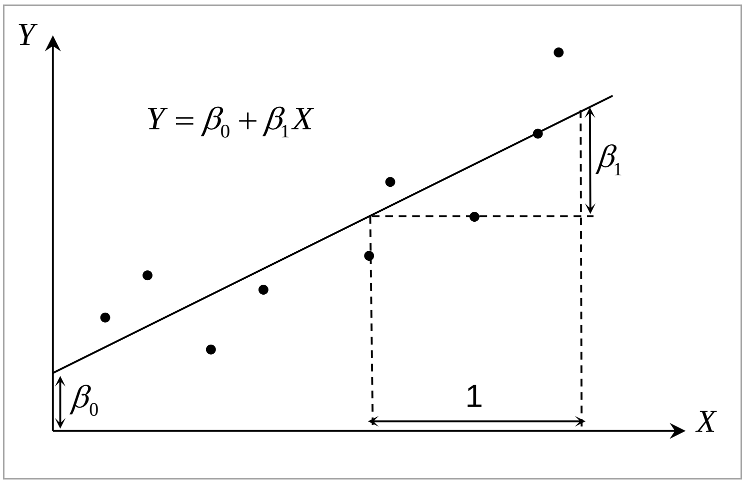
$$Y = \beta_0 + \beta_1 X \qquad \textbf{\textit{(1.1)}}$$

Esto es, una recta cuya pendiente es β_1 y cuya intersección con el eje vertical Y es el punto de coordenadas $(0, \beta_0)$. Véase *Figura 1.1*. Si esto es cierto, también es cierto que cada valor observado de Y, es decir y_i no siempre determina un punto que pertenece a la recta, esto es porque al efectuar la medida de Y cuando X es fijado x_i se comete un error aleatorio ε_i de tal manera que el valor y_i de Y está dado por la siguiente relación funcional:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad \textbf{\textit{(1.2)}}$$

ε_i es una variable aleatoria cuyo valor se ve influenciado por el observador, instrumento de medida o por el medio ambiente, mientras que X es una variable observable fijada por el investigador.

Figura 1.1
Regresión Lineal
"Relación lineal entre X y Y"



Fuente: ZURITA, G. [8]

El valor y_i dependerá del valor x_i, esto hace posible escribir el modelo condicional el cual trabaja bajo los siguientes supuestos:

$$E(y_i \mid X = x_i) = \beta_0 + \beta_1 x_i$$
$$E(\varepsilon_i) = 0;\ Var(\varepsilon_i) = \sigma^2 \text{ y; } Cov(\varepsilon_i, \varepsilon_j) = 0 \text{ para i} \neq \text{j} \qquad \textbf{\textit{(1.3)}}$$

$E(y_i \mid X = x_i)$ es llamado Función de Respuesta o Parte Sistemática o Determinística del modelo. Los valores β_0, β_1 y σ^2 son constantes desconocidas pero estadísticamente estimables.

TEOREMA DE GAUSS-MARKOV

De acuerdo con la expresión *(1.2)*, el valor observado y_i de Y es una variable aleatoria debido a la presencia del error. Dónde:

$$E(y_i) = E(\beta_0 + \beta_1 x_i + \varepsilon_i)$$

$$E(y_i) = E(\beta_0) + E(\beta_1 x_i) + E(\varepsilon_i)$$

$$E(y_i) = \beta_0 + \beta_1 x_i \qquad \textbf{(1.4)}$$

Se supone que $E(\varepsilon_i) = 0$ y el valor esperado de una constante es la misma constante. Por otra parte:

$$Var(y_i) = Var(\beta_0 + \beta_1 x_i + \varepsilon_i)$$

$$Var(y_i) = [(\beta_0 + \beta_1 x_i + \varepsilon_i) - (\beta_0 + \beta_1 x_i)]^2$$

$$Var(y_i) = [\varepsilon_i - 0]^2$$

$$Var(y_i) = [\varepsilon_i - E(\varepsilon_i)]^2$$

$$Var(y_i) = Var(\varepsilon_i)$$

$$Var(y_i) = \sigma^2 \qquad \textbf{(1.5)}$$

Si se supone la distribución de ε_i normal, entonces:

$$\varepsilon_i \sim N(0, \sigma^2) \qquad \textbf{(1.6)}$$

Esto implica que:

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2) \qquad \textbf{\textit{(1.7)}}$$

1.2.1. Estimación de los parámetros de Regresión Lineal Simple por Mínimos Cuadrados

Los valores β_0, β_1 y σ^2 son constantes desconocidas que deben ser estimadas estadísticamente. Los estimadores de β_0, β_1 serán llamados b_0 y b_1, y para encontrarlos utilizaremos el criterio de Mínimos cuadrados.

Un estimador de y_i denotado por $\hat{y}_i$ será:

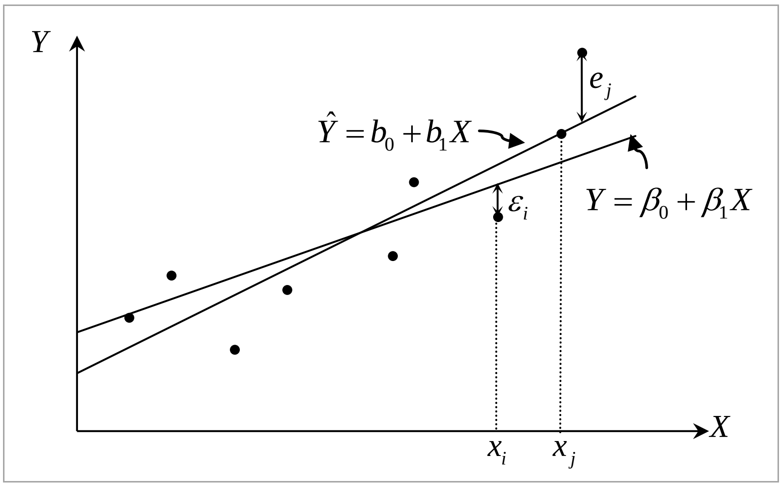
$$\hat{y}_i = b_0 + b_1 x_i \qquad \textbf{\textit{(1.8)}}$$

y la diferencia $e_i = (y_i - \hat{y}_i)$ estima el valor de ε_i cuando se lee y_i; en otras palabras:

$$\hat{\varepsilon}_i = e_i = y_i - \hat{y}_i \qquad \textbf{\textit{(1.9)}}$$

Considerando a Y como la variable a ser explicada y X, como la variable de explicación, el método de Mínimos Cuadrados busca la recta que minimice la Suma Cuadrática de los Errores. Véase *Figura 1.2*.

Figura 1.2
Regresión Lineal
"Criterio de Mínimos Cuadrados"



Los pasos a seguir para encontrar los estimadores de mínimos cuadrados en el modelo de Regresión Lineal Simple se detallan a continuación:

- Cálculo del error para cada observación y_i:

$$\varepsilon_i = y_i - E(y_i) = y_i - (\beta_0 + \beta_1 x_i) \qquad \textbf{(1.10)}$$

- Se define a Q como la suma de los errores al cuadrado o suma cuadrática de los errores:

$$Q = \sum_{i=1}^{n} \left[y_i - \left(\beta_0 + \beta_1 x_i \right) \right]^2$$
$$Q = \sum_{i=1}^{n} \varepsilon_i^2 \qquad \textbf{(1.11)}$$

- El método de Mínimos Cuadrados busca los valores de β_0 y β_1 que minimicen la suma cuadrática de los errores, es decir los valores de β_0 y β_1 que minimicen a Q. Para encontrar un mínimo relativo, derivamos respecto a cada uno de los parámetros beta e igualamos a cero, obteniendo así; un par de ecuaciones conocidas como Ecuaciones Normales:

$$\frac{\delta Q}{\delta \beta_0} = \frac{\delta Q}{\delta \beta_1} = 0$$

$$\frac{\delta Q}{\delta \beta_0} = -2\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)$$

$$\frac{\delta Q}{\delta \beta_1} = -2\sum_{i=1}^{n} x_i\left(y_i - \beta_0 - \beta_1 x_i\right)$$

$$\sum_{i=1}^{n} y_i - \beta_0 n - \beta_1 \sum_{i=1}^{n} x_i = 0$$

$$\sum_{i=1}^{n} x_i y_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0$$

- Reemplazamos β_0 y β_1 por sus estimadores y lo llevamos a la forma matricial:

$$\sum_{i=1}^{n} y_i = b_0 n + b_1 \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i y_i = b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2$$

$$\begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix} = \begin{pmatrix} b_0 n + b_1 \sum_{i=1}^{n} x_i \\ b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 \end{pmatrix}$$

$$\begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$$

$$\begin{pmatrix} y_1 + y_2 + ... + y_n \\ x_1 y_1 + x_2 y_2 + ... + x_n y_n \end{pmatrix} = \begin{pmatrix} 1+1+...+1 & x_1 + x_2 + ... + x_n \\ x_1 + x_2 + ... + x_n & x_1^2 + x_2^2 + ... + x_n^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & .. & 1 \\ x_1 & x_2 & .. & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ : \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & 1 & .. & 1 \\ x_1 & x_2 & .. & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ : & : \\ 1 & x_n \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$$

$$\mathbf{X}^T \mathbf{Y} = \mathbf{X}^T \mathbf{X} \mathbf{b}$$

Donde $\mathbf{b} \in R^2$, $\mathbf{Y} \in R^n$ y $\mathbf{X} \in M_{nx2}$, $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ es una matriz cuadrada y simétrica de dimensiones *pxp*. Si el rango de la matriz $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ es p, la matriz es invertible y por tanto:

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \qquad \textbf{\textit{(1.12)}}$$

Si la matriz $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ no es de rango p, entonces se usa la Inversa Generalizada. Véase [4] Zurita y Vera.

Otro parámetro del modelo es σ^2, que al mismo tiempo de ser la varianza de ε_i es también la de Y. Es natural estimar el valor de σ^2 con el mismo criterio que se utiliza para estimar la varianza de una población cualquiera considerando como denominador a n-2, ya que se pierden dos grados de libertad al estimar β_0 y β_1.

$$\hat{\sigma}^2 = s^2 = \sum_{i=1}^{n} \frac{\left(e_i - \overline{e}\right)^2}{n-2} = \sum_{i=1}^{n} \frac{e_i^2}{n-2} = \frac{SCE}{n-2} \qquad \textbf{\textit{(1.13)}}$$

Esta expresión es también conocida como Media Cuadrática del Error, la cual será discutida en la siguiente sección y es el estimador de mínimos cuadrados de σ^2.

1.2.2. Tabla de Análisis de Varianza, Coeficiente de Determinación y Determinación Ajustado Para el Modelo de Regresión Lineal Simple

Las sumas cuadráticas miden la dispersión de un grupo de observaciones, y no es una excepción la variable a ser explicada Y en un modelo de regresión. La expresión:

$$\sum_{i=1}^{n} \left(y_i - \overline{y}_i\right)^2 \qquad \textbf{\textit{(1.14)}}$$

Es llamada Suma Cuadrática Total (SCT), ya que ésta no se ve influenciada por el modelo de regresión lineal escogido, y su cálculo depende solamente de los valores observados de la variable Y.

Por otra parte $y_i - \overline{y}$ puede descomponerse de la siguiente manera:

$$y_i - \overline{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \overline{y})$$

Donde $\hat{y}_i$ es el valor estimado de y_i dado que $X = x_i$. Esto hace posible que la SCT sea expresada como:

$$\sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}_i\right)^2 = \sum_{i=1}^{n}\left[(y_i - \hat{y}_i) + (\hat{y}_i - \overline{y})\right]^2$$

$$\sum_{i=1}^{n}\left[\,(\hat{y}_i - \overline{y})^2 + (y_i - \hat{y}_i)^2 + 2(\hat{y}_i - \overline{y})(y_i - \hat{y}_i)\ \right]$$

Nótese que,

$$\sum_{i=1}^{n} 2(\hat{y}_i - \overline{y})(y_i - \hat{y}_i) = 0$$

Por lo que:

$$\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$SCT = SCR + SCE \qquad \textbf{(1.15)}$$

Entonces la variabilidad total definida por la SCT se supone explicada por dos fuentes que son la Regresión y el Error que también son sumas cuadráticas. La Suma Cuadrática de la

Regresión (SCR) y la Suma Cuadrática del Error (SCE) por su parte sí dependen del modelo escogido debido a la presencia de $\hat{y}_i$. Es importante destacar que la SCE mide la variabilidad de los valores observados alrededor de la recta cuya ecuación es $Y = b_0 + b_1 X$.

Lo deseable en un modelo de regresión lineal, es que la SCE sea lo más pequeña posible, ya que en la situación ideal la SCR=SCT. Esto hace posible definir una medida de la calidad del modelo que estamos utilizando, y esta medida será:

$$R^2 = \frac{SCR}{SCT} \qquad \textbf{\textit{(1.16)}}$$

A esta medida se la conoce como Coeficiente de Determinación R^2. Su valor estará en el intervalo cero y uno. Así, para que el modelo sea considerado “bueno” R^2 debe ser un valor lo más cercano posible a 1. Si multiplicamos el R^2 por 100%, obtenemos la Potencia de Explicación del Modelo, que es una medida del porcentaje de variabilidad de la variable de respuesta que es explicada por las variables de explicación.

Además, existe también el Coeficiente de Determinación Ajustado, el cual incorpora una corrección al R^2 en función del número de variables de explicación y el tamaño de la muestra:

$$R^2 ajustado = 1 - \frac{\left(1-R^2\right)\left(n-1\right)}{n-p} \qquad \textbf{\textit{(1.17)}}$$

Donde n, es el número de observaciones en la muestra y p es el número de betas a estimar en el modelo, en este caso $p=2$. El R^2 *ajustado* es una proporción del porcentaje de variabilidad de la variable de respuesta explicada por las variables de explicación, si se introduce una variable adicional al modelo.

La tabla de Análisis de Varianza (ANOVA) es utilizada en regresión para analizar estadísticamente la validez del modelo. Es una tabla compuesta por cuatro filas y cinco columnas. Véase *Tabla* 1.1

Tabla 1.1
Regresión Lineal
"Tabla de Análisis de Varianza (ANOVA)"

FUENTES DE VARIACIÓN	***GRADOS DE LIBERTAD***	***SUMAS CUADRÁTICAS***	***MEDIAS CUADRÁTUCAS***	***F***
REGRESIÓN	p-1	$\sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}_i\right)^2$	$\frac{SCR}{p-1}$	$\frac{MCR}{MCE}$
ERROR	n-p	$\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$	$\frac{SCE}{n-p}$	
TOTAL	n-1	$\sum_{i=1}^{n}\left(y_i - \overline{y}_i\right)^2$		

La tabla ANOVA sintetiza algunos resultados presentados anteriormente, pero también presenta algunos resultados nuevos. A cada fuente de variación se asocia un número entero que son sus “grados de libertad”, los cuales permiten calcular las Medias Cuadráticas y un estadístico de prueba F.

Dado el modelo de Regresión Lineal Simple $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, se desearía que el valor de la pendiente no sea cero, por lo que se postula el contraste de hipótesis:

$$H_0 : \beta_1 = 0 \text{ vs. } H_1 : \beta_1 \neq 0$$

Aplicando el Teorema de Cochran, SCR/σ^2 es una variable Aleatoria con distribución Ji-Cuadrado con *(p-1)* grados de libertad, mientras que SCE/σ^2 es una variable aleatoria con distribución Ji-Cuadrado con *(n-p)* grados de libertad. Esto permite afirmar que:

$$F = \frac{MCR}{MCT} = \frac{SCR/(p-1)}{SCT/(n-p)} \qquad \textbf{(1.18)}$$

Tiene una distribución F con *(p-1)* grado de libertad en el numerador y *n-p* grados de libertad en el denominador. Esto bajo el supuesto de que SCE y SCR son estocásticamente independientes. Por tanto:

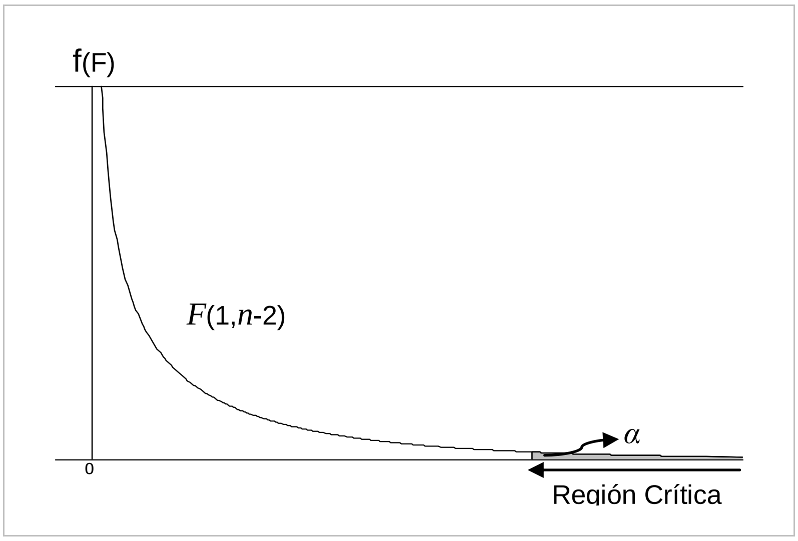
Con (1-α)100% de confianza se rechaza H_0 en favor de H_1 si:

$$F = \frac{MCR}{MCT} > F_{(\alpha, p-1, n-p)} \qquad \textbf{(1.19)}$$

Donde $F_{(\alpha, p-1, n-p)}$ es el percentil *(1-α)* de la distribución F con *(p-1)* grado de libertad en el numerador y n-p grados de libertad en el denominador. Véase *Figura 1.3*. En este caso, $p=2$ y por tanto la expresión *(1.19)* quedaría:

$$F = \frac{MCR}{MCT} > F_{(\alpha, 1, n-2)}$$

Figura 1.3
Regresión Lineal
"Distribución F(1,n-2)"



1.3. Regresión Lineal Múltiple

Para el caso en el que existen dos o más variables de explicación, el modelo es denominado modelo de Regresión Lineal Múltiple. En la regresión lineal múltiple es conveniente utilizar la notación matricial, ya que esta permite formular resultados generales en forma compacta y usar con gran ventaja muchos de los resultados de la teoría de matrices. Para ello veamos el siguiente ejemplo:

Se tienen las siguientes variables:

Y	15	12	16	13	14
X₁	20	13	18	14	16
X₂	14	15	18	13	11

Suponga que la variable Y es explicada por X_1 y X_2, es decir que el modelo postulado es:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \qquad \textbf{\textit{(1.20)}}$$

Por lo tanto:

$$15 = \beta_0 + \beta_1 20 + \beta_2 14 + \varepsilon_i$$

$$12 = \beta_0 + \beta_1 13 + \beta_2 15 + \varepsilon_i$$

$$16 = \beta_0 + \beta_1 18 + \beta_2 18 + \varepsilon_i$$

$$13 = \beta_0 + \beta_1 14 + \beta_2 13 + \varepsilon_i$$

$$14 = \beta_0 + \beta_1 16 + \beta_2 11 + \varepsilon_i$$

Lo cual en notación matricial será:

$$\begin{pmatrix} 15 \\ 12 \\ 16 \\ 13 \\ 14 \end{pmatrix} = \begin{pmatrix} 1 & 20 & 14 \\ 1 & 13 & 15 \\ 1 & 18 & 18 \\ 1 & 14 & 13 \\ 1 & 16 & 11 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_i \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}$$

Que sintéticamente puede representarse como:

$$\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} \text{ para } \mathbf{Y} \in R^5,\ \mathbf{X} \in M_{5x3},\ \boldsymbol{\beta} \in R^3 \text{ y } \boldsymbol{\varepsilon} \in R^5$$

De manera general, la expresión matemática para un modelo con (*p-1)* variables de explicación es:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_{(p-1)} x_{(p-1)i} + \varepsilon_i$$

Y expresado en forma matricial será:

$$\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} \qquad \textbf{\textit{(1.21)}}$$

$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$ es el vector de observaciones $\in R^n$

$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & .. & x_{1(p-1)} \\ 1 & x_{21} & x_{22} & .. & x_{2(p-1)} \\ 1 & x_{31} & x_{32} & .. & x_{3(p-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & .. & x_{n(p-1)} \end{pmatrix}$ es la Matriz de Diseño $\in M_{nxp}$

$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_p \end{pmatrix}$ es el vector de parámetros $\in R^p$

y,

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_p \end{pmatrix}$$ es el vector de errores $\in R^n$

El vector de observaciones $\mathbf{Y}$ dependerá de la matriz de diseño $\mathbf{X}$, es decir:

$$E(\mathbf{Y} \mid \mathbf{X}) = \mathbf{X}\boldsymbol{\beta} \qquad \textbf{\textit{(1.22)}}$$

Bajo los supuestos de:

$$E(\boldsymbol{\varepsilon}) = \mathbf{0} \text{ y } Var(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$$

Por tanto, el vector de observaciones $\mathbf{Y}$ en la expresión *(1.21)* es un vector aleatorio donde:

$$E(\mathbf{Y}) = E(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

$$E(\mathbf{Y}) = E(\mathbf{X}\boldsymbol{\beta}) + E(\boldsymbol{\varepsilon})$$

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} \qquad \textbf{\textit{(1.23)}}$$

Puesto que $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ y el valor esperado de una constante es la misma constante.

$$Var(\mathbf{Y}) = E\left[\left(\mathbf{Y} - E(\mathbf{Y})\right)\left(\mathbf{Y} - E(\mathbf{Y})\right)^T\right]$$

$$Var(\mathbf{Y}) = E\left[(\boldsymbol{\varepsilon})(\boldsymbol{\varepsilon})^T\right]$$

$$Var(\mathbf{Y}) = E\left[(\boldsymbol{\varepsilon} - \mathbf{0})(\boldsymbol{\varepsilon} - \mathbf{0})^T\right]$$

$$Var(\mathbf{Y}) = E\left[\left(\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon})\right)\left(\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon})\right)^T\right]$$

$$Var(\mathbf{Y}) = Var(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I} \qquad \textbf{\textit{(1.24)}}$$

Si se supone la distribución de $\boldsymbol{\varepsilon}$ normal, entonces:

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \qquad \textbf{\textit{(1.25)}}$$

Esto implica que:

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \qquad \textbf{\textit{(1.26)}}$$

1.3.1. Estimación de los parámetros de Regresión Lineal Múltiple por Mínimos Cuadrados

Cuando calculamos los estimadores de mínimos cuadrados de β_0 y β_1 para el caso de la Regresión Lineal Múltiple utilizando matrices, se obtuvo la expresión *(1.12)*:

$$\hat{\beta} = \mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

Se puede probar que ésta expresión es válida también para el caso de la Regresión Lineal Múltiple, cualquiera sea el número de observaciones n y el número de variables de explicación *(p-1):*

Partiendo de la expresión *(1.21)*, un estimador de $\mathbf{Y}$ será:

$$\hat{\mathbf{Y}} = \mathbf{Xb} \tag{1.27}$$

De la expresión *(1.22)* debemos determinar $\mathbf{b}$, pero la matriz de diseño $\mathbf{X}$ no es una matriz cuadrada, por lo que la multiplicamos por su transpuesta, la cual es cuadrada de dimensiones pxp y además es simétrica:

$$\mathbf{X}^T\hat{\mathbf{Y}} = \mathbf{X}^T\mathbf{Xb}$$
$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{Y}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Xb}$$
$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

Nótese que el conjunto solución es el mismo que se obtuvo en la regresión lineal simple, expresión *(1.12)*. La inversa de la matriz $\mathbf{X}^T\mathbf{X}$ existe siempre y cuando el rango de $\mathbf{X}$ sea p; caso contrario tendremos un problema de multicolinealidad el cual será discutido en el siguiente capítulo.

Si, en la expresión *(1.22)*, se reemplaza $\mathbf{b}$ por la expresión *(1.12)* se obtiene:

$$\hat{\mathbf{Y}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

O lo que es lo mismo:

$$\hat{\mathbf{Y}} = \mathbf{HY} \tag{1.28}$$

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \tag{1.29}$$

La matriz $\mathbf{H} \in \mathbf{S}_{nxn}$, es llamada *Matriz Hat* y permite escribir el vector de estimaciones de $\mathbf{Y}$ como una combinación lineal del vector de observaciones. Esta matriz es muy utilizada en regresión lineal, ya que permite simplificar una serie de expresiones, por ello continuaremos usándola a lo largo de este capítulo y los capítulos siguientes.

Se puede probar que $\mathbf{H}$ es idempotente; es decir $\mathbf{H}^2 = \mathbf{HH} = \mathbf{H}$ y en general $\mathbf{H}^n = \mathbf{H}$.

1.3.2. Tabla de Análisis de Varianza, Coeficiente de Determinación y Determinación Ajustado para el modelo de Regresión Lineal Múltiple

Las Sumas Cuadráticas presentadas en la *Tabla* 1.1 también pueden expresarse en notación matricial por medio de la antes mencionada Matriz Hat, expresión *(1.29):*

$$SCT = \mathbf{Y}^{\mathrm{T}}\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} \qquad \textbf{\textit{(1.30)}}$$

$$SCE = \mathbf{Y}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{H}\right)\mathbf{Y} \qquad \textbf{\textit{(1.31)}}$$

$$SCT = \mathbf{Y}^{\mathrm{T}}\left[\mathbf{H} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} \qquad \textbf{\textit{(1.32)}}$$

Donde $\mathbf{I}$ es la matriz identidad de nxn y $\mathbf{J}$ es una matriz nxn de unos. La estructura de la tabla ANOVA, es la misma que la presentada en la *Tabla* 1.1.

Una vez que se ha construido la tabla ANOVA, el siguiente paso será encontrar evidencia estadística de que al menos uno de los coeficientes betas que se ha propuesto en el modelo es distintos de cero, para ello se plantea el siguiente contraste de hipótesis:

$$H_0: \ \beta_1 = \beta_2 = ... = \beta_{p-1} = 0$$

vs.

H_0: Al menos uno de los betas no es cero

Nótese que β_0 no está incluido en el contraste. El estadístico de prueba es el mismo de la expresión *(1.19)*, el cual converge en distribución a una variable aleatoria F con p-1 grados de libertad en el numerador y n-p grados de libertad en el denominador. Esto es:

$$F = \frac{MCR}{MCT} \sim F_{(p-1, n-p)}$$

Por tanto, con (1-α)100% de confianza se rechaza H_0 en favor de H_1 si:

$$F = \frac{MCR}{MCT} > F_{(\alpha, p-1, n-p)}$$

Donde $F_{(\alpha, p-1, n-p)}$ es el percentil (1-α) de la distribución F con p-1 grado de libertad en el numerador y n-p grados de libertad en el denominador.

Los Coeficientes de Determinación y Determinación Ajustado son los mismos que en la Regresión Lineal Simple, expresiones *(1.16)* y *(1.17*) respectivamente.

1.4. Propiedades de los estimadores de Mínimos Cuadrados

Bajo el supuesto de que $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, el vector de estimaciones $\mathbf{b}$ tiene una distribución normal multivariada con:

$$E(\mathbf{b}) = E\left[\left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T Y}\right] = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T} E\left[\mathbf{Y}\right] = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T X} \boldsymbol{\beta}$$

$$E(\mathbf{b}) = \boldsymbol{\beta} \qquad \textbf{\textit{(1.33)}}$$

y

$$Var(\mathbf{b}) = \boldsymbol{\Sigma}_{\mathbf{b}} = E\left[\left(\mathbf{b} - E(\mathbf{b})\right)\left(\mathbf{b} - E(\mathbf{b})\right)^T\right]$$

Como:

$$\mathbf{b} - E(\mathbf{b}) = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T Y} - E\left[\left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T Y}\right]$$

$$= \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T Y} - \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T} E(\mathbf{Y})$$

$$= \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T}\left[\mathbf{Y} - E(\mathbf{Y})\right]$$

Entonces:

$$Var(\mathbf{b}) = \mathbf{\Sigma_b} = E\left[\left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T}\left[\mathbf{Y} - E(\mathbf{Y})\right]\left[\mathbf{Y} - E(\mathbf{Y})\right]^T \mathbf{X}\left(\mathbf{X^T X}\right)^{-1}\right]$$

$$Var(\mathbf{b}) = \mathbf{\Sigma_b} = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T} E\left[\left[\mathbf{Y} - E(\mathbf{Y})\right]\left[\mathbf{Y} - E(\mathbf{Y})\right]^T\right] \mathbf{X}\left(\mathbf{X^T X}\right)^{-1}$$

$$Var(\mathbf{b}) = \mathbf{\Sigma_b} = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T} \sigma^2 \mathbf{I} \mathbf{X}\left(\mathbf{X^T X}\right)^{-1}$$

$$Var(\mathbf{b}) = \mathbf{\Sigma_b} = \sigma^2 \left(\mathbf{X^T X}\right)^{-1} \qquad \textbf{\textit{(1.34)}}$$

En otras palabras:

$$\mathbf{b} \sim N(\boldsymbol{\beta},\ \sigma^2 (\mathbf{X^T X})^{-1}) \qquad \textbf{\textit{(1.35)}}$$

La expresión *(1.33)*, demuestra que los estimadores de mínimos cuadrados son insesgados. También se puede probar que estos estimadores son eficientes, consistentes y de máxima verosimilitud.

De acuerdo a la expresión *(1.34)*, un estimador de la matriz de varianzas y covarianzas de $\mathbf{b}$ será:

$$\mathbf{S}_b = s^2 (\mathbf{X^T X})^{-1} = MCE(\mathbf{X^T X})^{-1} \qquad \textbf{\textit{(1.36)}}$$

Al ser $\mathbf{b}$ un vector aleatorio con distribución normal p variada, entonces puede decirse que:

$$b_i \sim N(\beta_i, \sigma^2 c_{ii}) \qquad \textbf{\textit{(1.37)}}$$

Donde c_{ii} es el valor en la posición ii de la matriz simétrica $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$. Estos resultados permiten construir intervalos de confianza para β_i, así como proponer contrastes de hipótesis alrededor de los valores que pueden tomar.

Bajo los supuestos de normalidad e independencia de errores establecidos anteriormente,

$$\frac{b_i - \beta_i}{s_{b_i}} \sim t_{n-p} \qquad \textbf{\textit{(1.38)}}$$

En otras palabras este estadístico tiene una distribución t de Student con *(n-p)* grados de libertad.

Si la Hipótesis de que los b_i para $i>0$ son iguales a cero ha sido rechazada, el siguiente paso será determinar cuáles de los β_i es diferente de cero. Para ello se plantea el siguiente contraste:

$$H_0: \beta_i = 0 \text{ vs. } H_0: \beta_i \neq 0 \text{ para } i>0$$

Siendo el estadístico de prueba:

$$t = \frac{b_i}{s_{b_i}}$$

Por tanto, con (1-α)100% de confianza se rechaza H_0 a favor de H_1 si:

$$|t| > t_{(\alpha/2, n-p)} \qquad \textbf{\textit{(1.39)}}$$

Con la expresión *(1.37)* podemos construir los intervalos de confianza para los β_i:

$$b_i - t_{\alpha/2} s \sqrt{c_{ii}} - \leq \beta_i \leq b_i + t_{\alpha/2} s \sqrt{c_{ii}} \qquad \textbf{\textit{(1.40)}}$$

1.5. Sumas de Cuadrados Secuenciales

La Suma Cuadrática de Regresión puede ser particionada en tantas partes como variables de explicación X_i existan en el modelo. Cada una de estas particiones indica el incremento en la SCR cuando las variables X_i son incluidas en el modelo. A estas particiones de las conoce como *Sumas de Cuadrados Secuenciales* (SCS).

Si por ejemplo tenemos el siguiente modelo:

$$\hat{Y} = -603 - 0.422 X_1 + 3.94 X_2 + 9.92 X_3 \qquad \textbf{\textit{(1.41)}}$$

$$SCR=80469$$

La SCS de la variable X_1, denotada por $SC(\beta_1 | \beta_0)$ es la diferencia entre la SCR obtenida al incluir solo la variable X_1 en el modelo:

$$\hat{Y} = 358 + 1.08 X_1{}'$$

$$SCR=7349$$

y la SCR del modelo estimado con una constante:

$$\hat{Y} = 430.1 \qquad SCR=0$$

La cual se denotará como $SC(\beta_0)$, es fácil probar que esta suma cuadrática siempre es 0 ya que:

$$SC(\beta_0) = \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}_i\right)^2 = \sum_{i=1}^{n} \left(b_{0i} - b_{0i}\right)^2 = 0$$

Finalmente la SCS de la variable X_1 será:

$$SC(\beta_1 \mid \beta_0) = 7349 - 0 = 7349$$

De la misma manera, la SCS de la variable X_2, denotada por $SC(\beta_2 \mid \beta_0, \beta_1)$, es la diferencia entre la SCR obtenida al incluir en el modelo las variables X_1 y X_2:

$$\hat{Y} = 64 + 0.99X_1 + 2.13X_2 \qquad SCR=15485$$

Y la SCR obtenida al incluir solo la variable X_1 en el modelo:

$$SC(\beta_2 \mid \beta_0, \beta_1) = 15485 - 7349 = 8136$$

Finalmente, la SCS de la variable X_3, denotada por $SC(\beta_3 \mid \beta_0, \beta_1, \beta_2)$ es la diferencia entre la SCR obtenida al incluir en el modelo las variables X_1, X_2 y X_3; y la SCR obtenida al incluir las variables X_1 y X_2:

$$SC(\beta_3 \mid \beta_0, \beta_1, \beta_2) = 80469 - 15485 = 64984$$

Finalmente se tiene que:

$$SCR = SC(\beta_3 \mid \beta_0, \beta_1, \beta_3) + SC(\beta_2 \mid \beta_0, \beta_1) + SC(\beta_1 \mid \beta_0)$$

$$SCR = 7349 + 8136 + 64984 = 80469$$

Para el ejemplo que estamos considerando, la variable que más contribuye al modelo es la variable X_3 ya que es la de mayor SCS.

De manera general, dado el modelo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_{(p-1)} X_{(p-1)} + \varepsilon_i$$

$$SCR = SC(\beta_1 / \beta_0) + SC(\beta_2 / \beta_0, \beta_1) + ... + SC(\beta_p / \beta_0, \beta_1, \beta_2, ..., \beta_{p-1})$$

Dónde:

$$SC(\beta_1 / \beta_0) = SC(\beta_0, \beta_1) - SC(\beta_0)$$
$$SC(\beta_2 / \beta_0, \beta_1) = SC(\beta_0, \beta_1, \beta_2) - SC(\beta_0, \beta_1)$$
$$\vdots$$
$$SC(\beta_p / \beta_0, \beta_1, \beta_2, ..., \beta_{p-1}) = SC(\beta_0, \beta_1, \beta_2, \beta_3, ..., \beta_p) - SC(\beta_0, \beta_1, \beta_2, ..., \beta_{p-1})$$

La expresión $SC(\beta_p / \beta_0, \beta_1, \beta_2, ..., \beta_{p-1})$ es la SCS de la variable X_p y significa el incremento en la SCR cuando la variable X_p es incluida en el modelo, el cual ya contiene a las p-1 variables predictoras.

$SC(\beta_0, \beta_1, \beta_2, \beta_3, ..., \beta_p)$ es la SCR correspondiente al modelo de regresión que incluye las p variables.

CAPÍTULO 2

2. EVALUACIÓN DE LA CALIDAD DEL MODELO

2.1. Introducción

Para determinar si un modelo de Regresión Lineal es realmente "bueno" al momento de realizar estimaciones, no debe olvidarse la verificación del cumplimiento de los supuestos básicos efectuados. El no cumplimiento de alguno de estos supuestos podría llevarnos a modelos inestables; esto es muestras diferentes nos conducirían a modelos diferentes. Los supuestos que deben verificarse son principalmente acerca del error aleatorio ε_i, pero también existen otro tipo de eventos que pueden afectar la efectividad del modelo como la presencia de Valores Aberrantes, Puntos de Influencia, problemas de Enmascaramiento y Multicolinealidad. En este capítulo se presentan varios métodos de alta efectividad al momento de detectar este tipo de problemas y las posibles soluciones que pueden darse a cada uno de ellos.

El análisis de los residuales es una forma eficaz de descubrir diversos tipos de inadecuación del modelo:

$$\varepsilon_i = y_i - E[y_i], \; E[y_i] = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_{p-1} x_{(p-1)i} \qquad \textbf{\textit{(2.1)}}$$

y

$$e_i = \hat{\varepsilon}_i = y_i - \hat{y}_i \,, \; \hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \ldots + b_{p-1} x_{(p-1)i} \qquad \textbf{\textit{(2.2)}}$$

e_i es el estimador del error ε_i y se lo conoce como "*residual".*

2.2. Verificación de Supuestos acerca de $\boldsymbol{\varepsilon}_i$

Los supuestos que se hacen en un análisis de regresión respecto al error se resumen en la expresión:

$$\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \sigma^2 \mathbf{I}\right) \qquad \textbf{\textit{(2.3)}}$$

Donde $\boldsymbol{\varepsilon} \in R^n, \mathbf{0} \in R^n$ y $\mathbf{I}$ es la matriz identidad de dimensiones *nxn*; y *n* es el tamaño de la muestra. De esta expresión se deduce:

- La distribución del error ε_i es normal.
- La media del error ε_i es cero.
- La varianza del error ε_i es constante.

- Los errores no están correlacionados.

Si éstos supuestos no se cumplen, los estimadores de mínimos cuadrados siguen siendo insesgados pero dejan de ser eficientes; es decir, dejan de ser estimadores de mínima varianza y como consecuencia de ello, los intervalos de confianza y las pruebas de hipótesis basadas en las distribuciones t y F dejarían de ser confiables. En las siguientes secciones se presentan los métodos más utilizados al momento de verificar el cumplimiento de cada uno de estos supuestos.

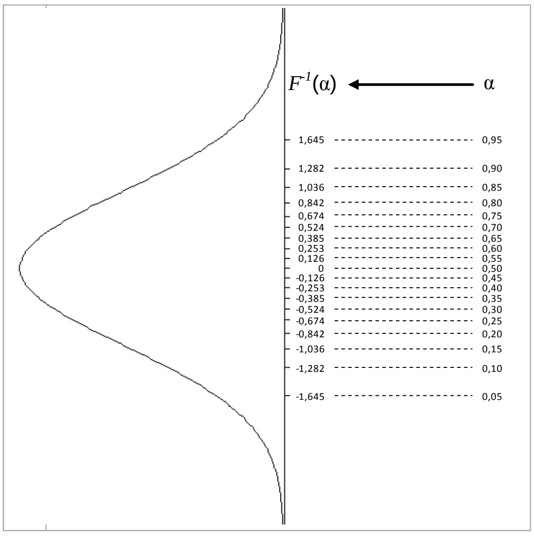
2.2.1. Normalidad

La no normalidad del error puede venir acompañada de varianza no constante. Existen varios métodos para detectar si los errores ε_i tienen una distribución normal, entre ellos los métodos de bondad de ajuste y el gráfico de probabilidad normal.

2.2.1.1. Gráfico de Probabilidad Normal

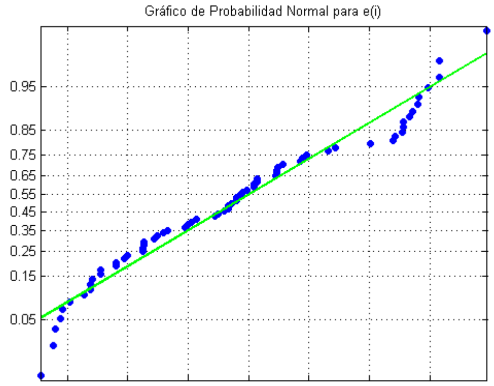
El Gráfico de Probabilidad Normal es un gráfico diseñado para que al graficarse la distribución normal acumulada se bosqueje una línea recta. Esto debido a que el eje vertical de este gráfico tiene una escala de Probabilidad Normal. Esta escala varía entre 0 y 1 y es construida a partir de la distribución inversa acumulada de la normal estándar. Véase *Figura 2.1*.

Figura 2.1
Evaluación de la Calidad del Modelo
"Escala de Probabilidad Normal"

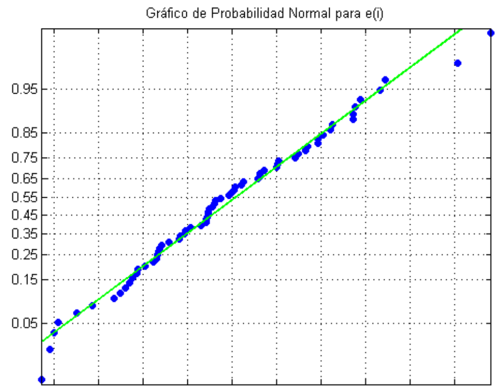


Sean $e_{[1]}, e_{[2]}, ..., e_{[n]}$, los residuales ordenados en forma ascendente. Si se grafica $e_{[i]}$ en función de la probabilidad acumulada $P_i = (i-0.5)/n$, para $i = 1, 2, ..., n$ en un gráfico de probabilidad normal, los puntos deberían aproximarse a una recta, si es que la distribución de probabilidad de los $e_{[i]}$ es normal. Véase *Figura 2.2*.

Figura 2.2
Evaluación de la Calidad del Modelo
"Gráfica de Probabilidad Normal"



(a) Distribución no Normal



(b) Distribución Normal

Los Gráficos de Probabilidad Normal también se trazan graficando el residual $e_{[i]}$ en función del valor esperado bajo el supuesto de normalidad, el cual es aproximado como:

$$E(e_{[i]}) \simeq F^{-1}[(i-0.5)/n]\sigma \qquad \textbf{(2.4)}$$

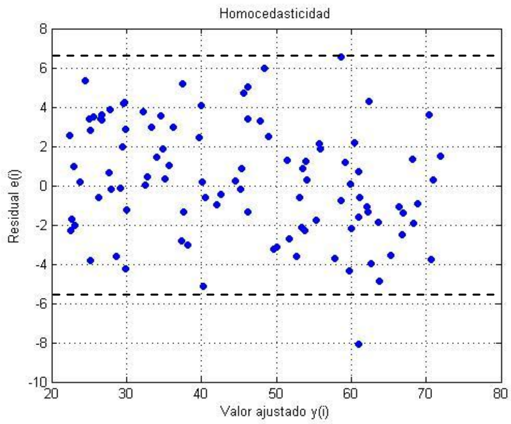
Donde F^{-1}, es la función inversa de la distribución acumulada normal estándar.

2.2.2. Homocedasticidad

2.2.2.1. Gráfico de los Residuos contra los Valores Ajustados $\hat{\mathbf{Y}}$

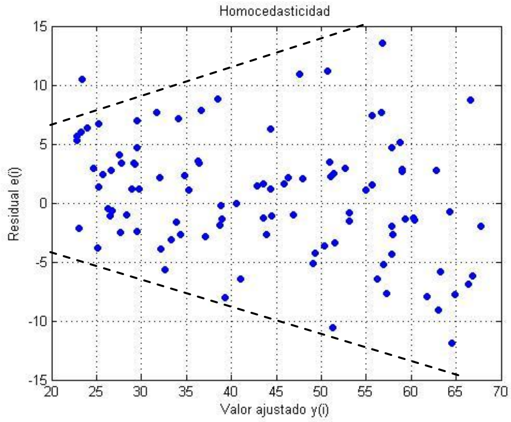
Una manera sencilla de verificar si la varianza del error es constante es realizando un gráfico de los residuales e_i contra los valores ajustados $\hat{y}_i$. Si la varianza es constante se esperaría que los errores fluctúen alrededor del eje horizontal, y que puedan ubicarse en una banda; caso contrario puede ser que la varianza no sea constante. Véase *Figura 2.3.*

Figura 2.3
Evaluación de la Calidad del Modelo
"Gráfico de los residuales vs. Los valores ajustados"



(a) Varianza no Constante

Figura 2.3
Evaluación de la Calidad del Modelo
"Gráfico de los residuales vs. Los valores ajustados"



(b) *Varianza Contante*

Para controlar la varianza no constante, se puede aplicar una transformación a la variable a ser explica Y o usar el método de mínimos cuadrados ponderados.

2.2.3. Correlación

Para detectar la presencia de correlación se pueden realizar diversas pruebas estadísticas, pero la más conocida es la desarrollada por Durbin y Watson en 1971.

2.2.3.1. Gráfico de los residuales en el tiempo t

Una forma sencilla de diagnosticar correlación entre los errores, es graficando los residuales de manera ordenada en el tiempo o espacio. Si estos muestran algún tipo de patrón lineal o cíclico por ejemplo, los errores podrían estar correlacionados. Véase *Figura 2.4.*

Figura 2.4
Evaluación de la Calidad del Modelo
"Gráfico de los residuales en el tiempo"

(a) No Correlacioón

(b) Correlación

2.2.3.2. Prueba de Durbin-Watson

La prueba de Durbin-Watson es utilizada en Series de Tiempo para detectar Correlación Serial. Esta prueba se basa en la Hipótesis de que los errores ε_i del modelo de regresión se generan en un proceso autorregresivo de primer orden, esto es:

$$\varepsilon_i = \rho\varepsilon_{i-1} + \delta_i \qquad \textbf{\textit{(2.5)}}$$

Donde δ_i es una variable aleatoria $N(0,\sigma^2)$ y ρ es el coeficiente de correlación. Además se puede probar que bajo estas condiciones:

$$E(\varepsilon_i) = 0 \qquad \textbf{\textit{(2.6)}}$$

$$Var(\varepsilon_i) = \sigma_a^2\left(\frac{1}{1-\rho^2}\right) \qquad \textbf{\textit{(2.7)}}$$

$$Cov(\varepsilon_i, \varepsilon_{i+u}) = \rho^{|u|}\sigma^2\left(\frac{1}{1-\rho^2}\right) \qquad \textbf{\textit{(2.8)}}$$

Donde los errores ε_i tienen media cero y varianza constante, pero están correlacionados, a menos que $\rho = 0$.

Ante esta situación Durbin y Watson plantearon la siguiente prueba unilateral:

$$H_o:\ \rho = 0 \ \text{ vs. } H_1:\ \rho > 0$$

Y determinaron la región crítica de la prueba en base al estadístico:

$$d = \frac{\sum_{i=2}^{n}\left(e_i - e_{i-1}\right)^2}{\sum_{i=2}^{n} e_i^{\,2}} = \frac{\mathbf{e^T A e}}{\mathbf{e^T e}} \qquad \textbf{\textit{(2.9)}}$$

Donde $\mathbf{e} = (e_1, e_2, ..., e_n)^T$ es el vector de residuales y $\mathbf{A}$ es la matriz correspondiente a la forma cuadrática que aparece en el numerador de la expresión (2.9). Si la Hipótesis Nula de la prueba es verdadera, la distribución del estadístico d dependería de la matriz de diseño $\mathbf{X}$. Sin embargo Durbin y Watson [1951] demostraron que d esta entre dos cotas d_L y d_U tales que si d esta entre estas dos cotas se puede llegar a una conclusión respecto a la hipótesis nula planteada:

Si $d < d_L$ Rechazar H_0

Si $d > d_U$ No Rechazar H_0

Si $d_L \leq d \leq d_U$ La prueba no es concluyente

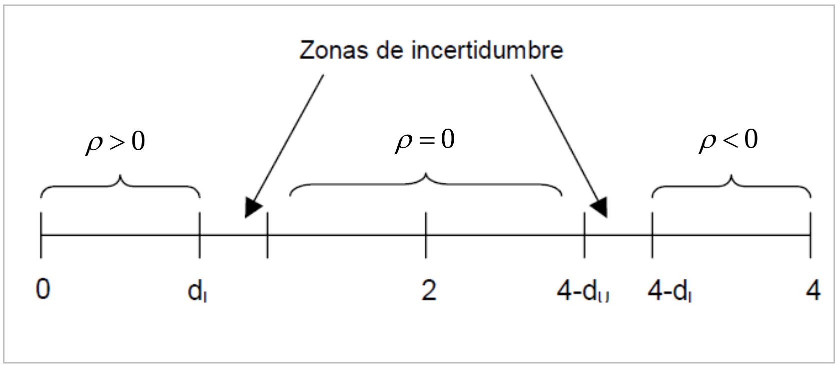
En el Anexo 1 se muestran los límites d_L y d_U para varios tamaños de muestra, diversas variables de explicación y dos tasas de error tipo I (α=0.05 y α=0.01)

[3] Maddala pudo probar que:

$$E(d / \rho = 0) \approx 2 + \frac{2(k-1)}{n-k} \qquad \textbf{\textit{(2.10)}}$$

Y que el Estadístico d es un valor comprendido entre 0 y 4. Véase *Figura 2.5*.

Figura 2.5
Evaluación de la Calidad del Modelo
"Región de Rechazo de la Prueba de Durbin-Watson"



Así, si el valor del estadístico d es próximo a 2, $\rho = 0$; si se aproxima a 4, $\rho < 0$ y si se aproxima a 0 $\rho > 0$.

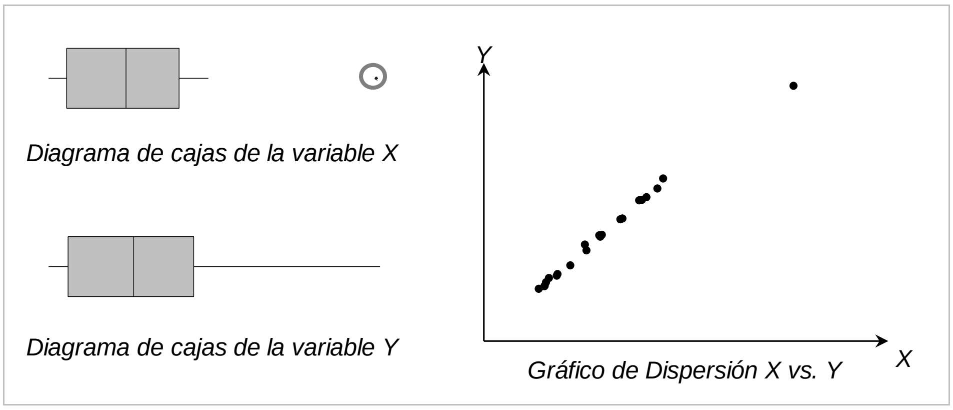
MATLAB tiene una función llamada dwtest que devuelve el estadístico d y el valor p de la prueba, sea esta de una cola o dos utilizando el algoritmo de PAN. Para conocer más sobre el funcionamiento de este algoritmo véase [1] Farebrother.

2.3. Valores aberrantes o atípicos

En todo análisis estadístico resulta importante detectar la presencia de valores aberrantes o atípicos, ya que éstos pueden afectar drásticamente a los estimadores. En el análisis de regresión, la presencia de un valor aberrante puede afectar la efectividad de un modelo como puede no hacerlo, por ello existen varios criterios para su identificación basados en el análisis de residuales.

Debemos considerar también que la presencia de valores aberrantes puede detectarse antes de la estimación del modelo, mediante un gráfico de dispersión entre la variable a ser explicada con cada una de las variables de explicación. Sin embargo no es recomendable un análisis univariado, pues los valores que se consideran aberrantes en una sola variable, pueden no serlo cuando se hace un análisis conjunto con otras variables. Véase *Figura 2.6.*

Figura 2.6
Evaluación de la Calidad del Modelo
"Valores aberrantes: Caso Univariado vs. Caso Bivariado"



El Gráfico de los Residuales e_i en función de $\hat{y}_i$ ¨y el Gráfico de Probabilidad Normal son también útiles para detectar valores atípicos potenciales.

2.3.1. Residuales

Ya se habían definido antes los residuales como:

$$e_i = \hat{\varepsilon}_i = y_i - \hat{y}_i \qquad \textbf{\textit{(2.11)}}$$

Siendo y_i una observación y $\hat{y}_i$ su valor ajustado. Además en el capítulo anterior se probó que:

$$E(e_i) = 0 \qquad \textbf{\textit{(2.12)}}$$

$$\hat{\sigma}^2_{e_i} = s^2_{e_i} = MCE \qquad \textbf{\textit{(2.13)}}$$

Los residuales cuyo valor absoluto es mayor a tres desviaciones estándar respecto de la media, indican que hay valores atípicos potenciales en el espacio de Y, sin embargo se recomienda analizarlos en conjunto con los residuales que se muestran a continuación.

2.3.2. Residuales Estandarizados

Ya que la varianza aproximada del error se estima con la MCE, los residuales estandarizados serán:

$$d_i = \frac{e_i}{\sqrt{MCE}} \qquad \textbf{\textit{(2.14)}}$$

Los residuales estandarizados tienen media cero y varianza aproximadamente unitaria. Un residual estandarizado d_i mayor que 3 indica que la observación i es un valor atípico potencial.

2.3.3. Residuales Estudentizados

Sea $\mathbf{H}$, la Matriz Hat definida en la expresión *(1.29)* como $\mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}$ y h_{ij} sus elementos; además $\mathbf{e}$ el vector de residuales:

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$$

En el capítulo anterior se probó que*:*

$$\hat{\mathbf{Y}} = \mathbf{HY}$$

Por tanto:

$$\mathbf{e} = \mathbf{Y} - \mathbf{HY}$$

$$\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y} \qquad \textbf{\textit{(2.15)}}$$

Entonces:

$$Var(\mathbf{e}) = Var[(\mathbf{I} - \mathbf{H})\mathbf{Y}]$$

$$Var(\mathbf{e}) = (\mathbf{I} - \mathbf{H})Var[\mathbf{Y}](\mathbf{I} - \mathbf{H})^{\mathbf{T}}$$

$$Var(\mathbf{e}) = (\mathbf{I} - \mathbf{H})\sigma^2\mathbf{I}(\mathbf{I} - \mathbf{H})^{\mathbf{T}}$$

Ya que $(\mathbf{I} - \mathbf{H})$ es una matriz simétrica e idempotente.

$$Var(\mathbf{e}) = \sigma^2 \left(\mathbf{I} - \mathbf{H}\right) \tag{2.16}$$

Esto quiere decir que:

$$Var(e_i) = \sigma^2(1 - h_{ii}) \quad \text{y} \qquad Cov(e_i, e_j) = -\sigma^2(h_{ij})$$

Entonces, los residuales estudentizados se definen dividiendo el i-ésimo residual entre su desviación estándar "exacta":

$$r_i = \frac{e_i}{\sqrt{MCE(1 - h_{ii})}} \tag{2.17}$$

Los residuales estandarizados y estudentizados aportan con frecuencia información equivalente. En conjuntos grandes de datos los residuales estandarizados no serán muy diferentes de los estudentizados. Un residual estudentizado r_i mayor que 3 indica la presencia de un valor atípico potencial. También son llamados residuales estudentizados internamente.

2.3.4. Residuales PRESS

Los residuales PRESS o residuales de predicción se definen como la diferencia entre el valor observado y_i para $i = 1, 2, ..., n$ y el valor estimado de esta observación basado en todas las observaciones excepto esta i-ésima $\hat{y}_{[i]}$:

$$e_{[i]} = y_i - \hat{y}_{[i]} \tag{2.18}$$

Es decir, se elimina la i-ésima observación y se ajusta el modelo de regresión a las *n*-1 observaciones restantes, para estimar y_i.

Se puede probar que existe una relación entre los residuales press y los residuales usuales. Véase [3] Hocking.

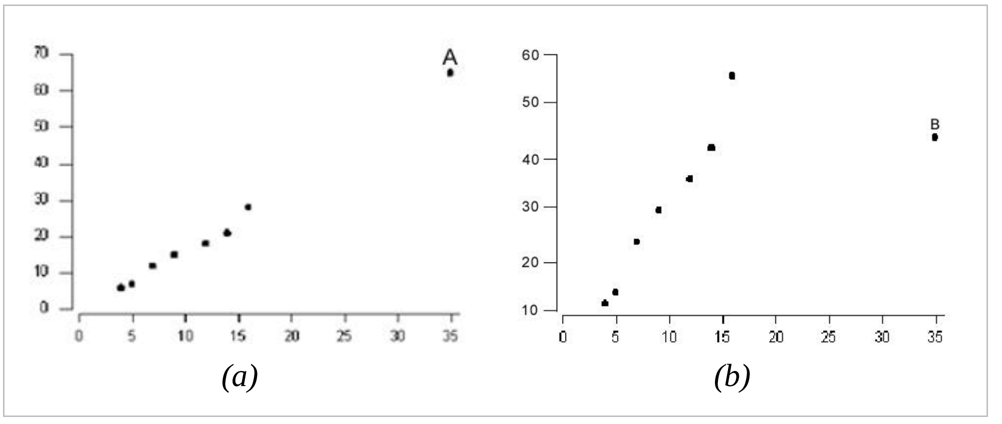
$$e_{[i]} = \frac{e_i}{1 - h_{ii}} \tag{2.19}$$

Una gran diferencia entre el residual ordinario y el residual PRESS indica un valor atípico potencial.

2.4. Puntos de Influencia

Los Puntos de Influencia o valores influyentes son aquellos que tienen un impacto notable sobre los coeficientes del modelo, por ello la importancia de localizarlos. La mayoría de los textos llaman "valores aberrantes" a un valor alejado solamente en la dirección vertical y Punto de influencia a una observación alejada en la dirección horizontal. Véase *Figura 2.7,* el punto A no afecta las estimaciones de los coeficientes de regresión, mientras el punto B si tiene un impacto notable en la estimación de estos coeficientes ya que atrae a la recta de regresión en su dirección.

Figura 2.7
Evaluación de la Calidad del Modelo
"Detección de Puntos de Influencia"



En esta sección se presentarán dos métodos para detectar Puntos de Influencia, éstos son Apalancamiento y la Distancia de Cook.

2.4.1. Apalancamiento

Sea $\mathbf{H}$, la matriz Hat definida en la expresión *(1.29)* como $\mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}$ y h_{ij} sus elementos. Además de acuerdo a la expresión *(1.28)* $\hat{\mathbf{Y}} = \mathbf{HY}$, entonces:

$$Var(\hat{\mathbf{Y}}) = Var(\mathbf{HY})$$

$$Var(\hat{\mathbf{Y}}) = \mathbf{H}Var(\mathbf{Y})\mathbf{H}^{\mathbf{T}}$$

$$Var(\hat{\mathbf{Y}}) = \mathbf{H}\sigma^2\mathbf{I}\mathbf{H}^{\mathbf{T}}$$

Ya que $\mathbf{H}$ es una matriz simétrica e idempotente:

$$Var(\hat{\mathbf{Y}}) = \sigma^2\mathbf{H} \qquad \textbf{\textit{(2.20)}}$$

Los elementos h_{ii} de la matriz $\mathbf{H}$ son una medida de lugar o ubicación del i-ésimo punto en el espacio de x, por lo tanto son vistos como la cantidad de “balanceo” o “palanqueo” de la i-ésima observación y_i sobre el i-ésimo valor ajustado $\hat{y}_i$. Los elementos grandes en la diagonal de la matriz $\mathbf{H}$ indican observaciones que son potencialmente influyentes. En general, todo elemento $h_{ii} > 2p/n$, es considerado un punto de apalancamiento o balanceo. Esto no se aplica para casos en los cuales $2p/n > 1$.

No todos los puntos de apalancamiento o balanceo serán influyentes. Véase *Figura 2.7*, el punto A es un punto de balanceo o apalancamiento, pero no es un punto influyente.

2.4.2. Distancia de Cook

La distancia de Cook mide el cambio que ocurriría en el vector de coeficientes estimado de regresión si la i-ésima observación fuera omitida. Está definida como:

$$CD_i^2 = \frac{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{[\mathbf{i}]})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{[\mathbf{i}]})}{pMCE} \qquad \textbf{\textit{(2.21)}}$$

Dónde:

- $\hat{\boldsymbol{\beta}}$ es el vector de coeficientes estimado con el modelo completo
- $\hat{\boldsymbol{\beta}}_{[\mathbf{i}]}$ es el vector de coeficientes estimado sin la i-ésima observación
- $\mathbf{X}$ es la matriz de Diseño
- MCE es el estimador de σ^2
- p es el número de parámetros en el modelo

Sea P_i el i-ésimo punto para $i=1,2,\ldots,n$ de p coordenadas. Dado el siguiente contraste de Hipótesis:

$$H_0: P_i \text{ no es un punto de influencia}$$

vs.

$$H_1: P_i \text{ es un punto de influencia}$$

Con (1-α)100% de confianza se debe rechazar H_0 a favor de H_1 si el estadístico definido en (2.21) es mayor que $F_{(\alpha,p,n-p)}$.

El estadístico definido en (2.21) no sigue una distribución $F_{(p,n-p)}$, pero funciona bien en la práctica compararla con una $F_{(\alpha,p,n-p)}$.

Se debe tener mucho cuidado si se quiere eliminar valores aberrantes y puntos de influencia, ya que estos no siempre provienen de un error de medición o digitación y en estos casos debe considerarse el uso de técnicas robustas de estimación que no sean tan sensibles a puntos influyentes como lo son los mínimos cuadrados.

2.5. Enmascaramiento

En la detección de valores aberrantes y puntos de influencia, es importante que el procedimiento no esté afectado por ninguno de los dos posibles efectos relacionados, conocidos como Efecto Enmascaramiento (Masking Effect) y Efecto Inundación (Swamping Effect).

El efecto enmascaramiento ocurre cuando no se detectan valores aberrantes o puntos de influencia que en realidad sí lo son, esto se da por la presencia de algún punto cercano que lo evita, mientras que el efecto inundación hace que se detecten puntos aberrantes o puntos de influencia que en realidad no lo son. Una forma de evitar este tipo de problemas es con la utilización de métodos de estimación robusta.

Por otra parte podemos detectar el efecto enmascaramiento e inundación con la utilización de un método llamado Mean-Shift Outliers Models.

2.5.1. Mean-Shift Outliers Models

Sea M un conjunto de datos de tamaño n con k valores aberrantes. Sea C un subconjunto de M sin los k valores aberrantes, es decir con c=n-k elementos. Si realizamos un modelo de regresión con todas las observaciones de C:

$$\mathbf{Y}_C = \mathbf{X}_C\boldsymbol{\beta}_C + \boldsymbol{\varepsilon}_C$$

Donde $\mathbf{X}_C \in M_{cxp}$ y $\mathbf{Y}_C \in R^c$. El estimador de mínimos cuadrados de $\boldsymbol{\beta}_C$ será:

$$\hat{\boldsymbol{\beta}}_C = (\mathbf{X}_C^{\mathbf{T}}\mathbf{X}_C)^{-1}\mathbf{X}_C^{\mathbf{T}}\mathbf{Y}_C$$

Este estimador no se verá afectado si introducimos las estimaciones correspondientes a los k valores aberrantes que no fueron considerados en la construcción del modelo y sus correspondientes valores en las variables de explicación:

$$\hat{\boldsymbol{\beta}}_C = (\mathbf{X}^{\mathbf{T}}\mathbf{X}\ )^{-1}\mathbf{X}^{\mathbf{T}}\tilde{\mathbf{Y}}$$

Donde $\mathbf{X} \in M_{nxp}$ y $\tilde{\mathbf{Y}} \in R^n$; $\tilde{\mathbf{Y}} = (y_1, ..., y_c, \hat{y}_{c+1}, ...\hat{y}_{c+k})$

Luego el vector de observaciones $\mathbf{Y}$ del conjunto M será:

$$\mathbf{Y} = \tilde{\mathbf{Y}} + \Delta\mathbf{Y}$$

Donde $\Delta\mathbf{Y} = (0,...,0,d_{c+1},...,d_{c+k})$. En secciones previas se probó que:

$$\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

Donde $\mathbf{H}$ es la antes mencionada Matriz Hat. Entonces:

$$\mathbf{e} = (\mathbf{I} - \mathbf{H})(\tilde{\mathbf{Y}} + \Delta\mathbf{Y})$$

$$\mathbf{e} = (\mathbf{I} - \mathbf{H})\tilde{\mathbf{Y}} + (\mathbf{I} - \mathbf{H})\Delta\mathbf{Y}$$

$$\mathbf{e} = (\mathbf{I} - \mathbf{H})\tilde{\mathbf{Y}} + \Delta\mathbf{Y} - \mathbf{H}\Delta\mathbf{Y}$$

$$\mathbf{e} = (e_1,...,e_c,0,...,0)^T + \Delta\mathbf{Y} - \mathbf{H}\Delta\mathbf{Y}$$

$$\mathbf{e} = (e_1,...,e_c,d_{c+1},...,d_{c+k})^T - \mathbf{H}\Delta\mathbf{Y}$$

$$\mathbf{e} = \mathbf{e}_C - \mathbf{H}\Delta\mathbf{Y}$$

Finalmente calculamos los residuales estudentizados, y si existe un residual estudentizado de entre los c primeros más grande que 3, entonces existe el “efecto inundación”. Si existe un residual estudentizado menor que 3 de entre los k últimos valores entonces existe el Efecto Enmascaramiento.

2.6. Multicolinealidad

Cuando existe una relación aproximadamente lineal entre las variables de explicación, es posible que los estimadores que se obtengan

tengan varianzas muy grandes aunque siguen conservando la propiedad de insesgados, además se puede aceptar con frecuencia la hipótesis nula de que un parámetro es cero, aun cuando la correspondiente variable sea relevante y por último puede que los coeficientes estimados serán muy sensibles a pequeños cambios en los datos.

Una forma de detectar multicolinealidad es calculando la matriz de correlación de las variables de explicación y ver qué pares de variables tienen correlación cercana a 1, sin embargo; se han desarrollado numerosas reglas que tratan de determinar en qué medida la multicolinealidad afecta a la estimación de un modelo, entre ellas el Factor de Agrandamiento de la Varianza (FAV) y el Número de Condición.

2.6.1. Factor de Inflación de la Varianza (FIV)

Si consideramos el modelo de regresión lineal múltiple:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_{(p-1)} X_{(p-1)} + \varepsilon \quad \text{para } i=1,2,\dots,n$$

Entonces se puede probar que la varianza del j-ésimo coeficiente de regresión estimado es:

$$s^2_{\hat{\beta}_j} = \frac{\hat{\sigma}^2}{(n-1)(1-R^2_j)s^2_j} \qquad \text{para } j=1,2,\dots,p-1$$

Donde R_j^2 es el coeficiente de determinación obtenido al hacer la regresión de X_j sobre el resto de las variables de explicación del modelo y s_j^2 es la varianza muestral de la variable X_j.

Si la correlación entre las variables de explicación fuera nula, la fórmula para estimar la varianza del j-ésimo coeficiente de regresión se reduciría a:

$$s_{\hat{\beta}_j} = \frac{\hat{\sigma}_2}{(n-1)s_j^2}$$

El FIV es la razón entre la varianza observada y la que habría sido en caso de que X_j no estuviera correlacionada con el resto de las variables de explicación del modelo:

$$FIV = \frac{1}{1-R_j^2}$$

Es decir que el FIV mide cuanto crece la varianza del j-ésimo coeficiente de regresión como consecuencia de que las variables estén altamente correlacionadas. Una variable de explicación con un FIV entre 5 y 10 puede causar multicolinealidad.

2.6.2. Número de Condición

El número de condición se define como la razón entre la raíz característica más grande $(\lambda_{\max})$ y la raíz característica más pequeña $(\lambda_{\min})$ de la matriz $\mathbf{X}^{\mathbf{T}}\mathbf{X}$:

$$k(\mathbf{X}) = \frac{(\lambda_{\max})}{(\lambda_{\min})}$$

donde $\mathbf{X}$ es la matriz de diseño sin la columna de unos. Si existe uno o más valores propios "pequeños" esto implicaría que hay dependencias casi lineales entre las columnas de la matriz $\mathbf{X}$. Por tanto, si el número de condición es menor que 100, no hay problema grave de multicolinealidad. Los números de condición de 100 a 1000 implican multicolinealidad de moderada a fuerte, y si es mayor que 1000 el problema de multicolinealidad es grave.

Entre las soluciones que pueden darse a la multicolinealidad están:

1. Eliminar del modelo las variables que tienen una correlación muy alta.
2. Incrementar el tamaño de la muestra
3. Regresión Ridge
4. Componentes principales
5. Mínimos Cuadrados Parciales

CAPÍTULO 3

3. INFORMACIÓN TÉCNICA ACERCA DEL DESARROLLO DE ERLA

3.1. Introducción

ERLA es una aplicación desarrollada para plataformas Microsoft Windows. Esta aplicación fue desarrollada en Visual Basic.NET, y por lo tanto su lenguaje y controles se basan en la arquitectura propia del sistema operativo Windows. La finalidad de orientar el desarrollo de ERLA a la plataforma Microsoft Windows es ofrecer una aplicación amigable para usuarios quienes en su mayoría hacen uso de este sistema operativo. Como primer requisito, se debe contar con el componente de Windows .Net Framework instalado en el equipo, aunque este componente puede omitirse en caso de ya haber sido instalado en el sistema operativo como ocurre en versiones más recientes de Microsoft Windows: Windows Server 2008, Windows Vista y Windows 7.

Por otra parte, ERLA también cuenta con un potente componente de MATLAB, para la realización exclusiva de los cálculos que el usuario ejecute, éste es el MCR (MATLAB Component Runtime) el cual es el segundo requisito que se debe instalar para la ejecución adecuada del software ERLA.

3.2. MATLAB

MATLAB que quiere decir MATrix LABoratory, es un software matemático que ofrece un entorno de desarrollo integrado (IDE) con un lenguaje de programación propio (lenguaje M). Está disponible para las plataformas Unix, Windows y Apple Mac OS X.

Durante mucho tiempo hubo críticas porque MATLAB es un producto propietario de The Mathworks, ya que los usuarios están sujetos a un vendor lock-in. Recientemente se ha proporcionado una herramienta adicional llamada MATLAB Builder bajo la sección de herramientas Application Deployment para utilizar funciones MATLAB como archivos de biblioteca que pueden ser usados con ambientes de construcción de aplicación .NET o Java. Pero la desventaja es que el computador donde la aplicación tiene que ser utilizada necesita el MCR (MATLAB Component Runtime) para que los archivos MATLAB funcionen correctamente. MCR puede ser distribuido libremente con archivos de biblioteca generados por el compilador MATLAB. Por lo tanto, el MCR es un requerimiento obligatorio para la ejecución del software ERLA.

3.3. Visual Studio 2010

Microsoft Visual Studio es un entorno de desarrollo integrado (IDE) para sistemas operativos Windows. Soporta varios lenguajes de programación tales como Visual C++, Visual C#, Visual J#, ASP.NET y Visual Basic .NET.

Visual Studio permite a los desarrolladores crear aplicaciones, sitios y aplicaciones web, así como servicios web en cualquier entorno que soporte la plataforma .NET. Así se pueden crear aplicaciones que se intercomuniquen entre estaciones de trabajo, páginas web y dispositivos móviles.

3.4. Integración de Visual Basic.NET y MATLAB

La integración de estas dos plataformas se lleva a cabo gracias a la factibilidad de integración que ofrece Visual Basic.NET con otras aplicaciones, y además, el componente gratuito de MATLAB (MCR), que permite el acceso desde la aplicación a sus más elementales funciones.

Para lograr la integración exitosa de estas aplicaciones, se debe empezar por la generación en MATLAB de una o varias librerías (archivos **.dll*). Estas librerías son un conjunto de funciones (archivos **.m*) que se agregan a un proyecto de implementación .NET en MATLAB. Finalmente el proyecto se compila para obtener un sólo archivo binario que luego será referenciado desde Visual Basic.NET. Véase *Figura 3.1.*

Adicionalmente, desde Visual Basic.NET se debe agregar una referencia hacia la librería *MWArray.dll*, la misma que usualmente se encuentra en la ruta:

- %MCR\toolbox\dotnetbuilder\bin\[win32/win64]\v2.0.

Donde %MCR es el directorio raíz en el que se ha instalado el componente MCR (MCR Installer) o una instalación MATLAB.

La librería MWArray.dll contiene la definición de variables elementales para la conversión de arreglos numéricos, cadenas y estructuras entre la aplicación .NET y el MCR. Véase ANEXO II.

El proyecto desarrollado en Visual Studio.NET es compilado para obtener los instaladores del mismo. Este aplicativo podrá ser instalado por el usuario final en sistemas operativos Windows que dispongan de .NET Framework y MCR. Véase *Figura 3.2*.

Figura 3.1
Información Técnica acerca del desarrollo de ERLA
"*Entorno de desarrollo de ERLA***"**

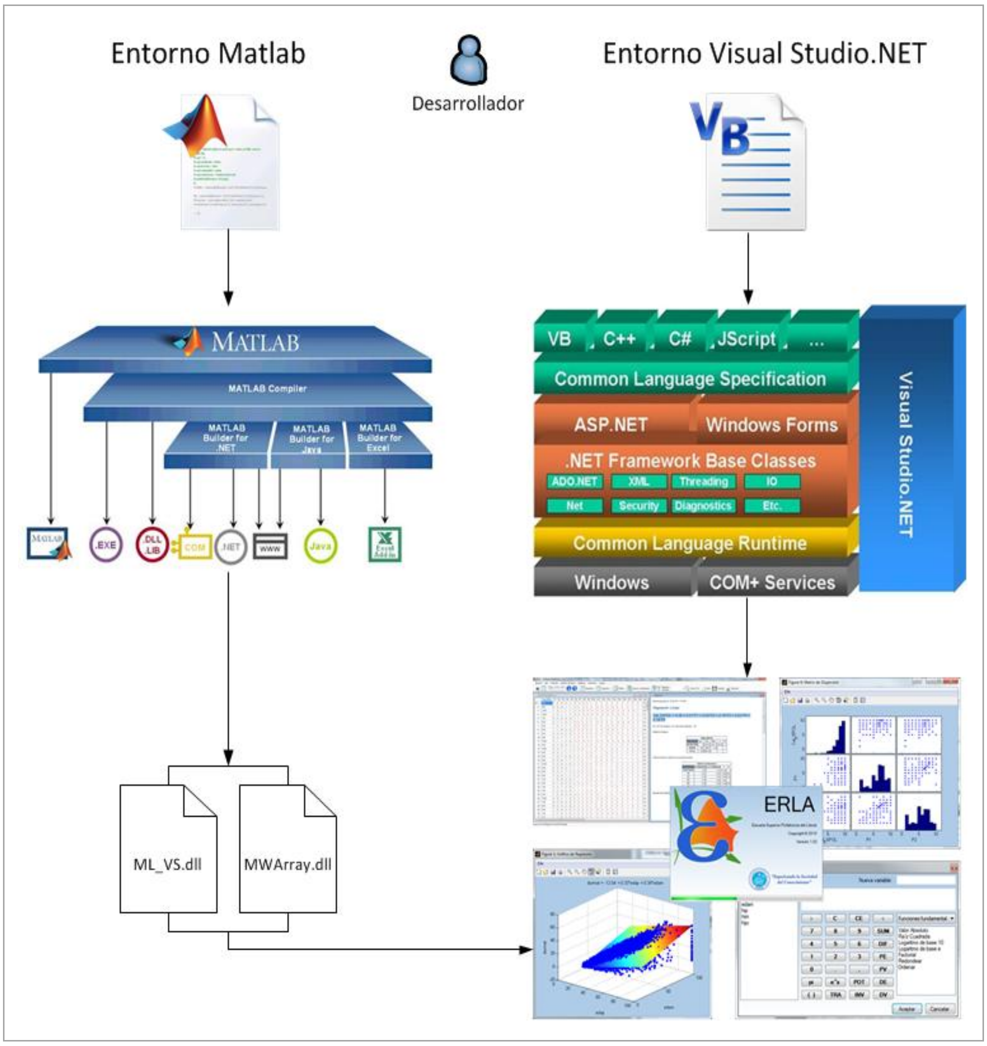
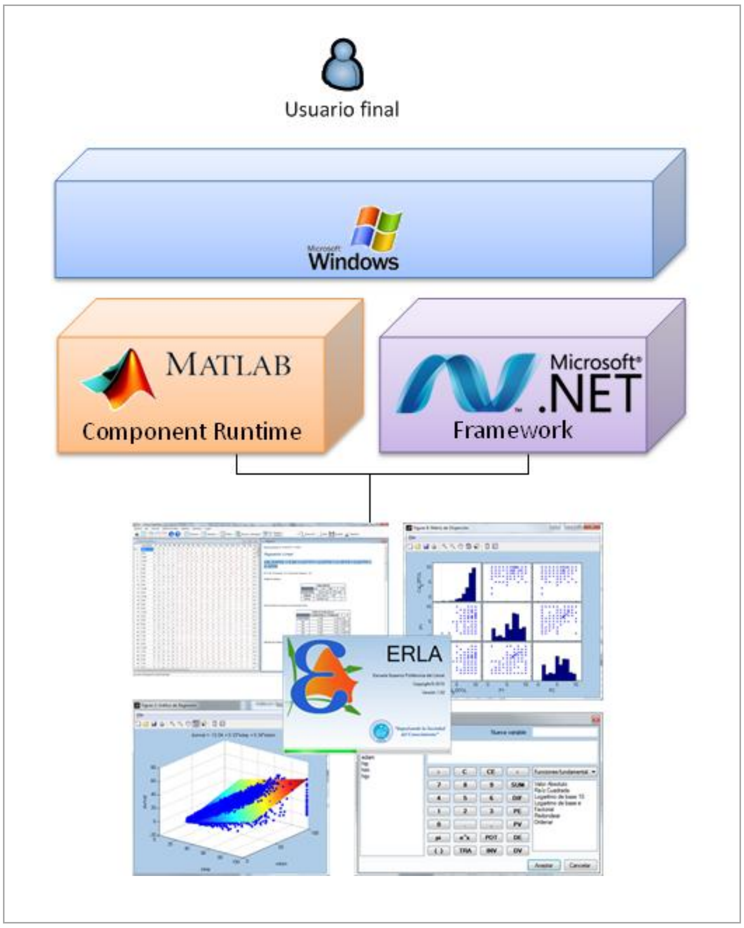


Figura 3.2
Información Técnica acerca del desarrollo de ERLA
"*Entorno de usuario de ERLA***"**



3.4.1. Programación ERLA

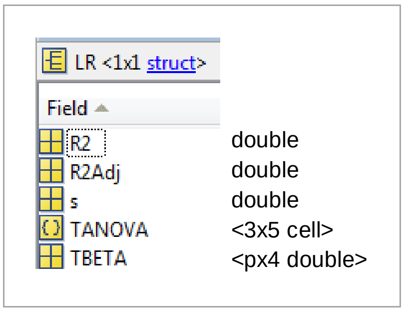
Para programar la función de Regresión Lineal en MATLAB, se usaron estructuras (MWStruct). Véase Figura 3.3. Las estructuras son un tipo de datos de MATLAB que almacena una cantidad numérica o una cadena de caracteres. La ventaja de usar una estructura, es que ésta puede almacenar cualquier tipo de dato de MATLAB, e incluso otra estructura.

Figura 3.3
Información Técnica acerca del desarrollo de ERLA
"*Función LienarRegression en lenguaje MATLAB***"**

```
function LR=LinearRegression(y,MX)
%Autor: Celia De La Cruz Cedeño & Juan Carlos Buenaño Cordero
%Fecha: 09/09/2010
%Contacto: adela@espol.edu.ec & jucabuen@espol.edu.ec
%y: Vector correspondiente a la variable a ser explicada.
%MX: Matriz con variables de explicación.
%LR: Estructura que contiene los resultados.
format long g;
d=size(MX);
n=d(1);
p=d(2)+1;
%Tabla ANOVA
GLR=p-1;
GLE=n-p;
GLT=n-1;
j=ones(n,1);
my=mean(y);
X=[j,MX];
B=(X'*X)\(X'*y);
SCE=(y-X*B)'*(y-X*B);
SCR=(X*B-my*j)'*(X*B-my*j);
SCT=(y-my*j)'*(y-my*j);
MCR=SCR/GLR;
MCE=SCE/GLE;
F=MCR/MCE;
VP=fcdf(F,GLR,GLE);
VP=1-VP;
TA=cell(3,5);
TA{1,1}=GLR;
TA{2,1}=GLE;
TA{3,1}=GLT;
TA{1,2}=SCR;
TA{2,2}=SCE;
TA{3,2}=SCT;
TA{1,3}=MCR;
TA{2,3}=MCE;
TA{1,4}=F;
TA{1,5}=VP;
%R2
R2=SCR/SCT;
%R2 Ajustado
R2adj=1-((1-R2)*(n-1)/(n-p));
%s
s=sqrt(MCE);
%Tabla de Inferencias para los beta
Sb=MCE*inv((X'*X));
TB=zeros(p,4);
for i=1:p
    TB(i,1)=B(i);
    TB(i,2)=sqrt(Sb(i,i));
    TB(i,3)=TB(i,1)/TB(i,2);
    TB(i,4)=abs(TB(i,3));
    TB(i,4)=tcdf(TB(i,4),n-p);
    TB(i,4)=(1-TB(i,4))*2;
End
%Mostrar resultados en la estructura LR
LR=struct('R2',{R2},'R2Adj',{R2adj},'s',{s},'TANOVA',{TA},'TBETA',{TB}
);
```

La función a la que se ha llamado LinearRegression devuelve una estructura con 5 tipos de datos:

Figura 3.4
Información Técnica acerca del desarrollo de ERLA
"*Resultados de la función LinearRegression*"



Esta función, junto con todas las demás funciones se compilan en un archivo (.ddl) que será referenciado desde Visual. Para ello se crea, en Visual Basic el formulario Regresión Lineal con el código que permite llamar a la función LinearRegression de MATLAB. Véase Figuras 3.5 y 3.6.

Figura 3.5
Información Técnica acerca del desarrollo de ERLA
"*Vista diseño del formulario Regresión Lineal*"

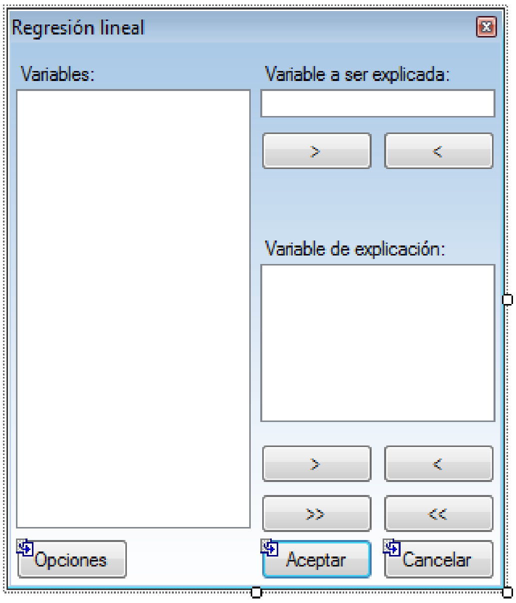


Figura 3.6

Información Técnica acerca del desarrollo de ERLA

"*Extracto del código del formulario Regresión Lineal"***"**

```
'Autor: Celia De La Cruz Cedeño & Juan Carlos Buenaño Cordero
'Fecha: 25/09/2010
'Contacto: adela@espol.edu.ec & jucabuen@espol.edu.ec

Public Class frmLinearRegression

Private Sub btnAceptar_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles btnAceptar.Click
If (csXY.lstX.Items.Count > 0 And csXY.lstY.Items.Count > 0) Then
    csXY.CreateVars()
    If (csXY.rows > csXY.columns) Then

        Dim oRS As New clsRegStats
        oWR.InsertTitle("Regresión Lineal")

        Dim struct As VSStruct
        oWR.InsertTitle("Análsis de Varianza de 1 Vía")
        struct = oRS.VSLinearRegression(csXY.yvar, csXY.xvar)


        'R2, R2 AJUSTADO Y DESVIACIÓN ESTANDAR
        oWR.InsertParagraph("R2: " & struct.str1 & " | R2 Ajustado: " &
        struct.str2 & " | Desviación Estándar: " & struct.str3)

        'TABLA ANOVA
        Dim tblrow(2), tblcol(4) As String
        tblcol(0) = "  G.L.  "
        tblcol(1) = "  S.C.  "
        tblcol(2) = "  M.C.  "
        tblcol(3) = "  F  "
        tblcol(4) = "  P  "
        tblrow(0) = "REGRESIÓN"
        tblrow(1) = "ERROR"
        tblrow(2) = "TOTAL"
        oWR.InsertParagraph("Análisis de Varianza:")
        oWR.InsertTable(oRS.VSGTHTMLtable("Tabla ANOVA", struct.arr4, tblcol,
        tblrow).ToString)

        'TABLA DE COEFICIENTES
        Dim tblrow1(csXY.columns + 1), tblcol1(3) As String
        tblcol1(0) = "ESTIMADOR"
        tblcol1(1) = "E.E.ESTIMADOR"
        tblcol1(2) = "T"
        tblcol1(3) = "P"
        tblrow1(0) = "CONSTANTE"
        For i As Integer = 1 To csXY.columns + 1
            tblrow1(i) = csXY.lstX.Items(i - 1).ToString
        Next
        oWR.InsertParagraph("Tabla de Inferencia respecto a los parámetros
        betas")
        oWR.InsertTable(oRS.VSGTHTMLtable("Tabla de Estimadores",
        struct.arr5, tblcol1, tblrow1).ToString)

        Me.Close()

   Else
        MessageBox.Show("Existen menos observaciones que variables.",
        Application.ProductName, MessageBoxButtons.OK,
        MessageBoxIcon.Exclamation)
      Me.Close()
      End If
Else
    MessageBox.Show("Seleccione las variables a utilizar.",
    Application.ProductName, MessageBoxButtons.OK, MessageBoxIcon.Exclamation)
   Me.Close()
End If
End Sub
```

Figura 3.7

Información Técnica acerca del desarrollo de ERLA

"*Extracto del código de la clase clsRegStats***"**

```
'Autor: Celia De La Cruz Cedeño & Juan Carlos Buenaño Cordero
'Fecha: 11/09/2010
'Contacto: adela@espol.edu.ec & jucabuen@espol.edu.ec

Imports MathWorks.MATLAB.NET.Utility
Imports MathWorks.MATLAB.NET.Arrays
Imports ML_VS
Public Class clsRegStats
    Dim mva As New MultivariateAnalysis
    Dim mf As New MathFunctions
    Dim da As New DescriptiveAnalysis
    Dim disp As New DisplayObjects
    Dim mwsa As MWStructArray = Nothing
    Dim mwa As MWArray = Nothing
    Dim prec As Integer = Precision

    Public Function VSLinearRegression(ByVal Y As MWNumericArray, ByVal X As
MWNumericArray)
        Dim strmat As New VSStruct
        mwsa = mva.LinearRegression(Y, X)
        strmat.str1 = mf.RoundTo(mwsa("R2"), prec).ToString
        strmat.str2 = mf.RoundTo(mwsa("R2Adj"), prec).ToString
        strmat.str3 = mf.RoundTo(mwsa("s"), prec).ToString
        strmat.arr1 = mf.RoundTo(mwsa("TANOVA"), prec)
        strmat.arr2 = mf.RoundTo(mwsa("TBETA"), prec)
        Return strmat
    End Function

    Public Function VSRoundTo(ByVal X As MWNumericArray, ByVal precision As
Integer) As MWArray
        Return mf.RoundTo(X, precision)
    End Function

    Public Function VSGTHTMLtable(ByVal VSTableName As String, ByVal Data As
MWArray, Optional ByVal VSColName As String() = Nothing, Optional ByVal
VSRowName As String() = Nothing) As String
        'Para presentar los resultados en tabla se requieren de los
argumentos:
        'VSTableName: Nombre de la tabla en el cuadro figure
        'Data: Datos que aparecerán en la tabla
        'VSColName (opcional): Etiqueta para los encabezados de columna
        'VSRowName (opcional): Etiqueta para los encabezados de fila
        Dim table As String = ""
        Dim TableName As New MWCharArray(VSTableName)
        If (VSColName Is Nothing And VSRowName Is Nothing) Then
            table = disp.GTHTMLtable(TableName, Data).ToString
        ElseIf (VSRowName Is Nothing And VSColName IsNot Nothing) Then
            Dim VarName As New MWCharArray(VSColName)
            Dim VarName2 As New MWCellArray(VarName)
            table = disp.GTHTMLtable(TableName, Data, "%" & prec & "g",
VarName2, "%s").ToString
        ElseIf (VSRowName IsNot Nothing And VSColName IsNot Nothing) Then
            Dim VarName As New MWCharArray(VSColName)
            Dim VarName2 As New MWCellArray(VarName)
            Dim VarName3 As New MWCharArray(VSRowName)
            Dim VarName4 As New MWCellArray(VarName3)
            table = disp.GTHTMLtable(TableName, Data, "%" & prec & "g",
VarName2, "%s", VarName4, "%s").ToString()
        End If
        Return table
    End Function

End Class
```

Para evitar repetir código, se ha creado un control personalizado llamado “ctrlSelectXY” el cual permite al usuario seleccionar las variables. Este control es usado en varios cuadros de diálogo de la aplicación. También se ha creado la clase llamada “clsRegStats”, en la cual se han definido los métodos “VSLinearRegression” y “VSGTHTMLtable”. Véase Figura 3.7. Finalmente, el formulario “frmWebReport” con los métodos “InsertParagraph” e “InsertTable”.

Cada función creada en MATLAB, es definida como un método de la clase “clsRegStats”, es decir que esta clase es la encargada de llamar y convertir a cada una de las funciones creadas en MATLAB para compilarlas en el archivo .dll.

El “frmWebReport” es el formulario que permite mostrar los resultados, y usa un control WebBrowser.

El ingreso de datos se lo hace a través del formulario “frmWorksheet”, el cual usa un control DataGridView.

CAPÍTULO 4

4. VALIDACIÓN ESTADÍSTICA DEL SOFTWARE ERLA

4.1. Introducción

En este capítulo se desarrolla la validación estadística del Software correspondiente a cada uno de los métodos y técnicas mencionados en los capítulos 1 y 2. Para el modelo de Regresión Lineal, se probará que los estimadores que se obtienen utilizando ERLA cumplen con las propiedades teóricas mencionadas en el capítulo 1. Los métodos y técnicas presentados en el capítulo 2, serán evaluados de acuerdo a la proporción de veces que su predicción es correcta.

4.2. Modelo de Regresión Lineal

4.2.1. Modelo de Regresión Lineal Simple

Suponga que la relación entre un par de variables llamadas X y Y está dada por la siguiente ecuación:

$$Y = 1 + 0.0125X$$

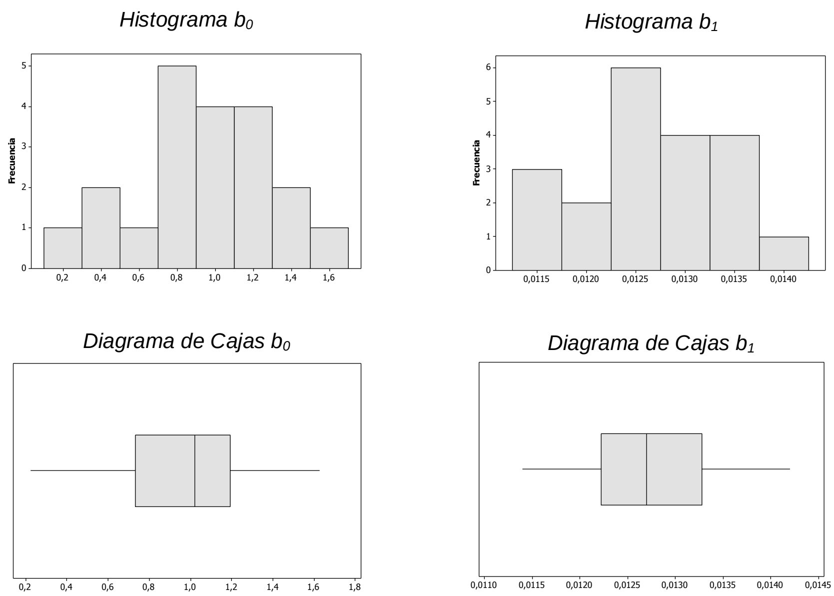
Para evaluar la efectividad del Software en cuanto al cálculo de los coeficientes de regresión lineal simple, se tomaron un total de 20 muestras con n=39, agregando un error aleatorio $\varepsilon \sim N(0,\sigma^2)$ donde $\sigma^2 = 1$. Los estimadores de los parámetros $\beta_0 = 1$ y $\beta_1 = 0.0125$ obtenidos en cada muestra se presentan en la Tabla 4.1.

Tabla 4.1
Validación Estadística del Software ERLA
"Estimadores de los Parámetros Betas en un modelo de Regresión Lineal Simple"

MUESTRA NO.	b_0	b_1	MUESTRA NO.	b_0	b_1
1	1,6265	0,0114	11	1,1128	0,0125
2	0,4949	0,0136	12	0,7163	0,0136
3	1,4362	0,0116	13	1,0942	0,0123
4	0,8987	0,0129	14	1,0855	0,0126
5	1,1409	0,0127	15	1,084	0,0117
6	0,826	0,0132	16	0,6222	0,0131
7	1,4225	0,0122	17	0,7766	0,0133
8	1,2115	0,0123	18	0,3776	0,0133
9	1,2489	0,0121	19	0,8267	0,013
10	0,9569	0,0127	20	0,2235	0,0142

En la Figura 4.1 se presenta, el Histograma de Frecuencias, el Diagrama de Cajas, y las estadísticas descriptivas de cada uno de los estimadores.

Figura 4.1
Validación Estadística del Software ERLA
"Estadísticas Descriptivas de los Estimadores de los Parámetros Betas en un modelo de Regresión Lineal Simple"



Estadísticas Descriptivas

Estimador	**Media**	**Error Estándar**	**Desviación Estándar**	**Varianza**	**Mínimo**
b_0	0,95912	0,08081	0,36138	0,13060	0,22350
b_1	0,01272	0,00016	0,00073	0,00000053	0,01140

Estimador	Q_1	**Mediana**	Q_3	**Máximo**	**Sesgo**
b_0	0,73137	1,02045	1,19385	1,62650	-0,24347
b_1	0,01223	0,01270	0,01328	0,01420	0,00500

Ambos estimadores tienen un sesgo cercano a cero, y el valor de la media es cercano al parámetro, en cada caso. Entonces se puede decir que los estimadores de mínimos cuadrados de los parámetros betas obtenidos en el Software cumplen con las propiedades presentadas en el capítulo 1.

4.2.2. Modelo de Regresión Lineal Múltiple

Para evaluar la efectividad del software en cuanto al cálculo de los coeficientes de regresión lineal múltiple, se ha considerado el siguiente modelo:

$$Y = -10 + 3X_1 + 2X_2$$

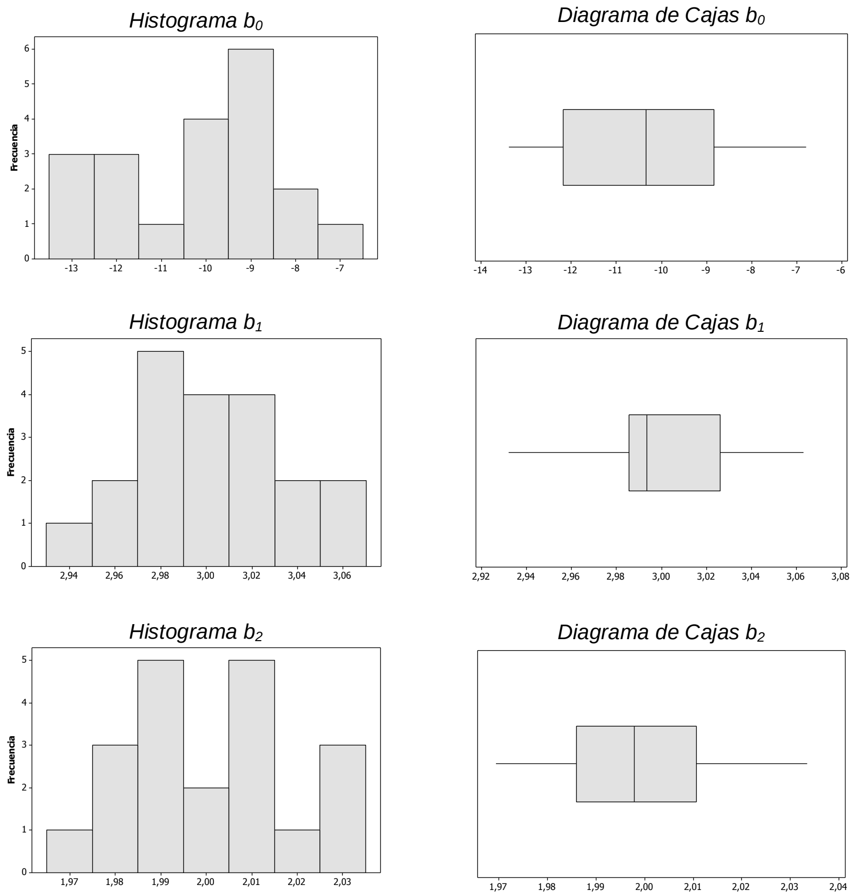
Se tomaron un total de 20 muestras, en este caso con n=100, agregando un error aleatorio $\varepsilon \sim N(0,\sigma^2)$ donde $\sigma^2 = 36$ y los valores de las variables X_1 y X_2 fueron fijados por el experimentador. Los estimadores de los parámetros $\beta_0 = -10$, $\beta_1 = 3$ y $\beta_2 = 2$ obtenidos en cada muestra se presentan en la Tabla 4.2 y las Estadísticas Descriptivas correspondientes se presentan en la Figura 4.2.

El sesgo correspondiente a cada uno de los estimadores es muy cercano a cero, el valor de la media es muy cercano al parámetro estimado, para cada caso; en los Histogramas y Diagramas de Cajas no se observa tendencia alguna. Por lo tanto se puede concluir que los estimadores de mínimos cuadrados de los parámetros betas obtenidos por medio del Software en un modelo de Regresión Lineal Múltiple cumplen con las propiedades teóricas presentadas en capítulos anteriores.

Tabla 4.2
Validación Estadística del Software ERLA
"Estimadores de los Parámetros Betas en un modelo de Regresión Lineal Múltiple"

MUESTRA NO.	b_0	b_1	b_3
1	-13,383	3,034	2,007
2	-10,457	2,991	1,992
3	-9,433	2,987	2,011
4	-12,885	3,032	2,017
5	-10,393	3,023	1,988
6	-10,529	3,014	2,010
7	-12,185	3,007	2,034
8	-8,354	2,992	1,978
9	-10,327	2,987	2,007
10	-8,578	2,932	2,027
11	-6,791	2,970	1,994
12	-13,029	3,027	2,031
13	-8,741	2,971	1,999
14	-12,291	3,063	1,982
15	-12,139	3,061	1,970
16	-9,207	2,956	1,997
17	-9,106	3,016	1,985
18	-9,433	2,987	2,011
19	-8,040	2,985	1,977
20	-10,342	2,996	1,990

Figura 4.2
Validación Estadística del Software ERLA
“Estadísticas Descriptivas de los Estimadores de los Parámetros Betas en un modelo de Regresión Lineal Múltiple”



Estadísticas Descriptivas

Estimador	***Media***	***Error Estándar***	***Desviación Estándar***	***Varianza***	***Mínimo***
b_0	-10,2820	0,4147	1,8545	3,4391	-13,3828
b_1	3,0016	0,0074	0,0332	0,0011	2,9319
b_3	2,0002	0,0041	0,0183	0,0003	1,9695

Estimador	Q_1	***Mediana***	Q_3	***Máximo***	***Sesgo***
b_0	-12,1734	-10,3349	-8,8325	-6,7908	-0,1456
b_1	2,9856	2,9936	3,0263	3,0634	0,0553
b_3	1,9861	1,9979	2,0106	2,0335	0,2446

4.3. Evaluación de la Calidad del Modelo

Las validaciones correspondiente a los métodos presentados en el Capítulo 2, se harán utilizando modelos de Regresión Lineal Simple, ya que; al trabajar en dos dimensiones será más fácil para el experimentador identificar las observaciones que ha colocado para la validación y la interpretación de los resultados.

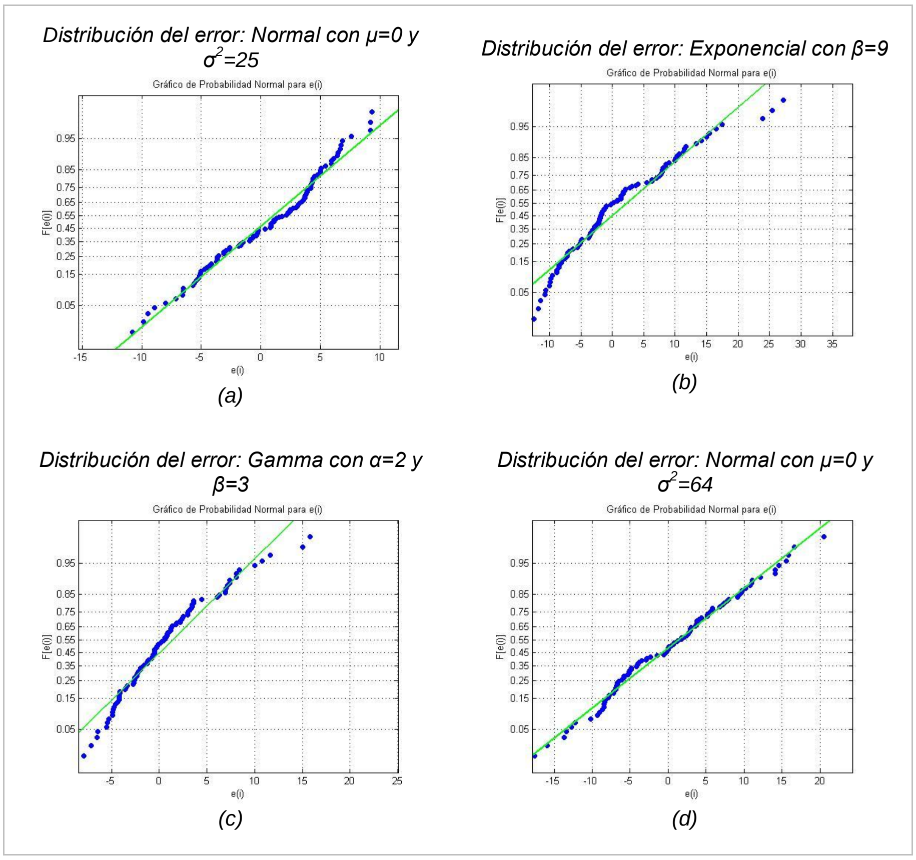
4.3.1. Gráfico de Probabilidad Normal

Para evaluar la validez de este método en el Software, se generaron cuatro muestras de tamaño 100 donde los errores agregados en cada muestra tienen las siguientes distribuciones:

1. Normal con $\mu=0$ y $\sigma^2=25$
2. Exponencial con $\beta=9$
3. Gamma con $\alpha=2$ y $\beta=3$
4. Normal con $\mu=0$ y $\sigma^2=64$

Los Gráficos de Probabilidad Normal obtenidos en cada caso se muestran en la Figura 4.3. En los gráficos correspondientes a los literales a y d se puede observar una tendencia lineal, ya que en estos casos la distribución del error es normal, mientras que en los literales b y c no puede observarse esta tendencia. Para cada una de las muestras, el Gráfico de Probabilidad Normal, nos condujo a la conclusión correcta.

Figura 4.3
Validación Estadística del Software ERLA
"Gráficos de Probabilidad Normal"



4.3.2. Gráfico de los Residuos Contra los Valores Estimados $\hat{y}_i$

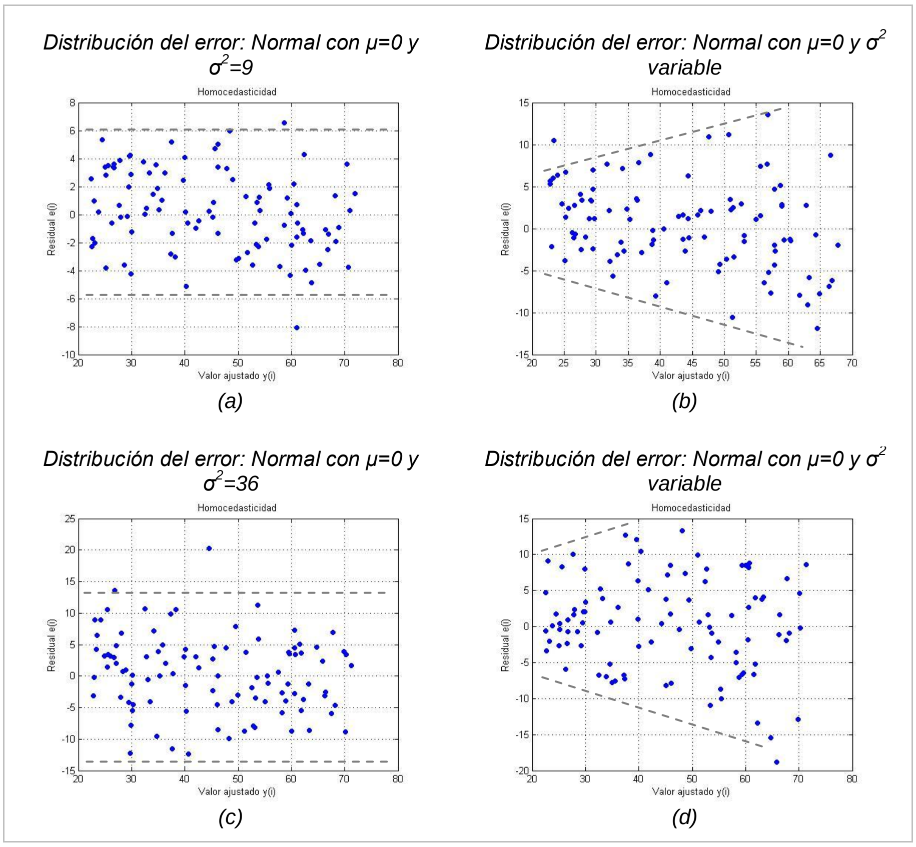
Para evaluar la validez de este método en el Software, fueron generadas cuatro muestras de tamaño 100 donde se colocaron errores con distribuciones normales de media cero y varianza:

1. Constante σ^2=25
2. Variable
3. Variable

4. Constante σ^2=64

Los gráficos obtenidos en cada caso se muestran en la Figura 4.4. En los literales a y c los puntos están cercanos entre ellos y pueden encerrarse en una banda horizontal, puesto que en estos casos la varianza es constante. Sin embargo en los literales b y d, existen puntos que están muy alejados de los demás y no pueden encerrarse en una banda horizontal.

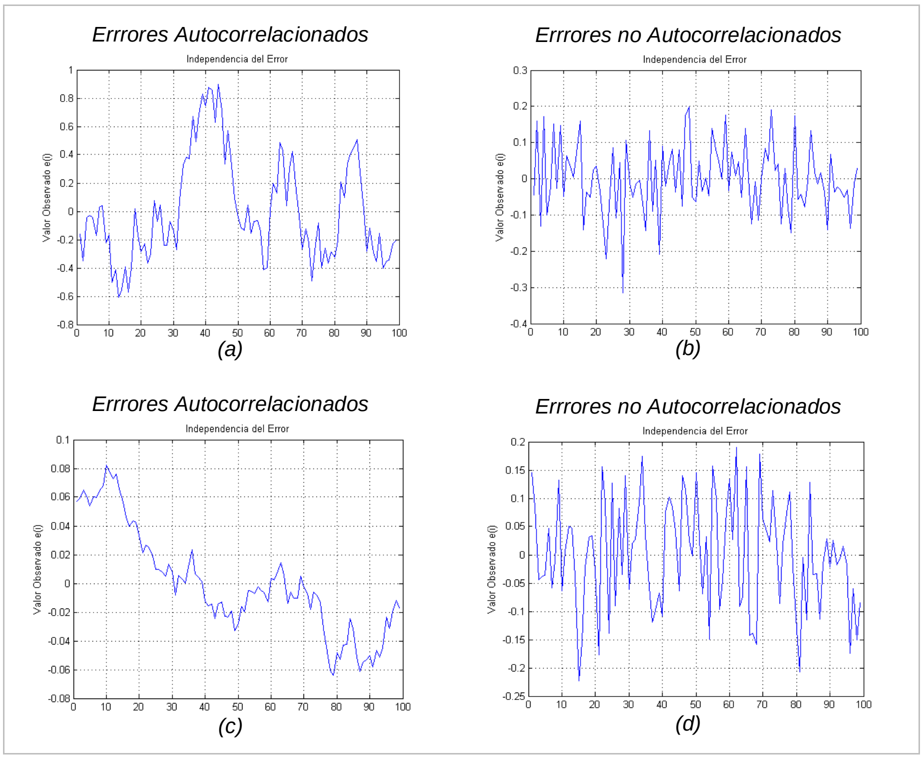
Figura 4.4
Validación Estadística del Software ERLA
"Gráficos de Residuales contra los Valores Estimados $\hat{y}_i$ "



4.3.3. Gráfico de los residuales en el tiempo t

Para verificar la validez de este método se generaron un total de 4 muestras de tamaño 100 donde en dos de ellas se agregaron errores con autocorrelación de orden 1; es decir que se generó un valor cualquiera para e_1 y se hizo que los e_i restantes dependan linealmente de este valor.

Figura 4.5
Validación Estadística del Software ERLA
"Gráficos de los residuales en el tiempo t"



Los Gráficos obtenidos en cada caso se muestran en la Figura 4.5. En los gráficos de los literales a y c, donde los errores están correlacionados, existe cierta tendencia; mientras que en los literales b y d donde los errores no están correlacionados, no hay una tendencia clara. Esto nos lleva a la conclusión correcta en cada caso.

4.3.4. Estadístico Durbin-Watson

Tabla 4.3
Validación Estadística del Software ERLA
"Estadístico Durbin-Watson α=0.01"

M. NO.	***CORR.***	***d***	***d_L***	***d_U***	***4-d_U***	***4-dl***	***ACIERTO***
1	Sí	0,168	1,502	1,582	2,418	2,498	Sí
2	Sí	0,060	1,502	1,582	2,418	2,498	Sí
3	Sí	0,077	1,502	1,582	2,418	2,498	Sí
4	Sí	0,099	1,502	1,582	2,418	2,498	Sí
5	Sí	0,072	,502	1,582	2,418	2,498	Sí
6	Sí	0,049	1,502	1,582	2,418	2,498	Sí
7	Sí	0,064	1,502	1,582	2,418	2,498	Sí
8	Sí	0,117	1,502	1,582	2,418	2,498	Sí
9	Sí	0,085	1,502	1,582	2,418	2,498	Sí
10	Sí	0,108	1,502	1,582	2,418	2,498	Sí
11	Sí	0,053	1,502	1,582	2,418	2,498	Sí
12	Sí	0,051	1,502	1,582	2,418	2,498	Sí
13	Sí	0,097	1,502	1,582	2,418	2,498	Sí
14	Sí	0,293	1,502	1,582	2,418	2,498	Sí
15	Sí	0,027	1,502	1,582	2,418	2,498	Sí
16	No	2,393	1,502	1,582	2,418	2,498	Sí
17	No	1,984	1,502	1,582	2,418	2,498	Sí
18	No	1,786	1,502	1,582	2,418	2,498	Sí
19	No	1,980	1,502	1,582	2,418	2,498	Sí
20	No	1,938	1,502	1,582	2,418	2,498	Sí

Para validar la efectividad de éste método, fueron generadas un total de 20 muestras de tamaño 100 donde se agregó errores correlacionados a 15 muestras y a las 5 restantes se agregaron errores no correlacionados.

En la Tabla 4.3 y 4.4 se presentan los resultados obtenidos para modelos de regresión lineal simple; y niveles de error tipo 1 α=0.01 y α=0.05. Para α=0.01 hubo un 100% de aciertos, mientras que para α=0.05 hubo un 95% de aciertos.

Tabla 4.4
Validación Estadística del Software ERLA
"Estadístico Durbin-Watson α=0.05"

M. NO.	*CORR.*	*d*	d_L	d_U	4-d_U	*4-dl*	*ACIERTO*
1	Sí	0,168	1,634	1,715	2,285	2,366	Sí
2	Sí	0,060	1,634	1,715	2,285	2,366	Sí
3	Sí	0,077	1,634	1,715	2,285	2,366	Sí
4	Sí	0,099	1,634	1,715	2,285	2,366	Sí
5	Sí	0,072	1,634	1,715	2,285	2,366	Sí
6	Sí	0,049	1,634	1,715	2,285	2,366	Sí
7	Sí	0,064	1,634	1,715	2,285	2,366	Sí
8	Sí	0,117	1,634	1,715	2,285	2,366	Sí
9	Sí	0,085	1,634	1,715	2,285	2,366	Sí
10	Sí	0,108	1,634	1,715	2,285	2,366	Sí
11	Sí	0,053	1,634	1,715	2,285	2,366	Sí
12	Sí	0,051	1,634	1,715	2,285	2,366	Sí
13	Sí	0,097	1,634	1,715	2,285	2,366	Sí
14	Sí	0,293	1,634	1,715	2,285	2,366	Sí
15	Sí	0,027	1,634	1,715	2,285	2,366	Sí
16	No	2,393	1,634	1,715	2,285	2,366	No
17	No	1,984	1,634	1,715	2,285	2,366	Sí
18	No	1,786	1,634	1,715	2,285	2,366	Sí
19	No	1,980	1,634	1,715	2,285	2,366	Sí
20	No	1,938	1,634	1,715	2,285	2,366	Sí

Por otra parte utilizando el valor p que se obtiene en el Software para una prueba de dos colas, hubo un 95% de aciertos ya que el valor p de la prueba en la muestra 16 está dentro de la zona de incertidumbre.

Tabla 4.5
Validación Estadística del Software ERLA
"Estadístico Durbin-Watson Valor p"

M. NO.	CORR.	d	VALOR P	ACIERTO
1	Sí	0,168	0,0000	Sí
2	Sí	0,060	0,0000	Sí
3	Sí	0,077	0,0000	Sí
4	Sí	0,099	0,0000	Sí
5	Sí	0,072	0,0000	Sí
6	Sí	0,049	0,0000	Sí
7	Sí	0,064	0,0000	Sí
8	Sí	0,117	0,0000	Sí
9	Sí	0,085	0,0000	Sí
10	Sí	0,108	0,0000	Sí
11	Sí	0,053	0,0000	Sí
12	Sí	0,051	0,0000	Sí
13	Sí	0,097	0,0000	Sí
14	Sí	0,293	0,0000	Sí
15	Sí	0,027	0,0000	Sí
16	No	2,393	0,0589	-
17	No	1,984	0,8548	Sí
18	No	1,786	0,2391	Sí
19	No	1,980	0,8387	Sí
20	No	1,938	0,6776	Sí

4.3.5. Detección de Valores Aberrantes

Para verificar la validez del Software al momento de detectar la presencia de valores aberrantes o atípicos se generaron un total de 15 muestras de tamaño 100. Todas ellas fueron contaminadas con valores aberrantes. En la Tabla 4.6 se presentan los resultados obtenidos.

Tabla 4.6
Validación Estadística del Software ERLA
"Detección de Valores Aberrantes"

MUESTRA NO.	***VALORES ABERRANTES COLOCADOS***		***VALORES ABERRANTES DETECTADOS***	
	No.	***Observación***	***No.***	***Observación***
1	1	10	1	10
2	1	18	1	18
3	1	31	1	31
4	1	5	2	5, 83
5	1	50	1	50
6	1	51	2	51, 20
7	1	66	1	66
8	1	70	1	70
9	1	92	1	92
10	2	10, 90	2	10, 90
11	2	1, 100	2	1, 100
12	2	1, 2	3	1, 2, 36
13	2	25, 75	2	25, 75
14	2	99, 100	2	99, 100
15	2	6, 7	2	6, 7

El Software detectó la presencia de valores aberrantes un 100% de las veces, y en tres ocasiones se detectaron más observaciones que pueden considerarse como atípicas potenciales; esto se debe a lo que se conoce como efecto enmascaramiento o inundación, los cuales se mencionaron en el Capítulo 2.

4.3.6. Detección de Puntos de Influencia

Para verificar la validez del Software al momento de detectar puntos de influencias se generó un total de 15 muestras de tamaño 100. Todas ellas fueron contaminadas con puntos de influencia.

Tabla 4.7
Validación Estadística del Software ERLA
"Detección de Puntos de Influencia: Apalancamiento"

MUESTRA NO.	***PUNTOS DE INFLUENCIA COLOCADOS***		***PUNTOS DE INFLUENCIA DETECTADOS***	
	No.	***Observación***	***No.***	***Observación***
1	1	29	1	84
2	1	5	1	84
3	1	55	1	84
4	1	78	1	84
5	1	99	1	84
6	2	7, 74	1	84
7	2	30, 47	1	84
8	2	30, 60	1	84
9	2	42, 85	1	84
10	3	15, 49, 87	1	84
11	1	17	2	17, 84
12	1	50	2	50, 84
13	2	13, 67	2	13, 67
14	2	33, 50	3	33, 50, 84
15	3	2, 21, 45	3	2, 21, 45

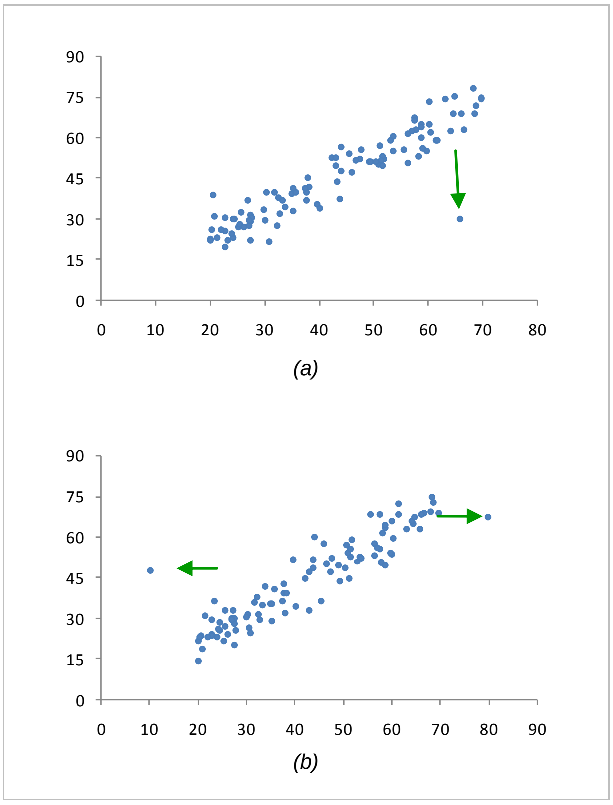
En la Tabla 4.7 se presentan los resultados obtenidos calculando el vector de apalancamiento y en la Tabla 4.8 se presentan los resultados obtenidos calculando el vector de distancias de Cook.

Tabla 4.8
Validación Estadística del Software ERLA
"Detección de Puntos de Influencia: Distancia de Cook"

MUESTRA NO.	PUNTOS DE INFLUENCIA COLOCADOS		PUNTOS DE INFLUENCIA DETECTADOS	
	No.	Observación	No.	Observación
1	1	29	1	29
2	1	5	1	5
3	1	55	1	55
4	1	78	1	78
5	1	99	1	99
6	2	7, 74	1	7
7	2	30, 47	2	30, 47
8	2	30, 60	2	30, 60
9	2	42, 85	1	85
10	3	15, 49, 87	2	49, 87
11	1	17	1	17
12	1	50	1	50
13	2	13, 67	2	13, 67
14	2	33, 50	2	33, 50
15	3	2, 21, 45	2	2, 21

El vector de apalancamientos, falla un 67% de las veces y 87% de las veces dice que la observación 84 es un punto influyente cuando en realidad no lo es. La explicación a esta situación puede observarse en la Figura 4.6. Nótese que las muestras en las cuales aparece la observación 84 como punto influyente, los puntos de influencia fueron colocados en la dirección vertical entre el mínimo y máximo valor de la variable X, mientras que en las muestras restantes éstos fueron colocados en la dirección horizontal.

Figura 4.6
Validación Estadística del Software ERLA
"Inserción de Puntos de Influencia"



La Distancia de Cook tuvo un 73% de éxitos al momento de detectar puntos de influencia. La Distancia de Cook es mucho mejor que el vector de Apalancamientos al momento de detectar Puntos de Influencia.

4.3.7. Multicolinealidad

Para verificar la validez del Software al momento de detectar problemas de multicolinealidad a través de los métodos mencionados en el capítulo 2, estos son el Factor de Inflación de la Varianza (FIV) y el Número de condición; se generó un total de 5 muestras de tamaño 50 para un modelo de Regresión Lineal Múltiple con dos variables de explicación. La correlación entre las variables de explicación X_1 y X_2 varía para cada muestra de la siguiente manera:

1. Muy alta: 0.99
2. Alta: 0.95
3. Media: 0.80
4. Baja: 0.46
5. Muy baja: 0.28

En la Tabla 4.9 puede observarse el FIV y el número de condición encontrado para cada una de las muestras.

Tabla 4.9
Validación Estadística del Software ERLA
"Diagnóstico de Multicolinealidad"

MUESTRA NO.	R^2 DEL MODELO DE REGRESIÓN ENTRE X_1 y X_2	FACTOR DE INFLACIÓN DE LA VARIANZA (FAV)		NÚMERO DE CONDICIÓN	
		VALOR	Multic.	VALOR	Multic.
1	0.99	71.35	Sí	4212,68	Sí
2	0.90	9.98	Sí	600,63	Sí
3	0.63	2.74	No	119,11	Sí
4	0.21	1.26	No	37,3	No
5	0.08	1.09	No	22,03	No

El diagnóstico no se aleja de la realidad ya que en las dos primeras muestras las variables X_1 y X_2 están altamente correlacionadas. Sin embargo puede observarse que estos métodos difieren en el diagnóstico de la muestra 3 donde, a pesar de que el coeficiente de correlación entre las variables X_1 y X_2 es de 0.80, el R^2 del modelo de regresión es 0.63. Esta diferencia depende de lo estricto que puede ser el método que se esté utilizando, finalmente quien tiene la última palabra es el investigador.

CAPÍTULO 5

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

1. El no cumplimiento de los supuestos en un análisis de regresión lineal hace que los estimadores de los coeficientes del modelo dejen de ser eficientes, los intervalos de confianza y las pruebas de hipótesis basadas en las distribuciones t y F dejan de ser confiables. El modelo se vuelve inestable, en el sentido de que muestras diferentes pueden conducir a modelos diferentes.

2. Un R^2 alto no siempre garantiza que el modelo se ajuste correctamente a los datos.

3. La presencia de valores aberrantes y puntos de influencia en un modelo de regresión lineal pueden disminuir la potencia de explicación del modelo.

4. Son muchos los métodos estadísticos que permiten evaluar la calidad del modelo de regresión lineal, algunos más potentes y efectivos que otros.

5. Los métodos utilizados para la verificación de la calidad del modelo de Regresión Lineal proporcionados por ERLA, han sido estadísticamente validados, lo cual garantiza la efectividad de sus resultados.

5.2. Recomendaciones

1. Verificar siempre que los supuestos o premisas bajo los cuales se trabaja en un análisis de regresión lineal se cumplan, puesto que la calidad del modelo encontrado puede verse afectada y las conclusiones finales pueden ser erradas.

2. Se debe tener mucho cuidado si desea eliminar valores aberrantes y puntos influyentes, ya que estos no siempre provienen de un error de

medición o digitación y en estos casos debe considerarse el uso de técnicas robustas de estimación que no sean tan sensibles a puntos influyentes como los mínimos cuadrados.

REFERENCIAS BIBLIOGRÁFICAS

[1] **FAREBROTHER R. W.** (1980), “Pan's Procedure for the Tail Probabilities of the Durbin-Watson Statistic”, *Applied Statistics* 29, 224 – 227.

[2] **HOCKING R.** (1996), “Methods and Applications of Linear Models: Regression and the Analysis of Variance”, Editorial Wiley, New York – U.S.A, 306 – 309.

[3] **MADDALA, G. S.** (1996), “Introducción a la Econometría”, Editorial PRENTICE-HALL HISPANOAMERICANA, México.

[4] **ZURITA G. Y VERA. F.** (1999), “Algoritmo para calcular la inversa generalizada de una matriz”, Acta Científica Ecuatoriana, 120 – 122.

[5] **ATKINSON, A. & RIANI, M.** (2000), **“*Robust Diagnostic Regression Analysis*”**, Editorial Springer, New York – U.S.A.

[6] **MONTGOMERY, D.** (2002), **“*Introducción al Análisis de Regresión Lineal*”**, Editorial Continental, México-México.

[7] **SEBER, A. & LEE, A.** (2003), **"*Linear Regression Analysis*"**, (2da Edición), Editorial Wiley, New York – U.S.A.

[8] **GUJARATI, D.** (2004), **"*Econometría Básica*"**, (4ta Edición), Editorial Mc Graw Hill, México-México.

[9] **BOVAS, A. & JOHANNES, L.** (2006), **"*Introduction to Regression Modeling*"**, Editorial Thomson, Belmont-U.S.A.

[10] **CHIANG, J.** (2007), **"*The Masking and Swamping Effects Using the Planted Mean-Shift Outliers Models*"**, Contemp. Math. Sciences, *Vol.2* No 7, 297-307, Taiwan-China.

[11] **MORILLAS, A. & DÍAZ, B.** (2007), **"*El Problema de los Outliers Multivariantes en el Análisis de Sectores Clave y Cluster Industrial*"**, Universidad de Málaga, España.

[12] **RODRÍGUEZ, L. (2007)**, **"*MATLAB Conceptos básicos y programación*"**, Escuela Superior Politécnica del Litoral, Guayaquil – Ecuador.

[13] **MATHWORKS**. (2008), **"*MATLAB - The Language Of Technical Computing*"**, Obtenido el: 12 de mayo de 2008, desde http://www.mathworks.com/products/MATLAB/?s_cid=global_nav.

[14] **ZURITA, G.** (2010), **"*Probabilidad y Estadística: Fundamentos y Aplicaciones*"**, (2da Edición), Talleres Gráficos ESPOL, Guayaquil-Ecuador.

[15] **ACUÑA FERNÁNDEZ, E. "*Diagnósticos de Regresión*"**, Universidad de Puerto Rico, obtenido en agosto de 2010 desde http://math.uprm.edu/~edgar/cap3sl.ppt

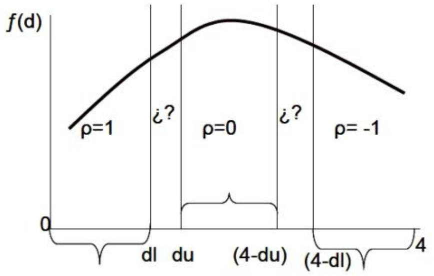
[16] **ACUÑA FERNÁNDEZ, E. "*Multicolinealidad*"**, Universidad de Puerto Rico, obtenido en agosto de 2010 desde http://math.uprm.edu/~edgar/cap7sl.ppt

[17] **ACUÑA FERNÁNDEZ, E. "*Regresión Lineal Múltiple*"**, Universidad de Puerto Rico, obtenido en agosto de 2010 desde http://math.uprm.edu/~edgar/cap2sl.pdf

[18] **RAMIREZ, D. "*Autocorrelación*"**, obtenido en septiembre de 2010 desde http://webdelprofesor.ula.ve/economia/dramirez/MICRO/FORMATO_PDF/Materialeconometria/Autocorrelacion.pdf

ANEXOS

ANEXO 1. TABLAS DURBIN WATSON

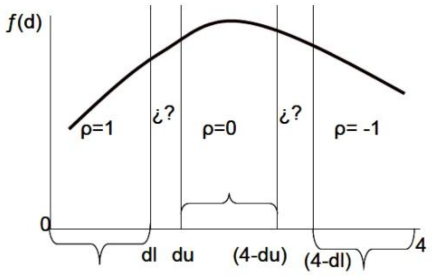


Estadístico Durbin-Watson α=0.01

K es el número de parámetros a estimar

	K=1		K=2		K=3		K=4		K=5		K=6		K=7		K=8		K=9		K=10	
n	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU
6	0.390	1.142	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
7	0.435	1.036	0.294	1.676	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
8	0.497	1.003	0.345	1.489	0.229	2.102	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9	0.554	0.998	0.408	1.389	0.279	1.875	0.183	2.433	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
10	0.604	1.001	0.466	1.333	0.340	1.733	0.230	2.193	0.150	2.690	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
11	0.653	1.010	0.519	1.297	0.396	1.640	0.286	2.030	0.193	2.453	0.124	2.892	-----	-----	-----	-----	-----	-----	-----	-----
12	0.697	1.023	0.569	1.274	0.449	1.575	0.339	1.913	0.244	2.280	0.164	2.665	0.105	3.053	-----	-----	-----	-----	-----	-----
13	0.738	1.038	0.616	1.261	0.499	1.526	0.391	1.826	0.294	2.150	0.211	2.490	0.140	2.838	0.090	3.182	-----	-----	-----	-----
14	0.776	1.054	0.660	1.254	0.547	1.490	0.441	1.757	0.343	2.049	0.257	2.354	0.183	2.667	0.122	2.981	0.078	3.287	-----	-----
15	0.811	1.070	0.700	1.252	0.591	1.465	0.487	1.705	0.390	1.967	0.303	2.244	0.226	2.530	0.161	2.817	0.107	3.101	0.068	3.374
16	0.844	1.086	0.738	1.253	0.633	1.447	0.532	1.664	0.437	1.901	0.349	2.153	0.269	2.416	0.200	2.681	0.142	2.944	0.094	3.201
17	0.873	1.102	0.773	1.255	0.672	1.432	0.574	1.631	0.481	1.847	0.393	2.078	0.313	2.319	0.241	2.566	0.179	2.811	0.127	3.053
18	0.902	1.118	0.805	1.259	0.708	1.422	0.614	1.604	0.522	1.803	0.435	2.015	0.355	2.238	0.282	2.467	0.216	2.697	0.160	2.925
19	0.928	1.133	0.835	1.264	0.742	1.416	0.650	1.583	0.561	1.767	0.476	1.963	0.396	2.169	0.322	2.381	0.255	2.597	0.196	2.813
20	0.952	1.147	0.862	1.270	0.774	1.410	0.684	1.567	0.598	1.736	0.515	1.918	0.436	2.110	0.362	2.308	0.294	2.510	0.232	2.174
21	0.975	1.161	0.889	1.276	0.803	1.408	0.718	1.554	0.634	1.712	0.552	1.881	0.474	2.059	0.400	2.244	0.331	2.434	0.268	2.625
22	0.997	1.174	0.915	1.284	0.832	1.407	0.748	1.543	0.666	1.691	0.587	1.849	0.510	2.015	0.437	2.188	0.368	2.367	0.304	2.548
23	1.017	1.186	0.938	1.290	0.858	1.407	0.777	1.535	0.699	1.674	0.620	1.821	0.545	1.977	0.473	2.140	0.404	2.308	0.340	2.479
24	1.037	1.199	0.959	1.298	0.881	1.407	0.805	1.527	0.728	1.659	0.652	1.797	0.578	1.944	0.507	2.097	0.439	2.255	0.375	2.417
25	1.055	1.210	0.981	1.305	0.906	1.408	0.832	1.521	0.756	1.645	0.682	1.776	0.610	1.915	0.540	2.059	0.473	2.209	0.409	2.362
26	1.072	1.222	1.000	1.311	0.928	1.410	0.855	1.517	0.782	1.635	0.711	1.759	0.640	1.889	0.572	2.026	0.505	2.168	0.441	2.313
27	1.088	1.232	1.019	1.318	0.948	1.413	0.878	1.514	0.808	1.625	0.738	1.743	0.669	1.867	0.602	1.997	0.536	2.131	0.473	2.269
28	1.104	1.244	1.036	1.325	0.969	1.414	0.901	1.512	0.832	1.618	0.764	1.729	0.696	1.847	0.630	1.970	0.566	2.098	0.504	2.229
29	1.119	1.254	1.053	1.332	0.988	1.418	0.921	1.511	0.855	1.611	0.788	1.718	0.723	1.830	0.658	1.947	0.595	2.068	0.533	2.193
30	1.134	1.264	1.070	1.339	1.006	1.421	0.941	1.510	0.877	1.606	0.812	1.707	0.748	1.814	0.684	1.925	0.622	2.041	0.562	2.160
31	1.147	1.274	1.085	1.345	1.022	1.425	0.960	1.509	0.897	1.601	0.834	1.698	0.772	1.800	0.710	1.906	0.649	2.017	0.589	2.131
32	1.160	1.283	1.100	1.351	1.039	1.428	0.978	1.509	0.917	1.597	0.856	1.690	0.794	1.788	0.734	1.889	0.674	1.995	0.615	2.104
33	1.171	1.291	1.114	1.358	1.055	1.432	0.995	1.510	0.935	1.594	0.876	1.683	0.816	1.776	0.757	1.874	0.698	1.975	0.641	2.080
34	1.184	1.298	1.128	1.364	1.070	1.436	1.012	1.511	0.954	1.591	0.896	1.677	0.837	1.766	0.779	1.860	0.722	1.957	0.665	2.057
35	1.195	1.307	1.141	1.370	1.085	1.439	1.028	1.512	0.971	1.589	0.914	1.671	0.857	1.757	0.800	1.847	0.744	1.940	0.689	2.037
36	1.205	1.315	1.153	1.376	1.098	1.442	1.043	1.513	0.987	1.587	0.932	1.666	0.877	1.749	0.821	1.836	0.766	1.925	0.711	2.018
37	1.217	1.322	1.164	1.383	1.112	1.446	1.058	1.514	1.004	1.585	0.950	1.662	0.895	1.742	0.841	1.825	0.787	1.911	0.733	2.001
38	1.227	1.330	1.176	1.388	1.124	1.449	1.072	1.515	1.019	1.584	0.966	1.658	0.913	1.735	0.860	1.816	0.807	1.899	0.754	1.985
39	1.237	1.337	1.187	1.392	1.137	1.452	1.085	1.517	1.033	1.583	0.982	1.655	0.930	1.729	0.878	1.807	0.826	1.887	0.774	1.970
40	1.246	1.344	1.197	1.398	1.149	1.456	1.098	1.518	1.047	1.583	0.997	1.652	0.946	1.724	0.895	1.799	0.844	1.876	0.749	1.956
45	1.288	1.376	1.245	1.424	1.201	1.474	1.156	1.528	1.111	1.583	1.065	1.643	1.019	1.704	0.974	1.768	0.927	1.834	0.881	1.902
50	1.324	1.403	1.285	1.445	1.245	1.491	1.206	1.537	1.164	1.587	1.123	1.639	1.081	1.692	1.039	1.748	0.997	1.805	0.955	1.864
55	1.356	1.428	1.320	1.466	1.284	1.505	1.246	1.548	1.209	1.592	1.172	1.638	1.134	1.685	1.095	1.734	1.057	1.785	1.018	1.837
60	1.382	1.449	1.351	1.484	1.317	1.520	1.283	1.559	1.248	1.598	1.214	1.639	1.179	1.682	1.144	1.726	1.108	1.771	1.072	1.817
65	1.407	1.467	1.377	1.500	1.346	1.534	1.314	1.568	1.283	1.604	1.251	1.642	1.218	1.680	1.186	1.720	1.153	1.761	1.120	1.802
70	1.429	1.485	1.400	1.514	1.372	1.546	1.343	1.577	1.313	1.611	1.283	1.645	1.253	1.680	1.223	1.716	1.192	1.754	1.162	1.792
75	1.448	1.501	1.422	1.529	1.395	1.557	1.368	1.586	1.340	1.617	1.313	1.649	1.284	1.682	1.256	1.714	1.227	1.748	1.199	1.783
80	1.465	1.514	1.440	1.541	1.416	1.568	1.390	1.595	1.364	1.624	1.338	1.653	1.312	1.683	1.285	1.714	1.259	1.745	1.232	1.777
85	1.481	1.529	1.458	1.553	1.434	1.577	1.411	1.603	1.386	1.630	1.362	1.657	1.337	1.685	1.312	1.714	1.287	1.743	1.262	1.773
90	1.496	1.541	1.474	1.563	1.452	1.587	1.429	1.611	1.406	1.636	1.383	1.661	1.360	1.687	1.336	1.714	1.312	1.741	1.288	1.769
95	1.510	1.552	1.489	1.573	1.468	1.596	1.446	1.618	1.425	1.641	1.403	1.666	1.381	1.690	1.358	1.715	1.336	1.741	1.313	1.767
100	1.522	1.562	1.502	1.582	1.482	1.604	1.461	1.625	1.441	1.647	1.421	1.670	1.400	1.693	1.378	1.717	1.357	1.741	1.335	1.765
150	1.611	1.637	1.598	1.651	1.584	1.665	1.571	1.679	1.557	1.693	1.543	1.708	1.530	1.722	1.515	1.737	1.501	1.752	1.486	1.767
200	1.664	1.684	1.653	1.693	1.643	1.704	1.633	1.715	1.623	1.725	1.613	1.735	1.603	1.746	1.592	1.757	1.582	1.768	1.571	1.779

VIENE… ANEXO 1. TABLAS DURBIN WATSON

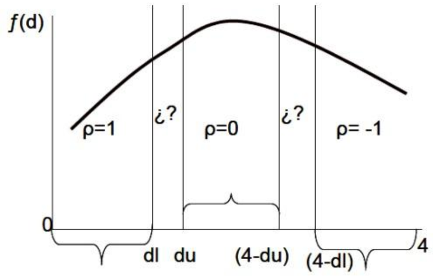


Estadístico Durbin-Watson α=0.01

K es el número de parámetros a estimar

	K=11		K=12		K=13		K=14		K=15		K=16		K=17		K=18		K=19		K=20	
n	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU
16	0.060	3.446	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
17	0.084	3.286	0.053	3.506	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
18	0.113	3.146	0.075	3.358	0.047	3.557	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
19	0.145	3023	0.102	3.227	0.067	3.420	0.043	3.601	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
20	0.178	2.914	0.131	3.109	0.092	3.297	0.061	3.474	0.038	3.639	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
21	0.212	2.817	0.162	3.004	0.119	3.185	0.084	3.358	0.055	3.521	0.035	3.671	-----	-----	-----	-----	-----	-----	-----	-----
22	0.246	2.729	0.194	2.909	0.148	3.084	0.109	3.252	0.077	3.412	0.050	3.562	0.032	3.700	-----	-----	-----	-----	-----	-----
23	0.281	2.651	0.227	2.822	0.178	2.991	0.136	3.155	0.100	3.311	0.070	3.459	0.046	3.597	0.029	3.725	-----	-----	-----	-----
24	0.315	2.580	0.260	2.744	0.209	2.906	0.165	3.065	0.125	3.218	0.092	3.363	0.065	3.501	0.043	3.629	0.027	3.747	-----	-----
25	0.348	2.517	0.292	2.674	0.240	2.829	0.194	2.982	0.152	3.131	0.116	3.274	0.085	3.410	0.060	3.538	0.039	3.657	0.025	3.766
26	0.381	2.460	0.324	2.610	0.272	2.758	0.224	2.906	0.180	3.050	0.141	3.191	0.107	3.325	0.079	3.452	0.055	3.572	0.036	3.682
27	0.413	2.409	0.356	2.552	0.303	2.694	0.253	2.836	0.208	2.976	0.167	3.113	0.131	3.245	0.100	3.371	0.073	3.490	0.051	3.602
28	0.444	2.363	0.387	2.499	0.333	2.635	0.283	2.772	0.237	2.907	0.194	3.040	0.156	3.169	0.122	3.294	0.093	3.412	0.068	3.524
29	0.474	2.321	0.417	2.451	0.363	2.582	0.313	2.713	0.266	2.843	0.222	2.972	0.182	3.098	0.146	3.220	0.114	3.338	0.087	3.450
30	0.503	2.283	0.447	2.407	0.393	2.533	0.342	2.659	0.294	2.785	0.249	2.909	0.208	3.032	0.171	3.152	0.137	3.267	0.107	3.379
31	0.531	2.248	0.475	2.367	0.422	2.487	0.371	2.609	0.322	2.730	0.277	2.851	0.234	2.970	0.193	3.087	0.160	3.201	0.128	3.311
32	0.558	2.216	0.503	2.330	0.450	2.446	0.399	2.563	0.350	2.680	0.304	2.797	0.261	2.912	0.221	3.026	0.184	3.137	0.151	3.246
33	0.585	2.187	0.530	2.296	0.477	2.408	0.426	2.520	0.377	2.633	0.331	2.746	0.287	2.858	0.246	2.969	0.209	3.078	0.174	3.184
34	0.610	2.160	0.556	2.266	0.503	2.373	0.452	2.481	0.404	2.590	0.357	2.699	0.313	2.808	0.272	2.915	0.233	3.022	0.197	3.126
35	0.634	2.136	0.581	2.237	0.529	2.340	0.478	2.444	0.430	2.550	0.383	2.655	0.339	2.761	0.297	2.865	0.257	2.969	0.221	3.071
36	0.658	2.113	0.605	2.210	0.554	2.310	0.504	2.410	0.455	2.512	0.409	2.614	0.364	2.717	0.322	2.818	0.282	2.919	0.244	3.019
37	0.680	2.092	0.628	2.186	0.578	2.282	0.528	2.379	0.480	2.477	0.434	2.576	0.389	2.675	0.347	2.774	0.306	2.872	0.268	2.969
38	0.702	2.073	0.651	2.164	0.601	2.256	0.552	2.350	0.504	2.445	0.458	2.540	0.414	2.637	0.371	2.733	0.330	2.828	0.291	2.923
39	0.723	2.055	0.673	2.143	0.623	2.232	0.575	2.323	0.528	2.414	0.482	2.507	0.438	2.600	0.395	2.694	0.354	2.787	0.315	2.879
40	0.744	2.039	0.694	2.123	0.645	2.210	0.597	2.297	0.551	2.386	0.505	2.476	0.461	2.566	0.418	2.657	0.377	2.748	0.338	2.838
45	0.835	1.972	0.790	2.044	0.744	2.118	0.700	2.193	0.655	2.269	0.612	2.346	0.570	2.424	0.528	2.503	0.488	2.582	0.448	2.661
50	0.913	1.925	0.871	1.987	0.829	2.051	0.787	2.116	0.746	2.182	0.705	2.250	0.665	2.318	0.625	2.387	0.586	2.456	0.548	2.526
55	0.979	1.891	0.940	1.945	0.902	2.002	0.863	2.059	0.825	2.117	0.786	2.176	0.748	2.237	0.711	2.298	0.674	2.359	0.637	2.421
60	1.037	1.865	1.001	1.914	0.965	1.964	0.929	2.015	0.893	2.067	0.857	2.120	0.822	2.173	0.786	2.227	0.751	2.283	0.716	2.338
65	1.087	1.845	1.053	1.889	1.020	1.934	0.986	1.980	0.953	2.027	0.919	2.075	0.886	2.123	0.852	2.172	0.819	2.221	0.789	2.272
70	1.131	1.831	1.099	1.870	1.068	1.911	1.037	1.953	1005	1.995	0.974	2.038	0.943	2.082	0.911	2.127	0.880	2.172	0.849	2.217
75	1.170	1.819	1.141	1.856	1.111	1.893	1.082	1.931	1.052	1.970	1023	2.009	0.993	2.049	0.964	2.090	0.934	2.131	0.905	2.172
80	1.205	1.810	1.177	1.844	1.150	1.878	1.122	1.913	1.094	1.949	1066	1.984	1039	2.022	1011	2.059	0.983	2.097	0.955	2.135
85	1.236	1.803	1.210	1.834	1.184	1.866	1.158	1.898	1.132	1.931	1106	1.965	1080	1.999	1053	2.033	1027	2.068	1000	2.104
90	1.264	1.798	1.240	1.827	1.215	1.856	1.191	1.886	1.166	1.917	1141	1.948	1116	1.979	1091	2.012	1066	2.044	1041	2.077
95	1.290	1.793	1.267	1.821	1.244	1.848	1.221	1.876	1.197	1.905	1.174	1.943	1.150	1.963	1126	1.993	1102	2.023	1079	2.054
100	1.314	1.790	1.292	1.816	1.270	1.841	1.248	1.868	1.225	1.895	1.203	1.922	1.181	1.949	1.158	1.977	1136	2.006	1113	2.034
150	1.473	1.783	1.458	1.799	1.444	1.814	1.429	1.830	1.414	1.847	1.400	1.863	1.385	1.880	1.370	1.897	1.355	1.913	1.340	1.931
200	1.561	1.791	1.550	1.801	1.539	1.813	1.528	1.824	1.518	1.836	1.507	1.847	1.495	1.860	1.484	1.871	1.474	1.883	1.462	1.896

VIENE… ANEXO 1. TABLAS DURBIN WATSON

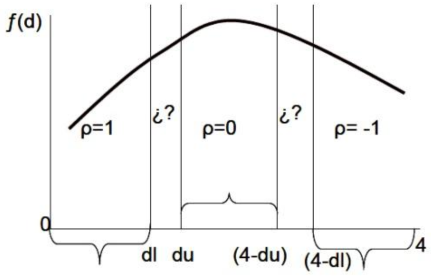


Estadístico Durbin-Watson α=0.05

K es el número de parámetros a estimar

	K=1		K=2		K=3		K=4		K=5		K=6		K=7		K=8		K=9		K=10	
n	dL		dL		dL		dL		dL		dL		dL		dL		dL		dL	
6	0.610	1400	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
7	0.700	1.356	0.467	1896	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
8	0.763	1.332	0.559	1.777	0.367	2287	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2588	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
10	0.879	1320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2822	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
11	0.927	1.324	0.758	1.604	0.595	1.928	0.444	2.283	0.315	2.645	0.203	3004	-----	-----	-----	-----	-----	-----	-----	-----
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.380	2.506	0.268	2.832	0.171	3149	-----	-----	-----	-----	-----	-----
13	1010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.444	2.390	0.328	2.692	0.230	2.985	0.147	3266	-----	-----	-----	-----
14	1045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3360	-----	-----
15	1077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.471	0.343	2.727	0.251	2.979	0.175	3.216	0.111	3438
16	1106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090	0.155	3.304
17	1133	1.381	1015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	2.537	0.356	2.757	0.272	2.975	0.198	3.184
18	1158	1.391	1046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.258	0.502	2.461	0.407	2.668	0.321	2.873	0.244	3.073
19	1180	1.401	1074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783	0.290	2.974
20	1201	1.411	1100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.691	2.162	0.595	2.339	0.502	2.521	0.416	2.704	0.336	2.885
21	1221	1.420	1125	1.538	1026	1.669	0.927	1.812	0.829	1.964	0.731	2.124	0.637	2.290	0.546	2.461	0.461	2.633	0.380	2.806
22	1239	1.429	1147	1.541	1053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571	0.424	2.735
23	1257	1.437	1168	1.543	1078	1.660	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514	0.465	2.670
24	1.273	1.446	1188	1.546	1101	1.656	1013	1.775	0.925	1.902	0.837	2.035	0.750	2.174	0.666	2.318	0.584	2.464	0.506	2.613
25	1.288	1.454	1206	1.550	1123	1.654	1038	1.767	0.953	1.886	0.868	2.013	0.784	2.144	0.702	2.280	0.621	2.419	0.544	2.560
26	1.302	1.461	1224	1.553	1143	1.652	1062	1.759	0.979	1.873	0.897	1.992	0.816	2.117	0.735	2.246	0.657	2.379	0.581	2.513
27	1.316	1.469	1.240	1.556	1162	1.651	1084	1.753	1004	1.861	0.925	1.974	0.845	2.093	0.767	2.216	0.691	2.342	0.616	2.470
28	1.328	1.476	1.255	1.560	1181	1.650	1104	1.747	1028	1.850	0.951	1.959	0.874	2.071	0.798	2.188	0.723	2.309	0.649	2.431
29	1.341	1.483	1.270	1.563	1198	1.650	1124	1.743	1050	1.841	0.975	1.944	0.900	2.052	0.826	2.164	0.753	2.278	0.681	2.396
30	1.352	1.489	1.284	1.567	1214	1.650	1143	1.739	1071	1.833	0.998	1.931	0.926	2.034	0.854	2.141	0.782	2.251	0.712	2.363
31	1.363	1.496	1.297	1.570	1.229	1.650	1160	1.735	1090	1.825	1020	1.920	0.950	2.018	0.879	2.120	0.810	2.226	0.741	2.333
32	1.373	1.502	1.309	1.574	1.244	1.650	1177	1.732	1109	1.819	1041	1.909	0.972	2.004	0.904	2.102	0.836	2.203	0.769	2.306
33	1.383	1.508	1.321	1.577	1.258	1.651	1193	1.730	1127	1.813	1061	1.900	0.994	1.991	0.927	2.085	0.861	2.181	0.796	2.281
34	1.393	1.514	1.333	1.580	1.271	1.652	1208	1.728	1144	1.808	1079	1.891	1015	1.978	0.950	2.069	0.885	2.162	0.821	2.257
35	1.402	1.519	1.343	1.584	1.283	1.653	1.222	1.726	1160	1.803	1097	1.884	1034	1.967	0.971	2.054	0.908	2.144	0.845	2.236
36	1.411	1.525	1.354	1.587	1.295	1.654	1.236	1.724	1175	1.799	1114	1.876	1053	1.957	0.991	2.041	0.930	2.127	0.868	2.216
37	1.419	1.530	1.364	1.590	1.307	1.655	1.249	1.723	1190	1.795	1131	1.870	1071	1.948	1011	2.029	0.951	2.112	0.891	2.197
38	1.427	1.535	1.373	1.594	1.318	1.656	1.261	1.722	1.204	1.792	1146	1.864	1088	1.939	1029	2.017	0.970	2.098	0.912	2.180
39	1.435	1.540	1.382	1.597	1.328	1.658	1.273	1.722	1.218	1.789	1161	1.859	1104	1.932	1047	2.007	0.990	2.085	0.932	2.164
40	1.442	1.544	1.391	1.600	1.338	1.659	1.285	1.721	1.230	1.786	1175	1.854	1120	1.924	1064	1.997	1008	2.072	0.952	2.149
45	1.475	1.566	1.430	1.615	1.383	1.666	1.336	1.720	1.287	1.776	1238	1.835	1189	1.895	1139	1.958	1089	2.022	1038	2.088
50	1.503	1.585	1.462	1.628	1.421	1.674	1.378	1.721	1.335	1.771	1.291	1.822	1.246	1.875	1201	1.930	1156	1.986	1110	2.044
55	1.528	1.601	1.490	1.641	1.452	1.681	1.414	1.724	1.374	1.768	1.334	1.814	1.294	1.861	1.253	1.909	1212	1.959	1170	2.010
60	1.549	1.616	1.514	1.652	1.480	1.689	1.444	1.727	1.408	1.767	1.372	1.808	1.335	1.850	1.298	1.894	1.260	1.939	1.222	1.984
65	1.567	1.629	1.536	1.662	1.503	1.696	1.471	1.731	1.438	1.767	1.404	1.805	1.370	1.843	1.336	1.882	1.301	1.923	1.266	1.964
70	1.583	1.641	1.554	1.672	1.525	1.703	1.494	1.735	1.464	1.768	1.433	1.802	1.401	1.838	1.369	1.874	1.337	1.910	1.305	1.948
75	1.598	1.652	1.571	1.680	1.543	1.709	1.515	1.739	1.487	1.770	1.458	1.801	1.428	1.834	1.399	1.867	1.369	1.901	1.339	1.935
80	1.611	1.662	1.586	1.688	1.560	1.715	1.534	1.743	1.507	1.772	1.480	1.801	1.453	1.831	1.425	1.861	1.397	1.893	1.369	1.925
85	1.624	1.671	1.600	1.696	1.575	1.721	1.550	1.747	1.525	1.774	1.500	1.801	1.474	1.829	1.448	1.857	1.422	1.886	1.396	1.916
90	1.635	1.679	1.612	1.703	1.589	1.726	1.566	1.751	1.542	1.776	1.518	1.801	1.494	1.827	1.469	1.854	1.445	1.881	1.420	1.909
95	1.645	1.687	1.623	1.709	1.602	1.732	1.579	1.755	1.557	1.778	1.535	1.802	1.512	1.827	1.489	1.852	1.465	1.877	1.442	1.903
100	1.654	1.694	1.634	1.715	1.613	1.736	1.592	1.758	1.571	1.780	1.550	1.803	1.528	1.826	1.506	1.850	1.484	1.874	1.462	1.898
150	1.720	1.747	1.706	1.760	1.693	1.774	1.679	1.788	1.665	1.802	1.651	1.817	1.637	1.832	1.622	1.846	1.608	1.862	1.593	1.877
200	1.758	1.779	1.748	1.789	1.738	1.799	1.728	1.809	1.718	1.820	1.707	1.831	1.697	1.841	1.686	1.852	1.675	1.863	1.665	1.874

VIENE… ANEXO 1. TABLAS DURBIN WATSON



Estadístico Durbin-Watson α=0.05

K es el número de parámetros a estimar

	K=11		K=12		K=13		K=14		K=15		K=16		K=17		K=18		K=19		K=20	
n	dL		dL		dL		dL		dL		dL		dL		dL		dL		dL	
16	0.098	3503	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
17	0.138	3.378	0.087	3557	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
18	0.177	3.265	0.123	3.441	0.078	3603	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
19	0.220	3.159	0.160	3.335	0.111	3.496	0.070	3642	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
20	0.263	3063	0.200	3.234	0.145	3.395	0.100	3.542	0.063	3676	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
21	0.307	2.976	0.240	3.141	0.182	3.300	0.132	3.448	0.091	3.583	0.058	3705	-----	-----	-----	-----	-----	-----	-----	-----
22	0.349	2.897	0.281	3.057	0.220	3.211	0.166	3.358	0.120	3.495	0.083	3.619	0.052	3731	-----	-----	-----	-----	-----	-----
23	0.391	2.826	0.322	2.979	0.259	3.128	0.202	3.272	0.153	3.409	0.110	3.535	0.076	3.650	0.048	3753	-----	-----	-----	-----
24	0.431	2.761	0.362	2.908	0.297	3.053	0.239	3.193	0.186	3.327	0.141	3.454	0.101	3.572	0.070	3.678	0.044	3773	-----	-----
25	0.470	2.702	0.400	2.844	0.335	2.983	0.275	3.119	0.221	3.251	0.172	3.376	0.130	3.494	0.094	3.604	0.065	3.702	0.041	3790
26	0.508	2.649	0.438	2.784	0.373	2.919	0.312	3.051	0.256	3.179	0.205	3.303	0.160	3.420	0.120	3.531	0.087	3.632	0.060	3.724
27	0.544	2.600	0.475	2.730	0.409	2.859	0.348	2.987	0.291	3.112	0.238	3.233	0.191	3.349	0.149	3.460	0.112	3.563	0.081	3.658
28	0.578	2.555	0.510	2.680	0.445	2.805	0.383	2.928	0.325	3.050	0.271	3.168	0.222	3.283	0.178	3.392	0.138	3.495	0.104	3.592
29	0.612	2.515	0.544	2.634	0.479	2.755	0.418	2.874	0.359	2.992	0.305	3.107	0.254	3.219	0.208	3.327	0.166	3.431	0.129	3.528
30	0.643	2.477	0.577	2.592	0.512	2.708	0.451	2.823	0.392	2.937	0.337	3.050	0.286	3.160	0.238	3.266	0.195	3.368	0.156	3.465
31	0.674	2.443	0.608	2.553	0.545	2.665	0.484	2.776	0.425	2.887	0.370	2.996	0.317	3.103	0.269	3.208	0.224	3.309	0.183	3.406
32	0.703	2.411	0.638	2.517	0.576	2.625	0.515	2.733	0.457	2.840	0.401	2.946	0.349	3.050	0.299	3.153	0.253	3.252	0.211	3.348
33	0.731	2.382	0.668	2.484	0.606	2.588	0.546	2.692	0.488	2.796	0.432	2.899	0.379	3.000	0.329	3.100	0.283	3.198	0.239	3.293
34	0.758	2.355	0.695	2.454	0.634	2.554	0.575	2.654	0.518	2.754	0.462	2.854	0.409	2.954	0.359	3.051	0.312	3.147	0.267	3.240
35	0.783	2.330	0.722	2.425	0.662	2.521	0.604	2.619	0.547	2.716	0.492	2.813	0.439	2.910	0.388	3.005	0.340	3.099	0.295	3.190
36	0.808	2.306	0.748	2.398	0.689	2.492	0.631	2.586	0.575	2.680	0.520	2.774	0.467	2.868	0.417	2.961	0.369	3.053	0.323	3.142
37	0.831	2.285	0.772	2.374	0.714	2.464	0.657	2.555	0.602	2.646	0.548	2.738	0.495	2.829	0.445	2.920	0.397	3.009	0.351	3.097
38	0.854	2.265	0.796	2.351	0.739	2.438	0.683	2.526	0.628	2.614	0.575	2.703	0.522	2.792	0.472	2.880	0.424	2.968	0.378	3.054
39	0.875	2.246	0.819	2.329	0.763	2.413	0.707	2.499	0.653	2.585	0.600	2.671	0.549	2.757	0.499	2.843	0.451	2.929	0.404	3.013
40	0.896	2.228	0.840	2.309	0.785	2.391	0.731	2.473	0.678	2.557	0.626	2.641	0.575	2.724	0.525	2.808	0.477	2.829	0.430	2.974
45	0.988	2.156	0.938	2.225	0.887	2.296	0.838	2.367	0.788	2.439	0.740	2.512	0.692	2.586	0.644	2.659	0.598	2.733	0.553	2.807
50	1.064	2.103	1.019	2.163	0.973	2.225	0.927	2.287	0.882	2.350	0.836	2.414	0.792	2.479	0.747	2.544	0.703	2.610	0.660	2.675
55	1.129	2.062	1.087	2.116	1.045	2.170	1003	2.225	0.961	2.281	0.919	2.338	0.877	2.396	0.836	2.454	0.795	2.512	0.754	2.571
60	1.184	2.031	1.145	2.079	1.106	2.127	1068	2.177	1029	2.227	0.990	2.278	0.951	2.330	0.913	2.382	0.874	2.434	0.836	2.487
65	1.231	2.006	1.195	2.049	1.160	2.093	1.124	2.138	1088	2.183	1052	2.229	1016	2.276	0.980	2.323	0.944	2.371	0.908	2.419
70	1.272	1.987	1.239	2.026	1.206	2.066	1.172	2.106	1139	2.148	1105	2.189	1072	2.232	1038	2.275	1005	2.318	0.971	2.362
75	1.308	1.970	1.277	2.006	1.247	2.043	1.215	2.080	1184	2.118	1153	2.156	1121	2.195	1090	2.235	1058	2.275	1027	2.315
80	1.340	1.957	1.311	1.991	1.283	2.024	1.253	2.059	1.224	2.093	1195	2.129	1165	2.165	1136	2.201	1106	2.238	1076	2.275
85	1.369	1.946	1.342	1.977	1.315	2.009	1.287	2.040	1.260	2.073	1232	2.105	1205	2.139	1177	2.172	1149	2.206	1121	2.241
90	1.395	1.937	1.369	1.966	1.344	1.995	1.318	2.025	1.292	2.055	1266	2.085	1240	2.116	1213	2.148	1187	2.179	1160	2.211
95	1.418	1.930	1.394	1.956	1.370	1.984	1.345	2.012	1.321	2.040	1296	2.068	1271	2.097	1247	2.126	1222	2.156	1197	2.186
100	1.439	1.923	1.416	1.948	1.393	1.974	1.371	2.000	1.347	2.026	1.324	2.053	1.301	2.080	1277	2.108	1253	2.135	1229	2.164
150	1.579	1.892	1.564	1.908	1.550	1.924	1.535	1.940	1.519	1.956	1.504	1.972	1.489	1.989	1.474	2.006	1458	2.023	1443	2.040
200	1.654	1.885	1.643	1.896	1.632	1.908	1.621	1.919	1.610	1.931	1.599	1.943	1.588	1.955	1.576	1.967	1.565	1.979	1.554	1.991

ANEXO 2. NORMAS DE ASIGNACIÓN PARA CONVERSIÓN DE .NET A MATLAB

En la siguiente tabla se muestran las normas de asignación de tipo de datos para la conversión. NET con los tipos de datos de MATLAB usando las clases derivadas de MWArray.

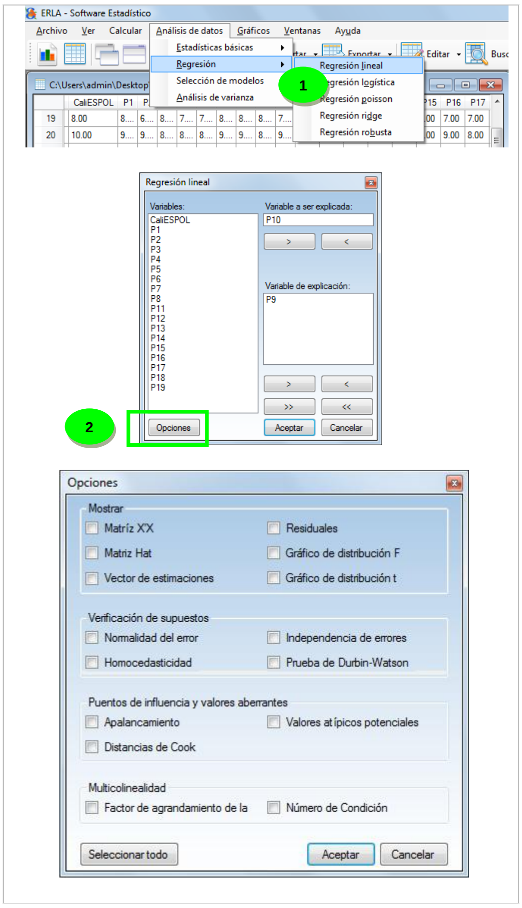
TIPO .NET	***TIPO MWARRAY***	***TIPO MATLAB***
System.Double	MWNumericArray	double
System.Number	MWNumericArray	double
System.Float	MWNumericArray	single
System.Byte	MWNumericArray	int8
System.Short	MWNumericArray	int16
System.Int32	MWNumericArray	int32
System.Int64	MWNumericArray	int64
System.Char	MWCharArray	char
System.String	MWCharArray	char
System.Boolean	MWLogicalArray	logical
N/A	MWStructArray	structure
N/A	MWCellArray	cell

ANEXO 3. VERIFICACIÓN DE CALIDAD DE MODELOS DE REGRESIÓN EN ERLA

Existen distintos métodos que permiten determinar la bondad de un modelo de regresión, ERLA incluye tales **opciones** en el cuadro de diálogo **Regresión lineal** al cual se puede acceder mediante la secuencia:

1. **Barra de menú➤ Análisis de datos➤ Regresión➤ Regresión lineal**
2. Pulse el botón **Opciones** del cuadro de diálogo **Regresión lineal**. Véase Figura 3.1.

“Medidas de calidad del modelo”



Es decir, una vez seleccionado el modelo de regresión lineal, debe pulsar el botón **Opciones** para visualizar los estadísticos que ofrece ERLA para verificar la validez del modelo. Las opciones relacionadas con la verificación de la calidad del modelo están principalmente en los grupos:

1. Verificación de supuestos.
2. Puntos de Influencia y valores aberrantes.
3. Multicolinealidad

Para más información véase Manual de usuario ERLA 2011. Capítulo 4 Sección 4. Verificación de Calidad en Modelos de Regresión.