



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

“INTRODUCCION A AJUSTES NO LINEALES EN EL ANALISIS DE REGRESION”

T E S I S

QUE PARA OBTENER EL TITULO DE

A C T U A R I A

P R E S E N T A :

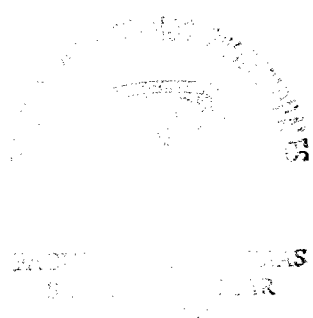
JUANA YOLANDA PERALTA MARTINEZ



DIRECTORA DE TESIS: M. EN A.P. MARIA DEL PILAR ALONSO REYES

2005

m. 343289





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL
AVENIDA DE
MEXICO

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo de tesis.

NOMBRE: Peralta Martínez Juana Yolanda

FECHA: 20 de abril del 2005

FIRMA: [Firma]

ACT. MAURICIO AGUILAR GONZÁLEZ
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:
Introducción a ajustes no lineales en el Análisis de Regresión

realizado por Peralta Martínez Juana Yolanda

con número de cuenta 09436135-5, quien cubrió los créditos de la carrera de: Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

- Director de Tesis
Propietario M. en A. P. María del Pilar Alonso Reyes
- Propietario M. en C. José Antonio Flores Díaz
- Propietario Dr. Luis Antonio Rincón Solís
- Suplente Act. Jaime Vázquez Alamilla
- Suplente Act. Marypaola Janett Maya López

Consejo Departamental de Matemáticas

Act. Jaime Vázquez Alamilla
FACULTAD DE CIENCIAS
CONSEJO DEPARTAMENTAL
DE
MATEMÁTICAS

Agradecimientos

A **Dios** por permitirme la vida y el tiempo para cumplir con mis metas.

A la Universidad por la gran oportunidad que me dio de poder estudiar, crecer académicamente, formando parte de esta institución, "**GRACIAS A MI UNIVERSIDAD**". De corazón soy "**PUMA**".

A mi Familia

Mis padres, **Cipriano Peralta** (El comprensible) y **Ofelia Martínez** (La incansable), gracias, por ser mis padres, por fomentar y apoyar decididamente mi formación profesional, por que sin eso y el tiempo que me permitieron estudiar, no hubiera podido terminar mi carrera y llegar a realizar este trabajo de investigación, gracias por el buen ejemplo que siempre me han dado, por su cariño, por su tiempo, por su paciencia, los adoro. Espero que Dios me dé la suficiente humildad, el suficiente tiempo y vida, para corresponder a todo lo que ustedes me han dado siempre.

Mi hermano, **Dieter Fernando**, que a pesar de lo que digas, desde lo más profundo de mi corazón te puedo decir que "Te amo", que agradezco al cielo que llegaras a mi vida, que eres el niño que me enseñó a querer como te quiero, a quien admiro como deportista y también por que a pesar de tus pocos años, en los momentos difíciles de mi vida has sido un gran apoyo, gracias, "Difer".

A mis sinodales:

Maestra **Ma. Pilar Alonso Reyes**, gracias, por la enorme paciencia que siempre me ha tenido, pero sobre todo por haberme apoyado, por haber notado mis carencias y simplemente ponerme a trabajar en ellas, sin más intención que hacerme sentir el compromiso de mejorar, por permitirme trabajar a su lado, por el tiempo que siempre se dio para ayudarme, por el que tal vez le hice perder, disculpas. Maestra, tendrá por siempre mi infinito agradecimiento, cariño y admiración como lo que es, una gran mujer y una excelente profesionista. Las más grandes bendiciones para usted, Alfredo y sus seres queridos.

Maestro **José Antonio** Flores, gracias, por su apoyo y orientación para hacer que mi trabajo quedara más presentable, por encontrar la forma de explicarme las cosas y hacerme ver mis errores de la manera más cordial, pero algo muy especial que tengo que agradecerle es su gran amabilidad y esa manera de hacer de cualquier momento el más agradable; de esa alegría que lo caracteriza, no la cambie siempre sea así. Por la confianza gracias.

Doctor **Luis Antonio** Rincón, gracias, por darme la oportunidad de trabajar con usted, por los consejos que me dio para mejorar mi redacción en beneficio de mi trabajo de investigación y por su tiempo.

Actuaria **Marypaola** Maya, gracias, por su apoyo a este trabajo desde el inicio, por los ánimos que siempre me dio, por las sugerencias para mejorar este trabajo, por los momentos que trabajamos juntas, así como el tiempo que dedico para escucharme y orientarme a pesar del enorme trabajo que tenía.

Maestro **Jaime** Vázquez Alamilla, gracias, por la confianza que deposito en mi trabajo, por la satisfacción que me dio de formar parte de mis sinodales y por el ejemplo que es usted como profesionista. Con cariño para usted mi admiración y respeto.

Maestra **Silvia** Alonso, gracias, por la oportunidad que me da de trabajar con usted.

A todos mis amigos que siempre estuvieron pendientes de mis avances e hicieron de los últimos años de la carrera algo muy agradable, **Alexandra, Araceli, Dalia, Edgar, Gabriela, Katia, Oscar, Ricardo** y a mis mejores amigas, en su momento, Leticia y Adriana, gracias.

Muy en especial a "La Latita", por que siempre me has guiado por el buen camino, por tu cariño, por ser el mejor amigo antes que cualquier cosa, por que siempre a cualquier hora y en cualquier lugar sé que tengo quien me escuche, por tu apoyo y ánimo para salir adelante, gracias.

Contenido

Introducción

i

1. Método de mínimos cuadrados

1.1.	Introducción al análisis de regresión	1
1.1.1.	Regresión	1
1.1.2.	Análisis de regresión lineal simple	2
1.1.3.	Supuestos de la regresión	3
1.1.4.	Tipos de modelos	4
1.2.	El criterio de mínimos cuadrados	6
1.2.1.	Objetivo	6
1.2.2.	Valores estimados	7
1.2.3.	Reducción de los estimadores	9
1.2.4.	Características de los estimadores	14
1.2.4.1.	Estimadores de máxima verosimilitud	17
1.2.5.	Estimación por intervalos	20
1.2.5.1.	Para β_0	21
1.2.5.2.	Para β_1	22
1.2.5.3.	Para σ^2	22
1.2.5.4.	Para la media	23
1.2.5.5.	De predicción	25
1.2.6.	Coefficiente de determinación	26
1.2.7.	Coefficiente de correlación	26
1.2.7.1.	Propiedades	27
1.2.8.	Identidad fundamental de los errores	28
1.2.9.	Prueba de hipótesis	29
1.2.9.1.	Para β_1	29

1.3.	Modelo de regresión lineal múltiple	33
1.3.1.	Método de mínimos cuadrados	34
1.3.2.	Propiedades del estimador de mínimos cuadrados de $\underline{\beta}$	36
1.3.3.	Estimadores máximo verosímiles	38
1.3.4.	Residuales	39
1.3.5.	Intervalos de confianza para cada β_i	40
2.	Regresión "no lineal"	
2.1.	Introducción	41
2.2.	Mínimos cuadrados no lineales	43
2.3.	Estimación de los parámetros de un sistema no lineal	46
2.3.1.	Transformaciones	47
2.3.2.	Linealización o método de Gauss-Newton	49
2.3.2.1.	La importancia de los valores iniciales	53
2.3.2.2.	Una interpretación geométrica de mínimos cuadrados no lineales	56
2.3.2.3.	Una interpretación geométrica de la linealización	59
2.3.3.	Descenso más pronunciado	61
2.3.4.	El acuerdo de Marquard	64
2.4.	El concepto del comportamiento de estimación no lineal	66
2.5.	Contornos de confianza	68
3.	Mínimos cuadrados generalizados	
3.1.	Introducción a los modelos lineales generalizados	70
3.1.1.	Funciones ligo y predictores lineales	72
3.1.2.	Residuales en el modelo lineal generalizado	74
3.2.	Mínimos cuadrados generalizados	75
3.2.1.	Supuestos	76
3.2.2.	Una matriz definida positiva, la transformación	77
3.2.3.	Valor estimado	79

3.2.4.	Características del estimador de mínimos cuadrados generalizados	80
3.2.4.1.	Estimadores máximo verosímiles	86
3.2.5.	Intervalo de confianza para β_i	89
3.2.5.1.	Intervalo de confianza para combinaciones lineales de las β_i	90
3.2.6.	Identidad de los errores para mínimos cuadrados generalizados	92
3.2.7.	Prueba de hipótesis	94
3.2.8.	Coefficiente de determinación	96
3.2.9.	Coefficiente de correlación r	96
3.2.10.	Devianza	97
3.3.	Geometría de mínimos cuadrados generalizados	99

4. Aplicación

4.1.	Introducción	102
4.2.	Mínimos cuadrados. Análisis simple	103
4.3.	Mínimos cuadrados. Análisis múltiple	110
4.4.	Mínimos cuadrados no lineales (linealización), ejemplo 1.	117
4.5.	Mínimos cuadrados no lineales (linealización), ejemplo 2.	122

Conclusiones

127

Bibliografía

129

Introducción

Este trabajo es una introducción al análisis de datos, a través de métodos de análisis de regresión, en los cuales se busca dar una ecuación que exprese la relación entre una variable dependiente y una o varias variables independientes, se usa específicamente el método de mínimos cuadrados. El enfoque del trabajo es la obtención de las formas básicas del método de mínimos cuadrados, para los casos lineal, no lineal y el generalizado, haciendo uso de procedimientos matemáticos y estadísticos.

En el primer capítulo se analiza el método de mínimos cuadrados en las formas lineal simple y múltiple, en ambos casos se obtienen las expresiones correspondientes a los coeficientes de regresión, así como también se considera el análisis de las propiedades de cada uno de los estimadores puntuales obtenidos y una estimación mediante intervalos de confianza.

En el segundo capítulo se presenta la regresión "no lineal", se analiza el caso cuando los mínimos cuadrados lineales ya no son aplicables para obtener las expresiones correspondientes a los coeficientes de regresión, dando a conocer los métodos bajo los cuales sí se puede dar solución a este problema, que son entre otros el de aplicar una transformación, hacer una linealización, usar el descenso más pronunciado, el acuerdo de Marquard, haciendo hincapié en el método de linealización (Gauss-Newton) el cual se basa en un proceso iterativo, auxiliándose de series de Taylor y el cual resulta ser el más usado en este trabajo.

El tercer capítulo hace una breve introducción de los modelos lineales generalizados para poder abordar el tema principal de esta parte del trabajo, que es el de mínimos cuadrados generalizados, donde se realiza la deducción de la forma general de las expresiones correspondientes a los coeficientes de regresión.

Este método de mínimos cuadrados generalizados se usa cuando se presenta principalmente varianza inconstante para los residuales. Así, el objetivo del método es la estabilización de la varianza.

Para lograrlo se realiza una transformación que permita obtener a la forma $\sigma^2 I$. En este capítulo se hace mención de un nuevo concepto, la devianza, ya que el coeficiente de determinación no es adecuado por diversos errores que genera.

El cuarto capítulo es una aplicación de la teoría vista en los anteriores, haciendo uso de una muestra de tamaño $n = 100$, de las variables peso, estatura, edad y el porcentaje de salud en los últimos cinco años. Los programas de cómputo con los que se realizan las regresiones son "Statística versión 6" y "Excel".

En Excel se puede realizar una regresión simple, por medio de Statística se realiza la regresión simple, múltiple, el método de linealización, para hacer la aplicación y ver cómo son los resultados prácticos de la teoría vista.

El trabajo aquí desarrollado se puede relacionar con otras materias tales como: economía, administración, ciencias sociales, medicina, biología, en física, química y la ingeniería.

Sobre la base de una muestra estadística se puede realizar, según el tema de interés, diferentes regresiones para encontrar un modelo que sea el que mejor se ajuste a la realidad del problema que se maneje.

La intención del trabajo es deducir las formas básicas del análisis de regresión, procurando desarrollar los procedimientos necesarios para poder llegar a los resultados que son conocidos, en algunos casos se hizo todo el desarrollo matemático para generar los estimadores de los parámetros.

Por último el trabajo muestra conclusiones sobre las aplicaciones, así como material bibliográfico que puede consultarse para revisar los temas vistos.

MÉTODO DE MÍNIMOS CUADRADOS

1.1. Introducción al análisis de regresión.

En muchas situaciones prácticas se presenta el problema de que dado un conjunto de datos se intenta determinar un modelo matemático que describa su comportamiento, el cual puede ser de dos tipos: el determinístico, en éste no concibe la posibilidad de algún error en el valor de una variable determinada que esté en función de otra, por otro lado se tiene el caso probabilístico, en el cual si se puede admitir el término error en la predicción de una variable determinada.

Cuando se necesita determinar una función que se ajuste a un conjunto de datos disponibles se puede hacer uso de distintos métodos los cuales pueden ser: el intuitivo, realiza el ajuste a simple vista, tomando como referencia dos puntos los cuales estén cerca de la función imaginaria y a los extremos del conjunto de puntos; el promedio por grupo, consiste en manejar los datos en dos partes iguales y calcular el valor promedio correspondiente en cada uno de ellos, sea x la variable independiente y y la dependiente, al determinar los valores promedio que serán las coordenadas de lo que se llama "el punto promedio" de cada grupo de puntos en el plano cartesiano; el de mínimos cuadrados, el cual asegura ofrecer la mejor precisión en el ajuste de una curva, dada la naturaleza de su enfoque y los principios que lo caracterizan.

1.1.1. Regresión

Uno de los objetivos en regresión es ver de qué manera se relacionan las variables que se quieren incluir en el modelo planteado, otro es ver si la variable incluida es la importante o la indispensable para realizar el modelo que explique la realidad que se desea modelar.

Al determinar las variables que se van a utilizar en el modelo al cual se le va aplicar el análisis de regresión para encontrar los valores de los coeficientes, basados en los datos que se estén manejando, donde se espera que el comportamiento de los datos pueda ser descrito por una relación funcional lineal en los coeficientes de regresión.

1.1.2. Análisis de Regresión Lineal Simple.

Aplicando el análisis de regresión a un sistema lineal simple, esto es, que el modelo de ajuste tenga una variable dependiente (y) y una variable explicativa o independiente (x) donde se ajusta una función a un conjunto de datos $(x_1, x_2, x_3, x_4, \dots, x_n)$, proporcionando una forma a través de la cual se puede establecer una asociación entre las dos variables de interés; se tiene la llamada curva de regresión. Ésta será una media condicional ya que " y " dependerá de los valores de los datos $(x_1, x_2, x_3, x_4, \dots, x_n)$, por lo tanto se tiene:

Se tiene $E(y/x)$ que se lee el valor esperado de y dado x .

La curva de regresión es calculada como:

$$E(y/x) = \int_{-\infty}^{\infty} yf(y/x)dy$$

Donde se tiene que la función $f(y/x)$ es:

$$f(y/x) = \frac{f(x, y)}{f(x)} \text{ y respectivamente para ésta se tiene:}$$

$$f(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

Lo que se desea es definir una relación que pueda existir entre las variables, de ahí que $E(y/x_i) = f(x_i)$ será la función deseada.

Como se habla de modelos lineales se definirá a $f(x_i) = \beta_0 + \beta_1 x_i$; donde β_0 y β_1 son los parámetros desconocidos y los cuales son llamados los coeficientes de regresión, además se considera que $f(x_i) = y_i$.

Este modelo no considera el error que sería indispensable tomar en cuenta, ya que el es propuesto para ajustar a la realidad *puede tener diferencias* y lo ideal sería que estas sean mínimas, preferentemente acercarlas tanto como se pueda o bien que sean cero, por lo tanto se requiere de la estimación del error para hacer la observación de la desviación de cada y_i a la línea de regresión alrededor de y_i , que es la distancia vertical que existe entre estos dos puntos.

1.1.3. Supuestos de la Regresión

En la regresión lineal simple considérese el modelo ahora con los siguientes supuestos:

- 1) $E[\varepsilon_i] = 0$ ya que se desea que el valor esperado de los errores sea cero.
- 2) $Var[\varepsilon_i] = \sigma^2$ con σ^2 constante común y desconocida, refleja que los factores no controlados influyen de la misma manera sobre cada respuesta y_i , además, se espera que las observaciones no se distribuyan de manera irregular alrededor de la línea media.
- 3) $E[\varepsilon_i, \varepsilon_j] = 0$
 $Cov[\varepsilon_i, \varepsilon_j] = E[\varepsilon_i, \varepsilon_j] - E[\varepsilon_i]E[\varepsilon_j] = 0$ esto es porque los y_i observados se consideran estadísticamente no relacionados.
- 4) $\varepsilon_i \sim N(0, \sigma^2)$ presenta una distribución normal la cual tiene media y varianza consecuencia de lo ya mencionado en 1) y 2).

De lo anterior se tiene que el i -ésimo error es $\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i)$ la suma de los cuadrados de los errores es $\sum \varepsilon_i^2 = \sum (y_i - \beta_0 - \beta_1 x_i)^2$.

1.1.4. Tipos de modelos

Por otro lado se pueden tener los siguientes modelos:

$y = \beta_0 + \beta_1 x$ es un Modelo Verdadero.

$y = \beta_0 + \beta_1 x + \varepsilon$ es un Modelo Empírico.

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ es un Modelo Estimado

Tipos de error

Entonces cualquier y_i que no se pueda explicar por la ecuación de regresión se deberá a un error. Se tiene que tomar en cuenta este término, ya que hay variables que tal vez no se incluyen en el modelo, pero, que de alguna manera afectan a éste.

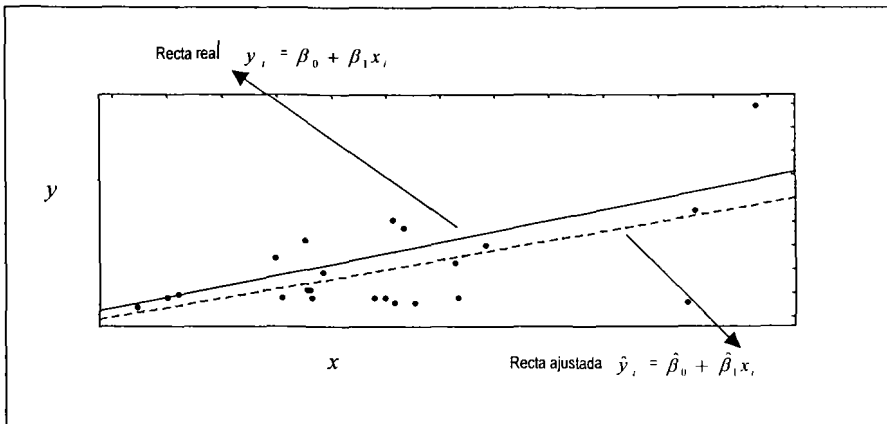
De manera abreviada así como se pueden tener tipos de modelos, se tienen diferentes errores, que son; los del modelo, suceden al no definirse de forma correcta éste, cuando se quiere explicar el fenómeno en estudio, es decir la falta de especificación en el modelo puede afectar a éste; el otro error es cuando no se puede controlar ciertos factores que afectan al experimento, en este caso el modelo se ve afectado por uno llamado aleatorio.

Sea $\varepsilon_i = y_i - E(y/x_i)$ donde x_i e y_i son los datos de los que el modelador se auxilia para llevar a cabo la regresión y $E(y/x_i)$ es el valor que se espera obtener.

De ahí que el término residual ε_i es igual a la diferencia del valor observado y el valor esperado de y_i , recordando que $E(y/x_i) = f(x_i) = \beta_0 + \beta_1 x_i$, entonces se define $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ donde ε_i será la variable aleatoria, la cual puede tomar valores positivos, negativos o nulos, además β_0 (ordenada al origen) y β_1 (la pendiente) son los parámetros desconocidos y donde

se tiene que si $\varepsilon_i = 0$, que es lo que difiere a este y_i definido y a $E(y/x_i)$, esto implica que la observación y_i se encuentra precisamente sobre la curva de regresión $E(y/x_i) = \beta_0 + \beta_1 x_i$.

Gráficamente



β_1 es la pendiente de la línea e indica el número de unidades y dirección que "y" cambia por cada unidad que aumente x.

Ejemplo:

Sea $\beta_0 = 5$ y $\beta_1 = 3$ y considérese $y_i = \beta_0 + \beta_1 x_i$.

Se tiene $y_i = \beta_0 + \beta_1 x_i = 5 + 3x_i$, tomando una x_i particular, sea $i = 0, 1, 2, 3$

$$y_1 = 5 + 3(0) = 5 \qquad y_3 = 5 + 3(2) = 11$$

$$y_2 = 5 + 3(1) = 8 \qquad y_4 = 5 + 3(3) = 14$$

Si se observa por cada unidad en que cambia x se tiene que y aumenta en 3 unidades, lo anterior debido a al valor de β_1 .

1.2. EL criterio de mínimos cuadrados

El modelo lineal simple:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, 2, 3, 4, 5, \dots, n$$

Tiene dos parámetros β_0 y β_1 que son desconocidos y serán estimados a partir de los datos en estudio o de interés para el modelador.

El criterio de mínimos cuadrados para obtener los estimadores está basado sobre los errores ajustados o residuales $\hat{\varepsilon}_i = y_i - \hat{y}_i$, donde $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ es el valor de la línea de ajuste y y_i el valor actual, claramente se puede considerar otras funciones de los datos basados en los errores verticales para derivar un criterio y escoger un estimador, pero los residuales son una buena elección porque ellos reflejan la inherente asimetría en las reglas de la respuesta y de predicción en un problema de regresión. A través del método de mínimos cuadrados para el modelo de regresión se busca:

- 1) Minimizar la $\sum \varepsilon_i^2$.
- 2) Tener la mejor recta, al obtener los mejores estimadores de β_0 y β_1 .
- 3) Tener el modelo matemático que explique la mayoría de los datos.

1.2.1. Objetivo

Lo que se busca es que $\sum (y_i - \hat{y}_i)^2$ sea minimizada y en la función \hat{y} lograr que esta diferencia sea tan cercana a cero como sea posible, dado esto será considerada la mejor aproximada. Como \hat{y}_i es la media de los valores " y_i " dados por un valor de x_i , lo ideal es que las diferencias entre

¹ El método más usado en regresión es el método de mínimos cuadrados. Este fue descubierto de manera independiente por Carl Friedrich Gauss en Alemania alrededor 1795 y por Adrien Marie Legendre en Francia alrededor 1805.

los valores que estén sobre y debajo de la recta se anulen mutuamente. Ya que se espera que la recta \hat{y} pase por los valores medios de los datos x_i y y_i .

1.2.2. Valores estimados

Distinguiendo a los estimadores de β_0 , β_1 , σ y σ^2 (valores desconocidos), como $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\sigma}$ y $\hat{\sigma}^2$ parámetros que se desean conocer a través del método de mínimos cuadrados [8].

Se define G como una función de x_i , y_i , $\hat{\beta}_0$ y $\hat{\beta}_1$ que se expresa de la siguiente forma $G = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$, los estimadores de mínimos cuadrados para β_0 y β_1 que minimizan a G , se obtienen mediante la diferenciación de G con respecto a $\hat{\beta}_0$ y $\hat{\beta}_1$, que se calcula enseguida:

$$\frac{\partial G}{\partial \hat{\beta}_1} = -2 \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(x_i) = 0$$

$$\sum (y_i x_i - \hat{\beta}_0 x_i - \hat{\beta}_1 x_i^2) = 0$$

$$\sum y_i x_i - \hat{\beta}_0 \sum x_i - \hat{\beta}_1 \sum x_i^2 = 0$$

$$\frac{\partial G}{\partial \hat{\beta}_0} = -2 \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\sum y_i - \sum \hat{\beta}_0 - \hat{\beta}_1 \sum x_i = 0$$

$$\sum y_i - n \hat{\beta}_0 - \hat{\beta}_1 \sum x_i = 0$$

De donde se obtiene el sistema de ecuaciones normales:

$$\sum y_i = n \hat{\beta}_0 + \hat{\beta}_1 \sum x_i, \dots\dots\dots(1)$$

$$\sum y_i x_i = \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2, \dots\dots\dots(2)$$

para las observaciones x_i y y_i .

Las ecuaciones normales se resuelven para $\hat{\beta}_0$ y $\hat{\beta}_1$, esto es, de (1) se despeja a $\hat{\beta}_0$ obteniendo:

$$\begin{aligned}\hat{\beta}_0 &= \frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n} \\ &= \frac{\sum y_i}{n} - \hat{\beta}_1 \frac{\sum x_i}{n} \\ &= \bar{Y} - \hat{\beta}_1 \bar{X}\end{aligned}$$

ahora como ya se conoce a $\hat{\beta}_0$ se sustituye en (2):

$$\begin{aligned}\sum y_i x_i &= \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 \\ \sum y_i x_i &= (\bar{Y} - \hat{\beta}_1 \bar{X}) \sum x_i + \hat{\beta}_1 \sum x_i^2 \\ \sum y_i x_i &= \sum x_i \bar{Y} - \sum x_i \hat{\beta}_1 \bar{X} + \hat{\beta}_1 \sum x_i^2 \\ \sum y_i x_i - \bar{Y} \sum x_i &= \hat{\beta}_1 (\sum x_i^2 - \bar{X} \sum x_i)\end{aligned}$$

de donde despejando a $\hat{\beta}_1$ se obtiene lo siguiente:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum y_i x_i - \bar{Y} \sum x_i}{\sum x_i^2 - \bar{X} \sum x_i} \\ &= \frac{\sum y_i x_i - \bar{Y} \sum x_i \left(\frac{n}{n}\right)}{\sum x_i^2 - \bar{X} \sum x_i \left(\frac{n}{n}\right)} \\ &= \frac{\sum y_i x_i - \frac{n\bar{Y} \sum x_i}{n}}{\sum x_i^2 - \frac{n\bar{X} \sum x_i}{n}} \\ &= \frac{\sum y_i x_i - n\bar{Y} \bar{X}}{\sum x_i^2 - n\bar{X}^2}\end{aligned}$$

Se puede expresar a $\hat{\beta}_1$ en términos de x_i , y_i y sus respectivas medias, esto es:

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2}$$

Por lo que los estimadores de β_0 y β_1 son:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2}$$

Como se observa $\hat{\beta}_0$ y $\hat{\beta}_1$ son calculados con los datos de una muestra de tamaño n de una determinada población, éstos dependerán de las características que sean deseables para analizar.

Al obtener los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ la recta de estimación para el modelo de regresión $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ es $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ donde \hat{y}_i es el estimador para medir y_i , de aquí que se tenga que la diferencia $\hat{y}_i - y_i$ sea un estimador del valor del error, para la i -ésima observación realizada, esto es $e_i = y_i - \hat{y}_i$.

1.2.3. Reducción de los estimadores

Los estimadores de la ecuación de regresión se pueden expresar como una combinación lineal de las observaciones y_i . A continuación se encontrará una expresión reducida del estimador $\hat{\beta}_1$, que servirá más adelante para calcular la $Var(\hat{\beta}_1)$.

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2} \\ &= \frac{\sum (x_i - \bar{X})y_i - \bar{Y} \sum (x_i - \bar{X})}{\sum (x_i - \bar{X})^2} \end{aligned}$$

$$\begin{aligned}
&= \frac{\sum (x_i - \bar{X}) y_i - \bar{Y} (\sum x_i - n\bar{X})}{\sum (x_i - \bar{X})^2} \\
&= \frac{\sum (x_i - \bar{X}) y_i - \bar{Y} [(n/n) \sum x_i - n\bar{X}]}{\sum (x_i - \bar{X})^2} \\
&= \frac{\sum (x_i - \bar{X}) y_i - \bar{Y} [n(\sum x_i / n) - n\bar{X}]}{\sum (x_i - \bar{X})^2} \\
&= \frac{\sum (x_i - \bar{X}) y_i - \bar{Y} [n\bar{X} - n\bar{X}]}{\sum (x_i - \bar{X})^2} \\
&= \frac{\sum (x_i - \bar{X}) y_i}{\sum (x_i - \bar{X})^2} \\
\hat{\beta}_1 &= \sum \left[\frac{(x_i - \bar{X})}{\sum (x_i - \bar{X})^2} \right] y_i \dots\dots\dots(3)
\end{aligned}$$

De la ecuación anterior, lo que esta entre corchetes es un arreglo de x_i , si se toma a $C_i = \frac{x_i - \bar{X}}{\sum (x_i - \bar{X})^2}$, que es la dispersión con respecto a la media y sustituyendo en (3) se tiene el término en forma lineal:

$$\hat{\beta}_1 = \sum C_i y_i$$

Se realiza lo mismo para $\hat{\beta}_0$, esto es dejarla expresada como combinación lineal de y_i , se realiza lo siguiente.

$$\begin{aligned}
\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\
&= \bar{Y} - \sum C_i y_i \bar{X} \\
&= \sum \frac{y_i}{n} - \bar{X} \sum C_i y_i \\
&= \sum \left(\frac{1}{n} - \bar{X} C_i \right) y_i \\
&= \sum \left(\frac{1}{n} - \bar{X} \left(\frac{x_i - \bar{X}}{\sum (x_i - \bar{X})^2} \right) \right) y_i
\end{aligned}$$

$$\begin{aligned}
&= \sum \left(\frac{\sum (x_i - \bar{X})^2 - n\bar{X}(x_i - \bar{X})}{n \sum (x_i - \bar{X})^2} \right) y_i \\
&= \sum \left(\frac{(\sum x_i^2 - n\bar{X}^2) - n\bar{X}x_i + n\bar{X}^2}{n \sum (x_i - \bar{X})^2} \right) y_i \\
\sum \left(\frac{(\sum x_i^2 - n\bar{X}x_i)}{n \sum (x_i - \bar{X})^2} \right) y_i &\text{ representando } \frac{\sum x_i^2 - n\bar{X}x_i}{n \sum (x_i - \bar{X})^2} = d_i = \frac{1}{n} - \bar{X}C_i, \text{ se llega a}
\end{aligned}$$

$$\boxed{\hat{\beta}_0 = \sum d_i y_i} \quad \text{tambi\u00e9n es de forma lineal.}$$

Con lo anterior se concluye que los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ son lineales en el sentido de que son combinaciones lineales de y_i , adem\u00e1s, $\hat{\beta}_0$ y $\hat{\beta}_1$ minimizan $\sum e_i^2$ por lo que tiene el mejor ajuste en ese sentido y donde C_i y d_i son constantes arbitrarias, ya que dependen de las x_i observadas en cada muestra.

A continuaci\u00f3n se deduce primero la varianza de y_i , la cual es necesaria para encontrar la varianza de $\hat{\beta}_1$.

$$\begin{aligned}
Var[y_i] &= Var[\beta_0 + \beta_1 x_i + \varepsilon_i] \\
&= Var[\varepsilon_i] \\
&= \sigma^2
\end{aligned}$$

haciendo uso del resultado anterior se calcula $Var(\hat{\beta}_1)$

$$\begin{aligned}
Var(\hat{\beta}_1) &= Var[\sum C_i y_i] \\
&= \sum C_i^2 Var[y_i] + \sum_i \sum_{i \neq j} C_i C_j Cov[y_i, y_j] \\
&= \sum C_i^2 \sigma^2
\end{aligned}$$

lo anterior se debe a que en los supuestos se dijo que $Cov[y_i, y_j] = 0$

sustituyendo a C_i se tiene

$$\begin{aligned}
Var(\hat{\beta}_1) &= \sigma^2 \sum \left(\frac{(x_i - \bar{X})}{\sum (x_i - \bar{X})^2} \right)^2 \\
&= \frac{\sigma^2 \sum (x_i - \bar{X})^2}{(\sum (x_i - \bar{X})^2)^2} = \frac{\sigma^2}{\sum (x_i - \bar{X})^2}
\end{aligned}$$

Para determinar la varianza de $\hat{\beta}_0$ se usa también el resultado $Var[y_i] = \sigma^2$.

$$\begin{aligned}
 Var[\hat{\beta}_0] &= Var\left[\sum d_i y_i\right] \\
 &= \sum d_i^2 Var[y_i] \\
 &= \sum d_i^2 \sigma^2 \\
 &= \sigma^2 \sum \left(\frac{1}{n} - \bar{X}C_i\right)^2 \\
 &= \sigma^2 \frac{\sum (1 - n\bar{X}C_i)^2}{n^2} \\
 &= \frac{\sigma^2}{n^2} \sum (1 - 2n\bar{X}C_i + n^2\bar{X}^2 C_i^2) \\
 &= \frac{\sigma^2}{n^2} (\sum 1 - 2n\bar{X} \sum C_i + n^2\bar{X}^2 \sum C_i^2) \\
 &= \frac{\sigma^2}{n^2} (n + n^2\bar{X}^2 \sum C_i^2) \\
 &= \frac{\sigma^2}{n^2} \left(n + \frac{n^2\bar{X}^2}{\sum (x_i - \bar{X})^2} \right) \\
 &= \frac{\sigma^2 n}{n^2} \left(1 + \frac{n\bar{X}^2}{\sum (x_i - \bar{X})^2} \right) \\
 &= \frac{\sigma^2}{n} \left(1 + \frac{n\bar{X}^2}{\sum (x_i - \bar{X})^2} \right) \\
 &= \frac{\sigma^2}{n} \left(\frac{\sum (x_i - \bar{X})^2 + n\bar{X}^2}{\sum (x_i - \bar{X})^2} \right) \\
 &= \frac{\sigma^2}{n} \left(\frac{\sum x_i^2 - n\bar{X}^2 + n\bar{X}^2}{\sum (x_i - \bar{X})^2} \right) \\
 &= \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \bar{X})^2}
 \end{aligned}$$

por lo tanto la varianza para cada estimador es:

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum (x_i - \bar{X})^2}$$

$$Var(\hat{\beta}_0) = \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \bar{X})^2}$$

Ya se hizo una reducción para $\hat{\beta}_0$ y $\hat{\beta}_1$ respectivamente, donde se mostraron como combinación lineal de y_i , esto es $\hat{\beta}_0 = \sum d_i y_i$, $\hat{\beta}_1 = \sum C_i y_i$, que se usan en el cálculo de la covarianza para ver como se comportan conjuntamente $\hat{\beta}_0$ y $\hat{\beta}_1$.

$$\begin{aligned}
 \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) &= E\left(\left(\hat{\beta}_0 - E(\hat{\beta}_0)\right)\left(\hat{\beta}_1 - E(\hat{\beta}_1)\right)\right) \\
 &= E\left(\left(\sum d_i y_i - E(\sum d_i y_i)\right) - \left(\sum C_i y_i - E(\sum C_i y_i)\right)\right) \\
 &= E\left(\left(\sum d_i y_i - \sum d_i E(y_i)\right)\left(\sum C_i y_i - \sum C_i E(y_i)\right)\right) \\
 &= E\left(\left(\sum d_i (y_i - E(y_i))\right)\left(\sum C_i (y_i - E(y_i))\right)\right) \\
 &= \sum d_i C_i E\left((y_i - E(y_i))^2\right) + \sum \sum d_i C_j E\left((y_i - E(y_i))(y_j - E(y_j))\right) \\
 &= \sum d_i C_i \text{Var}(y_i) + \sum \sum d_i C_j \text{Cov}(y_i, y_j) \\
 &= \sum d_i C_i \text{Var}(y_i)
 \end{aligned}$$

dado que $\text{Cov}(y_i, y_j) = 0$.

Sustituyendo los correspondientes valores de d_i y C_i se obtiene:

$$\begin{aligned}
 &= \sum \left(\frac{x_i - \bar{X}}{\sum (x_i - \bar{X})^2} \left(\frac{1}{n} - \frac{\bar{X}(x_i - \bar{X})}{\sum (x_i - \bar{X})^2} \right) \right) \text{Var}(y_i) \\
 &= \left(\frac{1}{n} \sum \frac{x_i - \bar{X}}{\sum (x_i - \bar{X})^2} - \bar{X} \frac{\sum (x_i - \bar{X})^2}{(\sum (x_i - \bar{X})^2)^2} \right) \sigma^2 \\
 &= \left(-\frac{\bar{X}}{\sum (x_i - \bar{X})^2} \right) \sigma^2 \neq 0
 \end{aligned}$$

con esto se tiene que $\hat{\beta}_0$ y $\hat{\beta}_1$ no son independientes.

1.2.4. Características de los estimadores

En la sección 1.2.3. se obtuvieron los estimadores puntuales de β_0 y β_1 , los cuales se desea que tengan un alto grado de ajuste, por lo que se presenta a continuación un análisis de las características de estos estimadores [8].

1) Insesgamiento²

2) Consistencia³

3) Suficiencia⁴

Análisis con respecto a $\hat{\beta}_1$:

$E[\hat{\beta}_1] = \beta_1$ indica que el estimador de $\hat{\beta}_1$ en promedio será β_1 [4].

Modelo completo.

Insesgamiento de $\hat{\beta}_1$:

Se desea obtener

$$\begin{aligned} E[\hat{\beta}_1] &= \beta_1 \\ &= E\left[\frac{\sum y_i(x_i - \bar{X})}{\sum (x_i - \bar{X})^2}\right] \\ &= \frac{1}{\sum (x_i - \bar{X})^2} \sum (x_i - \bar{X})E(y_i) \\ &= \frac{1}{\sum (x_i - \bar{X})^2} \sum (x_i - \bar{X})(\beta_0 + \beta_1 x_i) \\ &= \beta_1 \end{aligned}$$

dado que $\sum (x_i - \bar{X}) = 0$ y $\sum x_i(x_i - \bar{X}) = \sum (x_i - \bar{X})^2$.

² Cuando el valor esperado del estimador del parámetro es igual al parámetro que se está estimando.

³ Cuando el estimador es cada vez más cercano a el parámetro que se está estimando conforme va creciendo la muestra. En cuestión de convergencia en términos de probabilidades se espera que la varianza del estimador disminuya conforme n crece, y la media del estimador tiende hacia donde n crece.

⁴ cuando el estimador usa toda la información contenida en la muestra aleatoria con respecto al parámetro que se está estimando.

Insesgamiento de $\hat{\beta}_0$:

Por ver que

$$\begin{aligned} E[\hat{\beta}_0] &= \beta_0 \\ E(\hat{\beta}_0) &= E(\bar{Y} - \hat{\beta}_1 \bar{X}) \\ &= \frac{\sum E(y_i)}{n} - E(\hat{\beta}_1) \bar{X} \\ &= \frac{\sum (\beta_0 + \beta_1 x_i)}{n} - \beta_1 \bar{X} \\ &= \beta_0 + \beta_1 \bar{X} - \beta_1 \bar{X} \\ &= \beta_0 \end{aligned}$$

Bajo el modelo reducido se tiene.

Insesgamiento de $\hat{\beta}_1$:

Se tiene que

$$\begin{aligned} E[\hat{\beta}_1] &= E[\sum C_i y_i] \\ &= \sum C_i E[y_i] \end{aligned}$$

Por otra parte considerando $E[y_i]$

$$\begin{aligned} E[y_i] &= E[\beta_0 + \beta_1 x_i + \varepsilon_i] \\ &= \beta_0 + \beta_1 x_i + E[\varepsilon_i] \\ &= \beta_0 + \beta_1 x_i \end{aligned}$$

retomando la ecuación original

$$\begin{aligned} E[\hat{\beta}_1] &= \sum C_i (\beta_0 + \beta_1 x_i) \\ &= \sum C_i \beta_0 + \beta_1 \sum C_i x_i \\ &= \beta_0 \sum C_i + \beta_1 \sum C_i x_i \end{aligned}$$

y resolviendo lo siguiente

$$\begin{aligned} \sum C_i x_i &= \frac{\sum (x_i - \bar{X}) x_i}{\sum (x_i - \bar{X})^2} \\ &= \frac{\sum x_i^2 - \sum x_i \bar{X}}{\sum (x_i - \bar{X})^2} \\ &= \frac{\sum x_i^2 - n \bar{X}^2}{\sum x_i^2 - n \bar{X}^2} \\ &= 1 \end{aligned}$$

recordando la siguiente igualdad

$$\sum (x_i - \bar{X})^2 = \sum x_i^2 - n\bar{X}^2$$

y por otro lado se observa que

$$\begin{aligned} \sum C_i &= \frac{\sum (x_i - \bar{X})}{\sum (x_i - \bar{X})^2} \\ &= \frac{\sum x_i - \sum \bar{X}}{\sum (x_i - \bar{X})^2} \\ &= \frac{\sum x_i - n\bar{X}}{\sum (x_i - \bar{X})^2} \\ &= \frac{n\bar{X} - n\bar{X}}{\sum (x_i - \bar{X})^2} = 0 \end{aligned}$$

implica que

$$\begin{aligned} E[\hat{\beta}_1] &= \beta_0 \sum C_i + \beta_1 \sum C_i x_i \\ &= \beta_0(0) + \beta_1(1) \\ &= \beta_1 \end{aligned}$$

de donde el valor esperado de $\hat{\beta}_1$ es igual al parámetro que está siendo estimado β_1 , por lo que $\hat{\beta}_1$ es insesgado a β_1 .

Insesgamiento de $\hat{\beta}_0$:

$$\begin{aligned} E[\hat{\beta}_0] &= E[\bar{Y} - \hat{\beta}_1 \bar{X}] \\ &= E[\beta_0 + \beta_1 \bar{X} - \beta_1 \bar{X}] \\ &= E[\beta_0] \\ &= \beta_0 \end{aligned}$$

de ahí que el valor esperado de $\hat{\beta}_0$ es igual al parámetro β_0 , por lo tanto $\hat{\beta}_0$ es insesgado.

1.2.4.1. Estimadores máxima de verosimilitud

Si se supone que los ε_i son variables aleatorias (independientes), normalmente distribuidas, con media cero y varianza σ^2 para toda $i=1,2,3,4,\dots,n$, es posible obtener los estimadores de máxima verosimilitud de β_0 , β_1 y σ [12].

Como $\varepsilon_i \sim N(0, \sigma^2)$, lo que implica que $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, así

$$f(y) = \frac{e^{-\frac{1}{2\sigma^2}(y-(\beta_0+\beta_1 x_i))^2}}{\sqrt{2\pi}\sigma} \quad -\infty < y < \infty$$

Como se recordará los estimadores máximo verosímiles son generados a través de maximizar la función de verosimilitud, $L(y_1, y_2, y_3, \dots, y_n; \sigma^2)$, la cual se obtiene como el producto de las funciones marginales.

$$L(y_1, y_2, y_3, \dots, y_n; \sigma^2) = \prod \frac{e^{-\frac{1}{2\sigma^2}(y_i - (\beta_0 + \beta_1 x_i))^2}}{\sqrt{2\pi}\sigma}$$

$$L(y_1, y_2, y_3, \dots, y_n; \sigma^2) = \frac{e^{-\frac{1}{2\sigma^2} \sum (y_i - (\beta_0 + \beta_1 x_i))^2}}{\sqrt{(2\pi)^n \sigma^n}}$$

$$\ln L = -\frac{n}{2} \ln(2\pi) - n \ln \sigma - \frac{1}{2\sigma^2} \sum (y_i - \beta_0 - \beta_1 x_i)^2$$

$$1) \frac{\partial \ln L}{\partial \beta_0} = -\frac{1}{2\sigma^2} \sum 2(y_i - \beta_0 - \beta_1 x_i)(-1) = 0$$

$$2) \frac{\partial \ln L}{\partial \beta_1} = -\frac{1}{2\sigma^2} \sum 2(y_i - \beta_0 - \beta_1 x_i)(-x_i) = 0$$

$$3) \frac{\partial \ln L}{\partial \sigma} = \frac{1}{\sigma^3} \sum (y_i - \beta_0 - \beta_1 x_i)^2 - \frac{n}{\sigma} = 0$$

Se genera nuevamente el sistema de ecuaciones normales.

$$\begin{array}{l}
 1) \sum y_i - n\beta_0 - \beta_1 \sum x_i = 0 \\
 2) \sum x_i y_i = \beta_0 \sum x_i + \beta_1 \sum x_i^2 \\
 3) \frac{\sum (y_i - \beta_0 - \beta_1 x_i)^2}{n} = \sigma^2
 \end{array}
 \left. \vphantom{\begin{array}{l} 1) \\ 2) \\ 3) \end{array}} \right\} \begin{array}{l} \text{Ecuaciones} \\ \text{Normales} \end{array}$$

Al resolver las ecuaciones quedan las del método de mínimos cuadrados y de ahí que las soluciones sean las mismas.

$$\begin{aligned}
 \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\
 \hat{\beta}_1 &= \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2}
 \end{aligned}$$

el estimador de σ^2 es:

$$\hat{\sigma}^2 = \frac{\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n}$$

Todo estimador máximo verosímil es suficiente entonces $\hat{\sigma}^2$, $\hat{\beta}_0$ y $\hat{\beta}_1$ son suficientes respectivamente para σ^2 , β_0 y β_1 .

Propiedades del estimador de σ^2 .

Insesgamiento de $\hat{\sigma}^2$:

$$\begin{aligned}
 E[\hat{\sigma}^2] &= E\left(\frac{\sum (y_i - \hat{y}_i)^2}{n}\right) \\
 &= \frac{1}{n} E\left(\sum (y_i - \hat{y}_i)^2\right) \\
 &= \frac{1}{n} E\left(\sum (\beta_0 + \beta_1 x_i + \varepsilon_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2\right) \\
 &= \frac{1}{n} E\left(\sum ((\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)x_i + \varepsilon_i)^2\right) \\
 &= \frac{1}{n} E\left(\sum (-(\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)x_i + \varepsilon_i)^2\right)
 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} E \left(\sum \left((\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1) x_i - \varepsilon_i \right)^2 \right) \\
&= \frac{1}{n} E \left(\sum (\hat{\beta}_0 - \beta_0)^2 + (\hat{\beta}_1 - \beta_1)^2 \sum x_i^2 + \sum \varepsilon_i^2 + 2(\hat{\beta}_1 - \beta_1)(\hat{\beta}_0 - \beta_0) \sum x_i \right. \\
&\quad \left. - 2(\hat{\beta}_0 - \beta_0) \sum \varepsilon_i - 2(\hat{\beta}_1 - \beta_1) \sum x_i \varepsilon_i \right) \\
&= \frac{1}{n} \left(nE(\hat{\beta}_0 - \beta_0)^2 + E(\hat{\beta}_1 - \beta_1)^2 \sum x_i^2 + \sum E(\varepsilon_i^2) + 2E((\hat{\beta}_1 - \beta_1)(\hat{\beta}_0 - \beta_0)) \sum x_i \right. \\
&\quad \left. - 2E((\hat{\beta}_0 - \beta_0) \sum \varepsilon_i) - 2E((\hat{\beta}_1 - \beta_1) \sum x_i \varepsilon_i) \right) \\
&= \frac{1}{n} \left(n\text{Var}(\hat{\beta}_0) + \text{Var}(\hat{\beta}_1) \sum x_i^2 + n\sigma^2 + 2\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \sum x_i - 2E((\hat{\beta}_0 - \beta_0) \sum \varepsilon_i) \right. \\
&\quad \left. - 2E((\hat{\beta}_1 - \beta_1) \sum x_i \varepsilon_i) \right) \\
&= \frac{1}{n} \left(n \left(\frac{1}{n} + \frac{\bar{X}}{\sum (x_i - \bar{X})^2} \right) \sigma^2 + \sum x_i^2 \frac{\sigma^2}{\sum (x_i - \bar{X})^2} + n\sigma^2 - 2 \sum x_i \frac{\bar{X} \sigma^2}{\sum (x_i - \bar{X})^2} \right. \\
&\quad \left. - 2E \left(\sum \left(\frac{1}{n} - \frac{\bar{X}(x_i - \bar{X})}{\sum (x_i - \bar{X})^2} \right) \varepsilon_i \sum \varepsilon_i \right) - 2E \left(\frac{\sum (x_i - \bar{X}) \varepsilon_i}{\sum (x_i - \bar{X})^2} (\sum x_i \varepsilon_i) \right) \right) \\
&= \frac{1}{n} \left(\left(\frac{n}{n} + \frac{n\bar{X}}{\sum (x_i - \bar{X})^2} \right) \sigma^2 + \frac{\sum x_i^2}{\sum (x_i - \bar{X})^2} \sigma^2 + n\sigma^2 - 2 \sum x_i \left(\frac{n}{n} \right) \frac{\bar{X} \sigma^2}{\sum (x_i - \bar{X})^2} \right. \\
&\quad \left. - 2\sigma^2 \sum \left(\frac{1}{n} - \frac{\bar{X}(x_i - \bar{X})}{\sum (x_i - \bar{X})^2} \right) - 2 \frac{\sigma^2}{\sum (x_i - \bar{X})^2} \sum x_i (x_i - \bar{X}) \right) \\
&= \frac{1}{n} \left(\sigma^2 + \frac{n\bar{X}^2}{\sum (x_i - \bar{X})^2} \sigma^2 + \frac{\sum x_i^2}{\sum (x_i - \bar{X})^2} \sigma^2 + n\sigma^2 \frac{\bar{X}^2}{\sum (x_i - \bar{X})^2} \sigma^2 - 2\sigma^2 - 2\sigma^2 \right) \\
&= \frac{1}{n} \left((n-3)\sigma^2 + \frac{\sum x_i^2}{\sum (x_i - \bar{X})^2} \sigma^2 + \frac{n\bar{X}^2}{\sum (x_i - \bar{X})^2} \sigma^2 \right) \\
&= \frac{1}{n} \left(\frac{\sum x_i^2 - n\bar{X}^2}{\sum (x_i - \bar{X})^2} \sigma^2 + (n-3)\sigma^2 \right) \\
&= \frac{(n-2)\sigma^2}{n}
\end{aligned}$$

Resumiendo las propiedades se tiene el siguiente cuadro de las características de los estimadores, las cuales han sido mostradas.

$\hat{\beta}_0$	$\hat{\beta}_1$	σ^2
Insegamiento	Insegamiento	No es insegado
Consistencia	No es consistente	
Suficiente	Suficiente	Suficiente
Lineales	Lineales	
Mejores Lineales	Mejores Lineales	
Menor varianza	Menor varianza	

1.2.5. Estimación por intervalos

La estimación puntual con sus propiedades se desarrollo en los puntos anteriores, por lo que ahora será importante realizar la por intervalos para cada uno de los parámetros, así como también se introducirá el concepto de intervalo de predicción diferenciándolo de uno de confianza para un nuevo valor de la variable dependiente.

Como $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ y considerando que $\hat{\beta}_0$ y $\hat{\beta}_1$ son parámetros lineales de y_i , se tendría que estos estimadores, por ser combinaciones lineales de variables aleatorias normales, su distribución también se hereda con medias y varianzas ya obtenidas, lo cual se expresa como:

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2 \frac{\sum x_i^2}{n \sum (x_i - \bar{X})^2}\right)$$

$$\hat{\beta}_1 \sim N\left(\beta_1, \sigma^2 \frac{1}{\sum (x_i - \bar{X})^2}\right)$$

1.2.5.1. Intervalo para β_0

Para determinar el intervalo de confianza para β_0 se utiliza el método de la cantidad pivotal, por lo que hay que construir una cantidad que pueda ser usada para calcular la probabilidad fija en la que ésta se incluya entre dos valores.

$$\hat{\beta}_0 = \sum d_i y_i \sim N\left(\beta_0, \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \bar{X})^2}\right)$$

Estándarizando la variable $\hat{\beta}_0$ queda:

$$\frac{\hat{\beta}_0 - \beta_0}{\sigma \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{X})^2}}} \sim N(0,1). \text{ Como } \sigma \text{ es desconocida se tomará una variable Ji-cuadrada, y con}$$

ello poder construir una cantidad cuya distribución sea t , donde la que presenta una Ji-cuadrada es

$$\frac{\sum (y_i - \hat{y}_i)^2}{\sigma^2} \sim \chi^2_{(n-2)}, \text{ por lo tanto como } \frac{N(0,1)}{\chi^2} = t, \text{ resulta que:}$$

$$\frac{\frac{\hat{\beta}_0 - \beta_0}{\sigma \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{X})^2}}}}{\sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{\sigma^2 (n-2)}}} \sim t_{(n-2)}.$$

$$\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma} \sqrt{\frac{\sum x_i^2}{(n-2) \sum (x_i - \bar{X})^2}}} = \frac{\hat{\beta}_0 - \beta_0}{\bar{\sigma} \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{X})^2}}} \sim t_{(n-2)}, \text{ donde } \bar{\sigma}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-2}.$$

Así que para encontrar el intervalo de confianza $(1 - \alpha) \times 100\%$ para β_0 la cantidad anteriormente obtenida se encierra entre dos cuantiles y se calcula la probabilidad.

$$P\left(-a < \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma} \sqrt{\frac{\sum x_i^2}{(n-2)S_{xx}}}} < a\right) = 1 - \alpha, \quad \text{donde } S_{xx} = \sum (x_i - \bar{X})^2$$

Manipulando la doble desigualdad el intervalo requerido queda como:

$$\left(\beta_0 - a\hat{\sigma} \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{X})^2}}, \beta_0 + a\hat{\sigma} \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{X})^2}}\right)$$

donde a es el cuantil $1 - \alpha/2$ de una distribución t con $n - 2$ grados de libertad [4].

1.2.5.2. Intervalo para β_1

De igual manera que el intervalo para β_0 se construye el correspondiente para β_1 quedando:

$$\left(\hat{\beta}_1 - a\hat{\sigma} \sqrt{\frac{n}{(n-2)S_{xx}}}, \hat{\beta}_1 + a\hat{\sigma} \sqrt{\frac{n}{(n-2)S_{xx}}}\right)$$

con una confianza del $(1 - \alpha) \times 100\%$.

1.2.5.3. Intervalo de confianza para σ^2

A continuación se calcula el intervalo para la varianza del modelo y como se sabe que

$$\frac{\sum (y_i - \hat{y}_i)^2}{\sigma^2} \sim \chi^2_{(n-2)}, \text{ y esa cantidad puede encerrarse en un intervalo con probabilidad fija.}$$

$$P\left(a < \frac{n\hat{\sigma}^2}{\sigma^2} < b\right) = 1 - \alpha \text{ y por tanto el intervalo de confianza para } \sigma^2 \text{ al } (1 - \alpha) \times 100\% \text{ es:}$$

$$\left(\frac{n\hat{\sigma}^2}{b}, \frac{n\hat{\sigma}^2}{a}\right), \text{ donde } a \text{ y } b \text{ son cuantiles } \frac{\alpha}{2} \text{ y } 1 - \frac{\alpha}{2} \text{ respectivamente de una distribución } \chi^2 \text{ con}$$

$n - 2$ grados de libertad [4].

1.2.5.4. Intervalo de confianza para la media

De gran interés es calcular un intervalo de confianza para un valor de la variable dependiente que fue usada en el cálculo de los estimadores, a este se le conoce como intervalo de confianza para la media.

Se sabe que $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, como $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, se forma de una combinación lineal de las betas y éstas también son combinación lineal pero de las y_i , se tiene que \hat{y}_i se distribuye normal con una media y varianza que serán calculadas a continuación.

$$E[\hat{y}_i] = E[\hat{\beta}_0 + \hat{\beta}_1 x_i] = E[\hat{\beta}_0] + E[\hat{\beta}_1] x_i = \beta_0 + \beta_1 x_i$$

Para el caso de la varianza se tiene:

$$\begin{aligned} \text{Var}[\hat{y}_i] &= \text{Var}[\hat{\beta}_0 + \hat{\beta}_1 x_i] \\ &= \text{Var}[\hat{\beta}_0] + \text{Var}[\hat{\beta}_1] x_i^2 + 2\text{Cov}(\hat{\beta}_0, \hat{\beta}_1 x_i) \\ &= \frac{\sigma^2 \sum x_i^2}{nS_{xx}} + \frac{\sigma^2 x_i^2}{S_{xx}} - \frac{2x_i \bar{X} \sigma^2}{S_{xx}} \\ &= \frac{\sigma^2}{S_{xx}} \left(\frac{\sum x_i^2}{n} + x_i^2 - 2x_i \bar{X} \right) \\ &= \frac{\sigma^2}{S_{xx}} \left(\frac{\sum x_i^2}{n} + x_i^2 - 2x_i \bar{X} + \bar{X}^2 - \bar{X}^2 \right) \\ &= \frac{\sigma^2}{S_{xx}} \left(\frac{\sum x_i^2}{n} + (x_i - \bar{X})^2 - \bar{X}^2 \right) \\ &= \frac{\sigma^2}{S_{xx}} \left(\frac{\sum x_i^2 - n\bar{X}^2}{n} + (x_i - \bar{X})^2 \right) \\ &= \frac{\sigma^2}{S_{xx}} \left(\frac{S_{xx}}{n} + (x_i - \bar{X})^2 \right) \\ &= \sigma^2 \left(\frac{1}{n} + \frac{(x_i - \bar{X})^2}{S_{xx}} \right) \end{aligned}$$

La construcción de este intervalo no se da de manera general sino que se considera para una x_i concreta que se llamará en este trabajo x_i^* , por lo que haciendo el cambio para la estimación se tendrá $\hat{y}_i^* = \hat{\beta}_0 + \hat{\beta}_1 x_i^*$.

Procediendo nuevamente con la técnica de la cantidad pivotal se tendrá:

$$\frac{\hat{y}_i^* - (\beta_0 + \beta_1 x_i^*)}{\sigma \sqrt{\frac{1}{n} + \frac{(x_i^* - \bar{X})^2}{S_{xx}}}} = \frac{\hat{y}_i^* - E[y_i]}{\sigma \sqrt{\frac{1}{n} + \frac{(x_i^* - \bar{X})^2}{S_{xx}}}} \sim N(0,1),$$

como esta cantidad tiene en el denominador a

la varianza desconocida se divide con una Ji-cuadrada y sus respectivos grados de libertad para obtener, nuevamente una cantidad pivotal que se distribuye t .

$$\frac{\frac{\hat{y}_i^* - E[y_i]}{\sigma \sqrt{\frac{1}{n} + \frac{(x_i^* - \bar{X})^2}{S_{xx}}}}}{\frac{\sqrt{\frac{n\hat{\sigma}^2}{\sigma^2(n-2)}}}{\sqrt{\sigma^2(n-2)}}} \sim t_{(n-2)}$$

Siguiendo la técnica antes expuesta queda:

$$p \left(-a < \frac{\hat{y}_i^* - E[y_i]}{\hat{\sigma} \sqrt{\frac{n}{n-2} \left(\frac{1}{n} + \frac{(x_i^* - \bar{X})^2}{S_{xx}} \right)}} < a \right) = 1 - \alpha$$

Dejando el intervalo de confianza al $(1 - \alpha) \times 100\%$ como:

$$\left(\hat{y}_i^* - a\hat{\sigma} \sqrt{\frac{n}{n-2} \left(\frac{1}{n} + \frac{(x_i^* - \bar{X})^2}{S_{xx}} \right)}, \hat{y}_i^* + a\hat{\sigma} \sqrt{\frac{n}{n-2} \left(\frac{1}{n} + \frac{(x_i^* - \bar{X})^2}{S_{xx}} \right)} \right)$$

donde a es un cuantil de la distribución t al $1 - \frac{\alpha}{2}$.

1.2.5.5. Intervalo de predicción

Un intervalo que es muy parecido al mostrado en 1.2.5.4. es el de predicción, éste tiene como diferencia que su cálculo no va en referencia a una x_i observada sino a un nuevo elemento que no fue utilizado para determinar a los estimadores, a este valor se le llamará x_F , así que el éste será el correspondiente a y que se representará por y_F , que es desconocido pero se puede dar su estimación que es \hat{y}_F , el cual puede ser calculado por $\hat{y}_F = \hat{\beta}_0 + \hat{\beta}_1 x_F$.

El método de obtención de este intervalo se hará definiendo una nueva variable llamada Z .

$$Z = y_F - \hat{y}_F$$

A esta nueva variable se le calculará su esperanza y varianza con el fin de hacer una estandarización y proceder con la construcción de la cantidad pivotal.

$$\begin{aligned} E[Z] &= E[y_F - \hat{y}_F] \\ &= E[y_F] - E[\hat{y}_F] \\ &= E[\beta_0 + \beta_1 x_F + \varepsilon] - E[\hat{\beta}_0 + \hat{\beta}_1 x_F] \\ &= E[\beta_0] + E[\beta_1 x_F] + E[\varepsilon] - E[\hat{\beta}_0] - E[\hat{\beta}_1 x_F] \\ &= \beta_0 + \beta_1 x_F + 0 - \beta_0 - \beta_1 x_F = 0 \\ \text{Var}[Z] &= \text{var}[y_F - \hat{y}_F] \\ &= \text{Var}[y_F] + \text{Var}[\hat{y}_F] - 2\text{Cov}[y_F, \hat{y}_F] \\ &= \sigma^2 + \sigma^2 \left(\frac{1}{n} + \frac{(x_F - \bar{X})^2}{S_{xx}} \right) \end{aligned}$$

Como las y_i que fueron utilizadas para el cálculo de los estimadores no tienen nada que ver con y_F , da como resultado $\text{Cov}[y_F, \hat{y}_F] = 0$.

Usando nuevamente la técnica ya mencionada quedaría el intervalo de predicción para y_F al $(1 - \alpha) \times 100\%$ como:

$$\left(y_F - a\hat{\sigma} \sqrt{\frac{n}{n-2} \left(1 + \frac{1}{n} + \frac{(x_F - \bar{X})^2}{S_{xx}} \right)}, y_F + a\hat{\sigma} \sqrt{\frac{n}{n-2} \left(1 + \frac{1}{n} + \frac{(x_F - \bar{X})^2}{S_{xx}} \right)} \right)$$

donde a es el cuantil $1 - \frac{\alpha}{2}$ de una distribución $t_{(n-2)}$.

1.2.6. Coeficiente de determinación

Hasta el momento se han hecho cálculos para los estimadores, ya sea usando la estimación puntual o la de intervalos, pero no se ha dado una medida que permita decir qué tan bien el modelo de regresión lineal simple se ha ajustado a los datos, esta medida es conocida como el coeficiente de determinación.

El cual es denotado como r^2 , e indica el grado de ajuste de un modelo de regresión lineal, es decir es la medida que da la variabilidad explicada de los datos a través del modelo usado.

Este coeficiente está definido como:
$$r^2 = \frac{\sum(\hat{y}_i - \bar{Y})^2}{\sum(y_i - \bar{Y})^2}.$$

1.2.7. Coeficiente de correlación $\pm \sqrt{r^2} = r$

El coeficiente de correlación mide precisamente la correlación lineal entre dos variables y es definido como:

$$\rho_{xy} = \frac{Cov(x, y)}{\sigma_x \sigma_y} = \frac{E((x - E(x))(y - E(y)))}{\sigma_x \sigma_y} \quad [14].$$

El coeficiente de correlación, para el caso particular de la regresión no puede ser calculado por la definición dada anteriormente, ya que no se posee una función de densidad conjunta de las variables x y y . Por lo que se utiliza un estimador de éste llamado coeficiente de correlación de Pearson, su cálculo se hace por medio de:

$$r = \frac{\sum(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum(x_i - \bar{X})^2 \sum(y_i - \bar{Y})^2}}.$$

1.2.7.1. Propiedades

- 1) Puede ser positivo o negativo.
- 2) Varía entre -1 y 1 , es decir $-1 \leq r \leq 1$.
- 3) Es de naturaleza simétrica, es decir, el coeficiente de correlación entre x y y es igual al de y y x .
- 4) Si x y y son estadísticamente independientes, el coeficiente de correlación entre ellas es cero; pero por otro lado si $r = 0$, no necesariamente indica que las dos variables sean independientes.
- 5) Es una medida de asociación lineal y no tiene sentido para describir relaciones no lineales.
- 6) La correlación positiva indica que la relación entre las dos variables en cuestión es la asociación de valores grandes o pequeños de ambas; en cambio para el caso de la negatividad la relación es inversa, es decir, se da la correspondencia de valores grandes de x con valores pequeños de y o viceversa.
- 7) No implica necesariamente una relación causa y efecto.

Un resultado importante es que el coeficiente de determinación es el cuadrado del coeficiente de correlación, la demostración se deja al lector.

El hecho de introducir el coeficiente de determinación como medida de explicación del modelo, es porque se vuelve un cociente de dos medidas que intervienen en el punto siguiente.

1.2.8. Identidad fundamental de los errores

Cuando se tienen los datos de un fenómeno, considérese y_1, \dots, y_n algo importante que se debe calcular dentro de la estadística descriptiva es la dispersión de los datos, la cual puede obtenerse

como la raíz cuadrada de $\frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{n-1}$.

Se puede observar que de la fórmula anterior lo interesante es el numerador, por lo que a esa estructura se le incluirá el modelo de regresión.

$$\begin{aligned}\sum (y_i - \bar{Y})^2 &= \sum (y_i - \hat{y}_i + \hat{y}_i - \bar{Y})^2 \\ &= \sum ((y_i - \hat{y}_i) + (\hat{y}_i - \bar{Y}))^2 \\ &= \sum (y_i - \hat{y}_i)^2 + 2\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{Y}) + \sum (\hat{y}_i - \bar{Y})^2 \\ &= \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{Y})^2.\end{aligned}$$

Se deja al lector la demostración de que $\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{Y})$ es cero.

$\sum (y_i - \bar{Y})^2$ = Suma de cuadrados del modelo reducido SCE_{MR} (Suma de los cuadrados del error total)

$\sum (y_i - \hat{y}_i)^2$ = Suma de cuadrados del modelo completo SCE_{MC} (Suma de los cuadrados del error, residual).

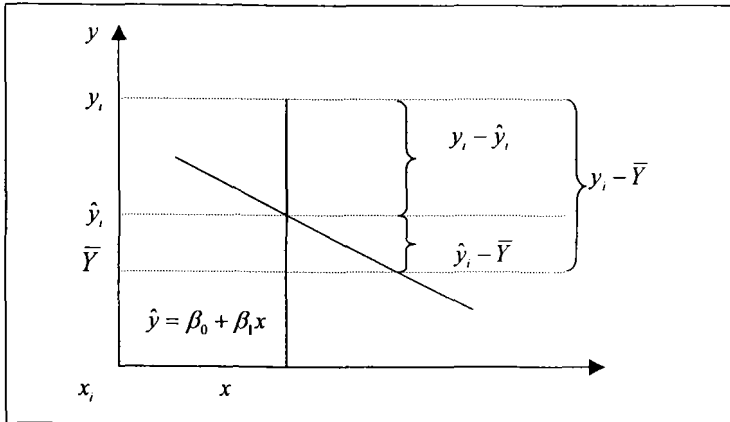
$\sum (\hat{y}_i - \bar{Y})^2$ = Suma de cuadrados del error bajo la hipótesis nula SCE_{H_0} (Suma de los cuadrados de la regresión) [17].

La distribución de cada termino se indica enseguida.

$$\frac{\sum (y_i - \bar{Y})^2}{\sigma^2} = \frac{\sum (y_i - \hat{y}_i)^2}{\sigma^2} + \frac{\sum (\hat{y}_i - \bar{Y})^2}{\sigma^2}$$

que se distribuyen respectivamente $\chi^2_{(n-1)}$ $\chi^2_{(n-2)}$ $\chi^2_{(1)}$

Lo interesante de esta igualdad es ver la gráficamente:



1.2.9. Prueba de Hipótesis

Dado un modelo de regresión lo importante sería saber ahora si se tiene un modelo sin ordenada al origen ($\beta_0 = 0$) o bien uno que es constante ($\beta_1 = 0$); por lo que el objetivo de esta sección es realizar una prueba de hipótesis para β_0 y β_1 .

1.2.9.1. Prueba de hipótesis para β_1

Si en la regresión se hace la suposición de que β_1 es cero, se tendrá como modelo una constante y esto en si mismo no es un modelo al cual se le puedan realizar análisis interesantes, por lo que probar que β_1 es igual a cero es de gran relevancia en el análisis de regresión.

La construcción de esta prueba de hipótesis $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ se realiza a través del cociente de verosimilitudes generalizado. Esta técnica implica la construcción de un forma en donde el numerador es el supremo de la función de verosimilitud bajo la hipótesis nula y en el denominador se encuentra el supremo de la verosimilitud en todo el espacio parametral.

Lo primero que se calculará es el denominador de ese cociente, recordando que

$$L(\beta_0, \beta_1, \sigma^2, y) = \frac{e^{-\frac{1}{2\sigma^2} \sum (y_i - (\beta_0 + \beta_1 x_i))^2}}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}}} \text{ es la función de verosimilitud donde el espacio}$$

parametral es $\{(\beta_0, \beta_1, \sigma^2) \mid \beta_0, \beta_1 \in \mathfrak{R}, 0 < \sigma^2 < \infty\}$.

Para obtener el supremo sólo es necesario encontrar los estimadores máximo verosímiles de los parámetros y evaluarlos en la función de verosimilitud. [12]

Recordando que los estimadores son:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad \hat{\beta}_1 = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2} \quad \hat{\sigma}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n}$$

Por lo que al sustituirlos en la función de verosimilitud queda:

$$\begin{aligned} \sup L(\hat{\theta}) &= \left(\frac{1}{2\pi\hat{\sigma}^2}\right)^{\frac{n}{2}} e^{-\frac{\frac{n}{2\sum (y_i - \hat{y}_i)^2} \sum (y_i - \hat{y}_i)^2}{2\sum (y_i - \hat{y}_i)^2} \sum (y_i - \hat{y}_i)^2} \\ &= \left(\frac{n}{2\pi \sum (y_i - \hat{y}_i)^2}\right)^{\frac{n}{2}} e^{-\frac{n}{2}} \end{aligned}$$

A continuación se calcula el numerador , bajo el supuesto de la hipótesis nula de que $\beta_1 = 0$.

$$\begin{aligned} L(\theta_{H_0}) &= \prod \frac{e^{-\frac{1}{2\sigma^2}(y_i - \mu_0)^2}}{(2\pi)^{\frac{1}{2}} (\sigma^2)^{\frac{1}{2}}} \dots\dots\dots(1) \\ &= \frac{e^{-\frac{1}{2\sigma^2} \sum (y_i - \beta_0)^2}}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}}} \end{aligned}$$

Para calcular el supremo bajo la hipótesis nula se usa el estimador máximo verosímil de σ^2 , sólo que ahora en ves de la media estimada se usa la media bajo la H_0 que en este caso es β_0

El estimador máximo verosímil de β_0 es $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$, al tomar la hipótesis nula queda $\hat{\beta}_0 = \bar{Y}$.

$$\begin{aligned} \sup L(\hat{\theta}_{H_0}) &= \left(\frac{n}{2\pi \sum (y_i - \bar{Y})^2} \right)^{\frac{n}{2}} e^{-\frac{n}{2 \sum (y_i - \bar{Y})^2} \sum (y_i - \bar{Y})^2} \\ &= \left(\frac{n}{2\pi \sum (y_i - \bar{Y})^2} \right)^{\frac{n}{2}} e^{-\frac{n}{2}} \end{aligned} \quad \dots\dots\dots (2)$$

de las expresiones (1) y (2) el cociente de verosimilitud queda:

$$\begin{aligned} \lambda &= \frac{\sup L(\theta_{H_0})}{\sup L(\theta)} \\ &= \frac{\left(\frac{n}{2\pi \sum (y_i - \bar{Y})^2} \right)^{\frac{n}{2}}}{\left(\frac{n}{2\pi \sum (y_i - \hat{y}_i)^2} \right)^{\frac{n}{2}}} \\ &= \left(\frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{Y})^2} \right)^{\frac{n}{2}} \leq \lambda_0 \end{aligned}$$

esto pasa si

$$\frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{Y})^2} \leq \lambda_0^{\frac{2}{n}}$$

usando la identidad fundamental de los errores se tiene

$$= \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{Y})^2} = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \hat{y}_i + \hat{y}_i - \bar{Y})^2} = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (\hat{y}_i - \bar{Y})^2 + \sum (y_i - \hat{y}_i)^2} = \frac{1}{1 + \frac{\sum (\hat{y}_i - \bar{Y})^2}{\sum (y_i - \hat{y}_i)^2}}$$

de ahí que se llega a

$$1 + \frac{1}{\frac{\sum (\hat{y}_i - \bar{Y})^2}{\sum (y_i - \hat{y}_i)^2}} \leq \lambda_0^{\frac{2}{n}} \quad \Leftrightarrow \quad 1 + \frac{\sum (\hat{y}_i - \bar{Y})}{\sum (y_i - \hat{y}_i)^2} \geq \frac{1}{\lambda_0^{\frac{2}{n}}} = \lambda_0^{-\frac{2}{n}}$$

$$\frac{\sum (\hat{y}_i - \bar{Y})^2}{\sum (y_i - \hat{y}_i)^2} \geq \lambda_0^{-\frac{2}{n}} - 1$$

$$\lambda_0^{-\frac{2}{n}} - 1 = \delta$$

$$\frac{\sum (\hat{y}_i - \bar{Y})^2}{\sum (y_i - \hat{y}_i)^2} \geq \delta$$

Esa cantidad construida es lo que constituye la región crítica, multiplicándola por una constante se forma una cantidad que se distribuye como la función **F**.

$$T = \frac{\frac{\sum (y_i - \bar{Y})^2}{(n-2)}}{\frac{\sum (y_i - \hat{y}_i)^2}{(n-2)}} = \frac{\hat{\beta}_1^2 \sum (x_i - \bar{X})^2}{\sum (y_i - \hat{y}_i)^2} \sim F_{(1, n-2)}$$

La prueba de medias genera un estadístico **F** en la cual la regla de decisión se establece como: rechace H_0 si $T > F_{(1, n-2)}^{1-\alpha}$.

Esta prueba de hipótesis puede mostrarse en una tabla conocida como ANOVA (Analysis of variance) o análisis de varianza [11].

ANOVA

Fuente de variación	Grados de Libertad	Suma de cuadrados	Cuadrados medios	F
Bajo H_0 Regresión	1	$\sum (\hat{y}_i - \bar{Y})^2$	$\sum (\hat{y}_i - \bar{Y})^2$	$\frac{\sum (\hat{y}_i - \bar{Y})^2}{\sum (y_i - \hat{y}_i)^2}$ $n - 2$
Modelo completo Residuales	$n - 2$	$\sum (y_i - \hat{y}_i)^2$	$\frac{\sum (y_i - \hat{y}_i)^2}{n - 2}$	
Modelo reducido Total	$n - 1$	$\sum (y_i - \bar{Y})^2$		

Un modelo de regresión puede considerarse bueno o malo dependiendo del tamaño de sus residuales. El modelo que presente el residual más pequeño será el mejor.

De una manera abreviada el modelo de regresión lineal múltiple se escribe como

$$Y_{n \times 1} = X_{n \times p} \underline{\beta}_{p \times 1} + \underline{\varepsilon}_{n \times 1}. \text{ A este modelo se le impondrán supuestos como:}$$

$$1. E(\underline{\varepsilon}) = \underline{0} \quad 2. Var - Cov(\underline{\varepsilon}) = \sigma^2 I$$

Todos los errores son independientes entre ellos $(\varepsilon_i \perp \varepsilon_j) \quad \forall i \neq j$ y por último aunque no necesario ε se distribuye normal multivariada con media el vector $\underline{0}$ y la matriz de varianza-covarianza $\sigma^2 I$

1.3.1. Método de mínimos cuadrados.

Estableciendo el modelo lineal múltiple se pretende como primer paso generar los estimadores de los parámetros. Al igual que el caso lineal simple se obtendrán los estimadores mínimos cuadrados

El método implica minimizar los residuales $e_i = Y - \hat{Y}$ que en el caso matricial se escribirán como:

$$\begin{aligned} \underline{e}' \underline{e} &= (Y - \hat{Y})'(Y - \hat{Y}) = G \\ G &= (Y - \hat{Y})'(Y - \hat{Y}) \\ &= (Y - X \hat{\beta})'(Y - X \hat{\beta}) \\ &= (Y' - \hat{\beta}' X')(Y - X \hat{\beta}) \\ &= Y'Y - Y'X \hat{\beta} - \hat{\beta}' X'Y + \hat{\beta}' X'X \hat{\beta} \\ &= Y'Y - 2 \hat{\beta}' X'Y + \hat{\beta}' X'X \hat{\beta} \\ &= \sum (e_i^2) \end{aligned}$$

Tomando en cuenta que $X'X$ es simétrica y sus columnas son linealmente independientes para poder encontrar los estimadores de los β_i .

Ahora derivando G con respecto a cada $\underline{\hat{\beta}}$

$$\frac{\partial G}{\partial \underline{\hat{\beta}}} = -2X'Y + 2X'X\underline{\hat{\beta}}$$

$$-2X'Y + 2X'X\underline{\hat{\beta}} = 0$$

$$2(X'X\underline{\hat{\beta}} - X'Y) = 0$$

$$X'X\underline{\hat{\beta}} = X'Y$$

$$\underline{\hat{\beta}} = (X'X)^{-1} X'Y \Leftrightarrow (X'X)^{-1}$$

El hiperplano ajustado quedará:

$$\hat{Y}_{n \times 1} = X_{n \times p} \underline{\hat{\beta}}_{p \times 1}$$

donde al sustituir a $\underline{\hat{\beta}}$ se obtiene:

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

tomando a $X(X'X)^{-1} X'$ como la matriz H de ahí que:

$$\hat{Y} = HY$$

H es una matriz importante para el análisis de regresión, la cual se caracteriza por tener la propiedad de ser una matriz idempotente⁵, para esto es necesario revisar la propiedad $H' = H$:

$$H' = H$$

$$= (X(X'X)^{-1} X')'$$

$$= (X')'(X'X)^{-1} X'$$

$$= X(X'X)^{-1} X'$$

Lo anterior se necesita para probar la idempotencia por lo tanto se tiene:

$$H^2 = HH = H$$

$$(X(X'X)^{-1} X')(X(X'X)^{-1} X') = X \underbrace{(X'X)^{-1} X'X}_{I} (X'X)^{-1} X'$$

$$= X(X'X)^{-1} X'$$

$$= H$$

⁵ idempotencia: al elevar a cualquier potencia una matriz el resultado es igual a la matriz original.

Otra matriz que es de igual importancia que H es $I - H$, la cual también es idempotente.

Recordando que:

$$\begin{aligned}(I - H)' &= I' - H' \\ &= I - H\end{aligned}$$

por lo tanto la idempotencia se muestra enseguida:

$$\begin{aligned}(I - H)^2 &= (I - H)'(I - H) \\ &= (I - H)(I - H) \\ &= I^2 - 2IH + H^2 \\ &= I^2 - 2H + H \\ &= I - H\end{aligned}$$

1.3.2. Propiedades del estimador de mínimos cuadrados de $\underline{\beta}$

Como primera cualidad se revisará el insesgamiento de $\underline{\hat{\beta}}$, por lo que se tomará la esperanza de ésta en seguida:

$$\begin{aligned}E(\underline{\hat{\beta}}) &= E\left(\left((X'X)^{-1}X'Y\right)\right) \\ &= (X'X)^{-1}X'E(Y) \\ &= \underbrace{(X'X)^{-1}X'X}_{I}\underline{\beta} \\ &= \underline{\beta}\end{aligned}$$

por lo tanto $\underline{\hat{\beta}}$ es insesgado a $\underline{\beta}$.

Por otra parte se puede calcular la esperanza de \hat{Y} que es:

$$\begin{aligned}E(\hat{Y}) &= E\left(X(X'X)^{-1}X'Y\right) \\ &= HE(Y) \\ &= HE(X\underline{\beta} + \varepsilon) \\ &= HE(X\underline{\beta}) + E(\varepsilon) \\ &= HE(X\underline{\beta}) + 0 \\ &= HX\underline{\beta} \\ &= X\underline{\beta}\end{aligned}$$

dado que $HX = \left(X(X'X)^{-1}X'\right)X = X \underbrace{\left((X'X)^{-1}X'X\right)}_I = X$

Para revisar otras propiedades se requiere obtener la matriz de varianzas-covarianzas de $\underline{\hat{\beta}}$, por lo que:

$$\begin{aligned}
 \text{Var} - \text{Cov}(\underline{\hat{\beta}}) &= \text{Var}(\underline{\hat{\beta}}) \\
 &= E\left(\left(\underline{\hat{\beta}} - E(\underline{\hat{\beta}})\right)\left(\underline{\hat{\beta}} - E(\underline{\hat{\beta}})\right)'\right) \\
 &= E\left[\left(\underline{\hat{\beta}} - \underline{\beta}\right)\left(\underline{\hat{\beta}} - \underline{\beta}\right)'\right] \\
 &= E\left(\underline{\hat{\beta}}\underline{\hat{\beta}}' - \underline{\hat{\beta}}\underline{\beta}' - \underline{\beta}\underline{\hat{\beta}}' + \underline{\beta}\underline{\beta}'\right) \\
 &= E\left(\underline{\hat{\beta}}\underline{\hat{\beta}}' - 2\left(\underline{\hat{\beta}}\underline{\beta}'\right) + \underline{\beta}\underline{\beta}'\right) \\
 &= E\left[\left(X'X\right)^{-1}X'Y\left(\left(X'X\right)^{-1}\left(X'Y\right)\right)' - 2\left(X'X\right)^{-1}X'Y\underline{\beta}' + \underline{\beta}\underline{\beta}'\right] \\
 &= E\left[\left(X'X\right)^{-1}X'Y\left(X'Y\right)\left(\left(X'X\right)^{-1}\right)'\right] - E\left[2\left(X'X\right)^{-1}X'Y\underline{\beta}'\right] + E\left[\underline{\beta}\underline{\beta}'\right] \\
 &= E\left[\left(X'X\right)^{-1}X'YY'X\left(X'X\right)^{-1}\right] - 2E\left[\left(X'X\right)^{-1}X'Y\underline{\beta}'\right] + E\left[\underline{\beta}\underline{\beta}'\right] \\
 &= \left(X'X\right)^{-1}X'E\left[YY'\right]X\left(X'X\right)^{-1} - 2\left(X'X\right)^{-1}X'E\left[Y\right]\underline{\beta}' + \underline{\beta}\underline{\beta}' \\
 &= \left(X'X\right)^{-1}X'\left(X\underline{\beta}\underline{\beta}'X' + \sigma^2I\right)X\left(X'X\right)^{-1} - 2\left(X'X\right)^{-1}X'X\underline{\beta}\underline{\beta}' + \underline{\beta}\underline{\beta}' \\
 &= \left(\left(X'X\right)^{-1}X'X\underline{\beta}\underline{\beta}'X' + \left(X'X\right)^{-1}X'\sigma^2I\right)X\left(X'X\right)^{-1} \\
 &\quad - 2\left(X'X\right)^{-1}X'X\underline{\beta}\underline{\beta}' + \underline{\beta}\underline{\beta}' \\
 &= \underbrace{\left(X'X\right)^{-1}X'X\underline{\beta}\underline{\beta}'X'X\left(X'X\right)^{-1}}_I + \underbrace{\left(X'X\right)^{-1}X'\sigma^2IX\left(X'X\right)^{-1}}_I \\
 &\quad - 2\left(X'X\right)^{-1}X'X\underline{\beta}\underline{\beta}' + \underline{\beta}\underline{\beta}' \\
 &= \underline{\beta}\underline{\beta}' - 2\underline{\beta}\underline{\beta}' + \underline{\beta}\underline{\beta}' + \sigma^2\left(X'X\right)^{-1} \\
 &\quad 2\underline{\beta}\underline{\beta}' - 2\underline{\beta}\underline{\beta}' + \sigma^2\left(X'X\right)^{-1} \\
 &= \sigma^2\left(X'X\right)^{-1}
 \end{aligned}$$

Hasta este momento lo que se puede obtener como cualidad es que los estimadores mínimo cuadrados de $\underline{\beta}$ son insesgados a ella, pero no se puede saber si son consistentes en error cuadrático medio ya que su matriz de varianzas-covarianzas no se da de manera explícita.

1.3.3. Estimadores Máximo Verosímiles

Para obtener los estimadores por este método, se requiere de la función de verosimilitud de y la cual por el supuesto de que $\underline{\varepsilon} \sim N(0, \sigma^2 I)$ y por ser $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$ entonces:

$$Y \sim N(E(X\underline{\beta} + \underline{\varepsilon}), \text{Var} - \text{Cov}(Y)) \quad Y \sim N(X\underline{\beta}, \sigma^2 I) \text{ asi}$$

$$f(\underline{Y}; \underline{\beta}, \sigma^2 I) = \frac{e^{-\frac{1}{2\sigma^2}(\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})}}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}}}$$

Al realizar la estimación máximo verosímil [12]

$$\begin{aligned} \ln f(\underline{\beta}, \sigma^2 I, Y) &= -\frac{1}{2\sigma^2}(\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta}) - \frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 \\ &= -\frac{1}{2\sigma^2}(Y'Y - Y'X\underline{\beta} - \underline{\beta}'X'Y + \underline{\beta}'X'X\underline{\beta}) - \frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 \\ \frac{\partial \ln f(\underline{\beta}, \sigma^2 I, Y)}{\partial \underline{\beta}} &= -\frac{1}{2\sigma^2}(-2X'Y + 2X'X\underline{\beta}) \\ &= -\frac{1}{\sigma^2}(-X'Y + X'X\underline{\beta}) = 0 \\ &= -X'Y + X'X\underline{\beta} = 0 \\ &= X'X\underline{\beta} = X'Y \end{aligned}$$

$$\boxed{\hat{\underline{\beta}} = (X'X)^{-1} X'Y} \quad \text{Estimador máximo verosímil}^6$$

$$\begin{aligned} \frac{\partial \ln f(\underline{\beta}, \sigma^2, Y)}{\partial \sigma} &= \frac{2(Y - X\underline{\beta})'(Y - X\underline{\beta})}{2\sigma^3} - \frac{n2\sigma}{2\sigma^2} \\ \frac{n}{\sigma} &= \frac{(Y - X\underline{\beta})'(Y - X\underline{\beta})}{\sigma^3} \\ \frac{n\sigma^3}{\sigma} &= (Y - X\underline{\beta})'(Y - X\underline{\beta}) \end{aligned}$$

$$\boxed{\hat{\sigma}^2 = \frac{(Y - X\hat{\underline{\beta}})'(Y - X\hat{\underline{\beta}})}{n}} \quad \text{Estimador máximo verosímil}$$

Como se sabe, de la teoría de la estimación se dice que todo estimador máximo verosímil es suficiente, por lo que $\hat{\underline{\beta}}$ y $\hat{\sigma}^2$ son suficientes para $\underline{\beta}$ y σ^2 respectivamente.

⁶ Si existe $(X'X)^{-1}$

1.3.4. Residuales

Dado un modelo teórico y su ajuste, es de gran interés ver cuánto difieren éstos, a la diferencia aritmética de los dos modelos se le llama residual, es decir:

$$\begin{aligned}\underline{e} &= Y - \hat{Y} \\ &= Y - X \hat{\underline{\beta}} \quad \text{usando } \hat{Y} = HY \\ &= Y - HY \\ &= (I - H)Y\end{aligned}$$

Considerando que los estimadores mínimos cuadrados de $\underline{\beta}$ surgen por que se quiere minimizar la suma de los cuadrados de los residuales, que en este tipo de notación quedarían:

$$\underline{e}'\underline{e} = (Y - \hat{Y})'(Y - \hat{Y})$$

Sustituyendo los valores correspondientes, considerando los supuestos ya definidos para el modelo, y tomando solo un factor del producto, se tiene:

$$\begin{aligned}Y - \hat{Y} &= X\underline{\beta} + \underline{\varepsilon} - X\underline{\hat{\beta}} \\ &= X\underline{\beta} + \underline{\varepsilon} - X(X'X)^{-1}X'Y \\ &= X\underline{\beta} + \underline{\varepsilon} - X(X'X)^{-1}X'(X\underline{\beta} + \underline{\varepsilon}) \\ &= X\underline{\beta} + \underline{\varepsilon} - X(X'X)^{-1}X'X\underline{\beta} - X(X'X)^{-1}X'\underline{\varepsilon} \\ &= \underline{\varepsilon} - X(X'X)^{-1}X'\underline{\varepsilon} \\ &= (I - X(X'X)^{-1}X')\underline{\varepsilon} \\ &= (I - H)\underline{\varepsilon} \\ \underline{e}'\underline{e} &= (Y - \hat{Y})'(Y - \hat{Y}) \\ &= ((I - H)\underline{\varepsilon})'(I - H)\underline{\varepsilon} \\ &= \underline{\varepsilon}'(I - H)'(I - H)\underline{\varepsilon} \\ &= \underline{\varepsilon}'(I - H)\underline{\varepsilon}\end{aligned}$$

por la idempotencia de $(I - H)$, este resultado muestra la relación existente entre residuales y errores.

Por otra lado si se toma el valor esperado de la suma de los residuales al cuadrado se observa que:

$$\begin{aligned}
 E(\underline{e}'\underline{e}) &= E(\underline{\varepsilon}'(I-H)\underline{\varepsilon}) \\
 &= E((\underline{\varepsilon}\underline{\varepsilon}')(I-H)) \\
 &= E(\underline{\varepsilon}\underline{\varepsilon}')(I-H) \\
 &= E(\underline{\varepsilon}\underline{\varepsilon}')\text{tr}(I-H) \\
 &= \sigma^2\text{tr}(I-H) \\
 &= \sigma^2(\text{tr}I - \text{tr}H) \\
 &= \sigma^2(n-p)
 \end{aligned}$$

se concluye

De aquí se puede deducir el estimador insesgado para σ^2 que resulta ser $\Rightarrow \hat{\sigma}^2 = \frac{\underline{e}'\underline{e}}{n-p}$.

1.3.5. Intervalos de confianza para cada β_i

El interés de estimar puntualmente al parámetro $\underline{\beta}$, puede ampliarse generando intervalos que indiquen cómo y cuánto varía cada β_i , por lo que se hace necesario dar a conocer como son los intervalos de confianza de cada una de las β_i . No es posible obtener una expresión de vectores de los intervalos, por lo que se reduce a un concepto univariado, es decir, el cálculo del intervalo para cada coeficiente $\underline{\beta}$ [14].

Como $E(\hat{\underline{\beta}}) = \underline{\beta}$ y su $\text{Var} - \text{Cov}(\hat{\underline{\beta}}) = \sigma^2(X'X)^{-1}$, por otro lado se puede tomar también como $E(\hat{\beta}_i) = \beta_i$ y la expresión de la varianza de $\hat{\beta}_i$ es $\text{Var}(\hat{\beta}_i) = \sigma^2 v_{ii}$

Considerando que $\hat{\beta}_i \sim N(\beta_i, \sigma^2 v_{ii})$ y procediendo como en la sección 1.2.5.1. se tiene:

$$P \left(-a < \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma} \sqrt{\frac{nv_{ii}}{n-p}}} < a \right) = 1 - \alpha$$

Así el intervalo de confianza al $(1-\alpha) \times 100\%$ para β_i es:

$$\left(\hat{\beta}_i - a\hat{\sigma} \sqrt{\frac{nv_{ii}}{n-p}} < \beta_i < \hat{\beta}_i + a\hat{\sigma} \sqrt{\frac{nv_{ii}}{n-p}} \right)$$

donde a es el cuantil $1 - \frac{\alpha}{2}$ de una distribución t .

REGRESIÓN “ NO LINEAL “

2.1. Introducción

En algunos casos reales al aplicar el método de mínimos cuadrados lineales, se pueden presentar características que se alejen de la posibilidad de manejar un modelo lineal, ya que algunas de las variables que se usan pueden estar relacionadas de una forma no lineal.

Se presenta la ecuación:

$$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \dots + \beta_p Z_p + \varepsilon$$

Z_i puede representar una función de variables de predicción básicas x_1, x_2, \dots, x_k .

Puede $Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \dots + \beta_p Z_p + \varepsilon$ representar una gran variedad de situaciones existentes en las cuales aplicar un modelo de regresión lineal no es apropiado.

Algunas veces la información conduce a diversas alternativas de modelos. Cuando se está manejando uno de forma no lineal, se debe generalmente preferir ajustar tal como sea posible. Aplicar el método de mínimos cuadrados en la regresión no lineal y usando las ecuaciones originales para encontrar los valores de β_i , se transforman en no lineales y el proceso para encontrar una solución se complica. Tal información puede mostrar directamente la forma actual del verdadero modelo, o puede ser representado por un conjunto de ecuaciones diferenciales, las cuales deben satisfacer a éste.

Se define como modelo no lineal al que no presenta la forma de la siguiente ecuación $Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \dots + \beta_p Z_p + \varepsilon$, es decir que es no lineal en parámetros, por ejemplo:

$$Y = \exp(\theta_1 + \theta_2 t^2 + \varepsilon)$$

$$Y = \frac{\theta_1}{\theta_1 - \theta_2} [e^{-\theta_2 t} - e^{-\theta_1 t}] + \varepsilon$$

En estos ejemplos los parámetros para ser estimados son denotados por θ más que β , t es la variable individual de predicción, y ε es el término de error aleatorio con $E(\varepsilon) = 0$ y $V(\varepsilon) = \sigma^2$ constante. Se podía también haber escrito estos modelos sin ε y reemplazar Y por η . Así que, los modelos deberían mostrarse como valores verdaderos de respuesta, η , que depende de t .

Al especificar los errores enteros del modelo, se da lo siguiente.

Los modelos en $Y = \exp(\theta_1 + \theta_2 t^2 + \varepsilon)$ y $Y = \frac{\theta_1}{\theta_1 - \theta_2} [e^{-\theta_2 t} - e^{-\theta_1 t}] + \varepsilon$ son ambos no lineales en el sentido de que ellos involucran a θ_1 y θ_2 en una forma no lineal, pero ellos son de características diferentes esencialmente $Y = \exp(\theta_1 + \theta_2 t^2 + \varepsilon)$.

Los parámetros de los modelos de regresión no lineal son diferentes [13], a los modelos de regresión lineal y es que cada uno presenta lo siguiente:

Modelo de regresión no lineal	Modelo de regresión lineal
No se distribuye normal	Se distribuye normal
Excede la mínima varianza posible	Tiene mínima varianza
El exceso de varianza varía de modelo a modelo	La mínima varianza varía de modelo a modelo.

Se busca una mejora en las propiedades que se presentan en el modelo no lineal, es decir, se hace una aproximación al modelo lineal y se toma el modelo que presente notables cambios en las propiedades del modelo original mejorado.

2.2. Mínimos cuadrados no lineales

La notación estándar para la situación de mínimos cuadrados no lineales es diferente para el caso mínimos cuadrados lineales.

Esto puede confundir a primera vista, pero la notación está bien establecida.

Suponga que la postulación del modelo es de la siguiente forma :

$$Y = f(x_1, x_2, x_3, \dots, x_k; \theta_1, \theta_2, \theta_3, \dots, \theta_p) + \underline{\varepsilon}$$

escribiendo:

$$X = (x_1, x_2, \dots, x_k)'$$

$$\theta = (\theta_1, \theta_2, \dots, \theta_p)'$$

se reduce la ecuación a:

$$Y = f(X, \theta) + \underline{\varepsilon}$$

$$\begin{aligned} E(Y) &= E(f(X, \theta) + \underline{\varepsilon}) \\ &= E(f(X, \theta)) + E(\underline{\varepsilon}) \\ &= f(X, \theta) \end{aligned}$$

Supuestos para los errores $E(\underline{\varepsilon}) = 0$, se considerará también que los errores están no están correlacionados, que $Var(\underline{\varepsilon}) = \sigma^2 I$ y generalmente $\underline{\varepsilon} \sim N(0, \sigma^2 I)$ así que los errores son independientes, esto es en donde existan "n" observaciones de la forma, $Y_u, x_{1u}, x_{2u}, \dots, x_{ku}$ se puede escribir el modelo en la forma alternativa, $Y_u = f(x_{1u}, x_{2u}, \dots, x_{ku}; \theta_1, \theta_2, \dots, \theta_p) + \varepsilon_u$ donde ε_u es el u-ésimo error y $u = 1, 2, \dots, n$ abreviado $Y_u = f(x_u, \theta) + \varepsilon_u$.

Los supuestos de normalidad e independencia de los errores pueden ser escritas como $\varepsilon_u \sim N(0, \sigma^2 I)$ donde $\varepsilon_u = (\varepsilon_{1u}, \varepsilon_{2u}, \dots, \varepsilon_{ku})'$, 0 es el vector de ceros e I la matriz identidad, ambos con sus apropiados tamaños. Se sabe que mínimos cuadrados busca minimizar la función suma de los cuadrados del error, se define para el modelo no lineal, esta función y los datos dados como:

$$S(\theta) = \sum_{u=1}^n \{Y_u - f(x_u, \theta)\}^2$$

Donde Y_u y x_u son observaciones, la suma de los cuadrados es una función de θ , se denotará por $\hat{\theta}$, a los estimadores mínimos cuadrados de éste, que es un valor con el cual se minimiza $S(\theta)$.

Se verá que, si $\varepsilon \sim N(0, \sigma^2 I)$ el estimador de mínimos cuadrados de θ es también el máximo verosímil de θ . Esto porque la función de verosimilitud para este problema puede ser escrita como :

$$L(\theta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{S(\theta)}{2\sigma^2}}$$

así que si σ^2 es conocida, maximizando $L(\theta, \sigma^2)$ con respecto a θ es equivalente a maximizar $e^{-\frac{S(\theta)}{2\sigma^2}}$, por las propiedades de una exponencial y recordando que $S(\theta) \geq 0$ es como minimizar $S(\theta)$ con respecto a θ .

Para encontrar el estimador de mínimos cuadrados $\hat{\theta}$ se necesita diferenciar la siguiente expresión $S(\theta) = \sum_{u=1}^n \{Y_u - f(x_u, \theta)\}^2$ con respecto a θ . Esto da la ecuación normal la cual se debe resolver para θ .

La ecuación normal toma la forma :

$$\sum_{u=1}^n \{Y_u - f(x_u, \hat{\theta})\} \left[\frac{\partial f(x_u, \theta)}{\partial \theta_i} \right]_{\theta=\hat{\theta}} = 0$$

para $i = 1, 2, \dots, p$ donde la ecuación denotada por los corchetes es la derivada de $f(x_u, \theta)$, con respecto a θ_i , con toda θ remplazada por la correspondiente $\hat{\theta}$ la cual tienen la misma suscripción. Al derivar la función $f(x_u, \theta)$ como en el caso lineal queda como una función de x_u , solamente y no involucra a $\hat{\theta}$ en todo.

es decir es independiente de θ . Esto sale de la ecuación normal, como una lineal en los términos $\theta_1, \theta_2, \dots, \theta_p$, como ejemplo; encontrar la forma para obtener el estimador mínimo cuadrado de $\hat{\theta}$ de θ para el modelo $Y = f(\theta, t) + \varepsilon$ donde $f(\theta, t) = e^{-\theta t}$ donde hay n-pares de observaciones de la siguiente manera y $(t_1, Y_1), (t_2, Y_2), \dots, (t_n, Y_n)$ están disponibles, encontrando que $\frac{\partial f}{\partial \theta} = -te^{-\theta t}$ y aplicando $\sum_{u=1}^n \{Y_u - f(x_u, \hat{\theta})\} \left[\frac{\partial f(x_u, \theta)}{\partial \theta_i} \right]_{\theta=\hat{\theta}} = 0$ conduce a una ecuación normal individual.

$$\sum_{u=1}^n [Y_u - e^{-\hat{\theta} t_u}] [-t_u e^{-\hat{\theta} t_u}] = 0$$

ó

$$\sum_{u=1}^n Y_u t_u e^{-\hat{\theta} t_u} - \sum_{u=1}^n t_u e^{-2\hat{\theta} t_u} = 0$$

Se observa que con un parámetro y un modelo no lineal simple al comparar con $\hat{\theta}$ y querer resolver la ecuación normal no es fácil.

Entre más parámetros estén involucrados, la solución de la ecuación normal, puede ser extremadamente difícil de obtener y los métodos iterativos pueden ser empleados en casi todos los casos.

Para arreglar las dificultades, puede pasar que la solución múltiple exista, correspondiendo a valores múltiples estacionarios de una función $S(\hat{\theta})$ (mínimos locales o mínimos globales).

En el caso de modelos no lineales, no se puede establecer un procedimiento o enunciar una regla general sobre las propiedades de los estimadores excepto en muestras grandes.

El insesgamiento y mínima varianza son aproximadas, solamente en los límites de muestras que llegan a ser grandes.

2.3. Estimación de los parámetros de un sistema no lineal

En algunos problemas no lineales es más conveniente escribir, la ecuación normal

$$\sum_{u=1}^n \{Y_u - f(x_u, \hat{\theta})\} \left[\frac{\partial f(x_u, \theta)}{\partial \theta_i} \right]_{\theta=\hat{\theta}} = 0$$

y desarrollar una técnica iterativa para resolverla, si esto trabaja satisfactoriamente o no depende de la forma de la ecuación y el método iterativo [15] entonces es usado. En suma para esta aproximación existen diversos métodos concurrentes empleados y disponibles para obtener estimadores por una acostumbrada forma de hacer cálculos, los métodos son los siguientes:

- 1.- Transformaciones (Proceso no iterativo)
- 2.- Linealización
- 3.- Descenso por la pendiente máxima
- 4.- Acuerdo Marquardt

Tomando en cuenta que se puede llevar a cabo una transformación en un modelo determinado se debe aclarar que ésta no pertenece a los métodos iterativos ya que se aplica cuando en la ecuación se presta para linealizarla y esto hace que el modelo se pueda manejar de una manera muy sencilla, sin embargo se considera como uno de los procedimientos para aproximar una forma de tipo no lineal a una lineal.

Linealización es la que más se desarrollará, por ser, la que generalmente se utiliza en estos casos comparándola con el uso del resto de los métodos que son descenso por la pendiente máxima y acuerdo Marquardt de los cuales sólo se harán breves menciones de su forma de trabajo, con el fin de considerarlos dentro de los procesos iterativos y como auxiliares para realizar una regresión cuando se presenten situaciones en donde no se puede aplicar una transformación o la linealización.

2.3.1. Transformaciones

Se presentan modelos en los que es fácil apreciar el problema de no linealidad, algunas veces es útil considerar una transformación que induzca a la linealidad.

Al llevar a cabo una transformación [2], se tiene una ventaja a considerar como buena en la regresión no lineal, ya que es fácil para obtener los valores iniciales de algunos parámetros en el modelo que se esté manejando .

Recordando el modelo ya mencionado $Y = \exp(\theta_1 + \theta_2 t^2 + \varepsilon)$ al cual se le puede aplicar una transformación, logaritmo natural (ln) ;

$$\ln Y = \theta_1 + \theta_2 t^2 + \varepsilon$$

quedando de la forma $Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \dots + \beta_p Z_p + \varepsilon$ que es lineal en parámetros, se puede por, así decir, que el modelo dado en la ecuación:

$$Y = \exp(\theta_1 + \theta_2 t^2 + \varepsilon)$$

es intrínsecamente lineal, por lo que puede ser transformada en una forma lineal.

Es importante decir que si en la ecuación $\ln Y = \theta_1 + \theta_2 t^2 + \varepsilon$ se sustituye, $\ln Y = Y^*$ y $t^2 = t^*$, los parámetros y la variable de predicción serían lineales, obteniéndose la ecuación $Y^* = \theta_1 + \theta_2 t^* + \varepsilon$, a la cual se le puede aplicar una regresión simple de mínimos cuadrados.

Por otro lado $Y = \frac{\theta_1}{\theta_1 - \theta_2} [e^{-\theta_1 t} - e^{-\theta_2 t}] + \varepsilon$ es complicado llevarlo a una forma lineal en los parámetros, a tal modelo se le llama intrínsecamente no lineal [2].

Un modelo intrínsecamente lineal es aquel que se puede hacer lineal por una transformación de parámetros, por ejemplo:

$Y = e^{\theta}x + \varepsilon$ es de este tipo, entonces, si se transforma $\beta = e^{\theta}$, el modelo llega a ser $Y^{**} = \beta x + \varepsilon$.

Y en este último modelo se usa la regresión lineal simple para estimar el parámetro β en $Y^{**} = \beta x + \varepsilon$. Sin embargo el estimador lineal de mínimos cuadrados del parámetro en $Y^{**} = \beta x + \varepsilon$ no sería en general equivalente al del estimador del parámetro no lineal en el modelo original:

$$Y = e^{\theta}x + \varepsilon.$$

La razón es que en el modelo no lineal $Y = e^{\theta}x + \varepsilon$, los mínimos cuadrados emplean la minimización de la suma de los residuales al cuadrado sobre Y , mientras en el modelo transformado se está minimizando la suma de los residuales al cuadrado sobre el Y^{**} , lo mismo pasa en el caso de Y^* .

Se debe tomar en cuenta que realizar una transformación de los datos involucra una transformación en la distribución de éstos, lo cual afecta a los supuestos esperados para tener una buena aproximación.

En caso de violar los supuestos en un alto grado, entonces será apropiado aplicar la parte de regresión no lineal mediante alguno de sus métodos iterativos, sobre los datos originales.

Recordando que lo que se busca al llevar acabo una regresión es poder tener un buen ajuste de los datos para una óptima interpretación de éstos y considerando también que se deben tener buenas propiedades, como varianza mínima, una buena distribución normal, etc.

Existen algunos modelos llamados condicionalmente lineales en parámetros los cuales presentan ventajas que pueden ser favorables o de mucha ayuda al estar trabajando en la regresión no lineal.

2.3.2. Linealización o Método Gauss-Newton

Este método usa los resultados mínimos cuadrados lineales haciendo una aproximación lineal de la función esperada en una sucesión de pasos, tratando de mejorar el parámetro. Supóngase que el modelo propuesto es de la forma:

$$Y_u = f(x_u, \theta) + \varepsilon_u$$

Sea $\theta_{10}, \theta_{20}, \dots, \theta_{p0}$ valores iniciales para los parámetros $\theta_1, \theta_2, \dots, \theta_p$. Estos valores iniciales pueden ser predeterminados o estimados, basados en cualquier información disponible. Pueden ser valores supuestos por la información obtenida en un ajuste a una ecuación similar en una prueba diferente [5].

Esos valores iniciales serán estrechamente mejorados sobre las iteraciones sucesivas descritas abajo. Si se lleva a una expansión de Taylor a $f(x_u, \theta)$ cerca del punto θ_0 donde $\theta_0 = (\theta_{10}, \theta_{20}, \dots, \theta_{p0})'$ un punto inicial y se reduce la expansión en la primera derivada, se puede decir que, en donde θ es cercano a θ_0 , y considerando el resultando

$$f(x_u, \theta) = f(x_u, \theta_0) + \sum_{i=1}^p \left[\frac{\partial f(x_u, \theta)}{\partial \theta_i} \right]_{\theta=\theta_0} (\theta_i - \theta_{i0})$$

se puede ver que la ecuación:

$$Y_u = f(x_u, \theta) + \varepsilon_u$$

es de la forma aproximada:

$$Y_u = f(x_u, \theta_0) + \sum_{i=1}^p \left[\frac{\partial f(x_u, \theta)}{\partial \theta_i} \right]_{\theta=\theta_0} (\theta_i - \theta_{i0}) + \varepsilon_u$$

$$Y_u - f(x_u, \theta_0) = f(x_u, \theta) - f(x_u, \theta_0) + \sum_{i=1}^p \left[\frac{\partial f(x_u, \theta)}{\partial \theta_i} \right]_{\theta=\theta_0} (\theta_i - \theta_{i0}) - f(x_u, \theta_0) + \varepsilon_u$$

$$Y_u - f(x_u, \theta_0) = \sum_{i=1}^p \left[\frac{\partial f(x_u, \theta)}{\partial \theta_i} \right]_{\theta=\theta_0} (\theta_i - \theta_{i0}) + \varepsilon_u$$

Si se escribe

$$f_u^0 = f(x_u, \theta_0)$$

$$\beta_i^0 = \theta_i - \theta_{i0}$$

$$Z_{iu}^0 = \left[\frac{\partial f(x_u, \theta)}{\partial \theta_i} \right]_{\theta=\theta_0}$$

Así f_u^0 es el valor de la función $f(x_u, \theta)$ sustituyendo el valor del vector que se desea estimar θ , por θ_{i0} , β_i^0 representa la diferencia entre la i -ésima componente del vector paramétrico por estimar y la correspondiente del previamente elegido. Es importante hacer notar que el estimar β_i^0 , equivale a estimar θ_i . Por último Z_{iu}^0 es el vector de derivadas parciales de la función $f(x_u, \theta)$ con respecto a θ evaluado en θ_0 . De donde se obtiene:

$$Y_u - f_u^0 = \sum_{i=1}^p \beta_i^0 Z_{iu}^0 + \varepsilon_u$$

en otras palabras es de la forma:

$$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \dots + \beta_p Z_p + \varepsilon$$

para el selecto orden de aproximación.

Se puede ahora estimar los parámetros β_i^0 , $i = 1, 2, \dots, p$ para aplicar la teoría de mínimos cuadrados lineales. En forma matricial se tiene:

$$Z_0 = \begin{bmatrix} Z_{11}^0 & Z_{21}^0 & \dots & Z_{p1}^0 \\ \vdots & \vdots & \dots & \vdots \\ Z_{1n}^0 & Z_{2n}^0 & \dots & Z_{pn}^0 \end{bmatrix} = \{Z_{iu}^0\}_{n \times p} \quad b_0 = \begin{bmatrix} b_1^0 \\ b_2^0 \\ \vdots \\ b_p^0 \end{bmatrix} \quad y_0 = \begin{bmatrix} Y_1 & - & f_1^0 \\ \vdots & \dots & \vdots \\ Y_n & - & f_n^0 \end{bmatrix} = Y - f^0$$

Donde $y_0 = b_0 Z_0$ es lineal en el parámetro b_0 de tal manera que Z_0 es de rango completo.

El estimador de $b_0 = (b_1^0, b_2^0, \dots, b_p^0)'$ es dado por $\hat{b}_0 = (Z_0' Z_0)^{-1} Z_0' (Y - f^0)$ el vector \hat{b}_0 por lo tanto minimiza la suma de los cuadrados:

$$SC(\theta) = \sum_{u=1}^n \left\{ Y_u - f(x_u, \theta_0) - \sum_{i=1}^p \beta_i^0 Z_{iu}^0 \right\}^2$$

con respecto a β^0_i $i = 1, 2, \dots, p$, donde $\beta^0_i = \theta_i - \theta_{i0}$ se escribe $b^0_i = \theta_{i1} - \theta_{i0}$ entonces θ_{i1} , $i = 1, 2, \dots, p$ se puede pensar como el mejor estimador de θ .

Se debe tener cuidado con la diferencia entre la suma de los cuadrados que es la siguiente

$S(\theta) = \sum_{u=1}^n (Y_u - f(x_u, \theta_0))^2$, donde el modelo apropiado no lineal es usado y la suma de los cuadrados $SC(\theta)$ al aproximar la expansión lineal del modelo es empleada.

Se puede ahora dar lugar a los valores θ_{i1} , los estimadores revisados en este trabajo fueron hechos sobre los valores θ_{i0} y se hará exactamente el mismo procedimiento descrito en la ecuación:

$$f(x_u, \theta) = f(x_u, \theta_0) + \sum_{i=1}^p \left[\frac{\partial f(x_u, \theta)}{\partial \theta_i} \right]_{\theta=\theta_0} (\theta_i - \theta_{i0})$$

$$f(x_u, \theta) = f(x_u, \theta_0) + Z_{1u} (\theta_1 - \theta_{10}) + Z_{2u} (\theta_2 - \theta_{20}) + \dots + Z_{pu} (\theta_p - \theta_{p0})$$

$$Z^0_{iu} = \left[\frac{\partial f(x_u, \theta)}{\partial \theta_i} \right]_{\theta=\theta_0} \quad i = 1, 2, \dots, p$$

a través de $SC(\theta)$, pero reemplazando todos los subíndices cero por uno. Esto llevará a otro conjunto de estimadores revisados θ_{i2} , y así sucesivamente.

En forma de vector, extendiendo la notación previa en una forma más simple, se puede escribir:

$$\begin{aligned} \theta_{j+1} &= \theta_j + b_j \\ &= \theta_j + (Z^j_j Z_j)^{-1} Z^j_j (Y - f^j) \end{aligned}$$

Donde :

$$Z_j = \{Z^j_{iu}\}$$

$$f^j = (f^j_1, f^j_2, \dots, f^j_n)'$$

$$\theta_j = \theta_{1j}, \theta_{2j}, \dots, \theta_{pj}$$

Este proceso iterativo se sigue hasta que la se converja a una solución, esto es hasta la j -ésima iteración sucesiva, es decir $j, j + 1$.

$$\left| \frac{\{\theta_{i(j+1)} - \theta_{ij}\}}{\theta_{ij}} \right| < \delta \quad i = 1, 2, \dots, p$$

donde δ es alguna cantidad preespecificada y suficientemente pequeña.

En cada estado del procedimiento iterativo, $S(\theta_j)$ puede ser evaluada para ver si va disminuyendo en sus valores.

Estimación de σ^2

Al lograr que el procedimiento de estimación converja hacia una solución final de los parámetros estimados, se logra también un estimado de la varianza σ^2 del error a partir del cuadrado medio residual.

$$\hat{\sigma}^2 = \frac{\sum (Y_u - \hat{Y}_u)^2}{n - p}$$

si se ve como

$$\hat{\sigma}^2 = \frac{\sum (Y_u - f(x_u, \hat{\theta}))^2}{n - p} = \frac{\delta(\hat{\theta})}{n - p}$$

recordando que p es el número de parámetros en el modelo de regresión no lineal.

Se puede estimar el valor de la matriz de covarianza asintótica (de muestras grandes) a través del vector de parámetros $\hat{\theta}$ que sería:

$$Var(\hat{\theta}) = \hat{\sigma}^2 (Z'Z)^{-1}$$

donde Z es compuesta por las derivadas parciales ya definidas y evaluada en el estimado de θ del paso final en mínimos cuadrados.

2.3.2.1. La importancia de buenos valores iniciales

Para un buen ajuste del modelo no lineal se requiere de unos buenos valores iniciales θ_0 ; es decir θ_0 tan cercano como sea posible a los valores reales de los parámetros [15].

Todos los procedimientos iterativos requieren de valores iniciales, toda la información prioritaria disponible debe ser usada para hacer esos valores iniciales tan confiables como se pueda.

Buenos valores iniciales a menudo permiten una técnica iterativa para que una solución converja mucho más rápido que de otra forma posible.

También si existe mínimo múltiple o si existen diversos mínimos locales en resumen da un mínimo absoluto.

Buenos valores iniciales, son los θ_0 más cercanos a los θ reales del parámetro y reducen la dificultad de convergencia.

Una elección pobre de los valores iniciales puede causar que converja a un indeseado punto estacionario de la superficie de suma de cuadrados, mínimo local de la función y puede ser inapropiada para la solución.

En el modelo de regresión no lineal, los parámetros a menudo tienen algún significado físico que puede ayudar a obtener valores iniciales, incluso al graficar la función esperada, para diversos valores de los parámetros, esto con la finalidad de familiarizarse con el comportamiento del modelo y ver como al cambiar los valores del parámetro afectan a éste.

Cuando un modelo no lineal conduce a un punto indeseable, éste puede tener valores paramétricos, los cuales son físicamente imposibles o bien no dan el mínimo valor verdadero de $S(\theta)$.

En algunos casos se puede transformar la función y obtener valores iniciales de una manera muy sencilla.

Pero en realidad no existe una guía estándar del mecanismo para obtener valores iniciales θ_0 , para cada problema de estimación no lineal se debe trabajar en varios problemas.

Si son p parámetros, sustituidos para p conjuntos de observaciones dentro del modelo postulado ignorando el error, resultando p ecuaciones por los parámetros. Ampliamente separados de ε a menudo se trabaja mejor.

Este método es iterativo y lo que se busca es que en cada paso se observen cambios en los valores que se están obteniendo con la finalidad de poder ver si está convergiendo a algún punto y esto se puede revisar midiendo la diferencia entre los valores de cada iteración que se esté realizando, de esta manera se observa si se está avanzando a algún punto, ya sea de una manera lenta o en definitiva, no se presentan progresos significativos y no se puede ver la convergencia.

Más adelante se tocará el tema de la geometría de la linealización por este método, por medio de la cual se puede analizar si el modelo converge o no, por otro lado si se sabe que el punto crítico es rechazado y de ahí que el vector de residuales sea ortogonal a la superficie esperada Y , por lo tanto para el plano tangente en la superficie esperada en $f(x_u, \theta)$.

Se puede adoptar la ortogonalidad de los residuales en el plano tangente como una medida de convergencia.

Si se utiliza esto se debe tomar en cuenta que es difícil encontrar la ortogonalidad exacta y por ello se debe de tolerar cierta inexactitud que se pueda presentar en los residuales al momento de hacer los cálculos y se debe tomar en cuenta la variabilidad estadística en el estimador mínimos cuadrados.

Si hay una buena aproximación del valor esperado $(\hat{\theta})$ cercano a (θ) , entonces existe una región para θ que corresponde a un disco, sobre el plano tangente con un radio de proporción a $\sqrt{S(\hat{\theta})}$.

El criterio para esto será el siguiente: si el radio es pequeño, la incertidumbre numérica del estimador mínimo cuadrado es significativamente comparable para la incertidumbre estadística de los parámetros, desafortunadamente este criterio involucra a $\hat{\theta}$ vector de mínimos cuadrados conocido.

De ahí que se modifica el criterio por sustitución del estimador θ , para $\hat{\theta}$ y mide la escala de extensión del plano tangente del vector componente de residuales para la escala de la componente ortogonal del vector de residuales en θ .

El procedimiento de linealización tiene algunos errores como:

1.- Converge muy lentamente, ya que un número muy grande de iteraciones es posible requerirlas antes de estabilizar la solución a través de la suma de los cuadrados $S(\theta_j)$, decrece consistentemente conforme j se incrementa. Esta clase de comportamiento no es común pero hay posibilidad de ocurrencia, sobre todo cuando no se tienen buenos valores iniciales.

2.- Oscila ampliamente, continuamente cambiando la dirección, y a menudo incrementando, así como decreciendo la suma de los cuadrados. No obstante la solución se establece eventualmente.

3.- No converge, así que la suma de cuadrados crece iteración tras iteración sin frontera.

Cabe aclarar que en teoría este método siempre converge.

La finalidad del método de linealización es encontrar el mínimo valor de $S(\theta)$ en este tipo de modelos no lineales.

2.3.2.2. Una interpretación geométrica de mínimos cuadrados no lineales

Se presenta un modelo de la siguiente forma:

$$Y = f(x, \theta) + \varepsilon$$

que es una función no lineal en θ y tomando una muestra de tamaño n , el espacio muestral será n -dimensional que contiene al vector Y , aunque puede hablarse de un espacio de estimación, éste no estará definido por los vectores columna de una matriz X en el sentido usado para el caso de modelos lineales.

Si se tienen p parámetros independientes $\theta = \theta_1, \theta_2, \dots, \theta_p$ el espacio de estimación de un modelo no lineal, es un subespacio del espacio muestral que consiste en el conjunto de puntos de la forma:

$$(f(x_1, \theta), f(x_2, \theta), \dots, f(x_n, \theta))$$

La suma de cuadrados $S(\theta)$ sigue representando la distancia cuadrada del vector $Y = (y_1, y_2, \dots, y_n)$ al vector sobre el espacio de estimación definido por $\theta = \theta_1, \theta_2, \dots, \theta_p$, de igual forma, la minimización de $S(\theta)$ corresponde a encontrar el vector sobre el espacio de estimación que menos diste de Y [13].

Los contornos para $S(\theta) = \text{constante}$ no tienen la forma de elipsoides, y suelen formar figuras muy irregulares [5].

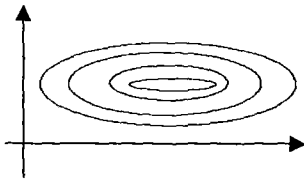
Por medio de la representación geométrica se puede ver o entender cual puede ser la diferencia entre un modelo no lineal y un modelo lineal, el cual es una forma alternativa de representar los datos de la muestra que se esté estudiando.

La función de la suma de los cuadrados de residuales $S(\theta)$ sólo depende de los elementos del parámetro θ en el modelo.

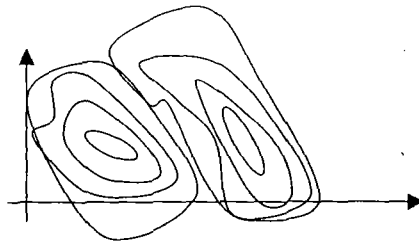
En el espacio p -dimensional, es el espacio geométrico de θ , parámetros: La función $S(\theta)$ puede ser representada por una gráfica de curvas de nivel, si el modelo fuera lineal en θ la gráfica de curvas de nivel se desearía fuera, elipsoidal y se esperaría que tuviera un sólo mínimo local, en la que cada curva de nivel en la superficie es una línea de suma constante de residuales al cuadrado.

Si el modelo no es lineal las curvas de nivel no son elipsoidales, son alargadas, tienden a ser irregulares y a menudo "presenta una superficie-aplatanada"

Contornos de un modelo lineal



Contornos de un modelo no lineal



La forma y la orientación precisa de las curvas de nivel de $S(\theta)$ dependen de la forma del modelo no lineal y de los datos que se hayan obtenido. Frecuentemente la superficie será muy alargada cercana al valor óptimo, por lo que muchas soluciones de θ producirán una suma de cuadrados de residuales cercana al óptimo global.

Quizás con diversos mínimos locales y con más de un mínimo global. Esto es, el mínimo más alto puede ser tomado en más de un θ punto de localización.

Si se hace mal el planteamiento del modelo, se obtiene como resultado un modelo sobreparamétrizado, esto es, que tiene más parámetros de los que necesita, es decir, datos inadecuados que no se usarán para estimar los parámetros propuestos.

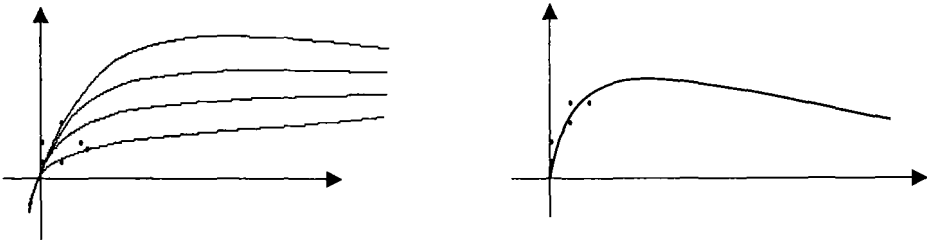
Considere $f(t, \theta_1, \theta_2)$ en la ecuación.

$$Y = \frac{\theta_1}{\theta_1 - \theta_2} [e^{-\theta_2 t} - e^{-\theta_1 t}] + \varepsilon$$

La cual representa una curva que comienza en $t = 0$ y termina en $t = \infty$ en el tamaño de la altura cero y ponerse a la altura de la curva en algún lugar entre la pendiente en $t = 0$ es θ_1 y el máximo está en:

$$t_{\text{maximo}} = \frac{\ln\left(\frac{\theta_1}{\theta_2}\right)}{(\theta_1 - \theta_2)}$$

Esto sigue de que si los datos cubren rápidamente parte de la curva se estará disponible para estimar bien a θ_1 , pero no θ_2 . Observe la siguiente figura:



En la figura de una sola curva se debe de obtener la información sobre el máximo.

La primera figura es adecuada para un modelo de un parámetro $Y = \theta t + \varepsilon$.

Pero inadecuado para estimar:

$$Y = \frac{\theta_1}{\theta_1 - \theta_2} [e^{-\theta_2 t} - e^{-\theta_1 t}] + \varepsilon$$

2.3.2.3. Una interpretación geométrica de la linealización

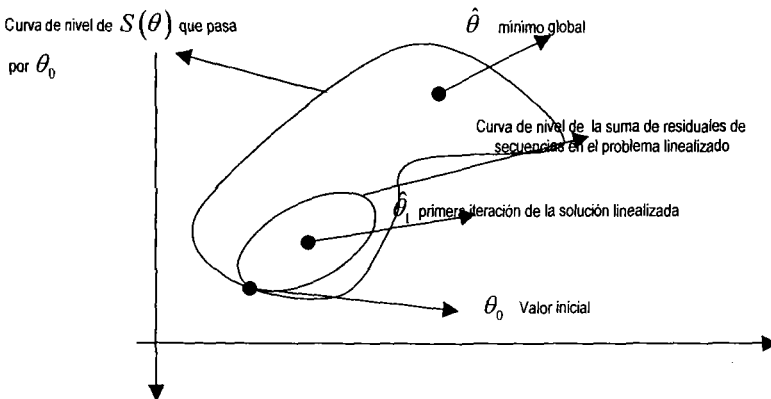
El método de linealización, reduce el problema a encontrar el mínimo más cercano a $S(\theta)$ para un modelo no lineal comenzando en θ_0 punto inicial, dentro de una serie de problemas de un modelo lineal.

La ecuación de linealización inicial:

$$f(X, \theta) = f(x_u, \theta_0) + \sum_{i=1}^p \left[\frac{\partial f(x_u, \theta)}{\partial \theta_i} \right] (\theta_i - \theta_{i0})$$

de $f(x_u, \theta)$ cerca de θ_0 reemplaza las curvas irregulares de nivel $S(\theta)$ por un conjunto de curvas elípticas $SC(\theta)$.

Las curvas irregulares de $S(\theta)$ pasan exactamente a través del punto inicial θ_0 :



Esto es, tienen la misma primera derivada respecto a la función del modelo correspondiente a la recta en θ_0 .

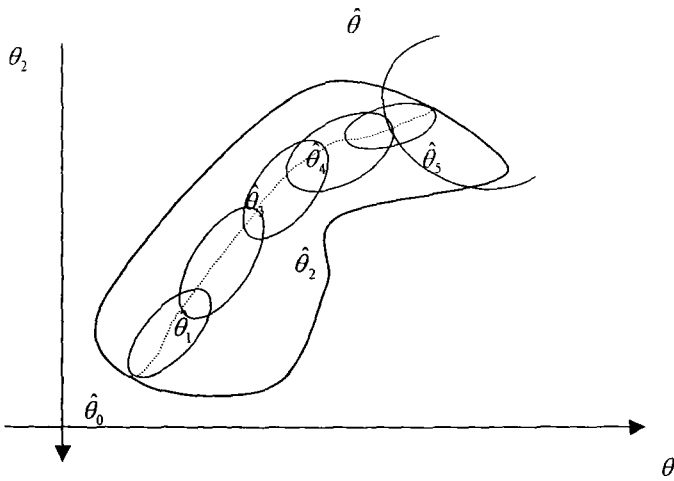
Al resolverse el problema linealizado, se va moviendo hacia el mínimo global, sobre el conjunto de curvas, esto lo hacen los mínimos cuadrados ordinarios [12].

El contorno $S(\theta)$ puede ser aproximado erróneamente o acertadamente, dependiendo de las circunstancias en que esté el supuesto modelo, de los datos disponibles, y la posición relativa de θ_0 y θ_1 en θ -espacio.

Se resuelve la linealización en θ_0 mínimo cuadrado, para alcanzar θ_1 .

Entonces en la siguiente iteración, se repite el proceso de linealización, comenzando en $\hat{\theta}_1$, lo que se espera es que, en las sucesivas iteraciones converja a $\hat{\theta}$, preferentemente a que diverja.

Gráficamente



La evolución definitiva de la linealización es una secuencia de problemas lineales para los cuales las soluciones se cierran hacia el mínimo global de la función no lineal.

La linealización se facilita con un buen valor inicial θ_0 , esto es, cuando al iniciar las iteraciones, θ_0 es un valor cercano a $\hat{\theta}$, ya que las curvas de nivel tienen una buena aproximación regularmente.

2.3.3. Descenso más pronunciado

Éste es uno de los métodos del cual se hará una breve mención con la finalidad de no dejar de tomar en cuenta que existen otros métodos de los que el modelador se puede auxiliar para llevar a cabo una regresión sobre un modelo no lineal.

Éste se concentra en la función suma de los cuadrados $S(\theta)$, ya definida y el uso de un proceso iterativo para encontrar el mínimo de la función. La idea básica es moverse de un vector inicial $\theta_0 = (\theta_{10}, \theta_{20}, \dots, \theta_{p0})$, un vector en el espacio paramétrico \mathfrak{R}^p a través del vector con componentes:

$$-\frac{\partial S(\theta)}{\partial \theta_1}, -\frac{\partial S(\theta)}{\partial \theta_2}, \dots, -\frac{\partial S(\theta)}{\partial \theta_p}$$

Cuyos valores cambian continuamente como en la trayectoria siguiente. Una manera de alcanzar esto en la práctica, sin evaluar la función derivada, es estimar el vector pendiente (sus componentes), en varios lugares de la superficie $S(\theta)$ por ajustar el plano a la función de aproximación.

El procedimiento es como sigue. Se inicia en una región particular de θ -espacio, (o el espacio paramétrico), diversos ensayos se hacen para seleccionar el mejor, se dicen combinación de niveles de $\theta_1, \theta_2, \dots, \theta_p$ y evaluando $S(\theta)$ en esas combinaciones.

Se define $Z^0 = (Z_{10}, Z_{20}, \dots, Z_{p0})$ con $Z_i^0 = -\lambda \left(\frac{\partial S}{\partial \theta_i} \Big|_{\theta=\theta_0} \right)$ donde λ , es un factor positivo de proporcionalidad [5].

Si se consideran los vectores Z^0 y θ_0 en la siguiente ecuación:

$$g(t) = S(\theta_0 + tZ^0) \text{ donde } t \in \mathfrak{R}$$

La función g toma los valores de la función S a lo largo de su vector, desde el punto θ_0 derivando esta función con respecto a t :

$$\frac{dg(t)}{dt} = \left[\frac{\delta S}{\delta \theta_i} \Big|_{\theta=\theta_0+tZ^0} \right] * Z^0$$

y evaluando esta ecuación en $t = 0$, resulta:

$$\begin{aligned} \frac{dg(0)}{dt} &= \left[\frac{\delta S}{\delta \theta_i} \Big|_{\theta=\theta_0} \right] * Z^0 \\ &= \left[\frac{\delta S}{\delta \theta_1} \Big|_{\theta=\theta_0}, \frac{\delta S}{\delta \theta_2} \Big|_{\theta=\theta_0}, \dots, \frac{\delta S}{\delta \theta_p} \Big|_{\theta=\theta_0} \right] * \left[-\lambda \frac{\delta S}{\delta \theta_1} \Big|_{\theta=\theta_0}, -\lambda \frac{\delta S}{\delta \theta_2} \Big|_{\theta=\theta_0}, \dots, -\lambda \frac{\delta S}{\delta \theta_p} \Big|_{\theta=\theta_0} \right] \end{aligned}$$

de ahí que

$$\frac{dg(0)}{dt} = -\lambda \left[\sum \frac{\delta S}{\delta \theta_i} \Big|_{\theta=\theta_0} \right]^2$$

La derivada de la función $g(t)$ en $t = 0$ resulta ser negativa, esto es, la función es decreciente en $t = 0$, así que, debe existir $t > 0$, tal que $g(t) < g(0)$.

Ahora se define $\theta_1 = \theta_0 + tZ^0$ como el nuevo punto para comenzar la iteración.

Recordando que $g(t) < g(0)$

$$S(\theta_0 + tZ^0) = g(t) < g(0) = S(\theta_0)$$

entonces

$$S(\theta_0 + tZ^0) < S(\theta_0)$$

y como $\theta_1 = \theta_0 + tZ^0$ se observa que

$$S(\theta_1) < S(\theta_0)$$

De esta manera siguiendo un proceso iterativo se obtendría una sucesión de puntos $\theta_0, \theta_1, \dots$, tal que $S(\theta_{k-1}) < S(\theta_k)$.

Y así bajo ciertas condiciones adecuadas esta sucesión tendería a un punto estacionario de la función $S(\theta)$, mismo que se tomaría como $\hat{\theta}$.

Como encontrar $t > 0$ tal que $g(t) < g(0)$; puede realizarse por medio de ensayos, observar primero el valor de la intersección de $g'(0)$ con el eje de t , si la desigualdad $g(t) < g(0)$ no se cumple, es demasiado grande t , se puede entonces tomar la mitad de este valor y observar la relación $g(t), g(0)$ y así se continúa en esta dirección hasta encontrar un punto sobre la recta $\theta_0 + t Z^0$ donde la función $S(\theta)$ toma un mínimo, entonces se detiene y se toma otra dirección de este método y se repite el procedimiento si el primer ensayo no funciona.

Ya que teóricamente, el método converge y esto se puede hacer en la práctica, después de algún rápido progreso inicial.

En resumen, el objetivo es pasar desde un punto de partida inicial θ_0 , en una dirección del vector cuyos componentes se determinan con las derivadas de la función suma de cuadrados de residuales con respecto a los elementos de θ , por lo general, esas derivadas se estiman ajustando una aproximación de primer orden, o plana, en torno al punto θ_0 .

Los coeficientes de regresión del modelo de primer orden se toman como aproximaciones de las primeras derivadas.

El método de descenso más pronunciado se usa mucho en la metodología de superficies de respuesta, para pasar de un estimado inicial de las condiciones óptimas de un proceso, a una región que contenga al óptimo con más probabilidad.

La desventaja principal es que converge lentamente, particularmente como cuando los $S(\theta)$ contornos son atenuantes de una forma aplanada.

Y esto sucede cuando la parte de descenso más pronunciado zigzaguea lentamente sobre una arista en cada iteración brindando solamente una ligera reducción en $S(\theta)$.

2.3.4. El acuerdo de Marquardt

El siguiente método es una combinación de los métodos anteriores pero de igual manera no se hará gran mención respecto al contenido y formalismo de éste, simplemente se reconocerá como otro método iterativo existente en los modelos no lineales.

El acuerdo de Marquardt, es un método para la solución de ciertos problemas no lineales en mínimos cuadrados. Este método representa una relación entre el método de linealización de Taylor y el método de descenso más pronunciado y aparece para combinar las características de ambos.

Esto es bueno, siempre y cuando converja y no lo haga bajando lentamente como en el método de descenso más pronunciado, que a menudo lo hace, sin embargo así se vuelve a enfatizar, que los otros trabajan perfectamente bien sobre algunos problemas prácticos.

El acuerdo de Marquardt, aparece para trabajar bien en muchas circunstancias y así hacer una sencilla práctica de modelos no lineales. Por razones establecidas los métodos no pueden ser determinados de manera general para todos los problemas no lineales.

La idea de este método puede ser explicada brevemente como sigue.

Suponga que se inicia de un cierto punto en el espacio paramétrico- θ si el método de descenso más pronunciado es aplicado, un cierto vector de dirección δ_x , donde δ_x puede ser el mejor vector de dirección local en la cual se mueve, para llegar a los valores más pequeños de $S(\theta)$ pero puede no ser la mejor de todas las direcciones.

Sin embargo la mejor dirección debe estar dentro de los 90° de δ_x , de otro modo, los valores de $S(\theta)$ tenderán a crecer más que a reducirse sobre dicho vector [3].

Marquardt propuso calcular el vector de incrementos en la k -ésima iteración dado por:

$$b_k = (Z_k' Z_k + \lambda I)^{-1} Z_k' (Y - f^k)$$

donde $\lambda > 0$, el procedimiento también implica reducir λ por un factor de 10 en cada iteración siempre que se satisfaga $S(\hat{\theta}_{k+1}) < S(\hat{\theta}_k)$. La estrategia es mantener a λ tan pequeña como sea posible y al mismo tiempo asegurar que la suma de cuadrados de residuales se reduce en cada iteración.

El algoritmo Marquardt da un método para interpolar valores entre vectores δ_x y δ_i y obtener un paso sustituible.

Redefiniendo como vector de corrección asociado al método de linealización δ_i , al definido como b_k en la ecuación $Y - f_u^k = Z_k \beta_k$, y vector de corrección asociado al método del descenso más pronunciado δ_g , Marquardt se percató de dos puntos importantes a saber:

- a) Cualquier método basado en la minimización de la función $S(\theta)$, deberá tener un vector asociado cuya dirección esté dentro de un rango $[0^\circ, 90^\circ]$ de $-\frac{\delta S}{\delta \theta}$, de otra forma los valores de $S(\theta)$ podrán tender a crecer más que a reducirse sobre dicho vector de corrección.
- b) Debido a la forma de la superficie $S(\theta)$, δ_i está casi 90° de δ_g , de hecho en un gran número de ensayos, Marquardt encontró que el ángulo Γ entre δ_i y δ_g era tal que: $80^\circ < \Gamma < 90^\circ$.

Con estas observaciones, Marquardt concluye que cualquier método que mejore tanto al de linealización como al de descenso más pronunciado, deberá tener un vector corrección que de alguna manera interpole entre δ_i y δ_g .

2.4. El concepto del comportamiento de estimación no lineal

A través de la representación geométrica de un modelo de regresión no lineal empleado anteriormente, éste es restringido en la práctica a muestras de tamaño $n \leq 3$ y número de parámetros $p \leq 2$. El principio conceptual es válido también para grandes dimensiones [13].

Un modelo de regresión lineal tiene una solución gráfica lineal, la cual es una línea recta para $p = 1$, un plano para $p = 2$ y un hiperplano para $p \geq 3$, y una no linealidad intrínseca cero respectivamente de los datos con los cuales el modelo está en combinación.

En resumen, las líneas sobre la solución gráfica, las cuales pueden ser llamadas "líneas paramétricas", representando valores constantes de θ_j , son elementos del vector parámetro θ , son rectas paralelas y son iguales en distancia para iguales incrementos para θ_j .

La solución gráfica es una superficie curva, y la línea parámetro sobre la solución, gráfica la proyección de esas líneas entre el plano tangente para la solución gráfica en $\hat{\theta}$, no son en general, ni rectas paralelas ni rectas de igual distancias.

Así las medidas tales como las anteriores cuantifican la curvatura de la línea parámetro, si existe carencia de paralelismo y la equidistancia, se puede dar al modelador una efectiva aproximación para que pueda ser descrito el comportamiento no lineal, de un modelo no lineal.

Existe, sin embargo, una segunda aproximación para ver el comportamiento no lineal. Esta se sigue de las propiedades del estimador de mínimos cuadrados de los parámetros del modelo lineal que se está discutiendo.

Suponiendo que los errores son independientes y están normalmente distribuidos.

El estimador de mínimos cuadrados $\hat{\theta}$ de θ en modelos lineales, es insesgado, se distribuye normal y alcanza la mínima varianza, para modelos no lineales de este estimador y tiene estas propiedades solamente de forma aproximada.

Sin embargo, se puede hablar del comportamiento no lineal de modelos no lineales en términos de las propiedades de muestreo, del estimador mínimo cuadrado.

Si el estimador mínimo cuadrado de un parámetro en un modelo no lineal es sólo ligeramente insesgado, con una distribución cercana a una normal y con una varianza ligeramente excedida de la mínima varianza, entonces es razonable decir que el comportamiento del estimador es cercano al lineal.

Si, por otro lado, el estimador de mínimos cuadrados es malamente calculado, cuando presenta una distribución lejana a la distribución normal, una varianza altamente excedida de la mínima varianza, por lo tanto, el modelo es no lineal en comportamiento, es decir, es lejano al comportamiento de un modelo lineal.

Un caso como el anterior será tomado como evidencia, ya que dado un modelo que se considere cercano en comportamiento al de un modelo lineal, pero, si este no sigue las propiedades deseables de modelos lineales, algunas de las cuales ya se conocen.

El modelo se considerará mejor, en la clase de modelos no lineales ya que su comportamiento está lejos de lo lineal.

Otras propiedades deseables del modelo cercano al modelo lineal, incluye la facilidad de obtener el estimador mínimo cuadrado y su relativamente directa interpretación.

Otros beneficios son que los valores predichos de la variable de respuesta Y será casi insesgada y que unido a la región de confianza para los parámetros serán cercanos a una forma elipsoidal características de modelos lineales.

2.5. Contornos de confianza

Una idea de no linealidad del modelo en estudio, puede ser obtenida, después de la estimación de θ , por la evaluación de la región de confianza elipsoidal, obtenida sobre la suposición de que la forma de linealización del modelo es válida alrededor de $\hat{\theta}$, el estimador final de θ . La forma de evaluación está dada por:

$$(\theta - \hat{\theta})' \hat{Z}' \hat{Z} (\theta - \hat{\theta}) \leq p S^2 F_{(p, n-p, 1-\alpha)}$$

donde \hat{Z} denota la matriz de la forma ya descrita, pero con $\hat{\theta}$ substituido en los elementos de θ_0 y donde:

$$S^2 = \frac{S(\hat{\theta})}{(n-p)}$$

Cuando la diferencia entre valores sucesivos θ_{j+1} y θ_j es suficientemente pequeño [5], la linealización termina en que $\theta_{j+1} = \hat{\theta} = \theta_j$, para propósitos prácticos, así $S(\hat{\theta})$ es un valor mínimo de $S(\theta)$ en:

$$S(\theta) = \sum_{u=1}^n \{Y_u - f(x_u, \theta)\}^2$$

Para la precisión impuesta por la terminación del procedimiento seleccionado.

Esto se puede examinar como:

$$SC(\theta) = \sum_{u=1}^n \left\{ Y_u - f(x_u, \theta_0) - \sum_{i=1}^p \beta_i^0 Z_{iu}^0 \right\}$$

Con $\hat{\theta}$, β_i^{j+1} y Z_i^{j+1} reemplazando θ_0 , β_i^0 y Z_{iu}^0 respectivamente y recordando que, para el orden de precisión impuesta por la terminación, precede $b_{j+1} = 0$ [5].

Acerca de la elipse no será una verdadera región de confianza cuando el modelo es no lineal.

Sin embargo se determina los puntos finales sobre el eje mayor de esta elipse por reducción canónica. El actual valor de $S(\theta)$ puede ser evaluado en esos puntos y comparado con cada uno. Bajo la teoría lineal los valores podrán ser todos iguales.

Un exacto contorno de confianza es definido con tomar $S(\theta) = \text{constante}$, pero no se conoce la correcta distribución de las propiedades en el caso no lineal general, no es capaz de obtener un nivel de probabilidad específica.

Sin embargo se elige el contorno tal que:

$$S(\theta) = S(\hat{\theta}) \left\{ 1 + \frac{p}{n-p} F_{(p, n-p, 1-\alpha)} \right\}$$

el cual, si el modelo es lineal da una elipse exacta $(1-\alpha) * 100$ límite y da un intervalo de confianza aproximado en el caso no lineal.

El contorno así determinado será un adecuado contorno de confianza, que no será elíptico en general, y éste es solamente el nivel de probabilidad al cual es aproximado. Cuando solamente dos parámetros son involucrados el contorno de confianza puede ser dibujado.

En resumen cuando una linealización, de un modelo no lineal es usado, todas las fórmulas y análisis de regresión no lineal pueden ser aplicados.

Algunos resultados obtenidos, son sin embargo válidos solamente hasta el punto en que la forma de linealizar da una buena aproximación a el modelo verdadero.

MÍNIMOS CUADRADOS GENERALIZADOS

3.1. Introducción. Modelos lineales generalizados

En ocasiones se puede realizar una transformación en la variable de respuesta para manejar fácilmente la no normalidad de la variable de respuesta y también la desigualdad de la varianza.

Ahora se verá un método alternativo para transformar datos, cuando no se satisfacen las condiciones de normalidad y de varianza constante.

El método se basa en el modelo lineal generalizado [10], el cual es la unión de los modelos de regresión lineal y no lineal, que también permite incorporar distribuciones de respuesta no normales.

En un modelo lineal generalizado la distribución de la variable de respuesta sólo necesita ser un miembro de la familia exponencial, que comprende las distribuciones normal, poisson, binomial, exponencial y gamma, entre sus miembros; además, el modelo lineal con error normal no es más que un caso especial del modelo lineal generalizado, por lo que en muchos aspectos se puede considerar que en éste método, se concentran aspectos del modelado, diseño de experimentos y el análisis de regresión de los datos.

Las distribuciones que son miembros de la familia exponencial tienen la forma general.

$$f(y_i, \theta_i, \phi) = e^{\left[\frac{1}{a(\phi)} (y_i \theta_i - b(\theta_i)) + h(y_i, \phi) \right]}$$

donde ϕ es un parámetro escalar, y θ_i se llama parámetro natural de localización.

Para los miembros de la familia exponencial [10],

$$\mu = E(y) = \frac{db(\theta_i)}{d\theta_i}$$

$$\begin{aligned} \text{Var}(y) &= \frac{d^2b(\theta_i)}{d\theta_i^2} a(\phi) \\ &= \frac{d\mu}{d\theta_i} a(\phi) \end{aligned}$$

Sea

$$\text{Var}(\mu) = \frac{\text{Var}(y)}{a(\phi)} = \frac{d\mu}{d\theta_i}$$

$$\frac{d\theta_i}{d\mu} = \frac{1}{\text{Var}(\mu)}$$

donde $\text{Var}(\mu)$ representa la dependencia de la varianza de la respuesta con su media. Es una característica de todas las distribuciones de la familia exponencial, a excepción de la distribución normal.

Observe en el caso de la normal:

$$\begin{aligned} f(y_i, \theta_i, \phi) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{1}{\sqrt{2\sigma^2}}(y_i - \mu)^2\right)} \\ &= (2\pi\sigma^2)^{-\frac{1}{2}} e^{\left(-\frac{1}{\sqrt{2\sigma^2}}(y_i - \mu)^2\right)} \\ &= e^{\left(-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{y_i^2}{2\sigma^2} + \frac{y_i\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right)} \\ &= e^{\left(\frac{1}{\sigma^2}\left(-\frac{y_i^2}{2} + y_i\mu - \frac{\mu^2}{2}\right) - \frac{1}{2}\ln(2\pi\sigma^2)\right)} \\ &= e^{\left(\frac{1}{\sigma^2}\left(y_i\mu - \frac{\mu^2}{2}\right) - \frac{y_i^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)\right)} \end{aligned}$$

Donde sustituyendo se tiene:

$$\begin{aligned}\theta_i &= \mu \\ b(\theta_i) &= \frac{\mu^2}{2} \\ a(\phi) &= \sigma^2 \\ h(y_i, \phi) &= -\frac{y_i^2}{2\sigma^2} - \frac{1}{2} \ln(2\pi\sigma^2) \\ E(y) &= \frac{db(\theta_i)}{d\theta_i} = \mu \\ Var(y) &= \frac{d^2b(\theta_i)}{d\theta_i^2} a(\phi) = \sigma^2\end{aligned}$$

$$f(y_i, \theta_i, \phi) = e^{\left[\frac{1}{\sigma^2} \left(y_i \mu - \frac{\mu^2}{2} \right) - \frac{y_i^2}{2\sigma^2} - \frac{1}{2} \ln(2\pi\sigma^2) \right]} = e^{\left[\frac{1}{a(\phi)} (y_i \theta_i - b(\theta_i)) + h(y_i, \phi) \right]}$$

es así como se exhibe que la función normal pertenece a la familia exponencial.

3.1.1. Funciones liga y predictores lineales

La finalidad del modelo lineal generalizado (MLG) es desarrollar un modelo lineal para una función adecuada del valor esperado de la variable respuesta [11].

Se define $\eta_i = g[E(y_i)] = g(\mu_i) = x'_i \beta$ como predictor lineal, de ahí que la respuesta esperada sea:

$$E(y_i) = g^{-1}(\eta_i) = g^{-1}(x'_i \beta)$$

g es la función liga⁷. La cual presenta muchas alternativas, tomando $\eta_i = \theta_i$, η_i es una liga canónica.

⁷ Función liga: es una función que relaciona la media de respuesta con un predictor lineal.

El siguiente cuadro muestra las ligas que se presentan para cada distribución

Distribución	Liga canónica	Nombre
Normal	$\eta_i = \mu_i$	Liga identidad
Binomial	$\eta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$	Liga logística
Poisson	$\eta_i = \ln(\lambda)$	Liga logarítmica
Exponencial	$\eta_i = \frac{1}{\lambda_i}$	Liga recíproca
Gamma	$\eta_i = \frac{1}{\lambda_i}$	Liga recíproca

Existen otro tipo de ligas que pueden ser usadas en un MLG y son:

1.- La liga probit $\eta_i = \phi[E(y_i)]$

donde ϕ se toma como la función normal estándar acumulada.

2.- La liga log-log complementaria

$$\eta_i = \ln(\ln[1 - E(y_i)])$$

3.- La liga de la familia de potencias

$$\eta_i = \begin{cases} E(y_i)^\lambda & \lambda \neq 0 \\ \ln[E(y_i)] & \lambda = 0 \end{cases}$$

En un sistema lineal general se presenta la distribución de la respuesta y la función liga, para decidir cual función liga se debe usar, es necesario tomar en cuenta como se ve afectada la función original en sus propiedades y la función liga [11], que mejor ajuste, será la apropiada para aplicar.

3.1.2. Residuales en el modelo lineal generalizado

Si existe algo que debe de tomarse en cuenta son los residuales que marcan la diferencia que se da entre el modelo original y el modelo lineal generalizado estimado, así se puede observar si sigue cumpliendo con las propiedades del término residual.

Los residuales originales y ordinarios del modelo lineal generalizado son las diferencias entre las observaciones y los valores ajustados.

$$\varepsilon_i = Y_i - \hat{Y}_i$$

En general en el análisis de residuales del modelo lineal generalizado se usan residuales de desviación. El i -ésimo residual de desviación se define como la raíz cuadrada de la contribución de la i -ésima observación a la desviación, multiplicada por el signo del residual original,

$$r_{d_i} = \sqrt{d_i} \text{signo}(Y_i - \hat{Y}_i)$$

donde d_i es la contribución de la i -ésima observación a la desviación.

Los residuales de desviación se comportan de manera similar a los residuales ordinarios en un modelo de regresión lineal con la teoría normal estándar. De esta manera al graficar los residuales de desviación, contra valores ajustados en una escala de probabilidad normal es un diagnóstico lógico. Cuando se grafican los residuales de desviación en función de valores ajustados se acostumbra transformar los primeros a una escala constante de información.

- a) Para respuestas normales, usar \hat{Y}_i
- b) Para respuestas binomiales, usar $2 \text{sen}^{-1} \sqrt{\hat{\pi}_i}$
- c) Para respuestas Poisson, usar $2\sqrt{\hat{Y}_i}$
- d) Para respuestas gamma, usar $2 \ln(\hat{Y}_i)$

3.2. Mínimos cuadrados generalizados

Se presenta casos donde son necesarios los mínimos cuadrados generalizados [6], para aplicar en una variable dependiente que no satisface la suposición de tener varianza constante.

En algunos casos usar una transformación es una buena opción para poder estabilizar a la varianza o mejorar las condiciones bajo las que se encuentre el modelo en estudio, pero existen casos en los que hacer esto no es lo más apropiado ya que probablemente se rompe la buena relación entre Y y X o destruye la suma y la distribución normal de los residuales.

Por otro lado las transformaciones logit y probit, no estabilizan a la varianza. La transformación arcoseno de la binomial estabiliza la varianza solamente si las muestras de tamaño, n_i , son iguales.

Por otro lado la varianza será proporcional a

$\frac{1}{n_i}$ y continuará la no estabilidad en la varianza aún después de la transformación.

Si el tratamiento adecuado está basado en que las muestras tienen un número desigual de observaciones, las varianzas serán diferentes incluso si las observaciones originales tenían igual varianza.

La estimación por mínimos cuadrados ordinarios no da en ocasiones la mínima varianza de los parámetros estimados cuando $Var - Cov(\varepsilon) \neq I\sigma^2$.

Se presenta el procedimiento de estimación que da mínima varianza lineal insesgada estimada, cuando la matriz de $Var - Cov$ de los errores es una matriz arbitraria simétrica definida positiva, $Var - Cov(\varepsilon) = V$.

Se considera entonces el método general, si la varianza es desigual y los errores no son independientes. Mínimos cuadrados generalizados amplía el usual modelo lineal para permitir a una arbitraria matriz $Var - Cov(\varepsilon) = V\sigma^2$ definida positiva.

3.2.1. Supuestos

Las suposiciones que se suelen hacer acerca del modelo de regresión lineal $Y = X\beta + \varepsilon$ son:

$$1) \quad E(\varepsilon) = 0$$

$$2) \quad \text{Var} - \text{Cov}(\varepsilon) = I\sigma^2$$

$$3) \quad E(\varepsilon_i, \varepsilon_j) = 0$$

$$3) \text{Var} - \text{Cov}(\varepsilon_i, \varepsilon_j) = 0$$

En ocasiones estos supuestos son poco razonables para el tipo de modelo que se esté manejando, por lo que se vera qué modificaciones se necesitan hacer al procedimiento de mínimos cuadrados ordinarios, cuando $\text{Var} - \text{Cov}(\varepsilon) \neq I\sigma^2$, es decir cuando se tiene algo de este estilo $\text{Var} - \text{Cov}(\varepsilon) = \sigma^2 V$, si se presenta el caso donde V es una matriz conocida de $n \times n$, se tiene una interpretación que no es complicada de manejar, por otro lado si se tiene que V es diagonal y esos elementos diagonales son distintos, las observaciones Y se observan no correlacionadas, pero tienen varianzas desiguales, mientras que si alguno de los elementos fuera de la diagonal de V aparecen distintos de cero, entonces se tienen observaciones correlacionadas.

La escala de la variable dependiente y el valor explicativo del modelo tiende a variar entre las distintas observaciones, aún si se controlan factores externos que puedan afectar al modelo.

El estimador de mínimos cuadrados ordinarios $\hat{\beta} = (X'X)^{-1} X'Y$ como en el caso de mínimos cuadrados no lineales, ya no son los más apropiados para manejarse, cuando $\text{Var} - \text{Cov}(\varepsilon) = \sigma^2 V$.

En estos casos es necesario realizar una transformación para corregir y mejorar los supuestos estándar de mínimos cuadrados que se vean gravemente afectados en el modelo.

3.2.2. Una matriz definida positiva, la transformación

Retomando del modelo $Y = X\beta + \varepsilon$ mínimos cuadrados ordinarios a los errores y si se presenta en este modelo $\sigma^2 V$ como la matriz de *Var - Cov* de los errores donde V debe ser no singular y definida positiva. Con un modelo en estas condiciones, lo recomendable es aplicar una transformación [15].

Se tiene que la matriz V es simétrica, de ahí que se pueda descomponer de la siguiente forma:

$$V = C K C'$$

donde las columnas de C son los vectores propios de V y los valores propios V están ordenados en la matriz diagonal K .

Como la matriz V es, además definida positiva, todos sus valores propios son positivos, entonces,

$K^{\frac{1}{2}}$ existe y viene dada por:

$$K^{\frac{1}{2}} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sqrt{\lambda_n} \end{pmatrix}$$

Se define la matriz $T' = CK^{\frac{1}{2}}$, para cualquier matriz V definida positiva es posible encontrar una matriz simétrica no singular T tal que:

$$T'T = TT = T^2 = V$$

Si T es no singular, ésta tiene una inversa T^{-1} .

Se presenta el caso donde σ^2 se desconoce y el modelo transformado es:

$$T^{-1}Y = T^{-1}X\beta + T^{-1}\varepsilon$$

definiendo las nuevas variables $Y^* = T^{-1}Y$, $X^* = T^{-1}X$ y $\varepsilon^* = T^{-1}\varepsilon$ queda:

$$Y^* = X^*\beta + \varepsilon^*$$

Hay que revisar cómo es el término error en este modelo transformado con respecto al valor esperado y la matriz de varianzas y covarianzas:

$$\begin{aligned} E(\varepsilon^*) &= E(T^{-1}\varepsilon) \\ &= T^{-1}E(\varepsilon) \end{aligned}$$

aplicando los supuestos mínimos cuadrados ordinarios

$$= T^{-1}(0) = 0$$

Para la matriz de *Var - Cov* se desarrolla:

$$Var - Cov(\varepsilon^*) = E\left[(\varepsilon^* - E(\varepsilon^*))(\varepsilon^* - E(\varepsilon^*))'\right]$$

dado que la esperanza de $E(\varepsilon^*) = 0$

$$\begin{aligned} &= E\left[(\varepsilon^* - 0)(\varepsilon^* - 0)'\right] \\ &= E\left[\varepsilon^* (\varepsilon^*)'\right] \\ &= E\left[T^{-1}\varepsilon (T^{-1}\varepsilon)'\right] \\ &= E\left[T^{-1}\varepsilon \varepsilon' T^{-1}\right] \\ &= T^{-1}E\left[\varepsilon \varepsilon'\right]T^{-1} \\ &= T^{-1}Var(\varepsilon)T^{-1} \\ &= \sigma^2 T^{-1}VT^{-1} \end{aligned}$$

Ahora bien dado que V es simétrica se obtiene $V = T T$

$$\begin{aligned} &= \sigma^2 \underbrace{T^{-1} T T T^{-1}}_I \\ &= \sigma^2 I \end{aligned}$$

Los errores del modelo transformado satisfacen los supuestos del modelo lineal clásico, entonces se puede decir que bajo dichos supuestos, el estimador mínimos cuadrados ordinarios es el más eficiente, por lo que, el estimador aplicado al modelo transformado también lo es.

3.2.3. Valor estimado

Descrita en los capítulos anteriores cual es la finalidad del método mínimos cuadrados y además considerando en el modelo transformado

$$Y^* = X^* \beta + \varepsilon^*$$

los errores ε^* , satisfacen los supuestos acostumbrados, por lo que se pueden aplicar los mínimos cuadrados ordinarios.

Mediante la función $S(\beta)$ y considerando que $V = T T' = T' T$, se obtiene, $V^{-1} = (T^{-1})' T^{-1}$ que sirve para desarrollar lo siguiente:

$$\begin{aligned} S(\beta) &= \varepsilon^{*'} \varepsilon^* \\ &= (T^{-1} \varepsilon)' T^{-1} \varepsilon \\ &= \varepsilon' (T^{-1})' T^{-1} \varepsilon \\ &= \varepsilon' V^{-1} \varepsilon \end{aligned}$$

por lo dicho anteriormente y recordando que $\varepsilon = Y - \hat{Y}$

$$\begin{aligned} &= (Y - \hat{Y})' V^{-1} (Y - \hat{Y}) \\ &= (Y - X \hat{\beta})' V^{-1} (Y - X \hat{\beta}) \\ &= (Y' - \hat{\beta}' X') (V^{-1} Y - V^{-1} X \hat{\beta}) \\ &= Y' V^{-1} Y - \hat{\beta}' X' V^{-1} Y - Y' V^{-1} X \hat{\beta} + \hat{\beta}' X' V^{-1} X \hat{\beta} \\ &= Y' V^{-1} Y - 2 \hat{\beta}' X' V^{-1} Y + \hat{\beta}' X' V^{-1} X \hat{\beta} \end{aligned}$$

El estimador del modelo mínimos cuadrados generalizados se obtiene mediante cálculo diferencial:

$$\begin{aligned}
 & Y'V^{-1}Y - 2\hat{\beta}'X'V^{-1}Y + \hat{\beta}'X'V^{-1}X\hat{\beta} \\
 \frac{\partial G}{\partial \hat{\beta}} &= -2X'V^{-1}Y + 2X'V^{-1}X\hat{\beta} \\
 &= 2(-X'V^{-1}Y + X'V^{-1}X\hat{\beta}) \\
 &\quad -X'V^{-1}Y + X'V^{-1}X\hat{\beta} = 0 \\
 X'V^{-1}X\hat{\beta} &= X'V^{-1}Y
 \end{aligned}$$

al despejar $\hat{\beta}$ se llega a la ecuación

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

donde $\hat{\beta}$ es el estimador de mínimos cuadrados generalizados ($\hat{\beta}_{MCG}$) de β .

El estimador mínimos cuadrados generalizados [16], puede tener varianzas grandes, cosa que no sucede con el estimador de mínimos cuadrados ordinarios

3.2.4. Características del estimador MCG

Se debe de revisar si $\hat{\beta}_{MCG}$ cumple con las características de un buen estimador, que son:

- 1) Insesgado
- 2) Consistente
- 3) Suficiente

Entonces se inicia con el incesgamiento, es decir, ver que se cumpla, $E[\hat{\beta}_{MCG}] = \beta$ esto es que el valor esperado sea igual al valor del parámetro que se esta estimando.

$$\begin{aligned}
 E[\hat{\beta}_{MCG}] &= E[(X'V^{-1}X)^{-1}X'V^{-1}Y] \\
 &= (X'V^{-1}X)^{-1}X'V^{-1}E(Y) \\
 &= (X'V^{-1}X)^{-1}X'V^{-1}E(X\beta + \varepsilon) \\
 &= (X'V^{-1}X)^{-1}X'V^{-1}(E(X\beta) + E(\varepsilon)) \\
 &= (X'V^{-1}X)^{-1}X'V^{-1}E(X\beta) \\
 &= (X'V^{-1}X)^{-1}X'V^{-1}X\beta \\
 &= \beta
 \end{aligned}$$

Si se cumple que:

$$E[\hat{\beta}_{MCG}] = \beta$$

por lo tanto, $\hat{\beta}_{MCG}$ es un estimador insesgado de β y se tiene la primera característica del estimador.

La varianza de este estimador se obtiene como:

$$\begin{aligned} Var - Cov[\hat{\beta}_{MCG}] &= E\left[\left(\hat{\beta}_{MCG} - E(\hat{\beta}_{MCG})\right)\left(\hat{\beta}_{MCG} - E(\hat{\beta}_{MCG})\right)'\right] \\ &= E\left[\left(\hat{\beta}_{MCG} - \beta\right)\left(\hat{\beta}_{MCG} - \beta\right)'\right] \\ &= E\left(\hat{\beta}_{MCG}\hat{\beta}_{MCG}' - \hat{\beta}_{MCG}\beta' - \beta\hat{\beta}_{MCG}' + \beta\beta'\right) \\ &= E\left(\hat{\beta}_{MCG}\hat{\beta}_{MCG}' - 2\left(\hat{\beta}_{MCG}\beta'\right) + \beta\beta'\right) \\ &= E\left[\left(X'V^{-1}X\right)^{-1}X'V^{-1}Y\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}Y\right)' - 2\left(X'V^{-1}X\right)^{-1}X'V^{-1}Y\beta' + \beta\beta'\right] \\ &= E\left[\left(X'V^{-1}X\right)^{-1}X'V^{-1}Y\left(X'V^{-1}Y\right)'\left(\left(X'V^{-1}X\right)^{-1}\right)'\right] \\ &\quad - E\left[2\left(X'V^{-1}X\right)^{-1}X'V^{-1}Y\beta'\right] + E[\beta\beta'] \\ &= E\left[\left(X'V^{-1}X\right)^{-1}X'V^{-1}YY'V^{-1}X\left(X'V^{-1}X\right)^{-1}\right] \\ &\quad - 2E\left[\left(X'V^{-1}X\right)^{-1}X'V^{-1}Y\beta'\right] + E[\beta\beta'] \\ &= \left(X'V^{-1}X\right)^{-1}X'V^{-1}E[YY']V^{-1}X\left(X'V^{-1}X\right)^{-1} \\ &\quad - 2\left(X'V^{-1}X\right)^{-1}X'V^{-1}E[Y]\beta' + \beta\beta' \\ &= \left(X'V^{-1}X\right)^{-1}X'V^{-1}\left(X\beta\beta'X' + \sigma^2V\right)V^{-1}X\left(X'V^{-1}X\right)^{-1} \\ &\quad - 2\left(X'V^{-1}X\right)^{-1}X'V^{-1}X\beta\beta' + \beta\beta' \end{aligned}$$

$$\begin{aligned}
&= \left((X'V^{-1}X)^{-1} X'V^{-1}X\beta\beta'X' + (X'V^{-1}X)^{-1} X'V^{-1}\sigma^2V \right) V^{-1}X(X'V^{-1}X)^{-1} \\
&\quad - 2(X'V^{-1}X)^{-1} X'V^{-1}X\beta\beta' + \beta\beta' \\
&= \underbrace{(X'V^{-1}X)^{-1} X'V^{-1}X\beta\beta'X'V^{-1}X(X'V^{-1}X)^{-1}}_I \\
&\quad + \sigma^2 \underbrace{(X'V^{-1}X)^{-1} X'V^{-1}V^{-1}X(X'V^{-1}X)^{-1}}_I - 2 \underbrace{(X'V^{-1}X)^{-1} X'V^{-1}X\beta\beta' + \beta\beta'}_I \\
&= \beta\beta' + \sigma^2 \underbrace{(X'V^{-1}X)^{-1} X'V^{-1}X(X'V^{-1}X)^{-1}}_I - 2\beta\beta' + \beta\beta' \\
&= \beta\beta' - 2\beta\beta' + \beta\beta' + \sigma^2 (X'V^{-1}X)^{-1} \\
&= 2\beta\beta' - 2\beta\beta' + \sigma^2 (X'V^{-1}X)^{-1} \\
&= \sigma^2 (X'V^{-1}X)^{-1}
\end{aligned}$$

por lo tanto

$$\boxed{\text{Var} - \text{Cov}(\hat{\beta}_{MCG}) = \sigma^2 (X'V^{-1}X)^{-1}}$$

Teorema de Gauss-Markov

El teorema de Gauss-Markov [12], establece que el estimador de mínimos cuadrados generalizados de β , que es $\hat{\beta}_{MCG} = (X'V^{-1}X)^{-1} X'V^{-1}Y$ es el mejor estimador lineal insesgado. Recordando que por "mejor" se quiere indicar que $\hat{\beta}_{MCG}$ minimiza la varianza para cualquier combinación lineal de los coeficientes estimados, la varianza de $\hat{\beta}_{MCG}$ es mínima dentro de la clase de estimadores insesgados y lineales en Y .

Para ver este comportamiento se propone el siguiente estimador $\tilde{\beta}$ otro estimador insesgado de β , $\tilde{\beta} \neq \hat{\beta}_{MCG}$ que satisfaga las condiciones de ser una combinación lineal de los datos, ser insesgado y de varianza mínima, si lo siguiente pasa:

$$\begin{aligned}
\text{Var} - \text{Cov}(\ell' \hat{\beta}_{MCG}) &= \ell' \text{Var} - \text{Cov}(\hat{\beta}_{MCG}) \ell \\
&= \ell' \sigma^2 (X'V^{-1}X)^{-1} \ell
\end{aligned}$$

Lo que se desea es ver que $Var - Cov(\ell' \tilde{\beta}) \geq \ell' \sigma^2 (X'V^{-1}X)^{-1} \ell$ con al menos una ℓ tal que $Var - Cov(\ell' \tilde{\beta}) > \ell' \sigma^2 (X'V^{-1}X)^{-1} \ell$.

Se puede escribir cualquier otro estimador de β que sea combinación lineal de los datos en la siguiente forma:

$$\tilde{\beta} = \left[(X'V^{-1}X)^{-1} X'V^{-1} + B \right] Y + b_0$$

sea B una matriz de $p \times n$, y b_0 es un vector de constantes de $p \times 1$, que ajusta en forma adecuada al estimador MCG, para formar el estimado alternativo.

Si el modelo es correcto debe ser insesgado, entonces

$$\begin{aligned} E(\tilde{\beta}) &= E \left[\left((X'V^{-1}X)^{-1} X'V^{-1} + B \right) Y + b_0 \right] \\ &= \left((X'V^{-1}X)^{-1} X'V^{-1} + B \right) E[Y] + b_0 \\ &= \left((X'V^{-1}X)^{-1} X'V^{-1} + B \right) X\beta + b_0 \\ &= \underbrace{(X'V^{-1}X)^{-1} X'V^{-1} X}_{I} \beta + BX\beta + b_0 \\ &= \beta + BX\beta + b_0 \end{aligned}$$

Para que $\tilde{\beta}$ sea insesgado se necesita que $BX = 0$ y $b_0 = 0$.

Ahora para el caso de la varianza de $\tilde{\beta}$

$$\begin{aligned} Var - Cov[\tilde{\beta}] &= E \left[(\tilde{\beta} - E(\tilde{\beta})) (\tilde{\beta} - E(\tilde{\beta}))' \right] \\ &= E \left[(\tilde{\beta} - \beta) (\tilde{\beta} - \beta)' \right] \\ &= E \left(\tilde{\beta} \tilde{\beta}' - \tilde{\beta} \beta' - \beta \tilde{\beta}' + \beta \beta' \right) \\ &= E \left(\tilde{\beta} \tilde{\beta}' - 2(\tilde{\beta} \beta') + \beta \beta' \right) \\ &= E \left[\left((X'V^{-1}X)^{-1} X'V^{-1} + B \right) Y \left((X'V^{-1}X)^{-1} X'V^{-1} + B \right) Y' \right. \\ &\quad \left. - 2 \left((X'V^{-1}X)^{-1} X'V^{-1} + B \right) Y \beta' + \beta \beta' \right] \end{aligned}$$

$$\begin{aligned}
&= E\left[\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}+B\right)Y\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}+B\right)Y\right] \\
&\quad -2E\left[\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}+B\right)Y\beta'\right]+E[\beta\beta'] \\
&=\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}+B\right)E[YY']\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}+B\right) \\
&\quad -2\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}+B\right)E[Y]\beta'+\beta\beta' \\
&=\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}+B\right)\left(X\beta\beta'X'+\sigma^2V\right)\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}+B\right) \\
&\quad -2\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}+B\right)X\beta\beta'+\beta\beta' \\
&=\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}\left(X\beta\beta'X'+\sigma^2V\right)+B\left(X\beta\beta'X'+\sigma^2V\right)\right) \\
&\quad \left(\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}\right)'+B'\right)-2\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}X\beta\beta'+BX\beta\beta'\right)+\beta\beta' \\
&=\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}X\beta\beta'X'+\sigma^2\left(X'V^{-1}X\right)^{-1}X'V^{-1}V+BX\beta\beta'X'+\sigma^2BV\right) \\
&\quad \left(\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}\right)'+B'\right)-2\left(X'V^{-1}X\right)^{-1}X'V^{-1}X\beta\beta'-2BX\beta\beta'+\beta\beta' \\
&=\left(\underbrace{\left(X'V^{-1}X\right)^{-1}X'V^{-1}X\beta\beta'X'+\sigma^2\left(X'V^{-1}X\right)^{-1}X'V^{-1}V}_{\gamma}+BX\beta\beta'X'+\sigma^2BV\right) \\
&\quad \left(\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}\right)'+B'\right)-2\underbrace{\left(X'V^{-1}X\right)^{-1}X'V^{-1}X\beta\beta'}_{\gamma}-2BX\beta\beta'+\beta\beta' \\
&=\left(\beta\beta'X'+\sigma^2\left(X'V^{-1}X\right)^{-1}X'+BX\beta\beta'X'+\sigma^2BV\right)\left(\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}\right)'+B'\right) \\
&\quad -2\beta\beta'-2BX\beta\beta'+\beta\beta' \\
&=\beta\beta'X'\left(\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}\right)'+B'\right)+\sigma^2\left(X'V^{-1}X\right)^{-1}X'\left(\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}\right)'+B'\right) \\
&\quad +BX\beta\beta'X'\left(\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}\right)'+B'\right)+\sigma^2BV\left(\left(\left(X'V^{-1}X\right)^{-1}X'V^{-1}\right)'+B'\right) \\
&\quad -2\beta\beta'-2BX\beta\beta'+\beta\beta'
\end{aligned}$$

$$\begin{aligned}
&= \beta\beta' \underbrace{X'V^{-1}X(X'V^{-1}X)^{-1}}_I + \beta\beta' X'B' + \sigma^2 (X'V^{-1}X)^{-1} X'V^{-1}X(X'V^{-1}X)^{-1} \\
&+ \sigma^2 (X'V^{-1}X)^{-1} X'B' + BX\beta\beta' X'V^{-1}X(X'V^{-1}X)^{-1} + BX\beta\beta' X'B' \\
&+ \sigma^2 BVV^{-1}X(X'V^{-1}X)^{-1} + \sigma^2 BVB' - 2\beta\beta' - 2BX\beta\beta' + \beta\beta' \\
&= \beta\beta' + \beta\beta' X'B' + \sigma^2 \underbrace{(X'V^{-1}X)^{-1} X'V^{-1}X(X'V^{-1}X)^{-1}}_I \\
&+ \sigma^2 (X'V^{-1}X)^{-1} X'B' + BX\beta\beta' X'V^{-1}X(X'V^{-1}X)^{-1} + BX\beta\beta' X'B' \\
&+ \sigma^2 \underbrace{BV^{-1}V}_I X(X'V^{-1}X)^{-1} + \sigma^2 BVB' - 2\beta\beta' - 2BX\beta\beta' + \beta\beta' \\
&= \beta\beta' + \beta\beta' X'B' + \sigma^2 (X'V^{-1}X)^{-1} + \sigma^2 (X'V^{-1}X)^{-1} X'B' + BX\beta\beta' + BX\beta\beta' X'B' \\
&+ \sigma^2 BX(X'V^{-1}X)^{-1} + \sigma^2 BVB' - 2\beta\beta' - 2BX\beta\beta' + \beta\beta' \\
&= \sigma^2 (X'V^{-1}X)^{-1} + 2\sigma^2 BX(X'V^{-1}X)^{-1} + BX\beta\beta' X'B' + \sigma^2 BVB'
\end{aligned}$$

como $BX = 0$, lo cual a su vez implica que $(BX)' = X'B' = 0$, se llega a

$$\begin{aligned}
&= \sigma^2 (X'V^{-1}X)^{-1} + \sigma^2 BVB' \\
&= \sigma^2 \left((X'V^{-1}X)^{-1} + BVB' \right)
\end{aligned}$$

por lo tanto

$$\boxed{Var - Cov(\tilde{\beta}) = \sigma^2 \left((X'V^{-1}X)^{-1} + BVB' \right)}$$

por otro lado

$$\begin{aligned}
Var - Cov(\ell\hat{\beta}) &= \ell' Var - Cov(\tilde{\beta}) \ell \\
&= \ell' \sigma^2 \left((X'V^{-1}X)^{-1} + BVB' \right) \ell \\
&= \ell' \sigma^2 (X'V^{-1}X)^{-1} \ell + \ell' BVB' \ell \sigma^2 \\
&= \ell' Var - Cov(\hat{\beta}) \ell + \ell' BVB' \ell \sigma^2 \\
&= Var - Cov(\ell' \hat{\beta}) + \ell' BVB' \ell \sigma^2
\end{aligned}$$

El resultado es que $BVB' = BT'TB'$ es, cuando menos una matriz semidefinida positiva, por lo anterior $\ell'BVB'\ell \sigma^2 \geq 0$, continuando, si se define $\ell^* = TB'\ell$, se tiene que

$$\ell'BVB'\ell = \ell^* \ell^* = \sum \ell^{*2}$$

debe ser estrictamente mayor que cero para una $\ell \neq 0$ a menos que $B = 0$, por todo lo anterior el estimador de mínimos cuadrados generalizados es el mejor estimador lineal insesgado.

3.2.4.1. Estimadores máximo verosímiles

Si, además, se asume que bajo el supuesto del modelo $\varepsilon \sim N(0, \sigma^2 V)$, lo que implica que $Y \sim N(X\beta, \sigma^2 V)$, se debe calcular el estimador máximo verosímil [12], y ver que es el mejor estimador insesgado.

Tomando la función de densidad de probabilidad de Y

$$f(\beta, \sigma^2, Y) = \frac{e^{-\frac{1}{2\sigma^2}(Y-X\beta)'V^{-1}(Y-X\beta)}}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}} |V|^{\frac{1}{2}}}$$

Al realizar la estimación máximo verosímil

$$\ln f(\beta, \sigma^2, Y) = -\frac{1}{2\sigma^2}(Y-X\beta)'V^{-1}(Y-X\beta) - \frac{n}{2} \ln(2\pi) - n \ln \sigma - \ln |V|^{\frac{1}{2}}$$

$$\ln f(\beta, \sigma^2, Y) = -\frac{1}{2\sigma^2}(Y' - \beta' X')(V^{-1}Y - V^{-1}X\beta) - \frac{n}{2} \ln(2\pi) - n \ln \sigma - \ln |V|^{\frac{1}{2}}$$

$$\ln f(\beta, \sigma^2, Y) = -\frac{1}{2\sigma^2}(Y'V^{-1}Y - \beta'X'V^{-1}Y - Y'V^{-1}X\beta + \beta'X'V^{-1}X\beta)$$

$$-\frac{n}{2} \ln(2\pi) - n \ln \sigma - \ln |V|^{\frac{1}{2}}$$

$$\ln f(\beta, \sigma^2, Y) = -\frac{1}{2\sigma^2}(Y'V^{-1}Y - 2\beta'X'V^{-1}Y + \beta'X'V^{-1}X\beta)$$

$$\ln f(\beta, \sigma^2, Y) = -\frac{1}{2\sigma^2}(Y'V^{-1}Y - 2\beta'X'V^{-1}Y + \beta'X'V^{-1}X\beta) - \frac{n}{2} \ln(2\pi) - n \ln \sigma - \ln |V|^{\frac{1}{2}}$$

$$\frac{\partial \ln f(\beta, \sigma^2, Y)}{\partial \beta} = -\frac{1}{2\sigma^2}(-2X'V^{-1}Y + 2X'V^{-1}X\beta)$$

$$= -\frac{1}{\sigma^2}(-X'V^{-1}Y + X'V^{-1}X\beta)$$

$$= (X'V^{-1}Y + X'V^{-1}X\beta)$$

$$X'V^{-1}X\beta = X'V^{-1}Y$$

$$\hat{\beta}_{MCG} = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

Estimador máximo verosímil

Ahora para calcular el estimador máximo verosimilitud de la varianza.

$$\ln f(\beta, \sigma^2, Y) = \frac{(Y'V^{-1}Y - 2\beta'X'V^{-1}Y + \beta'X'V^{-1}X\beta)}{\sigma^3} - \frac{n}{\sigma}$$

$$\frac{n}{\sigma} = \frac{(Y'V^{-1}Y - 2\beta'X'V^{-1}Y + X'V^{-1}X\beta^2)}{\sigma^3}$$

$$\frac{n\sigma^3}{\sigma} = (Y'V^{-1}Y - 2\beta'X'V^{-1}Y + X'V^{-1}X\beta^2)$$

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})'V^{-1}(Y - X\hat{\beta})}{n}$$

Estimador máximo verosímil

Dado que todo estimador máximo verosímil es suficiente entonces $\hat{\beta}_{MCG}$ y $\hat{\sigma}^2$ son suficientes.

Enseguida se analiza la propiedad de inesegamiento de $\hat{\sigma}_{MCG}$.

Inesegamiento de $\hat{\sigma}_{MCG}$:

$$\begin{aligned} E[\hat{\sigma}_{MCG}^2] &= E\left[\frac{(Y - X\hat{\beta})'V^{-1}(Y - X\hat{\beta})}{n}\right] \\ &= \frac{1}{n}E[(Y - X\hat{\beta})'V^{-1}(Y - X\hat{\beta})] \\ &= \frac{1}{n}E[(Y'V^{-1} - \hat{\beta}'X'V^{-1})(Y - X\hat{\beta})] \\ &= \frac{1}{n}E[Y'V^{-1}Y - \hat{\beta}'X'V^{-1}Y - Y'V^{-1}X\hat{\beta} + \hat{\beta}'X'V^{-1}X\hat{\beta}] \\ &= \frac{1}{n}E[Y'V^{-1}Y - 2\hat{\beta}'X'V^{-1}Y + \hat{\beta}'X'V^{-1}X\hat{\beta}] \\ &= \frac{1}{n}(E[Y'V^{-1}Y - 2\hat{\beta}'X'V^{-1}Y + \hat{\beta}'X'V^{-1}X\hat{\beta}]) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} \left(E \left[Y'V^{-1}Y - 2\hat{\beta}'X'V^{-1}Y + \hat{\beta}'X'V^{-1}X\hat{\beta} \right] \right) \\
&= \frac{1}{n} \left(E \left[Y'V^{-1}Y - 2\hat{\beta}'X'V^{-1}Y + \underbrace{\hat{\beta}'X'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y}_{H} \right] \right) \\
&= \frac{1}{n} \left(E \left[Y'V^{-1}Y - 2\hat{\beta}'X'V^{-1}Y + \hat{\beta}'X'V^{-1}Y \right] \right) \\
&= \frac{1}{n} \left(E \left[Y'V^{-1}Y - \hat{\beta}'X'V^{-1}Y \right] \right) \\
&= \frac{1}{n} \left(E \left[Y'V^{-1}Y - Y'V^{-1}X \underbrace{(X'V^{-1}X)^{-1}X'V^{-1}Y}_H \right] \right) \\
&= \frac{1}{n} E(Y'V^{-1}Y - HY'V^{-1}Y) \\
&= \frac{1}{n} E(Y'V^{-1}Y(I-H)) \\
&= \frac{1}{n} E(Y^2(I-H)) \\
&= \frac{1}{n} \text{tr}(\sigma^2(I-H)) \\
&= \frac{1}{n} \sigma^2(\text{tr}(I-H)) \\
&= \frac{1}{n} \sigma^2(\text{tr}(I) - \text{tr}(H)) \\
&= \frac{1}{n} \sigma^2(n-p)
\end{aligned}$$

por lo tanto

$$E[\hat{\sigma}^2] = \frac{1}{n} \sigma^2(n-p)$$

como se exhibe $E[\hat{\sigma}^2] \neq \sigma^2$, entonces, $\hat{\sigma}^2$ no es un estimador insesgado de σ^2 .

Se puede decir que $\frac{(Y - X\hat{\beta})'V^{-1}(Y - X\hat{\beta})}{n-p}$ si es un estimador insesgado de σ^2 .

3.2.5. Intervalo de confianza para β_i

Se tiene que al igual que en la sección 1.2.5. se construirá bajo el mismo método el intervalo de confianza para β_i , recordando que hay que considerar la distribución de el estimador del parámetro, del cual ya se obtuvieron su media y varianza por lo tanto se escribe que:

$$E[\hat{\beta}_{iMCG}] = \beta_i \quad \text{Var} - \text{Cov}(\hat{\beta}_{iMCG}) = \sigma^2 V_{ii}$$

si además $\varepsilon_i \sim N(0, \sigma^2 V_{ii})$ entonces $\hat{\beta}_{iMCG} \sim N(\beta_i, \sigma^2 V_{ii})$ de ahí que al estandarizar la

variable $\hat{\beta}_{iMCG}$ queda lo siguiente, $\frac{\hat{\beta}_{iMCG} - \beta_i}{\sqrt{\sigma^2 V_{ii}}} \sim N(0,1)$, retomando que σ es desconocida se

tomará una variable Ji-cuadrada y con ellas dos poder construir una cantidad que se distribuya t , esto queda como :

$$\frac{\hat{\beta}_{iMCG} - \beta_i}{\sqrt{\sigma^2 V_{ii}}} \sim t_{(n-p)}$$

$$\frac{\hat{\beta}_{iMCG} - \beta_i}{\sqrt{\frac{n\hat{\sigma}^2}{n-p}}} \sim t_{(n-p)}$$

y se realizan las operaciones correspondientes

$$\frac{1}{\sigma} \frac{\hat{\beta}_{iMCG} - \beta_i}{\sqrt{V_{ii}}} \sim t_{(n-p)}$$

$$\frac{1}{\sigma} \sqrt{\frac{n\hat{\sigma}^2}{n-p}} \sim t_{(n-p)}$$

$$\frac{\hat{\beta}_{iMCG} - \beta_i}{\sqrt{V_{ii}} \sqrt{\frac{n\hat{\sigma}^2}{n-p}}} \sim t_{(n-p)}$$

$$\frac{\hat{\beta}_{iMCG} - \beta_i}{\sqrt{V_{ii}} \hat{\sigma}} \sim t_{(n-p)}$$

para llegar a la siguiente distribución t

$$\frac{\hat{\beta}_{iMCG} - \beta_i}{\hat{\sigma} \sqrt{V_{ii}}} \sim t_{(n-p)}$$

por lo tanto se tiene
$$P\left(-a < \frac{\hat{\beta}_{MCG} - \beta_i}{\hat{\sigma}\sqrt{V_{ii}}} < a\right) = 1 - \alpha$$

obteniendo el intervalo de confianza al $(1 - \alpha) \times 100\%$ para β es:

$$\left(\hat{\beta}_{MCG} - a\hat{\sigma}\sqrt{V_{ii}} < \beta_i < \hat{\beta}_{MCG} + a\hat{\sigma}\sqrt{V_{ii}}\right)$$

recordando que a es el cuantil $1 - \frac{\alpha}{2}$ y $a \sim t_{\frac{1-\alpha}{2}}^{(n-p)}$.

3.2.5.1. Intervalo de confianza para combinaciones lineales de las β

Se pueden tener distintas combinaciones de las β_i a las cuales se les puede obtener un intervalo de confianza.

$$\begin{aligned} \ell' \beta &= \ell_0 \beta_0 + \ell_1 \beta_1 + \dots + \ell_{p-1} \beta_{p-1} \\ \ell' \hat{\beta}_{MCG} &= \ell_0 \hat{\beta}_0 + \ell_1 \hat{\beta}_1 + \dots + \ell_{p-1} \hat{\beta}_{p-1} \end{aligned}$$

se considera:

$$\begin{aligned} E(\ell' \hat{\beta}_{MCG}) &= \ell' \beta \\ \text{Var}(\ell' \hat{\beta}_{MCG}) &= \ell' \text{Var}(\hat{\beta}_{MCG}) \ell \\ \text{Var}(\ell' \hat{\beta}_{MCG}) &= \sigma^2 \ell' (X' V^{-1} X)^{-1} \ell \\ \ell' \hat{\beta}_{MCG} &\sim N\left(\ell' \beta, \sigma^2 \ell' (X' V^{-1} X)^{-1} \ell\right) \end{aligned}$$

Se procede a la estandarización del parámetro correspondiente obteniéndose:

$$\frac{\ell' \hat{\beta}_{MCG} - \ell' \beta}{\sqrt{\sigma^2 \ell' (X' V^{-1} X)^{-1} \ell}} \sim N(0,1)$$

de igual forma se pretende obtener una distribución t , esto es $\frac{\ell' \hat{\beta}_{MCG} - \ell' \beta}{\sqrt{\frac{\sigma^2 \ell' (X' V^{-1} X)^{-1} \ell}{\frac{(Y - X \hat{\beta})' V^{-1} (Y - X \hat{\beta})}{\sigma^2 (n-p)}}}} \sim t_{(n-p)}$

$$\frac{\frac{\hat{\beta}_{MCG} - \beta_i}{\sqrt{\sigma^2 \ell' (X' V^{-1} X)^{-1} \ell}}}{\sqrt{\frac{Y' V^{-1} Y - Y' V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1} Y}{\sigma^2 (n-p)}}} \sim t_{(n-p)}, \quad \frac{\hat{\beta}_{MCG} - \beta_i}{\sqrt{\ell' (X' V^{-1} X)^{-1} \ell}} \cdot \frac{1}{\sqrt{\hat{\sigma}^2}} \sim t_{(n-p)}$$

se obtiene finalmente $\frac{\ell' \hat{\beta}_{MCG} - \ell' \beta}{\sqrt{(X' V^{-1} X)^{-1} \ell' \hat{\sigma}^2}}$.

De ahí que se tenga

$$P \left(-a < \frac{\ell' \hat{\beta}_{MCG} - \ell' \beta}{\hat{\sigma} \sqrt{\ell' (X' V^{-1} X)^{-1} \ell}} < a \right) = 1 - \alpha$$

Intervalo para $\ell' \hat{\beta}_{MCG}$ al $(1 - \alpha) \times 100\%$

$$\left(\ell' \hat{\beta}_{MCG} - a \hat{\sigma} \sqrt{\ell' (X' V^{-1} X)^{-1} \ell}, \ell' \hat{\beta}_{MCG} + a \hat{\sigma} \sqrt{\ell' (X' V^{-1} X)^{-1} \ell} \right)$$

donde a es el cuantil $1 - \frac{\alpha}{2}$ y $a \sim t_{\frac{1-\alpha}{2}, (n-p)}$.

3.2.6. Identidad de los errores para mínimos cuadrados generalizados

Para mínimos cuadrados ordinarios se tiene la identidad siguiente:

$$\sum(Y - \bar{Y})^2 = \sum(Y - \hat{Y})^2 + \sum(\hat{Y} - \bar{Y})^2$$

esta es una identidad básica.

Donde:

$$\sum(Y - \bar{Y})^2 = \text{Suma de cuadrados del error del modelo reducido } SCE_{MR}$$

$$\sum(Y - \hat{Y})^2 = \text{Suma de cuadrados del error del modelo completo } SCE_{MC}$$

$$\sum(\hat{Y} - \bar{Y})^2 = \text{Suma de cuadrados del error bajo la Hipótesis nula } SCE_{H_0}$$

En términos de mínimos cuadrados generalizados se tiene

$$\begin{aligned} \sum(Y - \hat{Y})^2 &= \sum \varepsilon_i^2 \\ &= \varepsilon' \varepsilon \\ &= (Y - \hat{Y})' V^{-1} (Y - \hat{Y}) \\ &= (Y - X \hat{\beta}_{MCG})' V^{-1} (Y - X \hat{\beta}_{MCG}) \\ &= (Y' - \hat{\beta}'_{MCG} X') V^{-1} (Y - X \hat{\beta}_{MCG}) \\ &= (Y' V^{-1} - \hat{\beta}'_{MCG} X' V^{-1}) (Y - X \hat{\beta}_{MCG}) \\ &= Y' V^{-1} Y - \hat{\beta}'_{MCG} X' V^{-1} Y - Y' V^{-1} X \hat{\beta}_{MCG} + \hat{\beta}'_{MCG} X' V^{-1} X \hat{\beta}_{MCG} \\ &= Y' V^{-1} Y - \hat{\beta}'_{MCG} X' V^{-1} Y - \hat{\beta}'_{MCG} X' V^{-1} Y + \hat{\beta}'_{MCG} X' V^{-1} X \hat{\beta}_{MCG} \end{aligned}$$

$$\begin{aligned}
&= Y'V^{-1}Y - 2\hat{\beta}'_{MCG} X'V^{-1}Y + \hat{\beta}'_{MCG} X'V^{-1}X\hat{\beta}_{MCG} \\
&= Y'V^{-1}Y - 2\hat{\beta}'_{MCG} X'V^{-1}Y + \hat{\beta}'_{MCG} \underbrace{X'V^{-1}X(X'V^{-1}X)^{-1}}_I X'V^{-1}Y \\
&= Y'V^{-1}Y - 2\hat{\beta}'_{MCG} X'V^{-1}Y + \hat{\beta}'_{MCG} X'V^{-1}Y \\
&= Y'V^{-1}Y - \hat{\beta}'_{MCG} X'V^{-1}Y \\
&= Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^{-1} X'V^{-1}Y
\end{aligned}$$

por lo tanto $\sum(Y - \hat{Y})^2 = Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^{-1} X'V^{-1}Y$ es SCE_{MCG} para

el caso de mínimos cuadrados generalizados, de lo anterior se tiene la siguiente igualdad .

$$\sum(Y - \bar{Y})^2 = \sum(Y - \hat{Y})^2 + \sum(\hat{Y} - \bar{Y})^2$$

considerando que si siempre se cumple, entonces

$$\sum(Y - \hat{Y})^2 = \sum(Y - \bar{Y})^2 - \sum(\hat{Y} - \bar{Y})^2$$

de donde se obtendría el resto de los términos expresados así

$$\sum(\hat{Y} - \bar{Y})^2 = Y'V^{-1}X(X'V^{-1}X)^{-1} X'V^{-1}Y = \hat{\beta}'_{MCG} X'V^{-1}Y$$

$$\sum(Y - \bar{Y})^2 = Y'V^{-1}Y$$

los cuales se verán más adelante, estos son elementos fundamentales de la tabla de análisis de varianza.

3.2.7. Prueba de Hipótesis

Se desea realizar la siguiente prueba de hipótesis [8]:

$$H_0 : \beta = 0 \quad \text{vs.} \quad H : \beta \neq 0$$

Se tiene $\hat{\beta}_{MCG} \sim N\left(\beta, \sigma^2 (X'V^{-1}X)^{-1}\right)$

que al convertirla en una $N(0,1)$ se obtiene $\frac{\hat{\beta}_{MCG} - \beta}{\sqrt{\sigma^2 (X'V^{-1}X)^{-1}}} \sim N(0,1)$

al elevar al cuadrado todo $\frac{(\hat{\beta}_{MCG} - \beta)^2}{\sigma^2 (X'V^{-1}X)^{-1}} \sim \chi^2_{(n-p)}$

considerando $\frac{(Y - X\hat{\beta})'V^{-1}(Y - X\hat{\beta})}{\sigma^2} = \frac{Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y}{\sigma^2} \sim \chi^2_{(n-p)}$

$$\begin{aligned} \text{ahora} \quad & \frac{\frac{(\hat{\beta}_{MCG} - \beta)^2}{\sigma^2 (X'V^{-1}X)^{-1}}}{\frac{Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y}{\sigma^2 (n-p)}} = \frac{\frac{(\hat{\beta}_{MCG} - \beta)^2}{\sigma^2 (X'V^{-1}X)^{-1}}}{\frac{Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y}{(n-p)\sigma^2}} \\ & = \frac{\frac{(\hat{\beta}_{MCG} - \beta)^2}{(X'V^{-1}X)^{-1}}}{\frac{Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y}{(n-p)}} \sim F_{(p, n-p)} \end{aligned}$$

que al manejarla bajo H_0 resulta lo siguiente

$$\frac{\frac{\hat{\beta}_{MCG}\hat{\beta}'_{MCG}}{(X'V^{-1}X)^{-1}}}{\frac{Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y}{(n-p)}}$$

sustituyendo $\hat{\beta}'_{MCG}$ y $\hat{\beta}_{MCG}$ por sus identidades correspondientes se obtiene

$$\frac{\frac{\hat{\beta}'_{MCG} \hat{\beta}_{MCG}}{(X'V^{-1}X)^{-1}}}{\frac{Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y}{(n-p)}} = \frac{\frac{(X'V^{-1}X)^{-1}X'V^{-1}Y((X'V^{-1}X)^{-1}X'V^{-1}Y)'}{(X'V^{-1}X)^{-1}}}{\frac{Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y}{(n-p)}}$$

$$\frac{(X'V^{-1}X)(X'V^{-1}X)^{-1}X'V^{-1}Y((X'V^{-1}X)^{-1}X'V^{-1}Y)'}{Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y} = \frac{Y'V^{-1}X(X'V^{-1}X)^{-1}(X'V^{-1}Y)'}{Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y}$$

$$\boxed{\frac{\hat{\beta}'_{MCG}(X'V^{-1}Y)'}{Y'V^{-1}Y - \hat{\beta}'_{MCG}(X'V^{-1}Y)'} \sim F_{(p, n-p)}}$$

Se maneja como regla de decisión la siguiente desigualdad que de cumplirse se debe rechazar H_0 :

$$F > F^{1-\alpha}_{(1, n-2)}$$

Concluyendo el desarrollo anterior en la siguiente tabla

Análisis de varianza para mínimos cuadrados generalizados [11].

Fuente	Grados de libertad	Suma de cuadrados	Cuadrado medio	F
Regresión bajo H_0	p	$Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y$	$\frac{Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y}{p}$	$\frac{\hat{\beta}'_{MCG}(X'V^{-1}Y)'}{Y'V^{-1}Y - \hat{\beta}'_{MCG}(X'V^{-1}Y)'} \frac{p}{(n-p)}$
Modelo Completo	n-p	$Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y$	$\frac{Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y}{n-p}$	
Modelo Reducido	n	$Y'V^{-1}Y$		

3.2.8. Coeficiente de determinación

El coeficiente de determinación múltiple o medida del grado de ajuste del modelo en estudio [6], r^2 , no aporta muchas ventajas en modelo de mínimos cuadrados generalizados, ya que r^2 calculado al utilizar un programa estadístico mide el grado de ajuste del modelo que utiliza los datos transformados. Por lo que se sugiere utilizar:

$$r^2 = \frac{Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y}{Y'V^{-1}Y}$$

como medida de bondad de ajuste del modelo sin transformar. Sin embargo, no hay garantía de que r^2 se ubique en el intervalo $[0,1]$. Por eso es que no es una medida tan útil para comparar modelos.

3.2.9. Coeficiente de correlación r

Se sabe ya que es un valor relacionado con el r^2 . Se dice que r es el grado de asociación entre Y y X , entonces r

$$r = \frac{(X - \bar{X})(Y'V^{-1}Y)}{\sqrt{(X - \bar{X})^2 (Y'V^{-1}Y)}}$$

Se examinará un modelo geométrico, gracias al cual se puede presentar de una manera directa la teoría general, para el tipo de modelo que aquí se trata, la intuición geométrica aporta, a menudo, una gran ayuda.

3.2.10. Devianza

Se tiene un procedimiento para muestras grandes el cual conduce a una medida de variación del grado de ajuste que es llamada devianza.

Dadas n observaciones se puede ajustar el modelo el cual tiene n parámetros, éste se forma de todas las variaciones en las y_i , para el componente sistemático sin ningún componente aleatorio.

Es conveniente expresar la ecuación en términos del valor medio Y , que el parámetro canónico θ .

Sea $l(\hat{Y}, \phi; Y)$ el logaritmo de máxima verosimilitud sobre β para un valor ajustado de valores de los parámetros de dispersión ϕ .

La discrepancia de un ajuste es proporcional a dos veces la diferencia entre el máximo alcanzado por el logaritmo de verosimilitud y por el modelo en estudio. Si se denota $\hat{\theta} = \theta(\hat{Y})$ y $\bar{\theta} = \theta(Y)$

el estimado de el parámetro canónico bajo el modelo, la discrepancia, asumiendo $a_i(\phi) = \frac{\phi}{w_i}$,

puede ser escrito como:

$$\frac{\sum 2w_i (y_i (\bar{\theta}_i - \hat{\theta}_i) - b(\bar{\theta}_i) + b(\hat{\theta}_i))}{\phi} = \frac{D(Y; \hat{Y})}{\phi}$$

$D(Y; \hat{Y})$ es llamada devianza para el modelo que se esté manejando y ésta es una función de los datos solamente.

$D^*(Y; \hat{Y}) = \frac{D(Y; \hat{Y})}{\phi}$ así que el escalar devianza se expresa como un múltiplo del parámetro de dispersión.

Las formas de devianza según su tipo de distribución son las siguientes:

Normal	$\sum (Y - \hat{Y})^2$
Poisson	$2 \sum \left(Y \log \left(\frac{Y}{\hat{Y}} \right) - (Y - \hat{Y}) \right)$
Binomial	$2 \sum \left(y \log \left(\frac{Y}{\hat{Y}} \right) + (m - Y) \log \left(\frac{(m - Y)}{(m - \hat{Y})} \right) \right)$
Gamma	$2 \sum \left(-\log \left(\frac{Y}{\hat{Y}} \right) + \frac{(Y - \hat{Y})}{\hat{Y}} \right)$

Para la distribución normal la devianza es sólo la suma de residuales al cuadrado.

Recordando que se tiene otra medida de discrepancia que es el coeficiente de correlación de Pearson y la devianza que tiene una distribución χ^2 , para la teoría Normal en modelos lineales (considerando que el modelo es el verdadero) [10].

La devianza tiene una ventaja general sobre el coeficiente de correlación de Pearson, ya que la primera es aditiva para el conjunto de modelos si el estimador máximo verosimilitud es usado.

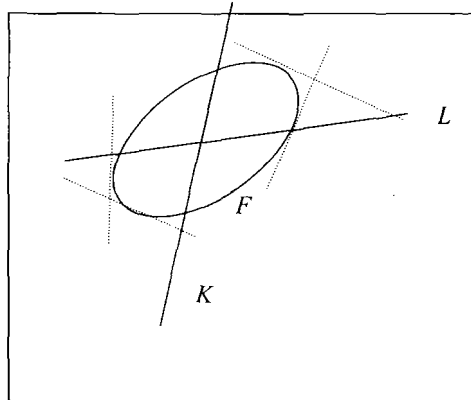
Para el caso de la función Normal se tendría:

$$r_{D_i} = \sqrt{\hat{Y}_i} \text{signo}(Y_i - \hat{Y}_i)$$

3.3. Geometría de mínimos cuadrados Generalizados

Se considera a L como una combinación lineal, donde la media de Y está contenida en S de $(Y - \hat{Y})$ (Siendo conocido por ser la matriz de covarianzas).

Se presenta una proyección del espacio n dimensional sobre la combinación lineal L , es decir por una transformación lineal que deja invariables a los vectores de L .



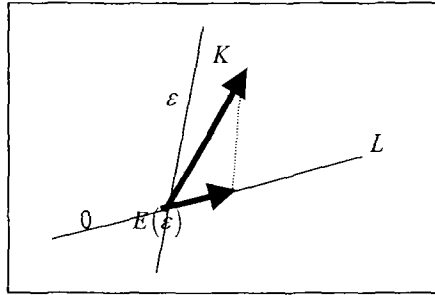
Estas proyecciones conducen, a estimaciones centradas. $E(Y)$ es igual a $E(\hat{Y})$, la cual por proyección es justamente igual a Y . Inversamente, para que un estimador lineal tenga una esperanza matemática igual a la esperanza Y de \hat{Y} , sea cual fuere \hat{Y} en L , es preciso que $E(\hat{Y})$ sea idéntico a \hat{Y} para todo \hat{Y} en L .

La ausencia de sesgo en la estimación de \hat{Y} implica una cierta justificación de los estimadores por proyección, el elipsoide indicador de $E(Y)$ da una imagen de la dispersión de los errores $E(Y) - \hat{Y}$.

Se observa que el elipsoide indicador de una proyección de Y sobre L , coincide con la imagen de F en esta proyección, este elipsoide contiene, necesariamente, la intersección de F y L .

Se reduce a esta intersección cuando la proyección de la imagen 0 (el origen) con la variedad lineal K , conjugada principal de L con relación a F .

Si se considerara que esto no pasa y que existe un vector ε de F cuya proyección sobre L paralelamente a K , no pertenece a la intersección de E y de L . Sea u un vector que hace máximo $u'E(\varepsilon)$ con la condición de $u'Vu \leq 1$. Dado que $E(\varepsilon)$ no pertenece a F , se tendrá $|u'E(\varepsilon)| \geq 1$.



Las derivadas de $u'E(\varepsilon) - ku'Vu$ respecto a las componentes de u , deben ser nulas.

De lo que resulta que $E(\varepsilon) = 2kVu$. Sea ahora un vector $\varepsilon^* = \varepsilon - E(\varepsilon)$. El vector ε pertenece a F formando parte del soporte $(Y - \hat{Y})$; lo mismo pasa con ε^* puesto que L toma parte de este soporte. Con lo considerado ε^* tiene una imagen reducida de vector nulo, está contenido en K , y por consiguiente, es conjugado de $E(\varepsilon)$. Sea entonces un vector u^* tal que $\varepsilon^* = Vu^*$. La conjugación entre ε^* y $E(\varepsilon)$ implica que $2ku'V = 0$, y además $u'\varepsilon^* = 0$.

Se tiene:

$$|u'\varepsilon^*| = |u'E(\varepsilon) + u'\varepsilon^*| = |u'E(\varepsilon)| > 1$$

lo que contradice la hipótesis según la cual ε pertenecería a F ya que $u'Vu \leq 1$.

Por tanto, en una proyección del espacio n dimensional sobre L , paralelamente a K , el elipsoide indicador de $E(Y)$ se reduce a la intersección de L con el elipsoide indicador de Y .

Si la matriz de covarianzas V tiene inversa, el estimador por proyección $Y^* = E(Y)$ que aplica sobre el origen la combinación lineal K , conjugada principal de L , es el que hace mínima la expresión $(Y - Y^*)'V^{-1}(Y - Y^*)$, siendo forzosamente el vector Y^* por parte de L .

Sea y_0 un vector cualquiera de L ; $y_0 - Y^*$ y $Y - Y^*$ son conjugadas. Al sustituir se tiene $(y_0 - Y^*)'V^{-1}(Y - Y^*) = 0$, de ahí que:

$$(Y - y_0)'V^{-1}(Y - y_0) = (Y - Y^*)'V(Y - Y^*) + (y_0 - Y^*)'V^{-1}(Y - Y^*)$$

La matriz V^{-1} se define positiva, como la matriz V , así que

$$(y_0 - Y^*)'V^{-1}(Y - Y^*) > 0 \quad \text{si } y_0 \neq Y^*$$

$$(Y - y_0)'V^{-1}(Y - y_0) > (Y - Y^*)'V(Y - Y^*)$$

Esto se deja ver cómo el estimador es único cuando V tiene una inversa. Si el soporte de $Y - Y^*$ no es el espacio m dimensional entero, sino una combinación lineal p dimensional, se deberá efectuar una transformación lineal del espacio de n dimensional en el espacio p dimensional, de tal manera que la imagen del soporte de $Y - Y^*$ sea el espacio p dimensional entero.

Cuando V es la matriz unidad, la expresión minimizada por Y^* se escribe $\sum (Y - Y^*)^2$, por lo tanto Y^* hace mínima la suma de los cuadrados de las desviaciones entre las componentes de Y y las componentes de Y^* de la media estimada. De ahí el nombre de "mínimos cuadrados generalizados" dada la estimación aquí considerada.

Lo anterior sugiere que un modelo lineal general debe ser mejor que el análisis estándar, usando transformaciones cuando un problema queda con varianza constante después de hacer la transformación.

Aplicación

4.1. Introducción

En este capítulo se da un ejemplo de regresión lineal simple, un modelo con sólo una variable independiente x y la variable dependiente y , con la finalidad de revisar la relación entre estas variables que por el tipo de regresión se espera que corresponda a un ajuste lineal. El modelo a estimar será:

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

Se da también un ejemplo en el caso del modelo de regresión donde interviene más de una variable independiente, es decir, el modelo de regresión múltiple

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$$

Los modelos de regresión lineal proporcionan gran información, para determinar los coeficientes de las variables, pero son insuficientes cuando se tiene un modelo no lineal.

Los ejemplos se desarrollaron con el programa de computo STATISTICA (versión 6.).

Se utilizó una muestra de tamaño 100, con las siguientes variables: Peso se toma como la variable dependiente y , estatura (x_1), edad (x_2), porcentaje de salud en 5 años (x_3), donde x_1, x_2, x_3 son variables independientes.

La finalidad de este último capítulo es dar un ejemplo de lo visto en la parte teórica de los capítulos anteriores de este trabajo.

4.2. Análisis Simple.

Se presenta la regresión con mínimos cuadrados lineales ordinarios, en el caso donde sólo se tiene una variable independiente, es decir $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ para este caso y_i representa el peso y x_i es la estatura.

Variable	Media y Dev. Est.		
	Media	Dev. Est.	N
Estatura	1.67890	0.035729	100
Peso	70.60000	9.316001	100

La tabla anterior tiene las medias correspondientes a la variable estatura e indica que en la mayoría de los datos se colocan alrededor de 1.67.

La tabla muestra la matriz de correlaciones.

Variable	Correlación	
	Estatura	Peso
Estatura	1.000000	0.422612
Peso	0.422612	1.000000

Gráficamente se puede observar que la relación entre las variables es positiva, la tendencia de cada una de las variables y su pendiente exhiben lo mencionado.

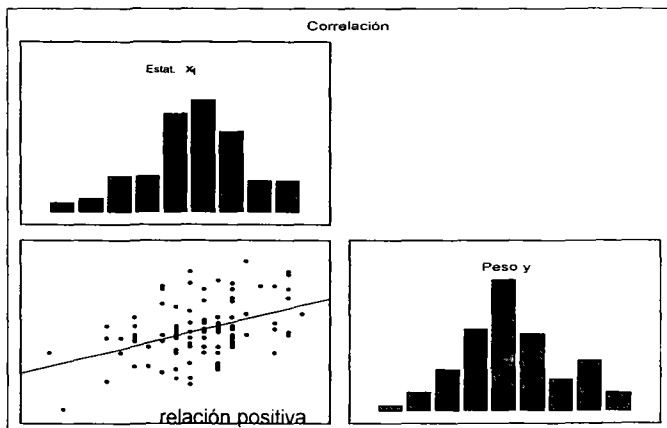


Tabla de resumen de la regresión

R= .42261198 R ² = .17860088 R ² Ajustada= .17021926 F(1,98)=21.309 p<.00001 Error est. estimado: 8.4862						
N=100	Beta	Err. Est. de Beta	B	Err. Est. de B	t(98)	p-level
Intersección			-114.402	40.08630	-2.85389	0.005270
Estatura	0.422612	0.091551	110.192	23.87117	4.61613	0.000012

De este cuadro se obtienen los estimadores de β_0 y β_1 , del modelo $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, de ahí:

$$\hat{y}_i = -114.402 + 110.192x_i$$

esto es que por cada unidad que cambie x_i se tendrá que y_i cambia o aumenta en 110.192 unidades.

Se obtuvo entonces:

$$R = .42261198$$

que significa que tan relacionadas están las variables de peso y estatura, las cuales se correlacionan positivamente, es decir que al tener una mayor estatura se tiene mayor peso, y al tener menos estatura se tiene un menor peso. Salvo sus excepciones que se presentan en la realidad.

Por otro lado se tiene:

$$R^2 = .17860088$$

que indica el nivel de explicación del modelo, significando que aproximadamente el 20% de la variación del peso está explicado por la estatura.

A continuación se da la tabla del análisis de varianza.

Fuente de Variación	Análisis de Varianza				
	Suma de Cuadrados	gl	Cuadrados Medios	F	p-level
Regresores	1534.539	1	1534.539	21.30863	0.000012
Residuales	7057.461	98	72.015		
Total	8592.000				

Esta tabla sirve para realizar la prueba de hipótesis

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

Se tiene que $F_{(1,98)}^* = 21.309$, $F_{(1,98)}^{.95} = 3.94$ y como la regla de decisión es rechazar H_0 si

$F_{(1,98)}^* > F_{(1,98)}^{.95}$, por lo tanto:

$F_{(1,98)}^* = 21.309 > F_{(1,98)}^{.95} = 3.94$ de ahí que se rechaza H_0 , entonces se puede decir que

$$\beta_1 \neq 0.$$

Ahora para revisar si se tiene problemas de autocorrelación se utiliza la prueba Durbin-Watson:

$$H_0 : \text{no hay autocorrelación} \quad \text{vs} \quad H_1 : \text{hay autocorrelación}$$

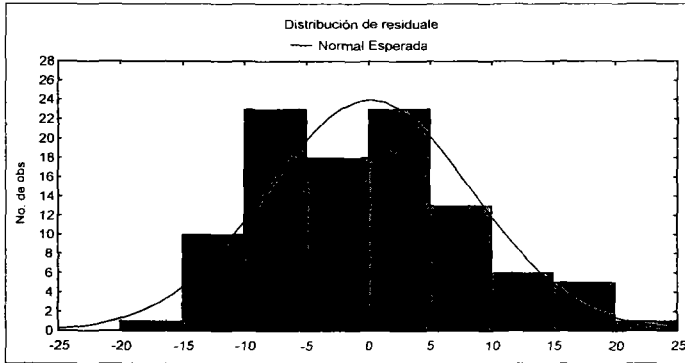
Regla de decisión, no rechace H_0 si $d > d_u$.

	Durbin-Watson d y correlación serial de re:	
	Durbin-Watson d	Corr. Serial
Estimado	1.692621	0.142572

Consultando la tabla correspondiente a Durbin-Watson se tiene para el limite inferior $d_u = 1.69$, al comparar $d = 1.692621 > 1.69 = d_u$ por lo tanto no se rechaza H_0 , de ahí que se afirma no tener autocorrelación.

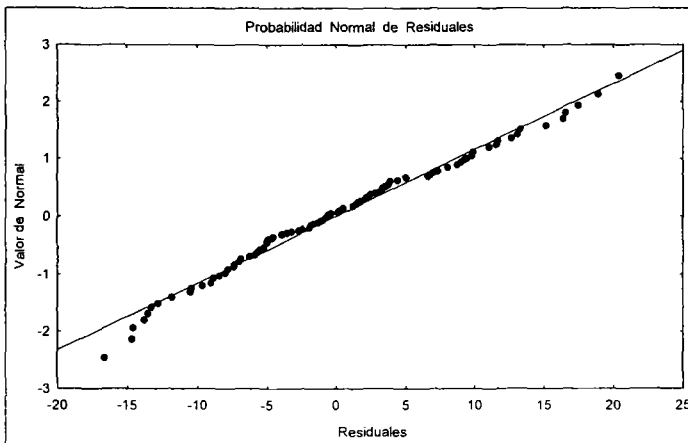
Análisis de normalidad.

En esta gráfica se analiza el comportamiento de los residuales, es decir se observa si cumplen tener una distribución normal.

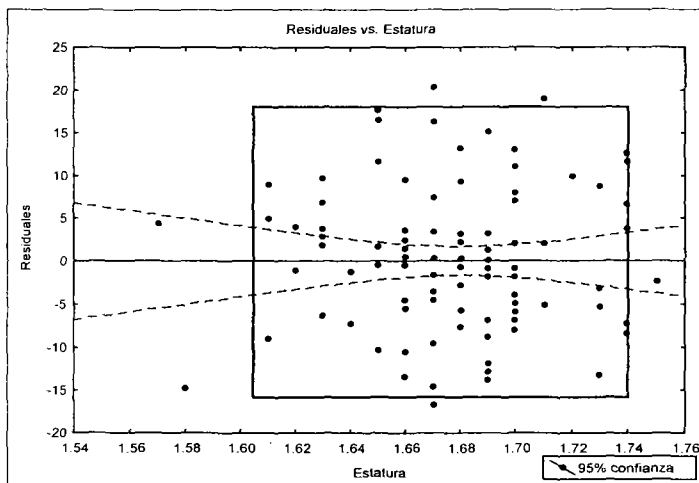


En la gráfica anterior se tiene normalidad, ya que casi se completa la campana de la distribución normal.

La siguiente gráfica es otra manera de analizar la normalidad de los residuales, si se ajustan los puntos a la línea recta, se dice que se tiene normalidad.

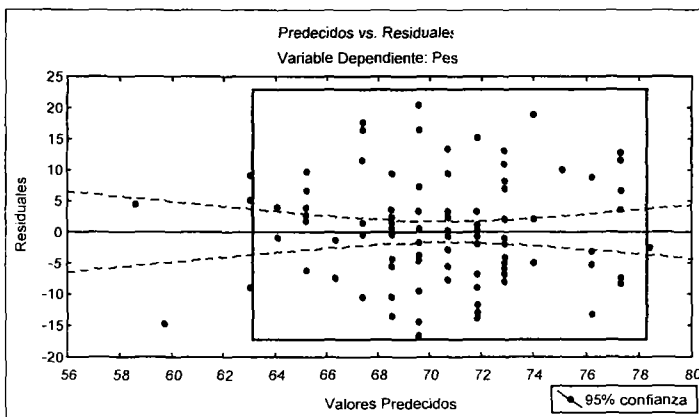


Análisis de la Varianza.

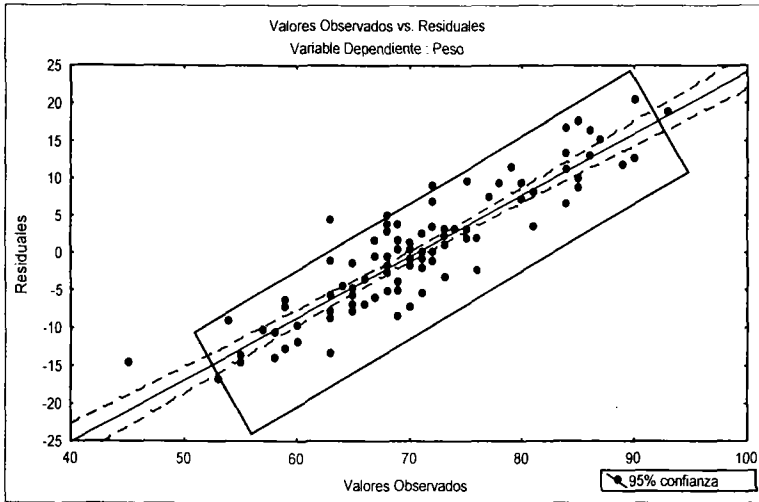


En la grafica anterior de residuales-estatura se tiene varianza constante ya que los puntos tienen una tendencia constante.

En la siguiente gráfica se tiene que los puntos al centro se distribuyen de manera constante, por lo tanto en general se puede decir que se presenta varianza constante.



En la siguiente gráfica la mayoría de los puntos se concentran dentro de un rectángulo, por lo que se puede decir que la varianza es constante.



Ahora bien, considerando un $\alpha = 0.05$ se obtiene el intervalo al 95% de confianza correspondiente a β_0 :

$$(-193.9518967, -34.8520026)$$

donde se puede exhibir que β_0 no pasa por el origen.

De igual forma el intervalo del 95% de confianza para β_1 es:

$$(62.82079343, 157.563922)$$

Los datos presentan tres outliers los cuales son:

Casos	Residuales Estandarizados						Residuales estandarizados: Peso		
	-5.	-4.	-3.	±2.	3.	4.	5.	Valores Observados	Valores Predicidos
3	90.00000	69.61929
24	85.00000	67.41544
25	93.00000	74.02699
Mínimo	85.00000	67.41544
Máximo	93.00000	74.02699
Media	89.33334	70.35390
Mediana	90.00000	69.61929

Los tres outliers resultaron no influyentes en el modelo.

Para analizar el supuesto de independencia se revisa la siguiente tabla y se puede observar que no se presenta ninguna tendencia periódica por lo se puede decir que los residuales cumplen con ser independientes.

Caso	Valores Predecidos				Valores Predecidos		
	58.6	.	.	78.4	Valores Observados	Valores Predecidos	Residuales
1	63.0	71.82314	-8.8231
2	70.0	68.51736	1.4826
3	90.0	69.61929	20.3807
4	65.0	72.92506	-7.9251
5	84.0	67.41544	16.5846
6	75.0	72.92506	2.0749
7	69.0	65.21159	3.7884
8	85.0	75.12891	9.8711
9	73.0	76.23083	-3.2308
10	60.0	71.82314	-11.8231
11	81.0	77.33276	3.6672
12	63.0	68.51736	-5.5174
13	81.0	72.92506	8.0749
14	69.0	74.02699	-5.0270
15	72.0	71.82314	0.1769
16	79.0	67.41544	11.5846
17	72.0	63.00775	8.9923
18	68.0	64.10967	3.8903
19	60.0	69.61929	-9.6193
20	63.0	76.23083	-13.2308
21	85.0	75.12891	9.8711
22	70.0	77.33276	-7.3328
23	84.0	70.72121	13.2788
24	85.0	67.41544	17.5846
25	93.0	74.02699	18.9730
26	90.0	77.33276	12.6672
27	86.0	69.61929	16.3807
28	68.0	68.51736	-0.5174
29	73.0	70.72121	2.2788
30	76.0	74.02699	1.9730
31	71.0	76.23083	-5.2308
32	60.0	71.82314	-11.8231
33	53.0	69.61929	-16.6193
34	63.0	68.51736	-5.5174
35	74.0	70.72121	3.2788
36	69.0	72.92506	-3.9251
37	55.0	69.61929	-14.6193
38	67.0	67.41544	-0.4154
39	65.0	70.72121	-5.7212
40	71.0	72.92506	-1.9251

4.3. Análisis Múltiple

Se consideran los siguientes datos Y (peso), x_1 (estatura), x_2 (edad) y x_3 (el porcentaje de salud en 5 años).

Se estimará el modelo

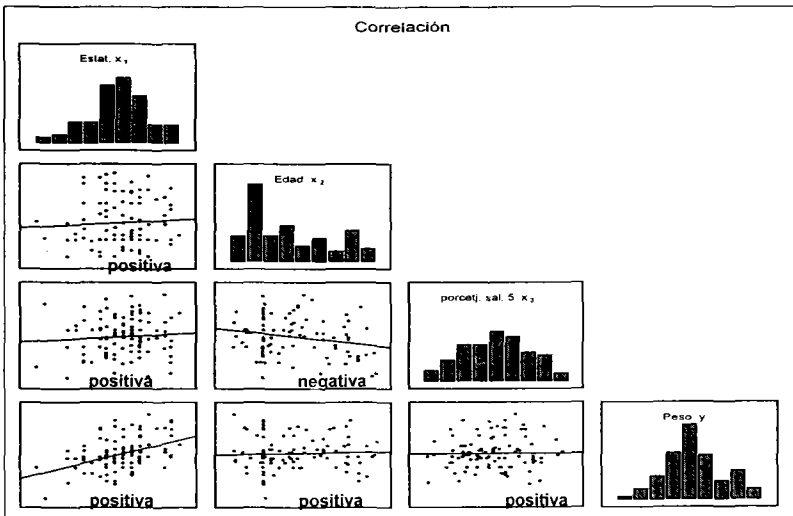
$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Ahora se tiene un cuadro de las correlaciones entre las variables.

Variable	Correlación			
	Estatura	Edad	% de salud en 5 años	Peso
Estatura	1.000000	0.0617388708	0.077852	0.422612
Edad	0.061739	1	-0.219874	0.043450
% de salud en 5 años	0.077852	-0.219873871	1.000000	0.007160
Peso	0.422612	0.043449882	0.007160	1.000000

Las variables estatura y peso se relacionan positivamente con el resto de las variables; la variable edad se relaciona positivamente con estatura y peso, pero se relaciona negativamente con el porcentaje de salud en 5 años y esta variable a su vez se relaciona positivamente con estatura y peso.

Gráfica donde se muestra la correlación entre variables



El coeficiente de determinación $R^2 = 0.17940838$, es decir, se explica un 17.94% de la variabilidad.

La regresión da:

Resumen de la Regresión						
R= .42356626 R ² = .17940838 R ² Ajustada= .15376489						
F(3,96)=6.9963 p<.00026 Error Est. de estim.: 8.5699						
N=100	Beta	Error Est de Beta	B	Error Est. de B	t(96)	p-level
Intersección			-114.070	40.50194	-2.81640	0.005895
Estatura	0.423660	0.093042	110.466	24.25995	4.55341	0.000015
Edad	0.012206	0.095087	0.007	0.05545	0.12837	0.898128
% de salud en 5 años	-0.023139	0.095194	-0.018	0.07479	-0.24307	0.808469

De ahí se tiene que la recta de regresión estimada es:

$$\hat{Y} = -114.070 + 110.466x_1 + 0.007x_2 + 0.07479x_3$$

De esta tabla se analiza también $r = .42356626$ que al igual que en el análisis simple es un coeficiente de correlación positivo, entonces las variables tienen una relación directamente proporcional

Análisis de Varianza

Effect	Suma de	gl	Cuadrados	F	p-level
	Cuadrados		Medios		
Regresores.	1541.477	3	513.8256	6.996255	0.000263
Residuales	7050.523	96	73.4429		
Total	8592.000				

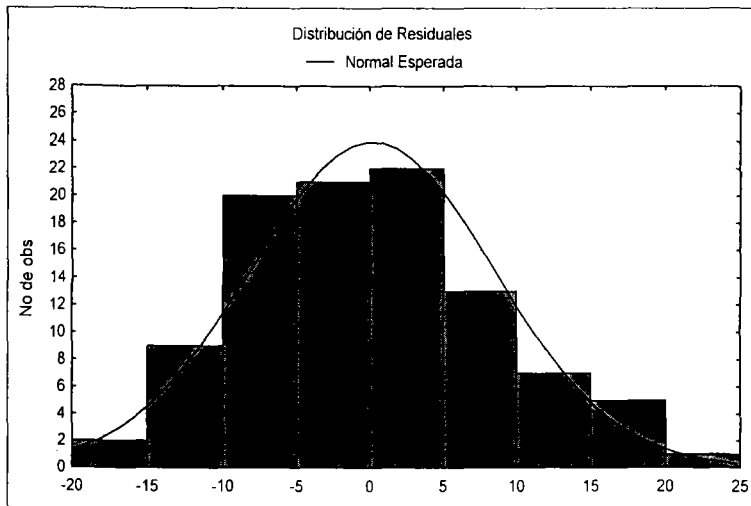
Se tiene la siguiente prueba de hipótesis para los coeficientes de la regresión pero no para el coeficiente β_0 , entonces se plantea.

$$\underline{\beta}^* = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \text{ de ahí que se tenga: } H_0: \underline{\beta}^* = 0 \text{ vs } H_1: \underline{\beta}^* \neq 0$$

donde se tiene que $F^* = 6.996 > 2.70 = F_{(3,96)}^{.95}$

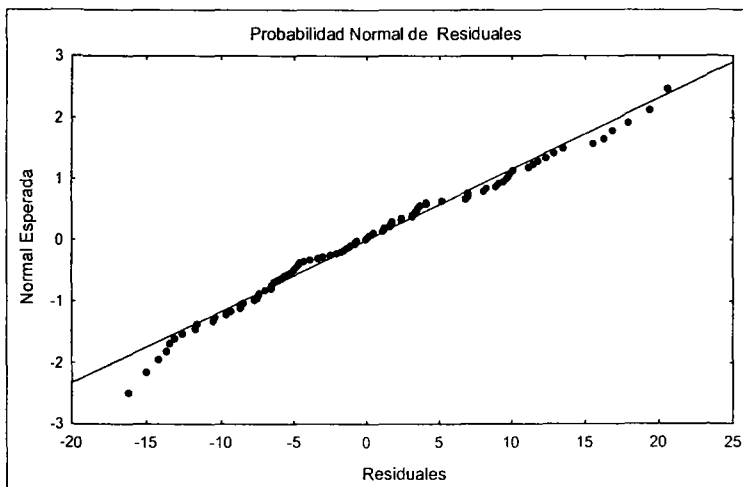
Se rechaza H_0 , por lo tanto $\beta_i \neq 0$ con $i = 1, 2, 3$.

Análisis de normalidad

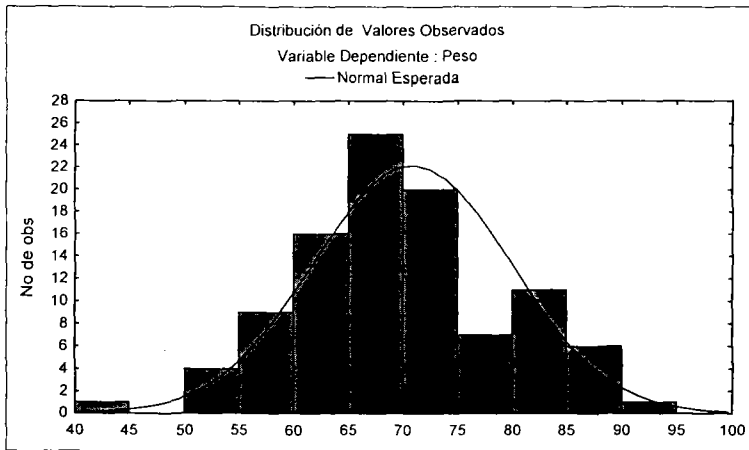


Como se ve en la gráfica, se tiene normalidad ya que casi se cubre el área de la campana.

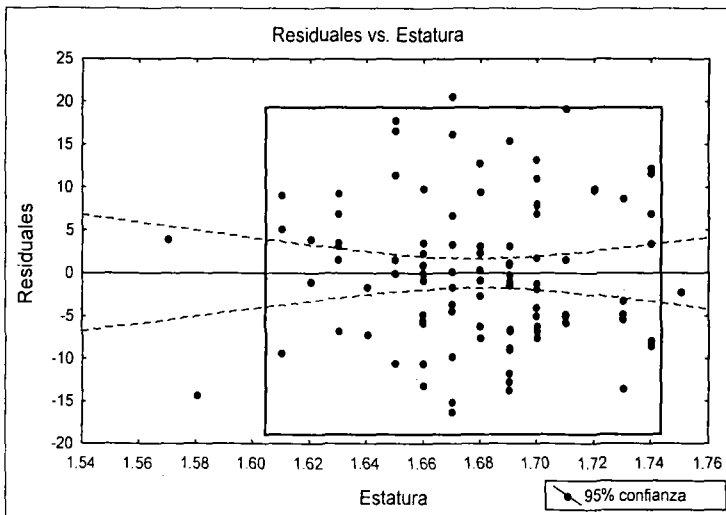
Por otro lado la siguiente gráfica no cambia mucho con respecto a la gráfica anterior, aclarando que no se tiene problemas de normalidad.



En el caso de los valores observados del peso se cumple la normalidad.

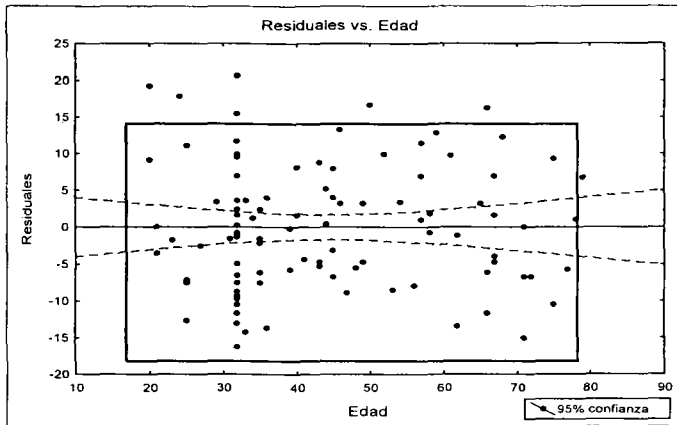


Análisis de la varianza.

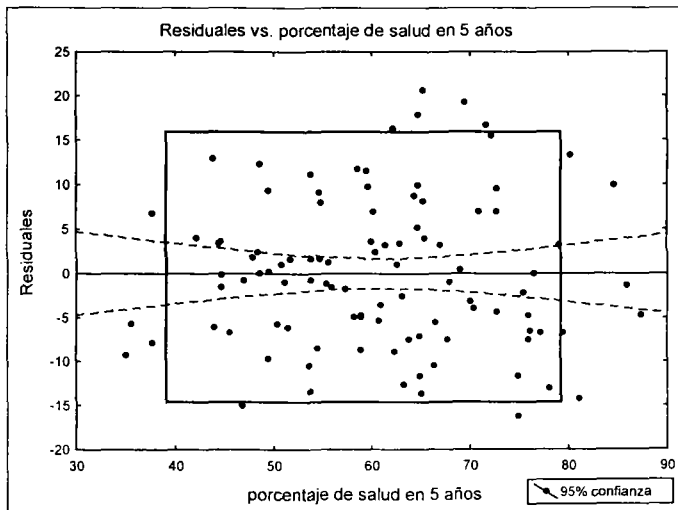


Se siguen teniendo varianza constante, ya que se pueden concentrar los puntos dentro de un rectángulo.

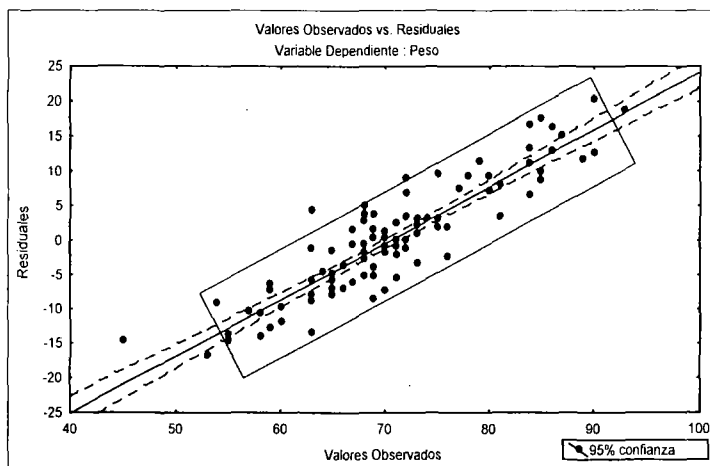
En la siguiente gráfica, se tiene varianza constante ya que la mayoría de los puntos se pueden encerrar en un rectángulo.



En la gráfica de porcentaje de salud se observa igual que la variable anterior esta es constante en su varianza con respecto a los residuales.



Aquí se observa que se puede considerar la presencia de varianza constante ya que a excepción de tres puntos la mayoría de estos pueden concentrarse dentro del rectángulo.



Prueba Durbin-Watson

H_0 : no hay autocorrelación vs. H_1 : hay autocorrelación

Durbin-Watson d y correlación serial		
	Durbin-Watson d	Corr. Serial
Estimado	1.698796	0.139444

El estadístico Durbin-Watson es de 1.69 y revisando los valores de la tabla se tiene.

$$d_L = 1.61 < \underset{\text{Durbin Watson}}{1.69} < 1.74 = d_U$$

por lo tanto no se puede concluir, es decir no se sabe que pasa con la autocorrelación bajo la prueba.

En la siguiente gráfica se puede observar que no existe ninguna tendencia en su comportamiento por lo se puede decir que se tiene independencia en los residuales.

Caso	58.9	Valores Predecidos		78.1	Valores Predecidos		
		Valores Observados	Valores Predecidos		Residuales		
1	63.00000	71.77258	-8.7726
2	70.00000	68.93508	1.0649
3	90.00000	69.45036	20.5496
4	65.00000	72.59001	-7.5900
5	84.00000	67.25159	16.7484
6	75.00000	73.26394	1.7361
7	69.00000	65.41362	3.5864
8	85.00000	75.28082	9.7192
9	73.00000	76.08338	-3.0834
10	60.00000	71.72530	-11.7253
11	81.00000	77.53825	3.4617
12	63.00000	68.66606	-5.6661
13	81.00000	73.04372	7.9563
14	69.00000	73.75290	-4.7529
15	72.00000	72.08219	-0.0822
16	79.00000	67.52445	11.4756
17	72.00000	62.92838	9.0716
18	68.00000	63.95266	4.0473
19	60.00000	69.73476	-9.7348
20	63.00000	76.50027	-13.5003
21	85.00000	74.98499	10.0150
22	70.00000	77.85282	-7.8528
23	84.00000	71.13585	12.8642
24	85.00000	67.19296	17.8070
25	93.00000	73.70710	19.2929
26	90.00000	77.73994	12.2601
27	86.00000	69.74758	16.2524
28	68.00000	68.86099	-0.8610
29	73.00000	70.66274	2.3373
30	76.00000	74.36455	1.6355
31	71.00000	75.79615	-4.7961
32	60.00000	71.66743	-11.6674
33	53.00000	69.27425	-16.2742
34	63.00000	68.43726	-5.4373
35	74.00000	70.74373	3.2563
36	69.00000	72.92010	-3.9201
37	55.00000	70.06325	-15.0633
38	67.00000	66.95863	0.0414
39	65.00000	71.18496	-6.1850
40	71.00000	72.84373	-1.8437
41	69.00000	74.72934	-5.7293
42	66.00000	72.63900	-6.6390
43	64.00000	68.70783	-4.7078
44	59.00000	66.09303	-7.0930
45	65.00000	69.37810	-4.3781
46	73.00000	71.84779	1.1522

4.4. Análisis de regresión no lineal, ejemplo 1.

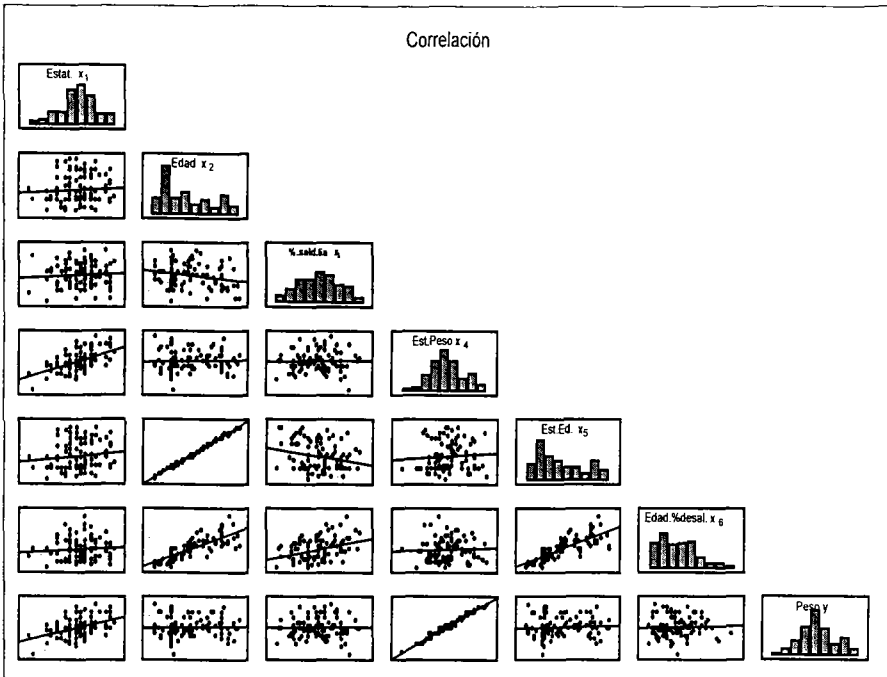
Se mencionó en la teoría que el método mínimos cuadrados lineales ya no es apropiado aquí, entonces se utiliza el método de linealización (Gauss-Newton).

Los datos para la realización de esta regresión son los mismos del caso anterior.

El modelo está dado por

$$Y_u = f(x_u, \theta) + \varepsilon_u$$

Gráfica de la correlación de las variables del modelo es:



Se puede observar una relación positiva en la mayoría de las variables, los puntos siguen la tendencia de su pendiente.

En este método las características se presentan de la siguiente forma:

Función Pérdida: Mínimos Cuadrados

La función anterior es suma de residuales ajustados y en el cuadro se muestran estos valores donde se va reduciendo los residuales en cada iteración que se realiza, así como los correspondientes valores estimados de los coeficientes de la regresión.

Proporción de conteo de varianza: 0.99942356, este valor es equivalente al coeficiente de determinación R².

El porcentaje de explicación del modelo es 99.94% que es muy alto a comparación del porcentaje obtenido en el ejercicio de mínimos cuadrados lineales.

El coeficiente de correlación es R = 0.99971174, expresa una relación positiva entre las variables.

Para este método se necesita un valor inicial para cada beta el cual se toma como 0.100000 para todas las betas, obteniéndose la tabla siguiente:

Modelo: $v_1 = \text{Intersección} + \beta_1 v_2 + \beta_2 v_3 + \beta_3 v_4 + \beta_4 v_5 + \beta_5 v_6$						
Var. Dep.: Peso						
Nivel de Confianza: 95.0% ($\alpha = 0.050$)						
	Estimado	Error Est.	t-valor gl = 93	p-level	Límite Inferior	Límite Superior
Intersección	66.1637	3.628509	18.2344	0.000000	58.9582	73.3692
beta1	-39.2558	2.202460	-17.8236	0.000000	-43.6294	-34.8821
beta2	0.0672	0.078833	0.8527	0.396004	-0.0893	0.2238
beta3	-0.0052	0.006198	-0.8336	0.406658	-0.0175	0.0071
beta4	0.5948	0.001647	361.1076	0.000000	0.5915	0.5980
beta5	-0.0431	0.046959	-0.9180	0.361001	-0.1364	0.0501
beta6	0.0001	0.000125	0.9201	0.359892	-0.0001	0.0004

Así el modelo estimado es:

$$\hat{Y} = 66.1637 - 39.2558x_1 + 0.0672x_2 - 0.0052x_3 + 0.5948x_4 - 0.0431x_5 + 0.0001x_6$$

La tabla que se presenta abajo muestra el total de iteraciones realizadas y los valores que toman los coeficientes en cada una de las iteraciones.

Modelo: $v_1 = \text{Intersección} + \beta_1 v_2 + \beta_2 v_3 + \beta_3 v_4 + \beta_4 v_5 + \beta_5 v_6 + \beta_6 v_7$ Var. Dep.: Peso								
	Función Pérdida	Intersección	beta1	beta2	beta3	beta4	beta5	beta6
1	2450.189	0.10000	0.1000	0.100000	0.100000	0.100000	0.100000	0.100000
2	2.225	66.14363	-39.2445	0.067740	-0.005150	0.594768	-0.043408	0.000115
3	2.225	66.16368	-39.2558	0.067223	-0.005166	0.594767	-0.043107	0.000115

Como se ve la primera iteración se maneja con los valores iniciales ya mencionados, el primer valor para la función pérdida (la suma de residuales ajustados) es muy grande, pero se van reduciendo en cada iteración, en este caso se realizaron tres iteraciones, pero se pueden presentar más, dependiendo del número necesario para alcanzar la linealización del modelo en estudio.

Análisis mediante la tabla Anova.

Modelo: $v_1 = \text{Intersección} + \beta_1 v_2 + \beta_2 v_3 + \beta_3 v_4 + \beta_4 v_5 + \beta_5 v_6 + \beta_6 v_7$ Var. Dep.: Peso					
Fuente de Variación	1	2	3	4	5
	Sum de Cuadrados	gl	Cuad. Med.	F-valor	p-valor
Regresión	507023.0	7.0000	72431.86	1360087	0.00
Residuales	5.0	93.0000	0.05		
Total	507028.0	100.0000			
Total Corregido	8592.0	99.0000			
Regresión vs. Total Corregido	507023.0	7.0000	72431.86	835	0.00

Al linealizar el modelo se puede proceder como en el caso de mínimos cuadrados lineales, por lo que se realiza la prueba F , sin considerar en ella a la intersección.

Prueba de hipótesis $H_0 : \beta_i^* = 0$ vs $H_1 : \beta_i^* \neq 0$

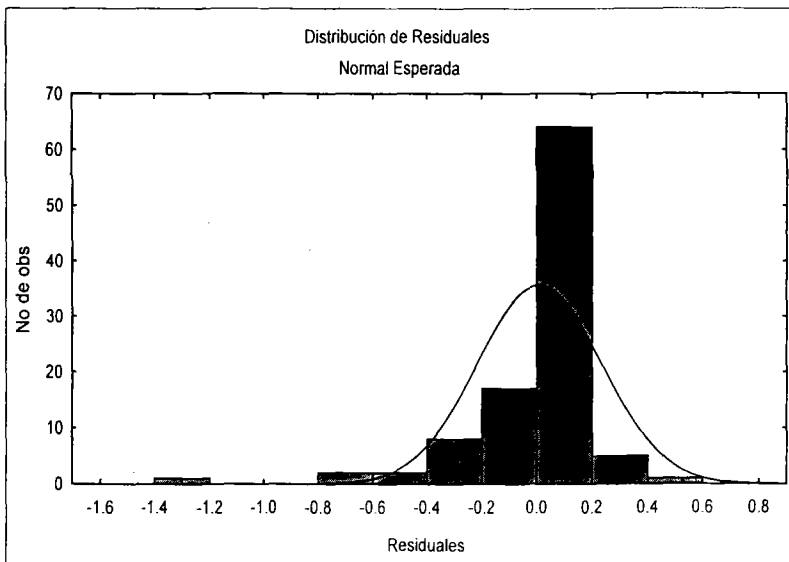
Bajo la regla de decisión, rechace H_0 si $F_{(7,93)}^* > F_{(7,93)}^{.95}$, con lo que al buscar los valores se tiene $F_{(7,93)}^* = 1360087 > 2.10 = F_{(7,93)}^{.95}$, entonces se rechaza H_0 y bajo este resultado, se puede decir que $\beta_i \neq 0$ $i = 1, 2, \dots, 7$, considerando una estimación puntual.

Al inicio de esta regresión se mostró un cuadro que contiene los intervalos de confianza de cada coeficiente, de donde se puede decir que hay cuatro coeficientes el de beta2, beta3, beta5 y beta6, pueden tomar cero como valor ya que pertenece al intervalo.

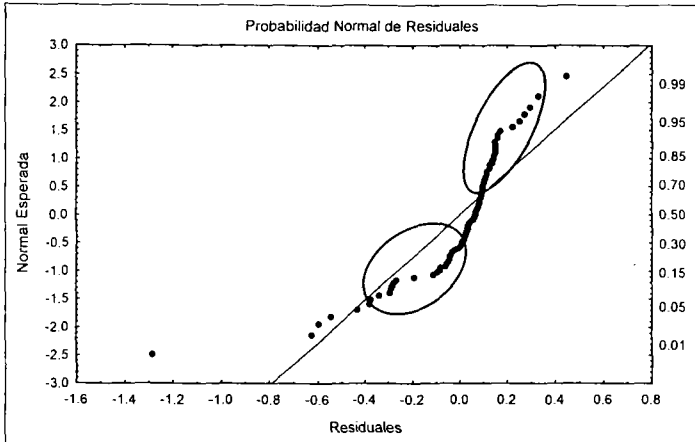
	Lím. Inf.	Lím. Sup.
beta2	-0.0893	0.2238
beta3	-0.0175	0.0071
beta5	-0.1364	0.0501
beta6	-0.0001	0.004

Análisis de normalidad

Es muy notable que las barras del histograma no se ajustan a la curva normal, además que sobrepasa una de ellas al área de normalidad estimada y se observan barras muy pequeñas que no alcanzan la campana.

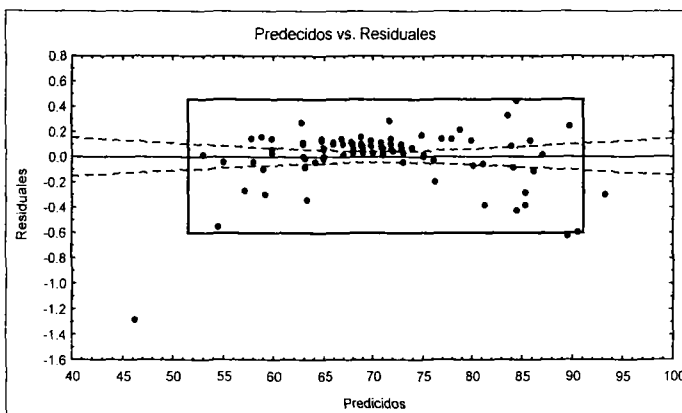


Se tienen problemas de normalidad en las regiones marcadas por una circunferencia, ya que los puntos no se ajustan a la recta, se tienen ondulamientos de los puntos que están por encima de la recta.



Análisis de la varianza.

En esta gráfica se tiene una varianza constante ya que la mayoría de los puntos se pueden encerrar en un rectángulo, quedando unos cuantos punto fuera de este rectángulo, los cuales se pueden considerar discrepantes ya que quedan muy alejados del resto de los puntos.



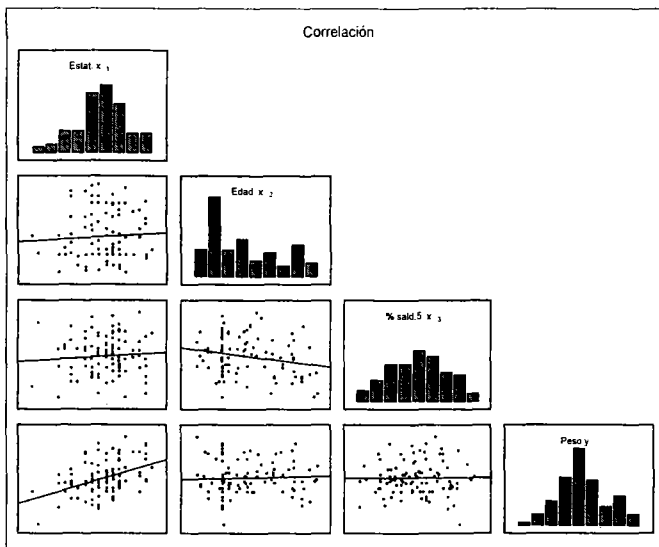
Se hace notar que al hacer la regresión mediante linealización, se presentó un problema en la normalidad, se puede intentar ver que modificaciones serían necesarias aplicar al modelo para poder mejorar esta situación siempre y cuando no se modifique el estado constante en que se encuentra la varianza.

La manera en que se manejó el modelo hace una diferencia, que se nota en R y R^2 que aumentaron considerablemente.

4.5. Análisis de regresión no lineal, ejemplo 2.

Los resultados al realizar la regresión son:

Gráfica de la correlación



La relación entre la mayoría de las variables es positiva salvo la relación entre Edad y el porcentaje de salud en 5 años la cual se puede decir que es negativa.

Resumen de la regresión

Proporción de conteo de varianza: 0.17983866 que es el coeficiente de determinación, que representa un 17.98 % de explicación del modelo.

El coeficiente de correlación $R = 0.42407389$ que indica una relación positiva entre la mayoría de sus variables .

El valor inicial es 0.100000 con el cual se inician las iteraciones para linealizar el modelo.

La tabla de resultados de la regresión es la siguiente:

Modelo : $v1=(intersección)*(exp(beta1*v2)-exp(beta2*v3)-exp(beta3*Var. Dep. : Peso$ Nivel de confianza: 95.0% (alpha=0.050)						
	Estimado	Error Est.	t-valor gl = 96	p-level	Limite Inferior	Limite Superior
intersección	4.988472	3.014	1.655043	0.101182	-0.99	10.971
beta1	1.578461	0.356	4.428371	0.000025	0.87	2.286
beta2	-0.501256	1170.181	-0.000428	0.999659	-2323.29	2322.290
beta3	-0.065630	0.240	-0.273267	0.785235	-0.54	0.411

Por lo tanto el modelo estimado es:

$$\hat{Y} = (4.988472)(\exp(1.578461x_1)) - \exp(-0.501256x_2) - \exp(-0.06563x_3)$$

En seguida se presenta la tabla de la Anova:

Modelo : $v1=(intersección)*(exp(beta1*v2)-exp(beta2*v3)-exp(beta3*v4)$ Var. Dep.: Peso					
Fuente de Variación	1 Suma de Cuadrados	2 gl	3 Cuadrados Med.	4 F-valor	5 p-level
Regresión	499981.2	4.0000	124995.3	1702.833	0.00
Residuales	7046.8	96.0000	73.4		
Total	507028.0	100.0000			
Total Corregido	8592.0	99.0000			
Regresión vs.Total Corregido	499981.2	4.0000	124995.3	1440.239	0.00

Ya obtenido el modelo se aplica la prueba **F** como en mínimos cuadrados lineales, recordando que la prueba deja fuera de ella a la intersección.

Prueba de hipótesis:

$$H_0 : \underline{\beta}^* = 0 \quad \text{vs.} \quad H_1 : \underline{\beta}^* \neq 0$$

Su regla de decisión es rechazar H_0 si $F^*_{(4,96)} > F_{(4,96)}^{.95}$

$F^*_{(4,96)} = 1702.833 > 2.46 = F_{(4,96)}^{.95}$, entonces se rechaza H_0 y bajo este resultado, se puede

decir que $\beta_i \neq 0 \quad i = 1, 2, \dots, 7$, considerando una estimación puntual.

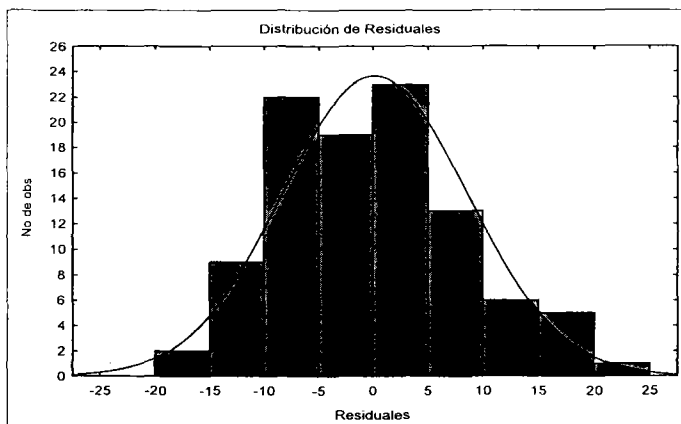
La siguiente tabla se conoce como historia de iteraciones ya que muestra el total de iteraciones realizadas y los valores que van tomando los coeficientes en cada una de ellas, en total se realizaron 87 aunque aquí solo se exhiben 28 por el tamaño de la tabla.

Modelo : $v1=(\text{intersección})*(\exp(\text{beta1}^*v2)-\exp(\text{beta2}^*v2))$					
Var. Dep.: Peso					
	Función Pérdida	intersección	beta1	beta2	beta3
1	2172.67	0.10000	0.10000	0.10000	0.10000
2	1796.85	0.09950	3.62241	0.09990	0.09993
3	1757.53	0.09866	3.64053	0.09970	0.09979
4	1683.77	0.09706	3.67494	0.09932	0.09953
5	1553.06	0.09415	3.73727	0.09864	0.09904
6	1343.05	0.08929	3.84104	0.09750	0.09821
7	1053.90	0.08221	3.99125	0.09584	0.09694
8	737.08	0.07363	4.16632	0.09378	0.09527
9	484.94	0.06423	4.31771	0.09164	0.09337
10	329.36	0.05281	4.42770	0.08957	0.09142
11	234.47	0.04051	4.53619	0.08758	0.08958
12	180.82	0.03086	4.65970	0.08573	0.08797
13	151.41	0.02488	4.77439	0.08393	0.08647
14	136.91	0.02044	4.87662	0.08200	0.08496
15	129.40	0.01728	4.96614	0.07980	0.08336
16	125.48	0.01535	5.03106	0.07715	0.08155
17	123.17	0.01490	5.04668	0.07375	0.07928
18	120.93	0.01673	4.97174	0.06915	0.07607
19	119.46	0.01901	4.89525	0.06647	0.07404
20	118.41	0.02395	4.74684	0.06187	0.07039
21	115.54	0.02936	4.62994	0.05801	0.06723
22	113.22	0.03488	4.53138	0.05466	0.06444
23	113.07	0.04546	4.36041	0.04900	0.05979
24	108.85	0.05049	4.31699	0.04664	0.05785
25	107.61	0.06002	4.20743	0.04252	0.05455
26	105.89	0.06863	4.13094	0.03905	0.05182
27	104.91	0.08422	4.00062	0.03338	0.04752
28	102.64	0.09745	3.91966	0.02892	0.04419

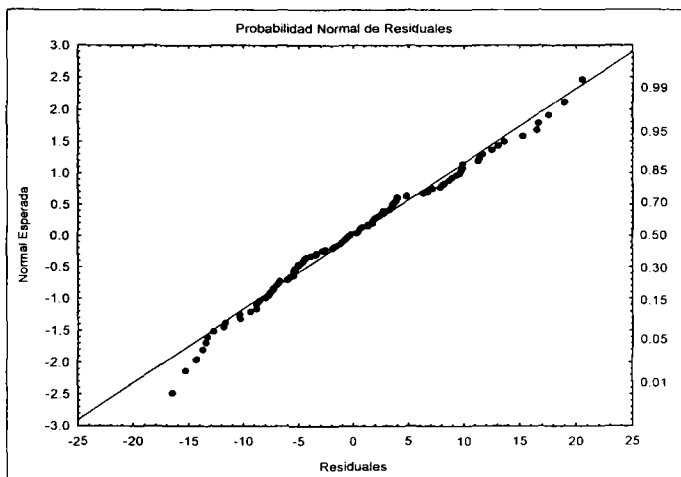
Como se puede observar cada iteración va disminuyendo, la cual llega a tener un valor final de 83.945, cumpliéndose así una de las principales características de este método de linealización.

Análisis de Normalidad

En la gráfica de abajo se cumple con la propiedad de la normalidad en los residuales.



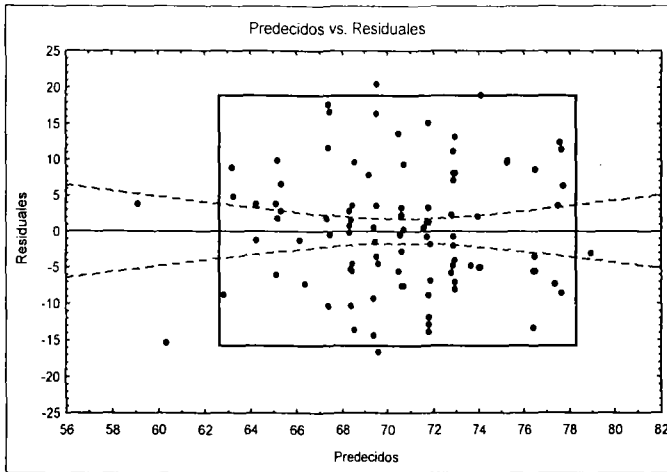
En la siguiente gráfica se puede observar como la mayoría de los puntos se ajustan a la recta con lo que se puede concluir que se tiene normalidad.



Análisis de la varianza

En la siguiente gráfica se presenta varianza constante ya que se pueden encerrar los puntos en un rectángulo.

En esta gráfica se tiene la presencia de varianza constante y se pueden observar ligeras alineaciones de los puntos y esto se debe a que hay repetición de datos o los valores entre los datos no difieren mucho.



Conclusiones

El trabajo expuesto tenía como objetivo explicitar el método de mínimos cuadrados en sus diversas acepciones, con el fin de tener en un compendio las particularidades del método.

Cabe decir que la variante más conocida es la de mínimos cuadrados ordinarios, debido a que es expuesta en el modelo de regresión lineal simple, pero las restantes variantes tienen su importancia y su aplicación tal vez en grado mayor que el ordinario.

De los métodos que se mencionaron en el capítulo dos, el modelo linealización por Gauss Newton es al que se le hizo mayor énfasis, ya que no es complicado en el momento de desarrollar las iteraciones. El problema que presenta es que puede converger muy lentamente a una solución, es decir no se garantiza la convergencia en todos los casos. Se tiene una debilidad fundamental que el de los valores iniciales ya que la buena elección de éstos puede facilitar el procedimiento de linealización de lo contrario los resultados no serán buenos.

La finalidad de mínimos cuadrados generalizados es estabilizar a la varianza, cuando se presenta este problema en los datos a analizar y cuando el modelo lineal no es aplicable simplifica a menudo el procedimiento de encontrar los valores estimados.

Se busca en general que el modelo cumpla con las propiedades del método que se maneje pretendiendo que sean las óptimas para hacer una buena interpretación al analizar los datos.

No se puede sugerir que uno de los tres métodos es el mejor para obtener un buen ajuste, debido a que cada problema presenta diferentes necesidades, y para el modelador será el que mejor resultados presente a ese problema.

Para hacer un buen análisis de datos y obtener un buen ajuste de estos, se debe al inicio del problema elegir las variables más aptas, las que se consideren lógicas en tener una relación con la variable que se considere como variable dependiente, o basarse en la experiencia y conocimiento del modelador.

El método a aplicar puede ser cualquiera de los vistos en este trabajo.

El modelo lineal simple/múltiple es de los más accesibles para poder trabajar ya que como es una función lineal facilita la obtención de los valores estimados de los coeficientes de la regresión. Presenta el problema de ser insuficiente para manejar formas no lineales.

En la parte final del trabajo se ha presentado una aplicación con el objetivo de ilustrar los métodos vistos.

Bibliografía

- [1] Bailey Trevor, Gatrell Tony, "*Interactive Spatial Data Analysis*", primera edición, Pretice Hall, 1996.
- [2] Bates Douglas M., Watts Donald G., "*Nonlinear Regression Analysis and Its Applications*", Wiley, Nueva York, 1988.
- [3] Borse Gerald J., "*Numerical Methods with MATLAB*", primera edición, Brooks Cole, 1996.
- [4] Canavos George C., "*Probabilidad y Estadística, aplicaciones y métodos*", primera edición, McGraw-Hill/ Interamericana de México, México, 1988.
- [5] Draper Norman, Smith Harry, "*Applied Regression Analysis*", segunda edición, Wiley, Nueva York, 1981.
- [6] Fernández V., "*Apuntes de Teoría Econométrica I.*", U.A.M., México, 2002.
- [7] Freund Rudolf Jakob, "*Regression Methods: A Tool for Data Analysis*", Marcel Dekker, 1979.
- [8] Gujarati Damodar, "*Econometria Básica*", primera edición, McGraw-Hill, México, 1981.
- [9] Kutner Michael H. , Neter John, Wasserman William, "*Applied Linear Regression Models*", segunda edición, Richard D Irwin, Homewood, 1989.
- [10] McCullagh P., Nelder J. A., "*Generalized Linear Models*", segunda edición, Chapman & Hall/CRC, Londres, 1989.
- [11] Montgomery, Peck, Vining, "*Introducción to liner regression analysis*", tercera edición, Wiley, Nueva York, 2002.
- [12] Mood Alexander McFarlane, Graybill Franklin A., Boes Duane C., "*Introduction to the Theory of Statistics*", tercera edición, McGraw-Hill, Nueva York, 1974.
- [13] Ratkowsky David A., "*Nonlinear Regression Modeling: A Unified Practical Approach*", segunda edición , Marcel Dekker Inc, 1983.
- [14] Ryan Thomas P., "*Modern Regression Methods*", segunda edición, 1978.
- [15] Seber George A. F., Wild C. J., "*Nonlinear Regression*", Wiley, Nueva York, 1977.
- [16] Weisberg S., "*Applied Linear Regression*", Segunda edición, Wiley , Nueva York, 1985.
- [17] Younger Mary Sue, "*A first course in linear regression*", segunda edición, Duxbury Press, North Scituate, Mass , 1985.