

## REGRESION NO LINEAL

GUILLERMO RIVAS M.

Estadístico, Universidad Nacional de Colombia

LUIS A. LÓPEZ P., ANTONIO VELASCO M.

Profesores Asociados, Universidad Nacional de Colombia

### 1. INTRODUCCIÓN

En el presente artículo se pretende dar una visión general de lo que es la regresión no lineal, para ésto, se tratan aspectos tales como: la no linealidad de un modelo, la forma de saber si un modelo es lineal o no, ¿cuándo usar la regresión no lineal?, también se discute la forma de estimar los parámetros en este tipo de modelos, así como sus propiedades asintóticas y la construcción asintótica de intervalos de confianza y pruebas de hipótesis. En la parte final se da una aplicación con datos reales, para el ajuste de un modelo no lineal haciendo uso del procedimiento NLIN del paquete SAS.

### 2. MODELOS NO LINEALES

**2.1. Introducción.** Un modelo se puede definir como una ecuación o conjunto de ecuaciones que describen el comportamiento de algún sistema, por ejemplo, el trabajo de un reactor químico o el crecimiento de un animal. La teoría estadística dedicada a los modelos lineales en los parámetros es muy amplia, debido en gran parte a sus múltiples aplicaciones y a la fácil interpretación de los resultados obtenidos de este tipo de análisis. Sin embargo, hay fenómenos observables que no pueden ser explicados por modelos lineales, por ejemplo, el desarrollo de una teoría en la química o la física, en tales situaciones un modelo no lineal en los parámetros se puede ajustar mejor.

Hoy en día es más viable, gracias al progresivo avance de los computadores, la aplicación de modelos no lineales en fenómenos donde el conjunto de parámetros no puede expresarse en forma lineal.

En matemática los fenómenos observables pueden ser como una ecuación de la forma:

$$(1) \quad Y = f(\theta, \zeta) + \epsilon$$

donde  $Y$  es el valor medido de una o más respuestas (las variables dependientes),  $\epsilon$  es el error experimental asociado con esta medida, y  $f(\theta, \zeta)$  es una función que contiene  $p$  parámetros  $\theta_1, \dots, \theta_p$  (notado por  $\theta$ ) y  $k$  variables  $\zeta_1, \dots, \zeta_k$ .

Si en  $Y$  se asocia una sola respuesta, por ejemplo, la concentración de producto de una reacción química de primer orden, la función  $f$  puede ser de la forma

$$f(\theta, \zeta) = \theta_1 [1 - \exp(-\theta_2 \zeta)],$$

donde  $\zeta$  indica el tiempo una vez que la reacción ha comenzado. Un segundo ejemplo, hace referencia al crecimiento de plantas u organismos el cual frecuentemente se representa por un modelo logístico de la forma

$$f(\theta, \zeta) = \frac{\theta_1}{1 + \theta_2 \exp(-\theta_3 \zeta)}$$

donde  $\zeta$  de nuevo representa el tiempo una vez que el crecimiento de la planta u organismo ha empezado.

**2.2 Supuestos sobre los errores.** A partir del modelo (1) si se supone:  $\epsilon_u \sim N(0, \sigma^2)$  para  $u = 1, \dots, n$ ; con  $\sigma^2$  desconocida y constante;  $\epsilon_u$  y  $\epsilon_s$  independientes para  $u \neq s$ ;  $u, s = 1, \dots, n$ ; una forma de validar estos supuestos es haciendo un análisis de los residuales, los cuales en su forma más general son obtenidos por:

$$r_u = y_u - f(\theta, \zeta_u) \quad u = 1, \dots, n.$$

Estos  $n$  valores se representan gráficamente contra los valores de predicción.

**2.3 Definición de No linealidad.** En esta sección, vamos a describir un procedimiento que permite diferenciar un modelo lineal de uno no lineal en los parámetros. Un modelo lineal se puede escribir en la forma

$$(2) \quad f(\theta, \zeta) = \sum_{i=1}^p \theta_i g_i(\zeta)$$

para alguna función  $g_i$  que depende sólo de los valores de  $\zeta$  pero no de los valores de  $\theta$  (parámetros). Los modelos que no se pueden escribir de esta forma son no lineales

en los parámetros o simplemente no lineales. Notese que la linealidad o no linealidad del modelo se determina por la forma en que entran los parámetros al modelo y no por la forma en que entran las variables  $\zeta_1, \dots, \zeta_k$ . Así por ejemplo, una ecuación cuadrática en  $\zeta$

$$(3) \quad f(\theta, \zeta) = \theta_1 + \theta_2\zeta + \theta_3\zeta^2$$

puede expresarse linealmente como en (2) haciendo  $g_1(\zeta) = 1$ ,  $g_2(\zeta) = \zeta$  y  $g_3(\zeta) = \zeta^2$ , en tanto que en el modelo

$$(4) \quad f(\theta, \zeta) = \theta_1 \exp(-\theta_2\zeta)$$

no puede hacerse esta representación. Una forma sencilla de ver si un modelo es lineal o no, es examinando las derivadas de  $f$  con respecto a cada uno de los parámetros  $\theta_i$ . Si  $\partial f / \partial \theta_i$  no depende de ninguno de los elementos de  $\theta = (\theta_1, \dots, \theta_p)^T$ , el modelo es lineal en  $\theta_i$ ; y si además es lineal en todos los  $p$  parámetros el modelo se dice lineal en los parámetros o simplemente lineal. Algunos modelos no lineales, mediante una apropiada transformación pueden describirse como en la ecuación (2), así por ejemplo el modelo caso (4) mediante la transformación logaritmo puede representarse como en (2). Las transformaciones para linealizar, sin embargo, tienen el efecto de transformar también los errores  $\epsilon$  alterando así la relación entre  $f$  y  $\epsilon$ . Generalmente los supuestos sobre  $\epsilon$  para el modelo original no tienen las mismas propiedades después de la transformación. Es bueno verificar la validez de los supuestos después de ajustar el modelo o usar modelos lineales generalizados en el análisis.

### 3. REGRESIÓN NO LINEAL

**3.1 Motivación.** La regresión no lineal se usa cuando se quieren estimar los parámetros de un modelo no lineal que relaciona una respuesta  $Y$  con algunas variables control o predictoras ( $\zeta_i, i = 1, \dots, k$ ). Esto es

$$Y_u = f(\theta, \zeta_u) + \epsilon_u \quad u = 1, \dots, n$$

Los modelos no lineales se originan cuando un investigador obtiene, por el desarrollo de una teoría o por otra situación, una relación funcional en la que los parámetros son no lineales. Un ejemplo es la función de Michaelis-Menten [7], la cual relaciona la tasa inicial o velocidad, de una reacción enzimática con la concentración de substrato  $\zeta$  por medio de la ecuación

$$f(\theta, \zeta) = \frac{\theta_1 \zeta}{\theta_2 + \zeta}$$

Este modelo es no lineal porque de acuerdo a 2.3 al obtener las derivadas

$$\frac{\partial f}{\partial \theta_1} = \frac{\zeta}{\theta_2 + \zeta}; \quad \frac{\partial f}{\partial \theta_2} = -\frac{\theta_1 \zeta}{(\theta_2 + \zeta)^2}$$

son función de al menos uno de los parámetros. Al tomar el recíproco de la función de Michaelis-Menten se tiene el modelo

$$\frac{1}{f} = \left(\frac{\theta_2}{\theta_1}\right) \left(\frac{1}{\zeta}\right) + \frac{1}{\theta_1}$$

el cual es lineal en los parámetros al hacer  $\beta_1 = \frac{1}{\theta_1}$  y  $\beta_2 = \frac{\theta_2}{\theta_1}$ . Se podría por consiguiente estimar  $\beta_1$  y  $\beta_2$  usando regresión lineal de los datos recíprocos de velocidad sobre los recíprocos de substrato y luego estimar  $\theta = (\theta_1, \theta_2)^T$ .

**3.2 Etapas Básicas en el Análisis de Regresión No Lineal.** En seguida, se presenta una síntesis de las etapas básicas en el análisis de la regresión no lineal:

1. Con los datos obtenidos de  $Y$  y  $\zeta_u$  y con la función de respuesta  $f(\theta, \zeta)$ , encontrar estimaciones iniciales para el vector de parámetros, o sea, obtener  $\theta_0$ .

2. Usando la información de 1., obtener las estimaciones mínimos cuadrados para  $\theta$  y producir un resumen de estadísticas por aproximación lineal.

3. Mirar si el modelo ajustado es adecuado, si las estimaciones de los parámetros son significativas examinando por ejemplo los residuales y las estimaciones de los parámetros como en regresión lineal.

#### 4. ESTIMACION DE PARÁMETROS

Supongamos que se tienen  $n$  observaciones  $\{Y_u, \zeta_u\}_{u=1}^n$  disponibles para ajustar el modelo de la forma

$$(5) \quad Y_u = f(\zeta_u, \theta) + \epsilon_u \quad u = 1, \dots, n$$

donde

$$\zeta_u = (\zeta_{1u}, \dots, \zeta_{ku})^T$$

$$\theta = (\theta_1, \dots, \theta_n)^T$$

$$\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$$

es decir, un modelo con  $p$  parámetros y  $k$  variables independientes. Bajo los supuestos de normalidad e independencia de los errores, de la forma  $\epsilon \sim N(0, I\sigma^2)$  donde  $0$  es un

vector de ceros e  $I$  es la matriz identidad ambos con tamaños apropiados. Definimos la suma de cuadrados del error para el modelo no lineal como:

$$(6) \quad S(\theta) = \sum_{u=1}^n [y_u - f(\zeta_u, \theta)]^2$$

Note que como  $y_u$  y  $\zeta_u$  son observaciones fijas, la suma de cuadrados es una función de  $\theta$ . Denotaremos por  $\hat{\theta}$ , una estimación mínimos cuadrados de  $\theta$ , esto es, un valor de  $\theta$  que minimiza  $S(\theta)$ . Se puede mostrar que, si  $\epsilon \sim N(0, I\sigma^2)$ , la estimación mínimos cuadrados de  $\theta$  es también la estimación de máxima verosimilitud de  $\theta$ . Debido al hecho que la función de verosimilitud para este problema se puede escribir como

$$L(\hat{\theta}, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{S(\hat{\theta})}{2\sigma^2}\right)$$

de modo que si  $\sigma^2$  es conocida, maximizar  $L$  con respecto a  $\theta$ , es equivalente a minimizar  $S(\theta)$  con respecto a  $\theta$ . Al diferenciar (6) con respecto a  $\theta$ , se obtienen  $p$  ecuaciones normales de la forma

$$(7) \quad \sum_{u=1}^n \left\{ Y_u - f(\zeta_u, \hat{\theta}) \right\} \left[ \frac{\partial f(\zeta_u, \theta)}{\partial \theta_i} \Big|_{\theta=\hat{\theta}} \right] = 0, \quad i = 1, \dots, \pi$$

Al resolver este sistema se obtiene la solución para  $\theta$ . Recordemos que si  $f$  es lineal entonces  $\frac{\partial f}{\partial \theta_i}$  para  $i = 1, \dots, p$  es una función de los  $\zeta_u$  independiente de  $\theta$ .

## 5. ESTIMACIÓN DE PARÁMETROS.

La estimación de parámetros en un modelo no lineal como se anoto anteriormente lleva implícito el uso de métodos numéricos, en esta sección se presentarán dos algoritmos de estimación, los cuales se encuentran implementados en el procedimiento NLIN del SAS; estos son:

1. Método de Gauss-Newton (Método de linealización).
2. Método de Marquardt

**5.1. Método de Gauss-Newton.** Este método usa los resultados de los mínimos cuadrados lineales en varias etapas; se supone que el modelo postulado es de la forma dada en la ecuación (5). Sean  $\theta_{10}, \dots, \theta_{p0}$  los valores iniciales para los parámetros  $\theta_1, \dots, \theta_p$ , los cuales se pueden conjeturar inteligentemente o estimar con anterioridad basados en cualquier información disponible para este propósito; al expandir en serie

de Taylor la función  $f(\zeta_u, \theta)$  alrededor del punto  $\theta_0$  donde  $\theta_0 = (\theta_{10}, \dots, \theta_{p0})^T$  y llevando en consideración sólo la primera derivada, (aproximación de primer orden),

$$(8) \quad f(\zeta_u, \theta) = f(\zeta_u, \theta_0) + \sum_{i=1}^p \left[ \frac{\partial f(\zeta_u, \theta)}{\partial \theta_i} \Big|_{\theta=\theta_0} \right] (\theta_i - \theta_{i0}).$$

Haciendo  $f_u^0 = f(\zeta_u, \theta_0)$ ;  $\beta_i^0 = \theta_i - \theta_{i0}$ ;  $F_{iu}^0 = \frac{\partial f(\zeta_u, \theta)}{\partial \theta_i} \Big|_{\theta=\theta_0}$  entonces (5) es aproximado por la forma

$$(9) \quad Y_u - f_u^0 = \sum_{i=1}^p \beta_i^0 f_{iu}^0 + \epsilon_u \quad u = 1, \dots, n$$

en otras palabras es lineal para el orden de aproximación seleccionado, y entonces se pueden estimar los parámetros  $\beta_i^0$ ,  $i = 1, \dots, p$ , aplicando la teoría de los mínimos cuadrados lineales.

Escribiendo,  $f_0 = \{f_{iu}^0\}$ , de orden  $n \times p$ ;  $b_0 = [b_i^0]$ , de orden  $P \times 1$  y  $Y - f^0 = [Y_1 - f_1^0]$ , de orden  $P \times 1$ , el estimador de  $\beta^0 = (\beta_1^0, \dots, \beta_p^0)^T$  está dado por

$$b_0 = (F_0^T F_0)^{-1} F_0^T (Y - f^0).$$

El vector  $b_0$  por consiguiente minimizará la suma de cuadrados

$$SS(\theta) = \sum_{u=1}^n \left[ Y_u - f(\zeta_u, \theta_0) - \sum_{i=1}^p \beta_i^0 f_{iu}^0 \right]^2$$

con respecto a los  $\beta_i^0$ ,  $i = 1, \dots, p$ , donde  $\beta_i^0 = \theta_i - \theta_{i0}$ . Si se escribe  $b_i^0 = \theta_{i1} - \theta_{i0}$ ; entonces los  $\theta_{i1}$ ,  $i = 1, \dots, p$ , se pueden pensar como las mejores estimaciones revisadas de  $\theta$ .

Luego se colocan los valores  $\theta_{i1}$ , las estimaciones revisadas, en los mismos lugares que fueron ubicados los  $\theta_{i0}$  y se sigue el mismo procedimiento descrito anteriormente pero reemplazando todos los ceros por unos. Esto nos llevará a otro conjunto de estimaciones revisadas  $\theta_{i2}$ , y así sucesivamente. En forma de vector podemos escribir,

$$\begin{aligned} \theta_{j+1} &= \theta_j + b_j \\ \theta_{j+1} &= \theta_j + (F_j^T F_j)^{-1} F_j^T (Y - f^j) \quad j = 0, 1, \dots \end{aligned}$$

donde

$$F_j = \{F_{j,u}^j\}; \quad f^j = (f_1^j, \dots, f_n^j)^T \quad \text{y} \quad \theta_j = (\theta_{1j}, \dots, \theta_{pj})^T.$$

Este proceso iterativo se repite hasta que la solución converja, esto es, hasta que en las iteraciones sucesivas,  $j, (j + 1)$ , se satisfaga

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

donde  $\delta$  es alguna cantidad positiva pequeña dada con anterioridad. En cada etapa del procedimiento iterativo, se calcula  $S(\theta_j)$  para ver si hay alguna reducción en la suma de cuadrados.

**5.2. Método de Marquardt (1963).** Este método realiza una interpolación óptima entre el método de linealización y el método del gradiente (o steepest-descent), la aproximación está basada en la máxima vecindad donde la aproximación de Taylor de primer orden da una adecuada representación del modelo no lineal. En contraste al método de linealización, el método del gradiente calcula el vector de correcciones  $b$  a partir del vector de estimaciones iniciales en la dirección del gradiente negativo de  $S(\theta)$ . Así,

$$b_g = - \left( \frac{\partial S(\theta)}{\partial \hat{\theta}_1}, \dots, \frac{\partial S(\theta)}{\partial \hat{\theta}_p} \right)^T$$

el subíndice  $g$ , indica que es el vector corrección dado por el método del gradiente. El problema del método del gradiente es que converge muy lentamente después de unas pocas iteraciones. Tanto con este método como con el método de linealización es necesario controlar cuidadosamente el tamaño del paso una vez establecida la dirección del vector corrección; el control consiste en tomar una fracción  $\lambda$ ,  $0 < \lambda < 1$ , del vector corrección en lugar de tomar  $\lambda = 1$ .

**5.3. Análisis cualitativo del problema.** En vista de las deficiencias de los métodos de linealización y del gradiente, es bueno revisar los principios fundamentales involucrados. Primero, cualquier método apropiado debe resultar en un vector corrección cuya dirección esté dentro de  $90^\circ$  del gradiente negativo de  $S(\theta)$ . De lo contrario, los valores de  $S(\theta)$  pueden aumentar en lugar de disminuir en puntos ubicados a lo largo del vector corrección. Segundo, debido a la forma alargada de la superficie  $S(\theta)$  en la mayoría de los problemas,  $b_l$  (vector corrección hallado por el método de linealización) está generalmente casi a  $90^\circ$  de  $b_g$ . "Realmente hemos encontrado que este ángulo, digamos  $\gamma$ , para una variedad de problemas cae en el rango  $80^\circ < \gamma < 90^\circ$ ", (Marquardt (1963)). Para una explicación más detallada del algoritmo ver (Marquardt (1963)).

Teniendo en cuenta estas consideraciones, Marquardt produce un método para interpolar los vectores  $b_g$  y  $b_l$ , para obtener un tamaño de paso apropiado. El vector

corrección descrito por él es el siguiente

$$b_M = (F_0^T F_0 + \lambda \text{diag}(F_0^T F_0))^{-1} F_0^T (Y - f^0)$$

Note que cuando  $\lambda \rightarrow 0$ , la dirección se aproxima a la del método de linealización. Cuando  $\lambda \rightarrow \infty$ , la dirección se aproxima a la del gradiente.

## 6. PROPIEDADES ESTADÍSTICAS DE LOS ESTIMADORES MÍNIMOS CUADRADOS

Para establecer propiedades estadísticas de los estimadores en muestras grandes, esto es, propiedades que se tienen sólo asintóticamente, es necesario que se cumplan ciertas condiciones de regularidad dadas por Gallant (1975, 1986), que se resumen en las siguientes etapas.

- Los vectores  $\zeta_u$ ,  $u = 1, \dots, n$ , se deben comportar apropiadamente cuando  $n \rightarrow \infty$ . Este comportamiento se obtiene cuando las componentes  $\zeta_{iu}$ ,  $i = 1, \dots, k$ , de  $\zeta_u$  se seleccionan por muestreo aleatorio de una distribución con momento de primer orden finito o por replicación de un conjunto fijo de puntos.

- La función de respuesta  $f(\zeta, \theta)$  debe ser continua en sus argumentos  $(\zeta, \theta)$ , así como también sus primeras y segundas derivadas, es decir,  $\frac{\partial f(\zeta, \theta)}{\partial \theta_i}$  y  $\frac{\partial^2 f(\zeta, \theta)}{\partial \theta_i \partial \theta_j}$  continuas en  $(\zeta, \theta)$ .

- El  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{u=1}^n [f(\zeta_u, \theta) - f(\zeta_u, \theta^*)]^2$  tiene un mínimo único en  $\theta^*$  (vector  $p$ -dimensional de parámetros desconocidos).

- El  $\lim_{n \rightarrow \infty} \frac{1}{n} F^T(\theta^*)F(\theta^*)$  es no singular.

Bajo las anteriores condiciones se tienen las siguientes propiedades asintóticas de los estimadores (Gallant 1986, Cháves et al 1984)

1.  $\hat{\theta}$  es consistente para  $\theta$ .
2.  $\hat{\sigma}^2 = \frac{S(\hat{\theta})}{n-p}$  es consistente para  $\sigma^2$ .
3.  $\hat{\theta} \sim AN_p(\theta, \sigma^2 C)$  ( $AN_p$ : Asintóticamente Normal  $p$ -variada) donde  $C = (F^T F)^{-1}$ .
4.  $\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \overset{A}{\sim} \chi^2(n-p)$ .
5.  $\hat{\theta}$  y  $\hat{\sigma}^2$  son asintóticamente independientes.
6. La matriz  $C = [F^T(\theta)F(\theta)]^{-1} = (F^T F)^{-1}$  se puede estimar por  $\hat{C} = [F^T(\hat{\theta})F(\hat{\theta})]^{-1}$ .

Un intervalo del 95% de confianza para  $\theta_i$  se puede encontrar a partir del valor crítico  $t_{0.025}$  de una variable  $t$  con  $n - p$  grados de libertad así:

$$\hat{\theta}_i \pm t_{0.025} \sqrt{\hat{\sigma}^2 \hat{c}_{ii}}$$



Similarmente, la hipótesis  $H_0 : \theta_i = \theta_{i0}$  se puede probar al nivel de significancia  $\alpha$ , comparando

$$|\tilde{t}_i| = \frac{|\hat{\theta}_i - \theta_{i0}|}{\sqrt{\hat{\sigma}^2 \hat{c}_{ii}}}$$

con  $|\tilde{t}_{\alpha/2}|$  y rechazando  $H_0$  cuando  $|\tilde{t}_i| > |\tilde{t}_{\alpha/2}|$ ;  $c_{ii}$  denota el  $i$ -ésimo elemento de la diagonal de  $\hat{C}$ .

Nota: Por la propiedad asintótica (3.) se tiene que  $Var(\hat{\theta}_i) = \sigma^2 c_{ii}$  y  $Cov(\hat{\theta}_i, \hat{\theta}_j) = \sigma^2 c_{ij}$ ; de donde se sigue que

$$c_{ii} = \frac{Var(\hat{\theta}_i)}{\sigma^2} \text{ y } c_{ij} = \frac{Cov(\hat{\theta}_i, \hat{\theta}_j)}{\sigma^2} \text{ para } i \neq j$$

de aquí se pueden estimar los componentes de  $C$  usando el cuadrado medio residual  $\hat{\sigma}^2$ , los errores estandar de los estimadores y la matriz de correlaciones de los estimadores. Esta información es dada por el paquete SAS (Cháves 1984; Velasquez et al 1986).

**6.1 Pruebas de hipótesis.** Sea la hipótesis de interés:

$$H_0 : \tau = \tau_0 \quad \text{vs} \quad H_1 : \tau \neq \tau_0$$

donde  $\tau$  es algún subvector del vector de parámetros  $\theta$ . Es decir el vector  $\theta$  se ha particionado de la forma  $\theta^T = (\rho^T, \tau^T)$  donde  $\tau$  es el subvector  $q \times 1$  de parámetros de interés y  $\rho$  es el subvector  $r \times 1$  de parámetros restantes. Note que  $p = q + r$ . Se distinguen tres estadísticas de prueba en este caso:

- La del cociente de verosimilitud ( $T$ )
- La prueba basada en la normalidad asintótica del estimador mínimos cuadrados
- La prueba de Wald.

-La estadística de prueba del cociente de verosimilitud  $T$ , es el cociente entre el estimador de máxima verosimilitud de la varianza para el modelo reducido por la hipótesis nula, digamos  $\check{\sigma}^2$  y el estimador de máxima verosimilitud de la varianza para el modelo completo, notado como antes por  $\hat{\sigma}^2$ . Es decir,  $\check{\sigma}^2$  es el estimador de la varianza del modelo  $Y_u = f(\zeta_u, (\rho^T, \tau_0^T)) + \epsilon_u$  que tiene como parámetros los componentes de  $\rho$  ya que  $\tau_0$  son conocidos bajo  $H_0$  y  $\hat{\sigma}^2$  resulta de estimar el modelo completo  $Y_u = f(\zeta_u, \theta) + \epsilon_u$ . Así, para calcular el cociente

$$T = \frac{\check{\sigma}^2}{\hat{\sigma}^2}$$

se hace necesario estimar dos modelos; el modelo reducido por la hipótesis nula y el modelo completo; esto hace que la aplicación de esta prueba sea un poco costosa y

difícil de realizar. La hipótesis  $H_0 : \tau = \tau_0$  se rechaza cuando  $T$  es mayor que un valor crítico  $c$ , donde  $P(t > c|H_0) = \alpha$ . El valor  $c$ , se puede aproximar por:

$$c \approx 1 + \frac{q}{n-p} F_\alpha$$

donde  $F_\alpha$  es el punto crítico  $\alpha$ -superior de una distribución  $F$  con grados de libertad del numerador  $q$  y del denominador  $n-p$ .

-Como se vió anteriormente,  $\hat{\theta} \sim AN_p(\theta, \sigma^2 C)$ , con  $\sigma^2 C$  estimado por  $\hat{\sigma}^2 \hat{C}$ . Entonces, una estadística de prueba para la hipótesis  $H_0 : \tau = \tau_0$  vs.  $H_1 : \tau \neq \tau_0$ , es la siguiente:

$$S = \frac{(\hat{\tau} - \tau_0)^T \hat{C}_{22}^{-1} (\hat{\tau} - \tau_0)}{q \hat{\sigma}^2}$$

donde  $\hat{C}_{22}$  es la submatriz de  $\hat{C}$  obtenida eliminando las primeras  $r$  filas y columnas y  $\hat{\tau}$  es el vector que resulta de eliminar las primeras  $r$  filas de  $\hat{\theta}$ ;  $\hat{\theta}^T = (\hat{\rho}^T, \hat{\tau}^T)$ . La hipótesis se rechaza cuando  $S > d$ , donde  $P(S > d|H_0) = \alpha$ . En la práctica  $d$  se aproxima por  $d \approx F_\alpha$  donde  $F_\alpha$  es como antes.

-La prueba de Wald: supongamos que se quiere probar

$$H_0 : h(\theta) = 0 \quad \text{vs.} \quad H_1 : h(\theta) \neq 0$$

donde  $h(\theta)$  es una función continua y diferenciable de  $\mathbb{R}^p \rightarrow \mathbb{R}^q$  ( $q < p$ ), con jacobiano

$$H(\theta) = \frac{\partial}{\partial \theta^T} h(\theta)$$

de orden  $q \times p$ . Cuando  $H(\theta)$  se evalúa en  $\theta = \hat{\theta}$  escribimos  $\hat{H} = H(\hat{\theta})$ . La estadística de Wald para probar  $H_0$  está dada (Cháves, 1982 y Gallant, 1986) por:

$$W = \frac{h^T(\hat{\theta})(\hat{H} \hat{C} \hat{H}^T)^{-1} h(\hat{\theta})}{q \hat{\sigma}^2}$$

La regla de decisión para este caso es: rechazar  $h(\theta) = 0$ , si  $W > F_\alpha$  con  $F_\alpha$  como antes.

## 7. EJEMPLO ILUSTRATIVO

Para mostrar una aplicación de los resultados teóricos, se usarán los datos obtenidos por Yesmith Santos en la elaboración de su tesis de magister en Ingeniería Química, U. Nal. de Colombia (1992). El estudio consistió en medir en una reacción química, la concentración de un sustrato en un determinado instante del tiempo. Las variables estudiadas fueron: concentración de un sustrato en un determinado tiempo, notada por  $Y$  y tiempo  $t$  en el cual se hizo la observación. Se obtuvieron un total de 10 observaciones. De lo anterior se tiene que el número de variables independientes es  $k = 1$ , el tiempo, y  $n = 10$ . Los datos observados aparecen en la tabla siguiente:

Tabla de datos.

$u$	$Y$	$t$
1	0.002	0.5
2	0.005	1.0
3	0.008	1.5
4	0.023	2.0
5	0.029	2.5
6	0.026	3.0
7	0.019	3.5
8	0.014	4.0
9	0.010	4.5
10	0.007	5.0

## IDENTIFICACIÓN DEL MODELO

Un análisis gráfico inicial permite concluir claramente que estas dos variables no están asociadas linealmente, ya que al comenzar la reacción química y hasta el tiempo  $t = 22.5$  min. hay un incremento en la concentración de sustrato  $Y$ , pero a partir de este punto y hasta el final de las observaciones, empieza a disminuir dicha concentración. Lo anterior sugiere un modelo de la forma:

$$Y_u = \theta_1 t_u^{\theta_2} \exp(-\theta_3 t_u) \quad u = 1, \dots, 10$$

donde  $\theta = (\theta_1, \theta_2, \theta_3)^T$  es el vector de parámetros a estimar; de aquí,  $p = 3$ . Notese que este modelo es intrínsecamente lineal puesto que al utilizar la transformación logaritmo natural ( $\ln$ ) se tiene

$$\ln Y_u = \ln \theta_1 + \theta_2 \ln t_u - \theta_3 t_u$$

el cual es lineal en los parámetros  $\beta_0 = \ln \theta_1$ ,  $\beta_1 = \theta_2$  y  $\beta_2 = -\theta_3$ .

## ESTIMACIÓN DEL MODELO

Para el presente caso, como el modelo es linealizable, se utilizó regresión lineal para encontrar las estimaciones de los parámetros  $\beta_0$ ,  $\beta_1$  y  $\beta_2$  y a partir de éstas los valores iniciales del vector  $\theta$  fueron:  $\theta_0 = (0.02494; 3.17171; 1.24044)^T$ .

Ya con el vector de estimaciones iniciales se puede usar el procedimiento NLIN del paquete SAS, para estimar los parámetros del modelo. El paquete realiza la estimación de los parámetros utilizando el algoritmo de Gauss-Newton, ya que se les están dando las derivadas parciales de la función de respuesta con respecto a cada uno de los parámetros y no se está utilizando la opción METHOD (ver programa). El programa para el ajuste del modelo en cuestión es:

```

DATA ILUSTR;
INFILE 'A:\datos.prn';
INPUT T Y;
PROC NLIN ITER=50 CONVERGENCE=1.0E-13;
  PARS T1=0.02494 T2=3.17171 T3=1.24044;          (Estimaciones iniciales de los parametros)
  MODEL Y=T1*T**T2*EXP(-T3*T);                  (Modelo a estimar)
  DER.T1=T**T2*EXP(-T3*T);
  DER.T2=T1*LOG(T)*T**T2*EXP(-T3*T);          (Derivadas parciales con respecto a cada parámetro)
  DER.T3=-T1*T**(T2+1)*EXP(-T3*T);
  OUTPUT OUT=SALIDA P=YHAT RESIDUAL=RES;
RUN;

```

La salida de resultados aparece a continuación:

#### Non-Linear Least Squares Summary Statistics Dependent Variable Y

Source	DF	Sum of Squares	Mean Square	
Regression	3	0.00281313659	0.00093771220	
Residual	7	0.00003186341	0.00000455192	
Uncorrected Total	10	0.00284500000		
(Corrected Total)	9	0.00080010000		
Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95% Confidence Interval	
			Lower	Upper
T1	0.032860700	0.00529487530	0.0203402090	0.0453811912
T2	6.715867384	0.86085322118	4.6802566386	8.7514781298
T3	2.532755590	0.31146501112	1.7962519590	3.2692592218

De los resultados obtenidos al ejecutar el programa, se tiene que después de diez iteraciones el vector de estimaciones queda dado por:  $\hat{\theta} = (0.032861, 6.715867, 2.532756)^T$ , con  $S(\hat{\theta}) = 0.0000319$ . Por consiguiente, el modelo ajustado queda:

$$\hat{Y}_u = 0.033t_u^{6.716} \exp(2.533t_u) \quad u = 1, \dots, 10$$

Las estimaciones de las varianzas son:

$$\hat{\sigma}^2 = \frac{0.000032}{7} = 0.00000455$$

$$\text{var}(\hat{\theta}_1) = (0.00529)^2 = 0.000028; \quad \text{var}(\hat{\theta}_2) = (0.86085)^2 = 0.741063$$

$$\text{var}(\hat{\theta}_3) = (0.31147)^2 = 0.097014$$

Nota: aquí  $\text{var}(\hat{\theta}_i)$ ,  $i = 1, 2, 3$ ; denota las respectivas estimaciones de  $\text{Var}(\hat{\theta}_i)$ . La matriz  $C$  se puede estimar de la siguiente manera:

$$\hat{c}_{11} = \frac{\text{var}(\hat{\theta}_1)}{\hat{\sigma}^2} = 6.1538, \quad \hat{c}_{22} = \frac{\text{var}(\hat{\theta}_2)}{\hat{\sigma}^2} = 162870.989$$

$$\hat{c}_{33} = \frac{\text{var}(\hat{\theta}_3)}{\hat{\sigma}^2} = 21321.7582, \quad \hat{c}_{12} = \frac{\text{cov}(\hat{\theta}_1, \hat{\theta}_2)}{\hat{\sigma}^2} = 9.070583,$$

$$\hat{c}_{13} = \frac{\text{cov}(\hat{\theta}_1, \hat{\theta}_3)}{\hat{\sigma}^2} = 180.086436, \quad \hat{c}_{23} = \frac{\text{cov}(\hat{\theta}_2, \hat{\theta}_3)}{\hat{\sigma}^2} = 58004.12481$$

Supongamos que se desea probar la hipótesis  $H_0 : \theta_1 = 0$  vs.  $H_1 : \theta_1 \neq 0$ , al nivel de significancia  $\alpha = 5\%$ , mediante la prueba de Wald.

En este caso,  $h(\theta) = \theta_1$  y por consiguiente  $h(\hat{\theta}) = \hat{\theta}_1 = 0.0329$ ; el jacobiano para este ejemplo queda

$$\hat{H}_{1 \times 3} = \frac{\partial h(\hat{\theta})}{\partial \theta^T} = \frac{\partial h(\hat{\theta}_1)}{\partial \theta^T} = (1, 0, 0).$$

Así el valor de la estadística de prueba de Wald es:

$$W_0 = \frac{(0.0329) \left\{ [1, 0, 0] \begin{bmatrix} 6.154 & 9.071 & 180.086 \\ 9.071 & 162870.989 & 58004.125 \\ 180.086 & 58004.125 & 21321.758 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}^{-1}}{1 \times (0.00000455)} \\ = \frac{0.00017589}{0.00000455} = 38.657$$

ahora bien,  $F_{.05}(1, 7) = 5.59$ . Como  $W_0 > 85.59$ , se rechaza  $H_0$ .

#### VERIFICACIÓN DE SUPUESTO

Como se dijo anteriormente, en el momento de ajustar un modelo, lineal o no, se hacen algunos supuestos sobre los errores  $\epsilon$ . Es bueno, una vez ajustado el modelo verificar la validez de tales supuestos. Con la ayuda del computador se hicieron las respectivas pruebas de homogeneidad de varianzas, promedio igual a cero, normalidad e independencia obteniéndose resultados satisfactorios; es decir, ningún supuesto deja de cumplirse.

#### BIBLIOGRAFIA

1. Burguete, H. F. (1982)., *El Uso de la Metodología Científica y la Estadística, con Especial Referencia a los Modelos no Lineales*, en las Ciencias Económicas., Conferencia Centro de Economía, C. P. Chapingo México..
2. Cháves, C.B., *Una prueba de Wald en regresión*, Tesis Colegio de Postgrado, México.
3. Cháves, C. B., Burguete, H. F. e Infante, G. S. (1984)., *Metodología Estadística en Modelos de Regresión no Lineal*, Comunicaciones en Estadística y Cómputo 3(1)., Centro de Estadística y Cálculo, C. P. Chapingo, México.

4. Draper, N. R. and Smith, H. (1981)., *Applied Regression Analysis 2nd ed*, vol. 10., John Wiley & Sons, New York.
5. Gallant, A. R. (1975), *Nonlinear Regresión*, The American Statistician 29 (2), 73-81.
6. Gallant, A. R. (1986), *Nonlinear Regression*, vol. 1, John Wiley & Sons, New York.
7. Marquardt, D. W. (1963), *An algorithm for least squares estimation of nonlinear parameters*, Journal of The Society For Industrial and Applied Mathematics 2, 431,441.
8. Michaelis, L. and Menten, M. L. (1913)., *Biochemische Zeit* 49, 333-339.
9. SAS. (1979)., *User's guide*, SAS Institute., Raleigh, N. C..
10. Velasquez, V. C. y Escarpetta G. J. (1986)., *Algunos Aspectos Teóricos y Aplicaciones de la Regresión no Lineal. Tesis de Estadísticos*, Universidad Nacional de Colombia, Santafé de Bogotá D. C..