Software Estadístico para Regresión. El caso de Regresión Logística y Regresión Poisson

Andrea Fuentes, Nathaly Rivera, Raúl Pinos, Gaudencio Zurita
Instituto de Ciencias Matemáticas
Escuela Superior Politécnica del Litoral (ESPOL)
Campus Gustavo Galindo, Km 30.5 vía Perimetral
Apartado 09-01-5863. Guayaquil-Ecuador
aefuente@espol.edu.ec, nrivera@espol.edu.ec, raalpino@espol.edu.ec, gzurita@espol.edu.ec

Resumen

El desarrollo del software de Análisis de Regresión Avanzada denominado ERLA está compuesto de diversos módulos específicos, entre los cuales está la Regresión Logística y la Regresión Poisson, que son métodos no lineales.
En los modelos de Regresión Lineal Simple se explica la variable de respuesta Y en función de solo una variable X y los valores de β_0 y β_1 lineales, además los parámetros β_0, β_1, y σ^2 del modelo de Regresión Lineal Simple pueden ser estimados a través de diferentes criterios tales como, Mínimos Cuadrados o Máxima Verosimilitud, la diferencia con la Regresión Logística y Poisson se basa en que son modelos de regresión no lineales, donde la regresión logística determina la relación entre una variable de respuesta Y que es binaria y una o más variables de explicación $X_1, X_2, \ldots, X_k$, que son variables continuas, mientras la regresión Poisson se utiliza un modelo no lineal que pertenece a la categoría del análisis de datos de recuento. Respecto a la estimación de parámetros para la regresión logística y Poisson se aplica el método de máxima verosimilitud , que parte de la función de verosimilutd de acuerdo a la regresión que se trabajará y debe recurrirse a métodos iterativos, como es el método de Newton–Raphson. Para aplicar ambos métodos se desarrollo el software estadístico ERLA (Estadística de Regresión Lineal Avanzada), desarrollado por estudiantes del Instituto de Ciencias Matemáticas.

Palabras Claves: *Logística, Poisson, Modelo de Regresión No Lineal*

Abstract

The development of the software called Advanced Regression Analysis, call ERLA is composed of several specific modules, among which is the Logistic Regression and Poisson Regression, which are nonlinear methods. In Simple Linear Regression models is explained the response variable Y depending on only one variable X and the values β_0 y β_1 are linear, also the parameters β_0, β_1, and σ^2 from the simple linear regression model can be estimated by different criteria such as least squares or maximum likelihood, the difference between Poisson and Logistic regression are based on nonlinear regression models, where the logistic regression determines the relationship between a response variable Y that is binary and one or more explanatory variables $X_1, X_2, \ldots, X_k$, which are continuous variables, while the Poisson regression using a nonlinear model which belongs to the category of analysis of count data. Respect to the estimation of parameters for logistic regression and Poisson applies the maximum likelihood method, which starts from the verosimilutd function according to the regression work and must use iterative methods, such as Newton-Raphson. To apply both methods we develop the statistical software ERLA (Advanced Linear Regression Statistics), developed by students of the Institute of Mathematical Sciences.

Keywords*: Logistic, Poisson, No Linear Regression*

1. Introducción

El presente trabajo es un proyecto previo a la obtención del título de Ingeniero en Estadística Informática, de la materia de Regresión Avanzada dictada por el Máster Gaudencio Zurita Herrera en el I término del 2010 y dirigida para la carrera de Ingeniería en Estadística Informática de la ESPOL.

Dentro de la materia se desarrollo el Software Estadístico de Regresión Lineal Avanzada denominado ERLA, citando modelos de Regresión No lineal como lo es Regresión Logística y regresión Poisson.

La regresión debido a su simplicidad es una de las técnicas más usadas que sirve para explicar valores de una o más variables de respuesta en términos de un grupo de variables predictoras.

Para el desarrollo de este proyecto consideramos fundamental mostrar el modelo de Regresión Lineal Simple que trabaja con una sola variable a ser explicada, Regresión Lineal Múltiple que considera más de una variable a ser explicada en términos de un grupo de variables de explicación, para dar paso a los modelos de regresión menos usados y que son no lineales.

2. Regresión Lineal

En Regresión Lineal Simple, tratamos de explicar Y en función de X con la asistencia de la ecuación de una recta con β_1 como la pendiente y β_0 como la intersección con el eje Y, pero una vez hecho el cálculo determinístico de Y, y tomado la lectura experimental de Y, se encuentra que no siempre coinciden, ya que hay la presencia de un error aleatorio ε_i, que nos hace reescribir la relación de Y con X: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

Dado este modelo donde Y es la variable a ser explicada condicionalmente por X, a quien llamaremos variable de explicación y ε_i una variable aleatoria que influencia en la observación del valor y_i de Y cuando X=x_i; vamos a trabajar con el siguiente modelo condicional y bajo los siguientes supuestos:

$$E(\mathrm{Y}|\mathrm{X_i} = \mathrm{x_i}) = \beta_0 + \beta_1 \mathrm{x_i} \quad \mathrm{i}=1, 2, \ldots, \mathrm{n}$$

$$\mathrm{E}(\varepsilon_i) = 0 \quad \mathrm{Var}(\varepsilon_i) = \sigma^2 ; \quad \mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \mathrm{i} \neq \mathrm{j}$$

Es un modelo de Regresión Lineal simple porque se explica la variable de respuesta Y en función de solo una variable X y los valores de β_0 y β_1 son lineales en la expresión $\mathrm{E}(\mathrm{Y}|\mathrm{X} = \mathrm{X_i}) = \beta_0 + \beta_1 \mathrm{x_i}$ que también es denominada Función de Respuesta o Parte Determinística del modelo; los valores de σ^2, β_0, β_1 son constantes desconocidas pero estadísticamente estimables; ε_i es una variable aleatoria como fuera Haciendo ε_i, que el valor observado $\mathrm{y_i}$de Y sea una Variable Aleatoria, expresamos el valor esperado del modelo:

$$\mathrm{E}[\mathrm{Y_i}] = \mathrm{E}[\mathrm{Y}|\mathrm{X} = \mathrm{x_i}] = \mu_{\mathrm{Y_i}} = \mathrm{E}[\beta_0 + \beta_1 \mathrm{x_i} + \varepsilon_i]$$

$$\mathrm{E}[\mathrm{Y_i}] = \mathrm{E}[\beta_0] + \mathrm{E}[\beta_1 \mathrm{x_i}] + \mathrm{E}[\cancel{\varepsilon_i}]^{0}$$

$$\mathrm{E}[\mathrm{Y_i}] = \beta_0 + \beta_1 \mathrm{x_i} \; = \mu_{\mathrm{Y_i}}$$

$$\mathrm{Var}(\mathrm{Y_i}) = \mathrm{Var}(\beta_0 + \beta_1 \mathrm{x_i} + \varepsilon_i) = \sigma^2$$

Si suponemos que el Error ε_i se distribuye normalmente, tenemos entonces que $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$, siendo σ^2 constante, supuesto de homocedasticidad, lo que implica que $\mathrm{Y_i} \sim \mathrm{N}(\beta_0 + \beta_1 \mathrm{x_i}, \sigma^2)$, que es conocido como el Teorema de Gauss-Markov

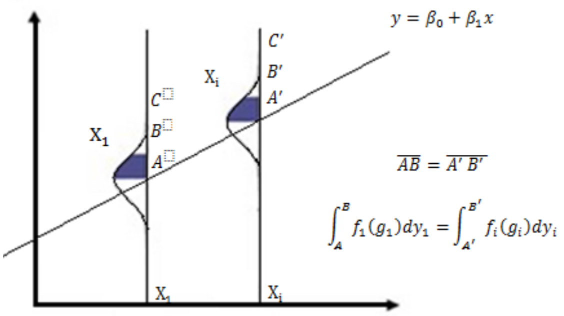

Figura 1Teorema de Gauss-Markov

2.1 Regresión Múltiple

Existe más de una Variable de Explicación, por lo que consideraremos un modelo con p términos y $(\mathrm{p}-1)$ variables de Explicación, suponiendo información de n casos, esto es: $\mathrm{i} = 1, 2, \ldots, \mathrm{n}$.

El Modelo Lineal para el i-ésimo caso es el siguiente, $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i,p-1} + \varepsilon_i$
Expresado el modelo de la forma matricial para n observaciones de Y y $\mathrm{X_1}, \mathrm{X_2}, \ldots, \mathrm{X_{p-1}}$ es:

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

Presentado de forma matricial es el siguiente

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1(p-1)} \\ 1 & x_{21} & x_{22} & \cdots & x_{2(p-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{n(p-1)} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Donde $\boldsymbol{Y} \in \mathrm{R^n}$, $\boldsymbol{X} \in \mathrm{M_{n\times p}}$ es la matriz de diseño, $\boldsymbol{\beta} \in \mathrm{R^p}$ es el vector de estimadores y $\boldsymbol{\varepsilon} \in \mathrm{R^n}$ denominado vector de errores. A demás téngase en cuenta que $\boldsymbol{E}[\boldsymbol{Y}] = [\boldsymbol{X\beta} + \boldsymbol{\varepsilon}] = \boldsymbol{X\beta}$ ya que;

$$\boldsymbol{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}; \boldsymbol{0} \in \mathrm{R^n}$$

Y la Matriz $\boldsymbol{\Sigma}_\varepsilon$ de Varianzas y Covarianzas del Error es: $\boldsymbol{\Sigma}_\varepsilon = \sigma^2 \mathbf{I}$**,** donde I es la Matriz identidad $n \times n$, y que los errores son independientes.

2.1.1 Estimación de parámetros

Debemos estimar los p coeficientes $\beta_0, \beta_1, \ldots, \beta_{p-1}$ para lo cual el Criterio de Mínimos Cuadrados propone que los Estimadores del modelo, sean los valores $\mathrm{b_0}, \mathrm{b_1}, \ldots, \mathrm{b_{p-1}}$ que minimizan **L**.
Así deseamos encontrar el un vector de los estimadores de Mínimos Cuadrados, $\boldsymbol{\beta}$, que minimice:

$$\mathrm{L} = \sum_{\mathrm{i}=0}^{\mathrm{n}} \boldsymbol{\varepsilon}^{\mathrm{T}} \boldsymbol{\varepsilon} = \; (\mathbf{y} - \mathbf{X\beta})^{\mathrm{T}} (\mathbf{y} - \mathbf{X\beta})$$

L se puede expresar como:

$$L = \; \mathbf{y}^{\mathrm{T}}\mathbf{y} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y} - \mathbf{y}^{\mathrm{T}}\mathbf{X\beta} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X\beta}$$

$$\mathrm{L} = \; \mathbf{y}^{\mathrm{T}}\mathbf{y} - 2\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X\beta}$$

Dado que $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$ es una matriz (1x1), o un escalar, y su transpuesta $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y} = \mathbf{y}^{\mathrm{T}}\mathbf{X\beta}$ es el mismo escalar. Los estimadores de Mínimos Cuadrados deben satisfacer

$$\left.\frac{\partial \mathrm{L}}{\partial \beta}\right|_{\widehat{\beta}} = -2\mathbf{X}^{\mathrm{T}}\mathbf{y} + 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\widehat{\boldsymbol{\beta}} = 0$$

Que se simplifica a:

$$\mathbf{X}^{\mathrm{T}}\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}^{\mathrm{T}}\mathbf{Y}$$

Ésta es la forma matricial de las Ecuaciones Normales de Mínimos Cuadrados; para resolver estas ecuaciones, multiplicamos a ambos lados por la inversa de $\boldsymbol{X}^T\boldsymbol{X}$, bajo el supuesto que $\boldsymbol{X}^T\boldsymbol{X}$ no es singular, esto es, que $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ existe, de tal modo que el estimador de Mínimos Cuadrados de $\boldsymbol{\beta}$ es:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

Da como resultado que sus estimadores obtenidos son Insesgados, lo que significa que $E\left[\widehat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$, puesto que: $E\left[\widehat{\boldsymbol{\beta}}\right] = E[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}] = \boldsymbol{\beta}$

Planteamos la ecuación estimada

$$\widehat{\boldsymbol{Y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$$

$$\widehat{\boldsymbol{Y}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

Y conociendo que $\mathbf{H} \in \mathrm{S}_{\mathrm{n}\times\mathrm{n}}$ es la denominada *Matriz Hat*:

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

La cual sirve para visualizar los valores estimados de **Y** como combinaciones lineales de los valores observados de **Y**, que se muestran en la siguiente forma:

$$\begin{bmatrix} \widehat{\mathrm{Y}}_1 \\ \widehat{\mathrm{Y}}_2 \\ \vdots \\ \widehat{\mathrm{Y}}_{\mathrm{n}} \end{bmatrix} = \widehat{\mathbf{Y}} = \mathbf{HY} = \begin{bmatrix} \mathrm{h}_{11} & \mathrm{h}_{12} & \mathrm{h}_{13} & \cdots & \mathrm{h}_{1\mathrm{n}} \\ \mathrm{h}_{21} & \mathrm{h}_{22} & \mathrm{h}_{23} & \cdots & \mathrm{h}_{2\mathrm{n}} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \mathrm{h}_{\mathrm{n}1} & \mathrm{h}_{\mathrm{n}2} & \mathrm{h}_{\mathrm{n}3} & \cdots & \mathrm{h}_{\mathrm{nn}} \end{bmatrix} \begin{bmatrix} \mathrm{Y}_1 \\ \mathrm{Y}_2 \\ \vdots \\ \mathrm{Y}_{\mathrm{n}} \end{bmatrix}$$

Además se puede probar que la matriz **H** es idempotente, es decir que: $\mathbf{HH} = \mathbf{H}^{\mathbf{2}} = \mathbf{H}$ **.**

3. Modelo de regresión No Lineal

El Modelo Lineal Generalizado, nace cuando las variables de Y y X no están relacionadas de una manera directa y utilizando las familias exponenciales que permiten descomponer las distribuciones exponenciales tales como Normal, Poisson, Binomial, en términos de funciones lineales crean una función de "enlace", la cual permite utilizar los mismos métodos que fueron aplicados para calcular los estimadores de beta, como mínimos cuadrados y máxima verosimilitud, pero en este caso las ecuaciones no tienen solución explicita, sino una solución implícita lo que hace que se necesite un método numérico.

3.1. Familias exponenciales

La familia exponencial es una clase de distribuciones de probabilidad cuya formulación matemática comparten cierta forma. Esta forma especial es escogida por interés matemático, que confiere a las distribuciones de esta familia una serie de propiedades algebraicas y estadísticas. Incluye distribuciones, sean estas continuas o discretas como la normal, binomial, etc..

Considérese una variable aleatoria Y cuya distribución de probabilidades depende de un parámetro θ. La distribución pertenece a las familias exponenciales si puede ser escrita de la forma.

$$f(x;\theta) = h(x)g(\theta)e^{[\eta(\theta)T(x)]}$$

Donde $h, g, \eta\ y\ T$ son funciones conocidas, Nótese la simetría entre x y θ. Esto se enfatiza si la ecuación $f(x;\theta)$ es reescrita como:

$$f(x;\theta) = \exp[\eta(\theta)T(x) + c(\theta) + d(x)]$$

Donde $h(x) = exp[d(x)]$ y $g(\theta) = \exp[c(\theta)]$.

Si $T(x) = x$, la distribución se dice que está en su Forma Canoníca (esto es, estándar), y $\eta(\theta)$ es llamada el parámetro natural de la distribución.

A $\eta(\theta)$ se lo conoce como Parámetro Natural, que nos proporciona en sí el "enlace" que se utilizará más adelante; $\eta(\theta)$ especifica los parámetros necesarios para dicha distribución.

$g(\theta)$ es el factor de "normalización", que asegura que $p(x|\theta)$ siga siendo una distribución de probabilidad

$T(x)$ es el estadístico suficiente de la "información".

$h(x)$ es una base de medida no negativa, que es generalmente 1.

Si hay otras variables en la función, además del parámetro de interés θ, son relegadas como parámetros ruido formando parte de las funciones η, T, c y d.

3.2. Modelo Lineal generalizado

Es una generalización de la Regresión Lineal para poder responder a otros tipos de modelos además de los lineales siempre y cuando la distribución de la respuesta sea miembro de las familias exponenciales.

La generalización se obtiene al suponer que $\mathrm{E}\big(\mathrm{Y}(\mathbf{X})\big)$ no es igual a la combinación lineal $\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}$, pero que está relacionado con este, por medio de una función de acuerdo a la naturaleza de Y. Formalmente el modelo lineal Generalizado consiste en 3 componentes:

1. El "componente aleatorio" Y(variable de respuesta), que tiene distribución de las familias exponenciales con un parámetro canónico η que determina la forma de la respuesta, por ejemplo, Poisson. Nótese que se necesita poder escribir la distribución de la familia exponencial en su forma canónica.
2. El ''componente sistemático'' que especifica que las covariables $\boldsymbol{X}$ sean parte del modelo por la combinación lineal $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta}$ y dado que estamos en la familia exponenciales, ellos definen el parámetro natural η.
3. Una función diferenciable y monótona g, que conecta el componente sistemático con el parámetro $\boldsymbol{\theta}$**.**

$$\mathrm{g}(\boldsymbol{\theta}) = \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta}$$

$$\mathrm{E}(\mathbf{Y}) = \boldsymbol{\theta} = \mathrm{g}^{-1}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta})$$

g es llamada la función de enlace y es la inversa de la función de respuesta. Dado $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta} = \eta$, la función de respuesta es la misma que la función de asignación entre el parámetro natural y el parámetro $\eta(\boldsymbol{\theta})$

3.3. Método de Newton-Raphson

Es un procedimiento numérico; se utiliza para encontrar raíces de una función o ecuaciones por aproximaciones sucesivas usando la tangente. Pensemos en una función f cuya regla de correspondencia es $f(x)$ y queremos hallar una de sus raíces, si existe. Para ello, escogemos un valor x_1, "cercano" a la raíz de la función, y trazamos una recta tangente que incluirá el punto x_2, Calculamos $f(x_2)$, este punto, nos dará un nuevo valor x_3, que es más cercano a la raíz que queremos calcular, *véase figura 2*.

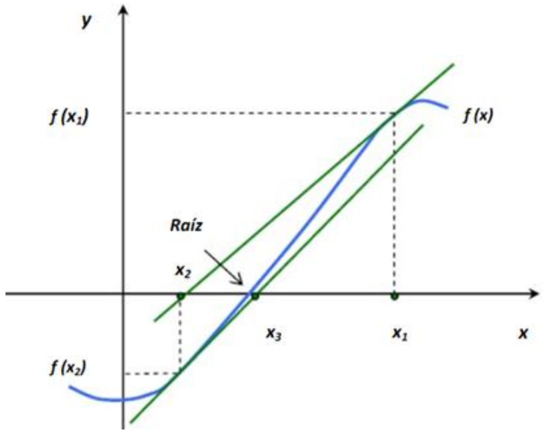

Figura 2 Newton-Raphson

Para encontrar el valor dex_2, se tomará la ecuación punto pendiente $f(x_2) - f(x_1) = m(x_2 - x_1)$.

Para que x_2 sea una raíz de$f$, $f(x_2)$ tendrá que ser igual a 0, para mayor comprensión, reemplazamos $f(x_2)$ por "y"; el enunciado quiere decir, hacemos $y = 0$ para poder hallar x_2:

$$-f(x_1) = m(x_2 - x_1)$$

Ahora tomamos "m" como $f'(x_2)$, al ser la pendiente de la recta tangente a la función en el punto $x_2, f(x_2)$ nos dará una mejor aproximación:

$$-f(x_1) = f'(x_1)(x_2 - x_1) \rightarrow \ x_2 - x_1 = \frac{f(x_1)}{f'(x_1)}$$

Ponemos la ecuación en función de x_2

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

Al generalizar, quedará la ecuación conocida como Newton Raphson:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

3.4. Función de enlace para Regresión Logística y Poisson

En Regresión Logística, considerando el caso en el cual $Y_1, Y_2, \ldots, Y_n$ son Bernoulli (Independientes con Probabilidad de éxito $\theta = \pi$.

$$E(Y_i) = \pi_i = P(Y_i = 1)$$

La relación entre x_i y π es: $\ln\frac{\pi(x)}{1-\pi(x)} = \boldsymbol{X^T\beta}$

$$f_p(\boldsymbol{x}) = p^y(1-p)^{1-y} = (1-p)exp\left(y\ln\left(\frac{\pi_i(x)}{1-\pi_i(x)}\right)\right)$$

Donde $\ln\frac{\pi(x)}{1-\pi(x)}$ es un parámetro "natural" de la familia exponencial y se lo usa como "enlace"

$$g(\pi) = \ln\frac{\pi}{1-\pi}$$

Entonces, $\pi(x) = \frac{e^{X^T\beta}}{1+e^{X^T\beta}}$; $\pi_i \in (0,1)$

En Regresión Poisson la función de enlace es es el logaritmo $g(\mu) = \ln\mu$, que responde a una distribución Poisson.

Si y es Poisson con parámetro μ, $E(y) = var(y) = \mu$ su función de enlace es:

$$g(\mu) = \ln(\mu) = \beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1}x_{p-1}$$

4. Regresión Logística

La regresión logística es un modelo no lineal mediante el cual se puede determinar la relación entre una variable de respuesta Y que es binaria y una o más variables de explicación $X_1, X_2, \ldots, X_k$, que son variables continuas.

Recordando las familias exponenciales, las que permiten que la distribución de Bernoulli sea definida en términos lineales:

$$p(x;\theta) = \theta^x(1-\theta)^{1-x}$$

$$p(x;\theta) = \exp\{x\, log\theta + (1-x)\log(1-x)\}$$

$$p(x;\theta) = \exp\left\{x\log\frac{\theta}{1-\theta} + \log(1-\theta)\right\}$$

Con este resultado y junto con lo que los modelos lineales generalizados definen, tomamos la "función de enlace" de la distribución.

$$g(\theta) = \log\left(\frac{\theta}{1-\theta}\right) = \boldsymbol{X^T\beta}$$

Y se obtiene la función de respuesta al invertir la función de enlace:

$$g^{-1}(\theta) = \frac{1}{1+\exp(\theta)}$$

Reemplazando $\theta = \mathbf{X}^T\boldsymbol{\beta}$, se obtiene la función de respuesta de la regresión logística

$$Y = f(\mathbf{X^T\beta}) = \frac{1}{1+\exp(\mathbf{X^T\beta})}$$

4.1. Estimación de los parámetros

Para la estimación de los coeficientes del modelo y de sus errores estándar se recurre al método de Máxima Verosimilitud, es decir, estimaciones que hagan máxima, la probabilidad de obtener Y proporcionados por los datos de la muestra. Pero no se obtienen expresiones explícitas para los valores de "los betas" incluidos en el modelo y por tanto debe recurrirse a Newton–Raphson.

Para aplicar el método de Máxima Verosimilitud en Regresión Logística se trabaja con que cada observación Y_i de la muestra sigue la distribución de Bernoulli, suponiendo independencia de las n observaciones, donde la densidad de probabilidades conjuntas, dado $\boldsymbol{\beta}$, de y_1, y_2, . . ., y_n está dada por:

$$p\,(y_1, y_2, \ldots, y_n|\boldsymbol{\beta}) = \prod_{i=1}^{n}[\pi_i{}^{y_i}]\,[1-\pi_i]^{1-y_i}$$

Bajo las siguientes condiciones: La variable Y, que es la variable dependiente, al ser n veces observada, condicionado a valores de X. Y la matriz **X** queda como una matriz con n filas y (p+1) columnas, por último un conjunto de coeficientes de regresión **β**, uno para cada variable de explicación, incluida la variable creada para la constante β_0, con 1 columna y (p+1) filas.

Derivando y usando el método de Newton Raphson para buscar la solución numérica del problema, antes encontrando la matriz Hassiana.
La cual se obtiene de la segunda derivada del vector de derivadas parciales de β;

$$\boldsymbol{H}(\boldsymbol{\beta}) = \left\| \frac{\partial^2 \boldsymbol{LL}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} \right\| = -\boldsymbol{X}^{\boldsymbol{T}} \cdot \boldsymbol{W} \cdot \boldsymbol{X}$$

Siendo $\boldsymbol{W}$ una matriz diagonal, que queda de la forma siguiente:

$$\boldsymbol{W} = \begin{bmatrix} \pi_1(1-\pi_1) & 0 & \cdots & 0 \\ 0 & \pi_2(1-\pi_2) & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \pi_n(1-\pi_n) \end{bmatrix}$$

Donde,

$$\pi_i = \frac{1}{1+e^{-\sum_{j=1}^{n+1}\beta_j x_{ij}}}, \qquad i = 1, 2, \dots n$$

Luego de obtener las derivadas de la función de verosimilitud de la ecuación de regresión, se llega a concluir que para las iteraciones se presenta lo siguiente:

$$\widehat{\boldsymbol{\beta}}_t = \widehat{\boldsymbol{\beta}}_{t-1} + (\boldsymbol{X}^T \cdot \boldsymbol{W}_{t-1} \cdot \boldsymbol{X})^{-1} \cdot \boldsymbol{X}^T \cdot (\boldsymbol{Y} - \pi_t)$$

5. Regresión Poisson

La Regresión Poisson es una técnica estadística en lo que se utiliza un modelo no lineal que pertenece a la categoría del análisis de datos de recuento. En estos casos, la variable dependiente toma más de dos valores discretos dígase: 0,1,2,3…, no negativos.

Partimos de una "Función de Enlace" para la Regresión Poisson.

$$\mu = e^{\beta_0+\beta_1 x_1+\cdots+\beta_{p-1}x_{p-1}}$$

Se tiene que y_i es la realización de una variable aleatoria Y_i, que sigue una distribución de Poisson, con parámetros λ_i, que está relacionada con las variables explicativas X,[3].

Así, $\text{Prob}(Y_i = y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$, donde y_i= 0,1,2…, al tiempo queλ_i = exp($\boldsymbol{\beta}^{\boldsymbol{T}}x_i$), y por lo tanto, ln λ_i= $\boldsymbol{\beta}^{\boldsymbol{T}}x_i$

Una característica de este tipo de distribución es:

$$E(y_i|x_i) = \text{Var}(y_i|x_i) = \lambda_i$$

Y sus efectos marginales, al igual que pasaba en el modelo de regresión logística depende de los valores de las variables explicativas, ya que:

$$\frac{\partial E(y_i|x_i)}{\partial x_i} = \lambda_i \boldsymbol{\beta}$$

5.1. Estimación de los parámetros

Para la estimación de parámetros utilizaremos el método de **Máxima Verosimilitud**, con la función de verosimilitud

$$L = \prod_{i=1}^{n} P[Y_i | X = x_i] = \prod_{i=1}^{n} \frac{\lambda_i^{y_i}}{y_i!} e^{-\lambda_i}$$

Y donde luego de tomando logaritmos y sustituyendo por un modelo logarítmico-lineal tenemos:
Donde tomando logaritmos:

$$\ln L = \sum_{i=1}^{n} [y_i(\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}) - \ln y_i! - \lambda_i]$$

Al derivar, se obtiene un sistema de ecuaciones implícitas, el cual no tienen solución explicita, por ello se utiliza el método numérico de Newton Raphson, como ya se explicó anteriormente, se muestra el cálculo directamente.
Como el vector de las primeras derivadas de
$g(\beta^*) = 0$ en el optimo β^*.
Dejando en términos de β^* nos lleva a:

$$\beta^* \cong \beta - [G(\beta^*)] - \frac{1}{2} g(\beta)$$

Para estos se aplicara el método Newton-Raphson.
De derivar el vector de las derivadas parciales de β, en forma matricial se tiene:

$$\boldsymbol{U}(\boldsymbol{\beta}) = \left[\frac{\partial LL(\beta)}{\partial \beta}\right] = \boldsymbol{X}^{\boldsymbol{T}} \cdot (\boldsymbol{Y} - \boldsymbol{\pi})$$

Y al derivar por segunda vez la función de verosimilitud se encuentra la matriz Hessiana:

$$\boldsymbol{H}(\boldsymbol{\beta}) = \left[\frac{\partial^2 LL(\beta)}{\partial\beta\partial\beta'}\right] = -\boldsymbol{X}^{\boldsymbol{T}} \cdot \boldsymbol{W} \cdot \boldsymbol{X}$$

Siendo:

$$\boldsymbol{W} = \begin{bmatrix} \pi_1 & 0 & \cdots & 0 \\ 0 & \pi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pi_n \end{bmatrix}$$

$$\boldsymbol{\pi_i} = \ \mathbf{e}^{\sum_{j=1}^{n+1}\beta_j x_{ij}} \qquad \text{i} = 1, 2, \dots \text{n}$$

Luego de obtener las derivadas de la función de verosimilitud de la ecuación de regresión, se concluye que para la i-esima iteración que:

$$\widehat{\boldsymbol{\beta}}_t = \widehat{\boldsymbol{\beta}}_{t-1} + (\boldsymbol{X}^T \cdot \boldsymbol{W}_{t-1} \cdot \boldsymbol{X})^{-1} \cdot \boldsymbol{X}^T \cdot (\boldsymbol{Y} - \boldsymbol{\pi})$$

14. Referencias

[1] ***Abraham, B. y Ledolter, J.*** (2006), Introduction to Regression Modeling,
Editorial Thomson Book/Cole.
[2] ***Cassella, G y Berger, R.*** (2002), Statistical Inference, Segunda Edición,
Editorial Thomson Book/Cole 2002.
[3] ***Freeman, H.*** (1979), Introducción a la inferencia estadística, Instituto Tecnológico de
Massachusetts, Editorial Trillas México.
[4] ***http://www.monografias.com/***, actualizado al 2005 y consultado a Enero del 2011
[5] ***http://www.scribd.com***. Actualizado el 4 de Diciembre del 2010 y consultado a Diciembre del 2010
[6] ***http://www.mathtools.net***. Actualizado a Marzo del 2010 y consultado a Junio del 2010.
[7]***http://www.maths.lth.se***. STIXBOX, Caja de Herramientas para Matlab, Versión 1.29, 10 de Mayo del 2000, consultado en Junio del 2010.
[8]***http://www.virtual.unal.edu.co/cursos/ciencias/2001091/html/capitulo_7/leccion-07-02.html***,
Universidad Nacional de Colombia. Consultado Febrero del 2011.
[9] http:// **es.wikipedia.org.** Mínimos Cuadrados, Categoría: Optimización | Análisis de Regresión |

Álgebra Lineal. Actualizado al 2010 y consultado en Julio del 2010.
[10] ***Montalvo, D.*** (2000), Tesis de Grado "Análisis estadístico de la producción arrocera en el Ecuador", Escuela Superior Politécnica del Litoral.
[11] ***Moral, I***. Modelos de Regresión: Lineal Simple y Regresión Logística, Capítulo 14.
[12]***Sosa, W***, Introducción a los Modelos de Regresión, Universidad de San Andrés, Argentina.
[13] ***Zurita, G.*** (2010) Probabilidad y Estadística, Fundamentos y Aplicaciones, Segunda Edición, Instituto de Ciencias Matemáticas ESPOL, Guayaquil, Ecuador.
[14] ***Andersen, E (September 1970).***≪Sufficiency and Exponential Families for Discrete Sample Spaces≫.Journal of the American Statistical
Association 65 (331): pp. 1248–1255.
[15]***Pitman, E. (1936).***≪Sufficient statistics and intrinsic accuracy≫.Proc. Camb. phil. Soc. 32: pp. 567–579.
[16]***Darmois, G. (1935).***≪Sur les lois de probabilites a estimationexhaustive≫. C.R. Acad. sci. Paris 200: pp. 1265–1266.
[17] ***Koopman, B (1936).***≪On distribution admitting a sufficient statistic≫.Trans. Amer. math. Soc. 39: pp. 399–409
.