UNIVERSIDAD NACIONAL AGRARIA LA MOLINA
Dpto. de Estadística e Informática

Capítulo II. Inferencia en los Modelos Lineales generalizados

Clase 2. Estimación en los MLG

Plan de aprendizaje

Inicio

- Motivación
- Logros
- Saberes previos

Desarrollo

- Estimación en los modelos lineales generalizados
- Ejercicios resueltos

Cierre

- Ejercicios propuestos
- Tarea

Motivación:

¿Para qué se estima los coeficientes en un MLG cuyo objetivo de predecir los ingresos?

- Predecir el ingreso por cada año de experiencia.
- ¿Cuál de las variables predictoras explica mejor los ingresos?.
- El ingreso es mayor para un hombre o para una mujer.

Logros:

Al término de la sesión, el estudiante estará en capacidad de:

- Comprender y aplicar los métodos de estimación de los coeficientes en los modelos lineales generalizados.
- Resolver ejercicios sobre los métodos de estimación los modelos lineales generalizados aplicando el R.
- Resolver ejercicios propuestos.

Saberes previos:

- ¿Cuáles son los métodos de estimación de parámetros?
- ¿Cuál es la diferencia de una estimación puntual y por intervalo de confianza?
- ¿Cuáles son los parámetros que se estiman en un MLG?

Estimación en los MLG

Los modelos mlg sirven para predecir la variable respuesta en contextos de no normalidad y heterocedasticidad.

1. Introducción
2. Estimación puntual
3. Estimación por intervalo de confianza
4. Ejercicios propuestos
5. Referencias bibliográficas

1. Introducción

El proceso de estimación puntual y por intervalo de confianza en los MLG, se basa en el método de máxima verosimilitud que presenta mejores propiedades de consistencia y de eficiencia asintótica.

El método de máxima verosimilitud, es muy similar al método de mínimos cuadrados ponderados iterativos; utiliza como algoritmo numérico el método de Newton-Rapson conocido también como el método Scoring.

2. Estimación puntual

La estimación se realiza por el método de máxima verosimilitud.
Considerando el vector aleatorio $Y'=(Y_1, Y_2, \ldots, Y_n)$ de variables aleatorias independientes, se desea estimar el vector de parámetros β y con predictor lineal $x'\beta$, el cual está relacionado con la variable respuesta a través de su valor esperado $E(Y)=\mu$ y la función de enlace $g(\mu)= x'\beta$.
Para cada V.A. Y_i la función de log-verosimilitud, con distribución que pertenece a la familia exponencial será:

$$\boxed{l_i = y_i b(\theta_i) + c(\theta_i) + d(y_i)}$$

También se tiene que:

$$E(Y_i) = \mu_i = -\frac{c'(\theta_i)}{b'(\theta_i)} \qquad V(Y_i) = \left[b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)\right] / \left[b'(\theta_i)\right]^3$$

$$g(\mu_i) = \eta_i = x'_i \beta$$

2. Estimación puntual

La función de log-verosimilitud para todos los Y_i , se define:

$$l = \sum_{i=1}^{n} l_i = \sum y_i b(\theta_i) + \sum c(\theta_i) + \sum d(y_i)$$

El estimador de máxima verosimilitud utiliza la función score para estimar los parámetros βj, y usando la regla de la cadena la variable score queda definida por:

$$U_j = \frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{n} \left[\frac{\partial l_i}{\partial \theta_i} x \frac{\partial \theta_i}{\partial \mu_i} x \frac{\partial \mu_i}{\partial \beta_j} \right]$$

Por lo tanto la VA score queda expresada:

$$U_j = \sum_{i=1}^{n} \left[\frac{(y_i - \mu_i)}{V(Y_i)} X_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right]$$

2. Estimación puntual

La matriz variancia-covarianca de la VA score (matriz de información) queda definida por:

$$\Im_{jk} = E(U_j U_k) = \sum_{i=1}^{N} \frac{x_{ij} - x_{ik}}{V(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

Sea **b** el vector de coeficientes estimados de **β**, entonces la ecuación de estimación por el método de score generalizado se expresa por:

$$b^{(m)} = b^{(m-1)} + \left[\Im^{(m-1)} \right]^{-1} U^{(m-1)}$$

Donde $b^{(m)}$ es el vector estimado del vector de parámetros para β en la m-iteración. Para m=1, se define $b^{(0)}$ un vector inicial de coeficientes de regresión.

2. Estimación puntual

En forma matricial se puede escribir:

La matriz de información: $\Im = X'WX$

La llamadas ***ecuaciones normales***: $X'WXb^{(m)} = X'Wz$

Por lo tanto, el vector estimado para la m-ésima iteración queda expresado por:

$$\underline{b}^{(m)} = (X'WX)^{-1} X'Wz$$

Donde:

W es una matriz diagonal (nxn) con elementos: $\boxed{w_{ii} = \frac{1}{V(Y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2}$

z es un vector (nx1) cuyos elementos: $\boxed{z_i = \eta_i + (y_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)}$

$Con: \eta_i = X'\underline{b} = b_0 + b_1 x_{1i} + ... + b_p x_{pi}$

El método por ser iterativo requiere un vector de coeficientes iniciales $b^{(0)}$.

Ejemplo 1. Para una regresión de Poisson.

Determine las expresiones para los estimadores de máxima verosimilitud por el método iterativo ***scoring.*** Presente las respectivas ecuaciones normales.

- **Formulación del MLG:**

Componente aleatorio: $Y_i \approx Poisson(\mu_i), \quad E(Y_i) = V(Y_i) = \mu_i \quad i = 1, 2, \ldots, n$

Componente sistemático: $\eta_i = X'\underline{\beta} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi}$

Función de enlace: $g(\mu_i) = \log(\mu_i) = \eta_i = X'\underline{\beta} \Rightarrow \mu_i = e^{\eta_i} = e^{X'\underline{\beta}}$

- **Hallando los elementos de la matiz W:**

$$Se\ sabe: \quad w_{ii} = \frac{1}{V(Y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \quad y \quad \mu_i = e^{\eta_i} = e^{X'\underline{\beta}}$$

$$Entonces: \quad \frac{\partial \mu_i}{\partial \eta_i} = e^{\eta_i} = e^{X'\underline{\beta}} \quad y \quad V(Y) = \mu_i = e^{X'\underline{\beta}}$$

$$Reemplazando: w_{ii} = \frac{1}{e^{x'\underline{b}}}\left(e^{x'\underline{b}}\right)^2 \Rightarrow \boxed{w_{ii} = e^{x'\underline{b}}}$$

- **Hallando los elementos del vector z:**

$$Se\ sabe: \quad z_i = \eta_i + (y_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right), \quad \mu_i = e^{\eta_i} = e^{X'\underline{\beta}}$$

$$Entonces: \quad \eta_i = \log(\mu_i) \quad \Rightarrow \quad \frac{\partial \eta_i}{\partial \mu_i} = \frac{1}{\mu_i} = \frac{1}{e^{X'\underline{\beta}}}$$

$$Reemplazando: \quad z_i = x'\underline{b} + (y_i - e^{x'\underline{b}})\left(\frac{1}{e^{x'\underline{b}}}\right) \Rightarrow \boxed{z_i = x'\underline{b} + \frac{y_i}{e^{x'\underline{b}}} - 1}$$

- **Las ecuaciones normales resultan:** $X'WXb^{(m)} = X'Wz$

- **El vector de coeficientes estimados:** $\underline{b}^{(m)} = (X'WX)^{-1} X'Wz$

Ejemplo 2. Para una regresión logística.

Determine los estimadores de máxima verosimilitud por el método iterativo ***scoring.*** Presente las respectivas ecuaciones normales.

- **Formulación del MLG:**

Componente aleatorio:

$$Y_i \approx Binomial(n_i, \pi_i), \quad con: \ \mu_i = E(Y_i) = n_i, \pi_i \ y \ V(Y_i) = n_i, \pi_i(1-\pi_i) \quad i = 1,2,\ldots,N$$

Componente sistemático: $\eta_i = X'\underline{\beta} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi}$

Función de enlace: $g(\mu_i) = \log it(\pi_i) = \log(\frac{\pi_i}{1-\pi_i}) = \eta_i \quad Pero: \mu_i = n_i \pi_i \quad \Rightarrow \pi_i = \frac{\mu_i}{n_i}$

$$\eta_i = \log(\frac{\pi_i}{1-\pi_i}) = \log(\frac{\mu_i}{n_i - \mu_i}) = \log(\mu_i) - \log(n_i - \mu_i)$$

- **Hallando los elementos de la matiz W:**

$$Se\ sabe\ que: \quad w_{ii} = \frac{1}{V(Y)_i}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$$

$$Se\ tiene: \quad \frac{\partial \mu_i}{\partial \eta_i} = (n_i \pi_i (1-\pi_i))^2 \quad y \quad V(Y) = n_i \pi_i (1-\pi_i)$$

$$Entonces\ reemplazando: \quad w_{ii} = \frac{1}{n_i \pi_i (1-\pi_i)}(n_i \pi_i (1-\pi_i))^2 \Rightarrow \boxed{w_{ii} = n_i \hat{\pi}_i (1-\hat{\pi}_i)}$$

- **Hallando los elementos del vector z:**

$$Se\ sabe\ que: \quad z_i = \eta_i + (y_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right), \qquad \eta_i = X'\underline{\beta}$$

$$Se\ tiene: \quad \mu_i = n_i\pi_i \quad y \quad \frac{\partial \eta_i}{\partial \mu_i} = \frac{1}{n_i\pi_i(1-\pi_i)}$$

$$Reemplazando: \quad z_i = \eta_i + (y_i - n_i\pi_i)\left(\frac{1}{n_i\pi_i(1-\pi_i)}\right) \Rightarrow \boxed{z_i = \eta_i + \left(\frac{y_i - n_i\hat{\pi}_i}{n_i\hat{\pi}_i(1-\hat{\pi}_i)}\right)}$$

$$Dónde: \quad \eta_i = x'\underline{b} = \log\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) \quad \Rightarrow \hat{\pi}_i = \frac{exp(\eta_i)}{1+exp(\eta_i)} = \frac{\exp(x'\underline{b})}{1+\exp(x'\underline{b})}$$

- **Las ecuaciones normales resultan:** $X'WXb^{(m)} = X'Wz$

- **El vector de coeficientes estimados:** $\underline{b}^{(m)} = (X'WX)^{-1}X'Wz$

Ejemplo 3. Para una regresión múltiple.

Determine los estimadores de máxima verosimilitud por el método iterativo ***scoring.*** Presente las respectivas ecuaciones normales.

- **Formulación del MLG:**

Componente aleatorio:

$$Y_i \sim Normal(\mu_i, \sigma^2), \quad con: \; E(Y_i) = \mu_i, V(Y_i) = \sigma^2 \quad i = 1, 2, \ldots, n$$

Componente sistemático: $\eta_i = X'\underline{\beta} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi}$

Función de enlace: $E(Y_i) = g(\mu_i) = \mu_i = \eta_i \Rightarrow \mu_i = X'\underline{\beta}$

- **Hallando los elementos de la matiz W:**

$$Se\ sabe\ que: \quad w_{ii} = \frac{1}{V(Y)_i}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \quad y \quad \mu_i = \eta_i = X'\underline{\beta}$$

$$Se\ tiene: \quad \frac{\partial \mu_i}{\partial \eta_i} = 1 \quad y \quad V(Y) = \sigma^2$$

$$Entonces\ reemplazando: \quad \boxed{w_{ii} = \frac{1}{\sigma^2}}$$

- **Hallando los elementos del vector z:**

$$Se\ sabe\ que: \quad z_i = \eta_i + (y_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right), \quad \mu_i = \eta_i = X'\underline{\beta} \quad \Rightarrow \mu_i = X'\underline{\beta}$$

$$Se\ tiene: \quad \eta_i = \mu_i \quad \Rightarrow \quad \frac{\partial \eta_i}{\partial \mu_i} = 1$$

$$Entonces\ reemplazando: \quad z_i = X'\underline{\beta} + (y_i - X'\underline{\beta}) \Rightarrow \boxed{\ z_i = y_i\ }$$

- **Las ecuaciones normales resultan:** $X'WXb^{(m)} = X'Wz$

$$Se\ tiene: X'WX = \frac{X'X}{\sigma^2} \quad y \quad X'Wz = \frac{X'y}{\sigma^2} \Rightarrow (X'X)b = X'y$$

$$El\ vector\ coeficientes\ estimado: b^{(m)} = (X'X)^{-1}X'y$$

$$Matriz\ de\ \inf ormación: \quad \Im = {(X'X)}\big/{\sigma^2}$$

$$Dónde: \sigma^2 = \frac{1}{n-p}\left(Y - Xb\right)'(Y + Xb)$$

3. Estimación por intervalo de confianza

Se pueden construir intervalos de confianza del (1-α)100% de nivel de confianza para los parámetros del modelo. Se tiene que la matriz de varianza-covarianza del vector estimado es:

$$Cov(\hat{\underline{\beta}}) = \Im^{-1} = (X'WX)^{-1}$$

Entonces para hallar un intervalo con un nivel de confianza del (1-α)100% para el parámetro j se tiene:

$$IC(\beta_j) = b_j \pm Z_{(1-\frac{\alpha}{2})}\sqrt{Var(b_j)}$$

Donde: $Var(b_j) = \Im_{jj}^{-1}$

Ejercicio 1.

Se tiene un experimento sobre la aplicación de dosis de un funguicida a sembríos de manzano. Se seleccionan 8 plantas de manzano y se recolectaron 15 hojas evaluando en cada hoja el número de hembras adultas de ácaros vivas. En la siguiente tabla se muestra los datos registrados por planta:

Número de hembras vivas	1	3	7	15	9	17	5	13
Dosis de funguicida (grs)	2	4	6	9	7	12	2	10

a. Formule el MLG, ajustando los datos a una regresión Poisson con función de enlace logaritmo.
b. Determine las expresiones de los estimadores de MV por el método iterativo ***scoring.*** Presente las ecuaciones normales.
c. Halle los coeficientes estimados, usando como valores iniciales $b^{(0)}$=[0.8, 0.1] y para obtener las 5 primeras iteraciones.
d. Halle las matrices de las ecuaciones normales, la matriz de información y los coeficientes estimados finales.
e. Halle intervalos del confianza del 95% para los coeficientes de regresión.

Solución:

a. Formule el MLG, ajustando los datos a una regresión Poisson con función de enlace logaritmo.

Variable dependiente: Y=Número de ácaros hembras
Variable predictora: X=Dosis de funguicidas (grs.)

Componente aleatorio: $Y_i \sim Poisson(\mu_i), con\ E[Y] = \mu$

Dónde: μ= Número promedio de ácaros hembras

Componente sistemático (predictor lineal): $\eta_i = X_i^{'} \underline{\beta} = \beta_0 + \beta_1 x_i$

Función de enlace (logaritmo):

$$E(Y) = g(\mu) = Log(\mu_i) = \eta_i \Rightarrow Log(\mu_i) = \beta_0 + \beta_1 x_i$$

$$Se\ tiene: \mu_i = e^{\beta_0 + \beta_1 x_i}$$

b. Determine las expresiones de los estimadores de MV por el método iterativo ***scoring.*** Presente las ecuaciones normales.

- **Hallando los elementos de la matiz W:**

$$Se\ tiene: \eta_i = x_i^{'}\underline{b} = b_0 + b_1 x_i$$

$$Se\ sabe\ que: \quad w_{ii} = e^{x_i^{'}\underline{b}} \Rightarrow \boxed{w_{ii} = e^{b_0 + b_1 x_i}}$$

- **Hallando los elementos del vector z:**

$$Se\ sabe\ que: \quad z_i = x_i^{'}\underline{b} + \frac{y_i}{e^{x_i^{'}\underline{b}}} - 1 \Rightarrow \boxed{z_i = b_0 + b_1 x_i + \frac{y_i}{e^{b_0 + b_1 x_i}} - 1}$$

- **Las ecuaciones normales resultan:** $X'WXb^{(m)} = X'Wz$

$$\Im = X'WX = \begin{bmatrix} \sum_{i=1}^{8} e^{b_0+b_1x_i} & \sum_{i=1}^{8} x_i e^{b_0+b_1x_i} \\ \sum_{i=1}^{8} x_i e^{b_0+b_1x_i} & \sum_{i=1}^{8} x_i^2 e^{b_0+b_1x_i} \end{bmatrix} \quad y \quad X'Wz = \begin{bmatrix} \sum_{i=1}^{8}\left[(b_0 + b_1x_i - 1)e^{b_0+b_1x_i} + y_i\right] \\ \sum_{i=1}^{8}\left[(b_0 + b_1x_i - 1)x_i e^{b_0+b_1x_i} + x_i y_i\right] \end{bmatrix}$$

- **El vector de coeficientes estimados:** $\underline{b}^{(m)} = (X'WX)^{-1} X'Wz$

c. Halle los estimadores use valores iniciales $b^{(0)}=[0.8, 0.1]$ y para obtener las 5 primeras iteraciones.

```
> # Entrada de datos
> y=c(1,3,7,15,9,17,5,13)
> X=cbind(X0=c(1,1,1,1,1,1,1,1), X1=c(2,4,6,9,7,12,2,10))
> n=dim(X)[1]; p=dim(X)[2]
> W=matrix(0,nrow=n,ncol=n)
> #  Valores iniciales
> b=c(0.8, 0.1)
> # Inicio de las iteraciones
> m=5
> for(i in 0:m)
+ {
+ cat("Iteracion = ",i," Coeficientes estimados:   ",b,"\n")
+ #  Estimacion MV con el algoritmo Scoring
+ Xb=b[1]+b[2]*X[,2]
+ wii=exp(Xb); diag(W)=wii
+ z=Xb+y/exp(Xb)-1
+ b=solve(t(X)%*%W%*%X)%*%(t(X)%*%W%*%z)
+ }
```

```
#  Estimacion MV con el algoritmo Scoring
Iteracion =  0  Coeficientes estimados:    0.8 0.1
Iteracion =  1  Coeficientes estimados:    0.6196843 0.2448946
Iteracion =  2  Coeficientes estimados:    0.7418862 0.1965832
Iteracion =  3  Coeficientes estimados:    0.8077255 0.1806436
Iteracion =  4  Coeficientes estimados:    0.8115629 0.1797246
Iteracion =  5  Coeficientes estimados:    0.8115743 0.1797218
```

Pr *imera iteración* : $m = 1$

$$b^{(1)} = ((X'WX)^{(0)})^{-1}(X'Wz)^{(0)} = \begin{bmatrix} 36.20622 & 278.2873 \\ 278.2873 & 2552.8367 \end{bmatrix} \begin{bmatrix} 90.5875 \\ 797.6262 \end{bmatrix} = \begin{bmatrix} 0.619684 \\ 0.244895 \end{bmatrix}$$

En la siguiente tabla se muestra los resultados de las primeras 5 *iteraciones* :

m	0	1	2	3		
$b_0^{(m)}$	0.8	0.619684	0.748862	0.807725	0.811563	0.811574
$b_1^{(m)}$	0.1	0.244895	0.196583	0.180643	0.179725	0.179722

d. Halle las matrices de las ecuaciones normales, la matriz de transición Y los coeficientes estimados finales.

```
> #  Ecuaciones normales
> XWX=t(X)%*%W%*%X; XWX
    X0       X1
X0  70  598.000
X1 598 5805.541
> XWz=t(X)%*%W%*%z; XWz
        [,1]
X0  164.2838
X1 1528.7038
```

$$X'WX = \begin{bmatrix} 70.0 & 598.00 \\ 598.00 & 5805.541 \end{bmatrix} \quad y \quad XWz' = \begin{bmatrix} 164.28 \\ 1528.73 \end{bmatrix}$$

$$Matriz\ de\quad \inf ormación: \quad \Im = X'WX = \begin{bmatrix} 70.0 & 598.00 \\ 598.00 & 5805.541 \end{bmatrix}$$

```
> #  Vector de coeficientes estimados
> b
        [,1]
X0 0.8115743
X1 0.1797218
```

$$\text{Los estimadores de máximo verosimilitud, por el método Scoring:}$$

$$\underline{\hat{\beta}} = \underline{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 0.81157 \\ 0.17972 \end{bmatrix}$$

$$El\ \text{mod}\,elo\ estimado: \quad \hat{y}_i = \log(\hat{\mu}_i) = 0.81157 + 0.17972x_i \Rightarrow \hat{\mu}_i = e^{0.81157+0.17972x_i}$$

e. Halle intervalos del confianza del 95% para los coeficientes de regresión.

```
> #  Intervalos de confianza de los coeficientes
> Cov_B=solve(t(X)%*%W%*%X); Cov_B
            X0          X1
X0  0.11900527 -0.01225814
X1 -0.01225814  0.00143490
> NC=0.95
> for(i in 1:p)  {
+ LI=b[i] - qnorm(1-((1-NC)/2))*sqrt(Cov_B[i,i])
+ LS=b[i] + qnorm(1-((1-NC)/2))*sqrt(Cov_B[i,i])
+ cat("IC Coeficiente :  b",i-1," LI = ",LI, "   LS = ",LS,"\n")
 }
IC Coeficiente :  b 0  LI =  0.1354428    LS =  1.487706
IC Coeficiente :  b 1  LI =  0.1054782    LS =  0.2539654
```

$$Se\ sabe\ que: \quad Cov(\underline{\beta}) = (X'WX)^{-1} = \Im^{-1} = \begin{bmatrix} 0.119005 & -0.012258 \\ -0.012258 & 0.001435 \end{bmatrix}$$

$$Erores\ estándar: S_{\beta_0} = \sqrt{0.119005} = \qquad S_{\beta_1} = \sqrt{0.001435} =$$

$$IC(\beta_0) = 0.81157 \mp 1.96\sqrt{0.119005} = (0.13543, \quad 1.48771)$$

$$IC(\beta_1) = 0.17972 \mp 1.96\sqrt{0.001435} = (0.10547, \quad 0.25397)$$

Ejercicio 2.

Se ha realizado un estudio con la finalidad de establece una relación entre la edad y la mortalidad por cardiopatía isquémica (CI) en pacientes diabéticos. Para el estudio se evalúo a 100 pacientes fallecidos, estableciéndose los rangos de edades y la causa del fallecimiento (CI: Cardiopatía isquémica u otra causa). Los datos se presentan en la siguiente tabla.

Edad	**Causa de muerte**		**n**
	Otra	**CI**	
20 - 29	9	1	10
30 – 34	13	2	15
35 – 39	9	3	12
40 – 44	10	5	15
45 – 49	7	6	13
50 – 54	3	5	8
55 – 59	4	13	17
60 – 69	2	8	10
Total	57	43	100

Ejercicio 2.

a. Determine las expresiones para los estimadores de máxima verosimilitud por el método iterativo ***scoring.*** Presente las respectivas ecuaciones normales.
b. Halle los coeficientes estimados de máxima verosimilitud por el método de Scoring. Use como valores iniciales los $b^{(0)}$=[-3.0, 0.1] para obtener las 5 primeras iteraciones***.***
c. Halle las matrices de las ecuaciones normales y la matriz de información y los coeficientes estimados finales
d. Halle e interprete intervalos del confianza del 95% para los coeficientes de regresión.

Solución:

a. Formule el MLG, ajustando los datos a una regresión logística binaria con función de enlace logit.

Variable dependiente: Y=Número de muertos cardiopatía isquémica
Variable predictora: X=Rango de edades (marca de clases)

Componente aleatorio:

$$Y_i \approx Binomial(n_i, \pi_i), \quad con: \mu_i = E(Y_i) = n_i, \pi_i \; y \; V(Y_i) = n_i, \pi_i(1-\pi_i) \quad i = 1, 2, \ldots, n$$

$$Dónde: \quad \pi_i = P(Muerte\ por\ CI)$$

Componente sistemático (predictor lineal): $\eta_i = X_i^{'}\underline{\beta} = \beta_0 + \beta_1 x_i$

Función de enlace (logaritmo):

$$E(Y_i) = g(\mu_i) = \log it(\pi_i) = \log(\frac{\pi_i}{1-\pi_i}) = \eta_i \quad Pero: \mu_i = n_i\pi_i \quad \Rightarrow \pi_i = \frac{\mu_i}{n_i}$$

$$\boxed{\eta_i = \log(\frac{\pi_i}{1-\pi_i}) = \beta_0 + \beta_1 x_i}$$

b. Determine las expresiones de los estimadores de MV por el método iterativo ***scoring.*** Presente las ecuaciones normales.

- **Hallando los elementos de la matiz W:**

$$Se\ sabe\ que: \quad w_{ii} = n_i\hat{\pi}_i(1-\hat{\pi}_i)$$

$$w_{ii} = \frac{1}{V(Y)_i}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$$

- **Hallando los elementos del vector z:**

$$Se\ sabe\ que: \quad z_i = b_0 + b_1 x_i + \left(\frac{y_i - n_i\hat{\pi}_i}{n_i\hat{\pi}_i(1-\hat{\pi}_i)}\right)$$

$g(\mu_i) = \eta_i$

$\eta_i = \log\left[\frac{\pi}{1-\pi}\right], \quad \pi = \frac{\mu}{n}$

$$Dónde: \quad \eta_i = \log\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) = b_0 + b_1 x_i \quad \Rightarrow \hat{\pi}_i = \frac{exp(b_0 + b_1 x_i)}{1 + exp(b_0 + b_1 x_i)}$$

- **Las ecuaciones normales resultan:** $X'WXb^{(m)} = X'Wz$

- **El vector de coeficientes estimados:** $\underline{b}^{(m)} = (X'WX)^{-1}X'Wz$

c. Halle los estimadores use valores iniciales $b^{(0)}=[-3.0,\ 0.1]$ para obtener las 5 primeras iteraciones.

```
> # Entrada de datos
> y=c(1,2,3,5,6,5,13,8)
> n=c(10,15,12,15,13,8,17,10)
> yy=cbind(y,n-y)
> X=cbind(X0=c(1,1,1,1,1,1,1,1), X1=c(25,32,37,42,47,52,57,65))
> nn=dim(X)[1]; p=dim(X)[2]
> W=matrix(0,nrow=nn,ncol=nn)
> #  Valores iniciales
> b=c(-3.0, 0.1)
> # Inicio de las iteraciones
> m=5
> for(i in 0:m) {
+ cat("Iteracion = ",i," Coeficientes estimados:   ",b,"\n")
+ #  Estimacion MV con el algoritmo Scoring
+ Xb=b[1]+b[2]*X[,2]
+ Pi=exp(Xb)/(1+exp(Xb))
+ wii=n*Pi*(1-Pi); diag(W)=wii
+ z=Xb+(y-n*Pi)/(n*Pi*(1-Pi))
+ b=solve(t(X)%*%W%*%X)%*%(t(X)%*%W%*%z)
+ }
```

c. Halle los estimadores use valores iniciales $b^{(0)}=[-3.0,\ 0.1]$ para obtener las 5 primeras iteraciones.

```
#  Estimacion MV con el algoritmo Scoring
Iteracion =  0  Coeficientes estimados:    -3 0.1
Iteracion =  1  Coeficientes estimados:    -2.181847 0.02296492
Iteracion =  2  Coeficientes estimados:    -4.901556 0.1043407
Iteracion =  3  Coeficientes estimados:    -5.016856 0.1044396
Iteracion =  4  Coeficientes estimados:    -5.029769 0.1047098
Iteracion =  5  Coeficientes estimados:    -5.029787 0.1047102
```

d. Halle los coeficientes estimados y las matrices de las ecuaciones normales, la matriz de información finales.

```
> #  Vector de coeficientes estimados
> b
           [,1]
[1,] -5.0297872
[2,]  0.1047102
```

$$Ecuación\ estimada: \quad \log\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) = -5.0298 + 0.1047x_i \quad \Rightarrow \hat{\pi}_i = \frac{exp(-5.0298+0.1047x_i)}{1+exp(-5.0298+0.1047x_i)}$$

d. Halle los coeficientes estimados y las matrices de las ecuaciones normales, la matriz de información finales.

```
> #  Ecuaciones normales
> XWX=t(X)%*%W%*%X; XWX
          [,1]       [,2]
[1,]  18.15779   834.3667
[2,] 834.36671 40228.6357
> XWz=t(X)%*%W%*%z; XWz
          [,1]
[1,] -3.963087
[2,] 15.662292
```

$$\boxed{z_i = x'\underline{b} + \frac{y_i}{e^{x'\underline{b}}} - 1}$$

$$X'WX = \begin{bmatrix} 18.158 & 834.367 \\ 834.367 & 40228.636 \end{bmatrix} \quad y \quad XWz' = \begin{bmatrix} -3.963 \\ 15.662 \end{bmatrix}$$

$$Matriz\ de \quad \inf ormación: \quad \Im = X'WX = \begin{bmatrix} 18.158 & 834.367 \\ 834.367 & 40228.636 \end{bmatrix}$$

e. Halle intervalos del confianza del 95% para los coeficientes de regresión.

```
> #  Intervalos de confianza de los coeficientes
> Cov_B=solve(t(X)%*%W%*%X); Cov_B
            [,1]          [,2]
[1,]  1.17300985 -0.0243289477
[2,] -0.02432895  0.0005294553
> NC=0.95
> for(i in 1:p) {
+ LI=b[i] - qnorm(1-((1-NC)/2))*sqrt(Cov_B[i,i])
+ LS=b[i] + qnorm(1-((1-NC)/2))*sqrt(Cov_B[i,i])
+ cat("IC Coeficiente :  b",i-1," LI = ",LI, "   LS = ",LS,"\n")
+ }
IC Coeficiente :  b 0  LI =  -7.152538    LS =  -2.907037
IC Coeficiente :  b 1  LI =  0.05961165    LS =  0.1498088
```

$$Se\ sabe\ que: \quad Cov(\underline{\beta}) = (X'WX)^{-1} = \mathfrak{I}^{-1} = \begin{bmatrix} 1.17301 & -0.02433 \\ -0.02433 & 0.00053 \end{bmatrix}$$

$$Errores\ est\acute{a}ndar: S_{\beta_0} = \sqrt{1.17301} \qquad S_{\beta_1} = \sqrt{0.00053}$$

$$IC(\beta_0) = -5.02979 \mp 1.96\sqrt{1.17301} = (-7.1525, \quad -2.9070)$$

$$IC(\beta_1) = 0.10471 \mp 1.96\sqrt{0.00053} = (0.05961, \quad 0.14981)$$

4. Ejercicios propuestos.

Una empresa de especialista en análisis de encuestas con la finalidad de establecer los costos de aplicar una encuesta de opinión en Lima, desea modelar el número de encuestas realizadas por día, asumiendo que es una variable aleatoria con distribución de Poisson, en función del tiempo (horas) de su aplicación. Se extrae una muestra de 10 digitadoras y se evalúa el número de encuestas digitadas y el tiempo (minutos) por día.

Número de encuestas	20	16	15	17	24	27	21	16	22	20
Tiempo (minutos)	10	8	6.5	8	12	12	10	7	9	8

a. Determine las expresiones para los estimadores MV con el método iterativo ***scoring.*** Presente las respectivas ecuaciones normales.
b. Halle los estimadores de MV con el método Scoring. Use como valores iniciales los $b^{(0)}=[0.6, 0.2]$ y con 5 primeras iteraciones.
c. Halle las matrices de las ecuaciones normales y la matriz de información y los coeficientes estimados finales
d. Halle e interprete intervalos del confianza del 95% para los coeficientes de regresión.

5. Referencias bibliográficas.

1. Annette J. Dobson, (2002) An Introducction to Generalized Linear Models. Second Edition, Editorial Chapman & May.
2. Peter K. Dunn and Gordon K. Smyth (2018). Generalized Linear Models with examples in R. Springer Texts in Statistics.
3. McCullagh, Meter – Nelder, J.A, (1989) Generalized Linear Models. Second Edition. Editorial Chapman & Hall.