

# Análisis de medias para datos categóricos \*

Robert J. Flowers \*\*

Luz María Silva López \*\*\*

Universidad Juárez Autónoma de Tabasco, DACB

---

El procedimiento principal para definir modelos lineales para el análisis de una tabla de contingencia es el de Grizzle, Starmer y Koch. Este procedimiento usa estimadores de cuadrados mínimos ponderados para estimar los parámetros del modelo y supone una distribución multinomial para los datos de una tabla de contingencia. En este artículo se considera una alternativa al procedimiento de Grizzle, Starmer, y Koch que usa el método de máxima verosimilitud. Se obtienen los estimadores de máxima verosimilitud usando un algoritmo de Newton y Raphson. Para hacer los cálculos pertinentes, se proporciona un programa hecho en el lenguaje de Matlab.

The Grizzle, Starmer, and Koch procedure is currently the main procedure for analyzing linear models for categorical data. This procedure uses weighted least squares in order to estimate the parameters of the regression model of interest and supposes a multinomial distribution for the data contained in the contingency table. In this article, maximum likelihood estimates are obtained using a Newton Raphson algorithm. A Matlab program is provided for doing the calculations.

*Palabras clave: Estimación por máxima verosimilitud, análisis de medias, análisis de preferencias de jerarquización.*

*Keywords: Maximum likelihood estimation, means analysis, rank choice analysis*

---

## 1. Introducción

En Grizzle, Starmer y Koch (1969) [1], se define un procedimiento de cuadrados mínimos ponderados que permite la definición de modelos lineales de la forma  $Cp = X\beta$  donde  $p$  es el vector de proporciones de una distribución multinomial y  $C$  es una matriz de transformación. Forthofer y Lehnen (1981) describen en [5], éste procedimiento detalladamente y dan muchos ejemplos de como aplicarlo. Flowers [1] y Flowers, López, y Alavez [3] definen un algoritmo iterativo de cuadrados mínimos ponderados para obtener estimadores de máxima verosimilitud para modelos del mismo tipo de los definidos por Grizzle, Starmer y Koch. Izquierdo [7] también considera modelos de ésta forma pero usa estimadores de información discriminante mínima. Los tres procedimientos dan estimadores que son óptimos asintóticamente normales.

Flowers [2] definió un procedimiento de varios algoritmos de un forma análoga a la de Grizzle, Starmer, y Koch [6]), pero usó el supuesto de unas distribuciones de Poisson para los datos y obtuvo los estimadores empleando el método de máxima verosimilitud. El primero de los algoritmos usa una matriz de transformación y permite la evaluación de los modelos de simetría y homogeneidad marginal para una sola po-

---

\*Recibido el 10 junio de 2009 y aceptado el 12 de octubre de 2009

\*\***Dirección postal:** Carr. Cunduacán-Jalpa Km 1, Cunduacán Tabasco, México. A.P. 24 C.P. 86690. Tel. (+52)914 336-0928. **Correo electrónico:** robert.flowers@basicas.ujat.mx

\*\*\***Correo electrónico:** luz.silva@basicas.ujat.mx

blación. Cada una de los modelos del párrafo anterior permiten manejar más de una población. El tercero de los algoritmos considerado por Flowers [2] usa tres matrices de transformaciones y permite la definición de los mismos modelos que se pueden analizar usando los algoritmos del párrafo anterior. El primero de los algoritmos usa un algoritmo iterativo de cuadrados mínimos ponderados, pero los demás algoritmos usan un algoritmo de Newton y Raphson. En éste artículo, consideramos solamente el tercer algoritmo.

## 2. Metodología

En esta sección, definiremos una clase de modelos lineales que pueden ser expresados en la forma

$$C_3 \exp[C_2 \ln(C_1 m)] = X\beta.$$

Para la mayoría de los modelos de interés se obtienen los mismos estimadores bajo los supuestos de una distribución multinomial, una distribución producto-multinomial, o una distribución de Poisson. Aquí, supondremos que  $Y_i$  es una variable que sigue una distribución de Poisson con media  $m_i$ .

Definimos

$$\begin{aligned} y' &= (y_1, y_2, \dots, y_r), \\ m' &= (m_1, m_2, \dots, m_r). \end{aligned}$$

y obtenemos los estimadores del modelo lineal al maximizar el logaritmo de la función de verosimilitud sujeto a la siguiente restricción

$$C_3 \exp[C_2 \ln(C_1 m)] = X\beta.$$

Se pueden obtener los estimadores de máxima verosimilitud al diferenciar el Lagrangiano

$$y' \ln(m) - \iota' m - \tau' (C_3 \exp[C_2 \ln(C_1 m)] - X\beta)$$

con respecto a  $m$ , donde  $\iota$  es un vector  $n \times 1$  de unidades y  $\tau$  es un vector de multiplicadores de Lagrange. Luego, igualando la derivada a cero se obtiene

$$D_m^{-1}(y - m) - C_1' D_{a_1}^{-1} C_2' D_{a_2} C_3' \tau = 0$$

donde  $D_m$  es una matriz diagonal compuesta de los elementos del vector  $m$ ,  $a_1 = C_1 m$ ,  $a_2 = \exp[C_2 \ln(C_1 m)]$ , y  $D_{a_1}$  y  $D_{a_2}$  son matrices diagonales compuestas de los elementos de los vectores  $a_1$  y  $a_2$  respectivamente. Ahora se define

$$H = C_1' D_{a_1}^{-1} C_2' D_{a_2} C_3'$$

lo cual nos da

$$D_m^{-1}(y - m) = H\tau.$$

Después, aplicando una formula obtenida de un desarrollo en serie de Taylor se llega al siguiente resultado

$$C_3 \exp[C_2 \ln(C_1 y)] = C_3 \exp[C_2 \ln(C_1 m)] + C_3 D_{a_2} C_2 D_{a_1}^{-1} C_1 (y - m) + \gamma$$

donde  $\gamma$  representa el término de error en la serie de Taylor. Posteriormente se puede sustituir la definición de  $H$  para obtener

$$C_3 \exp[C_2 \ln(C_1 y)] = C_3 \exp[C_2 \ln(C_1 m)] + H'(y - m) + \gamma$$

Luego, se define

$$u = C_3 \exp[C_2 \ln(C_1 y)] - \gamma$$

y se aplica la restricción

$$C_3 \exp[C_2 \ln(C_1 y)] = X\beta$$

para obtener

$$u = X\beta + H'(y - m).$$

Así, combinando la ecuación anterior con la ecuación

$$D_m^{-1}(y - m) = H\tau$$

se tiene que

$$u = X\beta + H'D_m H\tau.$$

Ahora, se define

$$V = H'D_m H.$$

y entonces

$$u - X\beta = V\tau.$$

Por tanto, si derivamos el Lagrangiano con respecto a  $\beta$  obtenemos  $X'\tau = 0$ . Entonces

$$X'V^{-1}(u - X\beta) = X'\tau = 0.$$

Finalmente despejando  $\beta$  da

$$\beta = (X'V^{-1}X)^{-1}X'V^{-1}u.$$

Y se puede dar el algoritmo como sigue:

$$\begin{aligned} \beta^{(s+1)} &= (X'[V^{(s)}]^{-1}X)^{-1}X'[V^{(s)}]^{-1}u^s \\ m^{(s+1)} &= y - D_m^{(s)}H^{(s)}[V^{(s)}]^{-1}[u^{(s)} - X\beta^{(s+1)}] \\ u^{(s+1)} &= C_3 \exp[C_2 \ln(C_1 m^{(s+1)})] + (H^{(s+1)})'(y - m^{(s+1)}) \end{aligned}$$

cuyas condiciones iniciales son

$$\begin{aligned} m^{(0)} &= y + \frac{1}{2}t \\ u^{(0)} &= C_3 \exp[C_2 \ln(C_1 m^{(0)})]. \end{aligned}$$

### 3. Aplicaciones

#### 3.1 Análisis estadístico de preferencias

Los datos de la tabla 1 son de Flowers y Navarrete [4]. Ésta tabla presenta los resultados de una encuesta donde se solicita la jerarquización de cuatro opciones de

capacitación. Esta encuesta fue hecha por el Sistema de Desarrollo Integral de la Familia a sus asesores comunitarios.

Se desea calcular la media de los rangos para cada opción y determinar si hay una diferencia significativa entre las opciones. En esta encuesta, la opción de mayor importancia recibe un 4 y la de menor importancia un 1. Además, se supone una distribución multinomial para el número de personas que eligieron cada jerarquización.

**Cuadro 1. Opciones de capacitación**

Opción 1	Opción 2	Opción 3	Opción 4	Frecuencia
1	2	3	4	23
1	2	4	3	13
1	3	2	4	25
1	3	4	2	15
1	4	2	3	5
1	4	3	2	11
2	1	3	4	31
2	1	4	3	22
2	3	1	4	36
2	3	4	1	18
2	4	1	3	9
2	4	3	1	3
3	1	2	4	31
3	1	4	2	9
3	2	1	4	32
3	2	4	1	15
3	4	1	2	10
3	4	2	1	7
4	1	2	3	16
4	1	3	2	12
4	2	1	3	18
4	2	3	1	11
4	3	1	2	22
4	3	2	1	6

Para poder definir el vector de proporciones  $p$ , se necesita definir  $C_1$  y  $C_2$  como

$$C_1 = \begin{bmatrix} I_{24} \\ \iota'_{24} \end{bmatrix}$$

y

$$C_2 = [ I_{24} \quad -\iota_{24} ] .$$

Aquí,  $I_n$  representa una matriz identidad de dimensión  $n$  y  $\iota_n$  representa un vector

columna de dimensión  $n$  con todos los elementos iguales a la unidad. Con éstas definiciones de  $C_1$  y  $C_2$  se obtiene  $p = \exp[C_2 \ln(C_1 m)]$  donde  $p$  es un vector de proporciones. Se quiere definir la media de los rangos para cada opción, pero como la suma de cuatro medias deben ser diez, solamente es necesario estimar tres medias. Para comparar las medias de las primeras tres opciones, se debe definir  $C_3$  como

$$C_3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 3 & 3 & 3 & 4 & 4 & 4 & 4 & 4 & 4 \\ 2 & 2 & 3 & 3 & 4 & 4 & 1 & 1 & 3 & 3 & 4 & 4 & 1 & 1 & 2 & 2 & 4 & 4 & 1 & 1 & 2 & 2 & 3 & 3 \\ 3 & 4 & 2 & 4 & 2 & 3 & 3 & 4 & 1 & 4 & 1 & 3 & 2 & 4 & 1 & 4 & 1 & 2 & 2 & 3 & 1 & 3 & 1 & 2 \end{bmatrix}$$

esto da

$$\mu = C_3 p = C_3 \exp[C_2 \ln(C_1 m)] = X\beta$$

donde

$$\mu' = [\mu_1 \quad \mu_2 \quad \mu_3].$$

Se puede probar la hipótesis de que éstas tres medias son iguales al definir la matriz de diseño como

$$X = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Para este modelo  $G^2 = 7.768$  con dos grados de libertad. Ya que este valor es significativo en el nivel  $\alpha = 0.05$ , concluimos que las primeras tres medias no son iguales.

Otra forma en que se puede definir  $C_1$ ,  $C_2$  y  $C_3$  es la siguiente:

$$C_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 3 & 3 & 3 & 4 & 4 & 4 & 4 & 4 & 4 \\ 2 & 2 & 3 & 3 & 4 & 4 & 1 & 1 & 3 & 3 & 4 & 4 & 1 & 1 & 2 & 2 & 4 & 4 & 1 & 1 & 2 & 2 & 3 & 3 \\ 3 & 4 & 2 & 4 & 2 & 3 & 3 & 4 & 1 & 4 & 1 & 3 & 2 & 4 & 1 & 4 & 1 & 2 & 2 & 3 & 1 & 3 & 1 & 2 \\ 1 & 1 \end{bmatrix},$$

$$C_2 = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

y

$$C_3 = I_3$$

donde  $I_3$  es la matriz identidad de  $3 \times 3$ . En ambos casos, los resultados son los mismos; pero es recomendable usar esta última definición ya que son matrices más pequeñas.

Para probar que las primeras dos medias son iguales se define la matriz de diseño como

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

Como se tiene que  $G^2 = 7.475$  con 1 grado de libertad, se concluye que éste modelo no hace un buen ajuste a los datos y por lo tanto,  $\mu_1 \neq \mu_2$ . Luego, se obtiene el siguiente modelo de interés al definir

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Éste tercer modelo tiene  $G^2 = 0.852$  con un grado de libertad y sí hace un buen ajuste a los datos. Así, después de hacer todas las comparaciones, se encuentra que éste es el mejor modelo. A continuación, en la tabla 2, se muestran los estimadores de máxima verosimilitud para las  $m_i$  para el modelo 3.

**Cuadro 2. Valores esperados**

Opción 1	Opción 2	Opción 3	Opción 4	Frecuencia
1	2	3	4	24.21
1	2	4	3	14.06
1	3	2	4	25.64
1	3	4	2	16.22
1	4	2	3	5.13
1	4	3	2	11.58
2	1	3	4	31.80
2	1	4	3	23.16
2	3	1	4	35.12
2	3	4	1	18.95
2	4	1	3	8.78
2	4	3	1	3.08
3	1	2	4	30.24
3	1	4	2	9.23
3	2	1	4	30.47
3	2	4	1	15.39
3	4	1	2	9.52
3	4	2	1	6.83
4	1	2	3	15.24
4	1	3	2	11.70
4	2	1	3	16.74
4	2	3	1	10.73
4	3	1	2	20.46
4	3	2	1	5.71

Para éste modelo  $\widehat{\beta}_1 = 2.226$  y  $\widehat{\beta}_2 = 0.189$ . Entonces  $\mu_1 = \mu_3 = 2.415$ ,  $\mu_2 = 2.226$

y  $\mu_4 = 10 - 2.415 - 2.226 - 2.415 = 2.944$ . Ahora, en la tabla 3 se presentan los valores de  $G^2$  para las comparaciones entre pares de medias. Estos son los mismos resultados obtenidos por Flowers y Navarrete (2007) usando un modelo de la forma  $Cm = \theta$ .

**Cuadro 3. Comparaciones entre pares de medias**

$H_0$	$G^2$
$\mu_1 = \mu_2$	7.475
$\mu_1 = \mu_3$	0.852
$\mu_1 = \mu_4$	29.506
$\mu_2 = \mu_3$	2.665
$\mu_2 = \mu_4$	63.697
$\mu_3 = \mu_4$	34.411

### 3.2 Análisis de medias correlacionadas

Los datos de la tabla 4 son de Flowers, López y Alávez [3]; fueron analizados por Izquierdo [7] usando modelos de simetría y por Flowers y Toledo [8] usando modelos de asociación. Estos datos incluyen jugadores de béisbol con por lo menos 30 veces al bate en los años 2000 y 2001.

**Cuadro 4. Clasificación de los bateadores según sus promedios de bateo**

2000	2001				total
	$\leq .260$	$(.260, .280]$	$(.280, .300]$	$> .300$	
$\leq .260$	35	15	9	7	66
$(.260, .280]$	13	22	11	4	50
$(.280, .300]$	16	19	12	14	61
$> .300$	13	16	7	31	67
Total	77	72	39	56	244

Ya que los niveles de los factores que representan a los factores para el año 2000 y el año 2001 son ordinales, uno puede probar la hipótesis de que hubo un cambio en el promedio de bateo en los dos años. Para hacer esto, se puede definir

$$C_1 = \begin{bmatrix} I_{16} \\ \iota'_{16} \end{bmatrix}$$

y

$$C_2 = [ I_{16} \quad -\iota_{16} ] .$$

Con estas definiciones de  $C_1$  y  $C_2$  se obtiene  $p = \exp[C_2 \ln(C_1 m)]$  donde  $p$  es un vector de proporciones. Aquí se quiere definir la media de los promedios de bateo para los años 2000 y 2001. Así el vector de medias se obtiene al definir  $C_3$  como

$$C_3 = \begin{bmatrix} .25 & .27 & .29 & .31 & .25 & .27 & .29 & .31 & .25 & .27 & .29 & .31 & .25 & .27 & .29 & .31 \\ .25 & .25 & .25 & .25 & .27 & .27 & .27 & .27 & .29 & .29 & .29 & .29 & .31 & .31 & .31 & .31 \end{bmatrix}$$

Esto da

$$\mu = C_3 p = C_3 \exp[C_2 \ln(C_1 m)] = X \beta$$

donde

$$\mu' = [ \mu_1 \quad \mu_2 ] .$$

Se puede probar la hipótesis de que las dos medias son iguales al definir la matriz de diseño como

$$X = \begin{bmatrix} 1 \\ 1 \end{bmatrix} .$$

En la tabla 5, se muestran los estimadores de máxima verosimilitud para las  $m_i$ . Ya que  $G^2 = 7.006$  con un grado de libertad, se concluye que este modelo no hace un buen ajuste a los datos.

**Cuadro 5.** Valores esperados para el modelo 1

2000	2001				total
	≤.260	(.260,.280]	(.280,.300]	>.300	
≤.260	32.00	17.15	12.01	11.23	66
(.260,.280]	11.55	22.00	12.58	5.34	50
(.280,.300]	12.79	16.88	12.00	16.01	61
>.300	9.44	12.79	6.22	31.00	67
Total	77	72	39	56	244

Al igual que en el ejemplo anterior, se puede dar una definición distinta para  $C_1$ ,  $C_2$  y  $C_3$

$$C_1 = \begin{bmatrix} .25 & .27 & .29 & .31 & .25 & .27 & .29 & .31 & .25 & .27 & .29 & .31 & .25 & .27 & .29 & .31 \\ .25 & .25 & .25 & .25 & .27 & .27 & .27 & .27 & .29 & .29 & .29 & .29 & .31 & .31 & .31 & .31 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} ,$$

$$C_2 = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$$

y

$$C_3 = I_2$$

donde  $I_2$  es la matriz identidad de 2x2. Ambas definiciones de  $C_1$ ,  $C_2$  y  $C_3$  dan los mismos resultados, pero debido a que la segunda forma presenta matrices más pequeñas, hace que esta sea una mejor opción que la primera.

### 3.3 Accidentes fatales durante la lluvia

Los datos presentados en la tabla 6 fueron analizados por Flowers [1]. Estos datos registran los accidentes provocados por la carretera mojada en el estado de Texas en los años 1978 y 1979.

**Cuadro 6.** Accidentes fatales durante la lluvia en área urbana y rural

Área	Inter.	Carretera	Número de fatalidades		
			1	2	3+
Urbano	Si	Express	9	0	0
Urbano	Si	Dividido	20	2	0
Urbano	Si	No Dividido	87	5	0
Urbano	No	Express	70	6	1
Urbano	No	Dividido	25	3	2
Urbano	No	No Dividido	232	17	3
Rural	Si	Express	2	0	0
Rural	Si	Dividido	2	0	0
Rural	Si	No Dividido	31	7	0
Rural	No	Express	49	4	0
Rural	No	Dividido	30	5	1
Rural	No	No Dividido	248	49	14

La variable  $y$  se define de las observaciones de la tabla 6 fila por fila. Se quiere definir el número promedio de fatalidades por accidente para cada fila de la tabla 6. Para hacer esto, se puede definir  $C_1$ ,  $C_2$  y  $C_3$  como

$$C_1 = I_{12} \otimes \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

$$C_2 = I_{12} \otimes \begin{bmatrix} 1 & -1 \end{bmatrix}$$

$$C_3 = I_{12}$$

La matriz de diseño es la siguiente:

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & 0 \\ 1 & 1 & -1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 & 0 \\ 1 & -1 & 1 & 0 & 1 \\ 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 0 \\ 1 & -1 & -1 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 \end{bmatrix}$$

En este caso la  $G^2 = 9.785$  con 7 grados de libertad hace un buen ajuste de los datos. En la tabla 7 se muestran los estimadores de máxima verosimilitud para los coeficientes de regresión.

**Cuadro 7.** Estimaciones de los coeficientes de regresión

Variable	Coficiente	Error estándar	Estadística Z
1	1.1379	0.0191	59.5319
2	-0.0554	0.0140	-3.9694
3	-0.0304	0.0124	-2.4534
4	-0.0441	0.0220	-2.0043
5	0.0429	0.0305	1.4066

En la tabla 8 se muestran los estimadores de máxima verosimilitud para los  $m_i$ .

**Cuadro 8.** Valores esperados para el número de accidentes fatales durante la lluvia

Area	Inter.	Carretera	Número de fatalidades		
			1	2	3+
Urbano	Si	Express	8.9646	0	0.0354
Urbano	Si	Dividido	19.9123	2.0877	0
Urbano	Si	No Dividido	87.1072	4.8928	0
Urbano	No	Express	72.2391	4.2266	0.5344
Urbano	No	Dividido	26.5180	2.2898	1.1922
Urbano	No	No Dividido	227.6388	19.9708	4.3903
Rural	Si	Express	1.8813	0	0.1187
Rural	Si	Dividido	1.7942	0	0.2058
Rural	Si	No Dividido	31.7661	6.2339	0
Rural	No	Express	44.5980	7.2831	1.1189
Rural	No	Dividido	28.1473	6.1050	1.7477
Rural	No	No Dividido	254.3449	45.1220	12.5331

Los resultados obtenidos mediante el algoritmo aquí presentado, coinciden con los que obtuvo Flowers (1984).

#### 4. Sumario

En este artículo se presenta una alternativa al procedimiento de Grizzle, Starmer, y Koch para el análisis de modelos lineales de la forma . Dicha alternativa es el método de máxima verosimilitud, el cual supone que cada observación de la tabla sigue una distribución de Poisson, mientras que el procedimiento de Grizzle, Starmer, y Koch supone distribución multinomial. Es importante mencionar que si la suma de los valores estimados por el algoritmo es igual a la suma de los observados para cada población, entonces la distribución condicional dada la suma es multinomial. En el presente artículo, se usa dicho resultado para obtener lo estimadores de los parámetros de una distribución multinomial.

Aquí se obtuvieron los estimadores de máxima verosimilitud, empleando un algoritmo de Newton y Raphson, en el apéndice se proporciona un programa hecho en Matlab, con el cual fueron analizados los ejemplos aquí mostrados. Cabe mencionar que como se pudo ver en algunos ejemplos, es posible tener más de una forma de definir las matrices de transformación, en tal caso es conveniente usar aquellas matrices que simplifiquen el análisis, es decir, usar las matrices más pequeñas.

#### Apéndice

A continuación se presenta el programa realizado en Matlab, con el cual se obtuvieron los resultados mostrados en el presente artículo.

Es importante señalar que para ejecutar el programa, en la ventana de comandos se debe dar la matriz de observaciones "y", la matriz de modelo "X", y las matrices  $C_1$ ,  $C_2$  y  $C_3$ . Posteriormente se corre el programa.

El programa empleado para el análisis de los datos, es el siguiente:

```

1  % Este programa sirve para probar la hipótesis de que las medias son
2  % iguales en un modelo de la forma:  $C3 = \exp[C2 \ln(C1m)] = XB$ . Asimismo
3  % calcula los valores de beta (b), estimadores (m), G cuadrada
4  % ( $G_{cuad}$ ), los grados de libertad (gl), desviación estándar
5  % (SB) y Z.
6
7  n=length(y);
8  Id=eye(n);
9  r=size(X,1);
10 s=size(X,2);
11 iota=ones(n,1);

```

```

12
13 % Las condiciones iniciales son:
14 m=y + (1/2)*iota;
15 u=C3*exp(C2*log(C1*m));
16 Dm=diag(m);
17
18 % Consideremos la siguiente notación:
19 a1=C1*m;
20 a2=exp(C2*log(a1));
21 Da1=diag(a1);
22 Da2=diag(a2);
23
24 H=C1'*inv(Da1)*C2'*Da2*C3'; V=H'*Dm*H;
25
26 % Iteraciones para obtener los valores de beta y m:
27 for i=1:30
28     b=inv(X'*inv(V)*X)*X'*inv(V)*u;
29     m=y-Dm*H*inv(V)*(u-X*b);
30     for l=1:n
31         if abs(m(l))< 0.001
32             m(l)=0;
33         end
34     end
35     Dm=diag(m);
36     a1=C1*m;
37     a2=exp(C2*log(a1));
38     Da1=diag(a1);
39     Da2=diag(a2);
40     H=C1'*inv(Da1)*C2'*Da2*C3';
41     V=H'*Dm*H;
42     u=C3*a2+H'*(y-m);
43 end
44
45 % Resultados finales del vector beta y m:
46 fprintf('El valor de beta es: ')
47 b=[inv(X'*inv(V)*X)]*X'*inv(V)*u
48
49 fprintf('El valor de m es: ')
50 m=y-Dm*H*inv(V)*(u-X*b)
51
52 % Cálculo de la G2:
53 Dy=diag(y);
54 G_cuad=0;
55
56 for j=1:n
57     if y(j)~=0
58         G_cuad=G_cuad-2*y(j)*log(m(j)/y(j));
59     else
60         G_cuad=G_cuad;
61     end

```

```

62 end
63
64 %Resultado de la  $G^2$  y los grados de libertad:
65 fprintf ( 'El valor de G cuadrada es: ')
66 G_cuad=G_cuad
67
68 fprintf ( 'Los grados de libertad de G cuadrada son: ')
69 gl=r-s
70
71 % Cálculo de la varianza de beta:
72 Var_b=diag( inv (X' * inv (V) * X) );
73
74 % Cálculo de la desviación estándar de beta:
75 fprintf ( 'La Desviacion Estandar de beta es: ')
76 SB=sqrt( Var_b)
77
78 % Cálculo de las Z(i):
79 DSB=diag( SB );
80 Z=0;
81
82 for k=1:s
83     if SB(k)~=0
84         Z=inv( DSB) * b;
85     else
86         Z=Z;
87     end
88 end
89
90 % Resultado de Z(i):
91 fprintf ( 'El valor de Z es: ')
92 Z=inv( DSB) * b
93
94 % fin del programa

```

### Referencias

- [1] Flowers, R.J. *Análisis of discrete data using linear regresión*. Universidad y Ciencia, **1**(2), 75-85. (1984).
- [2] Flowers, R.J. *A maximum likelihood procedure for defining regression models for categorical data*. *Estadística*, **39**(132): 37-54. (1987).
- [3] Flowers, R.J., L. López Segovia, y J. Alavez Ramírez. *Pruebas estadística para homogeneidad marginal o simetría*. *Revista de Ciencias Básicas*, **3**(1): 29-44. (2004).
- [4] Flowers, R.J. y M. Navarrete Martínez. *Análisis estadístico de las preferencias de jerarquización de varias opciones*. *Revista de Ciencias Básicas*, **6**(1): 3-9. (2007)
- [5] Forthofer, R.N. y R.G. Lehnen. *Public Program Analysis: A New Categorical Data Approach*. Belmont Learning Publications. (1981).

- [6] Grizzle, J.E., C.F. Starmer, y G.G. Koch. *Análisis of categorical data by linear models*. Biometrics, **25**:489-504. (1969).
- [7] Izquierdo Vallardares, O.M. *Estimadores de Información Discriminante Mínima*. Tesis no publicada. Cunduacán, Tabasco: División Académica de Ciencias Básicas, Universidad Juárez Autónoma de Tabasco. (2006).
- [8] Toledo Cruz, E. y Flowers, R. *Modelos loglineales*. Revista Ciencias Básicas UJAT **5**(1) octubre 2006.