## ESTIMACION DE LOS PARAMETROS DEL MODELO

Para estimar P $(E \mid x)$ es necesario estimar el factor de ponderaciones $\beta$, y la constante $\alpha$ del modelo. Dos métodos de estimación de $\alpha$ y $\beta$ se consideran a continuación: a) el método basado en la función lineal discriminante y b) el método de máxima verosimilitud.

### 1.- Estimación de los Parámetros $\alpha$ y $\beta$ a través de la función lineal discriminante

Si se supone que tanto $h(x \mid E,\theta)$ son funciones de densidad normales multivariadas con la misma matriz de covarianzas $\Sigma$ pero con diferente vector de medias, podemos identificar al escalar $x'\beta$, donde $\beta = \Sigma^{-1} .(\theta_1-\theta_2)$, como la función lineal discriminante. Es claro que un estimador de $\beta$ puede obtenerse usando los estimadores consistentes de $\Sigma$, $\theta_2$ y $\theta_1$. Además dado que

$$\alpha = \frac{-1}{2} \; (\theta_2 + \theta_1)'\beta \; -\ln \left[ \frac{1-P(E)}{P(E)} \right]$$

podemos obtener un estimador de $\alpha$ usando el estimador de $\beta$, los estimadores consistentes de $\theta_1$ y $\theta_2$ y como un estimador de P(E) a la razón del número de casos en la muestra con la característica que se estudia (empleado) entre el tamaño de la muestra.

Los estimadores consistentes de $\Sigma$, $\theta_1$ y $\theta_2$ pueden obtenerse de la siguiente manera. Se forman dos grupos de observaciones. En el primer grupo se incluyen las observaciones de los individuos que poseen la característica bajo estudio (empleado) y en el segundo los restantes. Suponiendo que cada grupo consiste de $n_1$ y $n_2$ observaciones respectivamente y que cada observación consiste de p variables podemos representar las observaciones por grupo y las medias por variables de la siguiente manera:

$$\begin{array}{cccc}
X'_{11}, & X'_{12}, & \ldots, & X'_{1p} \\
X'_{21}, & X'_{22}, & \ldots, & X'_{2p} \\
\vdots & \vdots & & \vdots \\
X'_{n_1 1}, & X'_{n_1 2}, & \ldots, & X'_{n_1 p} \\
\hline
\bar{X}'_1 & \bar{X}'_2 & \ldots & \bar{X}'_p
\end{array}
\qquad\qquad
\begin{array}{cccc}
X^{\circ}_{11}, & X^{\circ}_{12}, & \ldots, & X^{\circ}_{1p} \\
X^{\circ}_{21}, & X^{\circ}_{22}, & \ldots, & X^{\circ}_{2p} \\
\vdots & \vdots & & \vdots \\
X^{\circ}_{n_2 1}, & X^{\circ}_{n_2 2}, & \ldots, & X^{\circ}_{n_2 p} \\
\hline
\bar{X}^{\circ}_1 & \bar{X}^{\circ}_2 & \ldots & \bar{X}^{\circ}_p
\end{array}$$

Dentro de cada grupo se pueden obtener las observaciones en forma de desviaciones con respecto a la media para cada variable y formar matrices de la forma:

$$x' = \begin{bmatrix} x'_{11} & x'_{12} & \ldots & x'_{1p} \\ x'_{21} & x'_{22} & \ldots & x'_{2p} \\ \vdots & \vdots & & \vdots \\ x'_{n_1 1} & x'_{n_1 2} & \ldots & x'_{n_1 p} \end{bmatrix}$$

$$x^{\circ} = \begin{bmatrix} x^{\circ}_{11} & x^{\circ}_{12} & \cdots & x^{\circ}_{1p} \\ x^{\circ}_{21} & x^{\circ}_{22} & \cdots & x^{\circ}_{2p} \\ \vdots & \vdots & & \vdots \\ x^{\circ}_{n_2 1} & x^{\circ}_{n_2 2} & \cdots & x^{\circ}_{n_2 p} \end{bmatrix}$$

donde:

$$x'_{11} = (X'_{11} - \bar{X}_1) \ \ldots \ x'_{n_1 p} = (X'_{n_1 p} - \bar{X}'_p)$$

y

$$x^{\circ}_{11} = (X^{\circ}_{11} - \bar{X}^{\circ}_1) \ldots \ x^{\circ}_{n_2 p} = (X^{\circ}_{n_2 p} - \bar{X}^{\circ}_p)$$

Usando las matrices de las observaciones en forma de desviaciones podemos obtener las matrices de productos cruzados para cada grupo, esto es:

$$S_1^2 = \frac{1}{n_1 - 1} (x^1)'(x^1)$$

$$S_2^2 = \frac{1}{n_2 - 1} (x^{\circ})'(x^{\circ})$$

Finalmente el estimador consistente de $\Sigma$ (matriz de covarianzas) está dado por:

$$\hat{\Sigma} = \frac{1}{n_1 + n_2 - 2} \left[ (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 \right]$$

Los estimadores consistentes de $\theta_1$ y $\theta_2$ están dados por:

$$\theta'_1 = (\bar{X}'_1, \bar{X}'_2, \ldots, \bar{X}'_p)$$

$$\theta'_2 = (\bar{X}^{\circ}_1, \bar{X}^{\circ}_2, \ldots, \bar{X}^{\circ}_p)$$

Definiendo a los vectores:

$$D' = (\bar{X}'_1 - \bar{X}^{\circ}_1, \bar{X}'_2 - \bar{X}^{\circ}_2, \ldots, \bar{X}'_p - \bar{X}^{\circ}_p) \quad y$$

$$v' = (\bar{X}'_1 + \bar{X}^{\circ}_1, \bar{X}'_2 + \bar{X}^{\circ}_2, \ldots, \bar{X}'_p + \bar{X}^{\circ}_p)$$

Los estimadores de $\beta$ y $\alpha$ se expresan:

$$\hat{\beta} = \Sigma^{-1} D \quad y \quad \hat{\alpha} = -\frac{1}{2}\hat{\beta}' v - \ln\left(\frac{n_2}{n_1}\right)$$

Para hacer pruebas de significancia acerca de la diferencia entre el valor estimado de los parámetros con otro valor preestablecido, se puede usar la matriz de covarianzas asintóticas de los estimadores la cual está dada por:

$$V(\hat{\beta}) = \left[\frac{1}{n_0} + \frac{1}{n_1}\right](\Sigma)^{-1}$$

2.- <u>Estimación de los parámetros $\alpha$ y $\beta$ a través del método de Máxima Verosimilitud.</u>

La estimación de los parámetros de la función de distribución de la logística a través del método de máxima verosimilitud, ha sido formalmente establecida en varios artículos. (Ver ejemplo a Nerlove y Press, 1973, y las referencias contenidas ahí).

De manera general, el método de estimación de esta sección puede establecerse como sigue:

Sea $y = (y_1, y_2, \ldots, y_n)$ y $x'_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$

donde $x_{i1} = 1$ para toda i.

La función de verosimilitud para la i-ésima observación está dada por:

$$L(y_i \mid x_i) = \left[P(y_i = 1 \mid x_i)\right]^{y_i} \left[P(y_i = 0 \mid x_i)\right]^{1-y_i}$$

La función de verosimilitud para una muestra aleatoria de n observaciones está dada por:

$$L(y \mid x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} \left[P(y_i = 1 \mid x_i)\right]^{y_i} \left[P(y_i = 0 \mid x_i)\right]^{1-y_i}$$

En el supuesto caso que P siga la función de distribución de la logística, puede demostrarse a través de manipulaciones algebraicas sencillas que la función de verosimilitud de las "n" observaciones está dada por:

$$L(y \mid x_1, x_2, \ldots, x_n) = \frac{\exp\left[\beta' \sum_{i=1}^{n} x_i \, y_i\right]}{\prod_{i=1}^{n} \left[1 + \exp(x'_i \beta)\right]}$$

donde: $\beta' = (\alpha, \beta_1, \ldots, \beta_{p-1})$

Se puede notar que para dados $x_1, x_2, \ldots, x_n$, $t = \sum_{i=1}^{n} x_i y_i$ es suficiente para $\beta$. Esto es una ventaja teórica importante, ya que el estimador de $\beta$ es una función de t y por tanto tendrá menor error cuadrático medio que cualquier otro estimador.

El estimador de máxima verosimilitud de $\beta$ ($\hat{\beta}$), se obtiene derivando el logaritmo natural de la función de verosimilitud con respecto a $\beta$ e igualando el vector de derivadas a cero. Es fácil probar que la función de verosimilitud es cóncava y que por tanto la función tiene un máximo global cuando $\beta = \hat{\beta}$.

El estimador de $\beta$ ($\hat{\beta}$) debe satisfacer las ecuaciones:

$$\sum_{i=1}^{n} \left[ \frac{1}{1+\exp(-x'_i \hat{\beta})} \right] x_i = \sum_{i=1}^{n} x_i y_i$$

Dado que las ecuaciones dadas arriba no son lineales en $\beta$, éstas pueden resolverse por cualquier método iterativo (por ejemplo a través del método de Newton-Raphson).

Actualmente existen programas para computadora muy eficientes que facilitan mucho la solución de las ecuaciones descritas arriba.

Una característica general del método de estimación de máxima vero similitud es que el estimador de $\beta$ ($\hat{\beta}$) se distribuye asintóticamente - normal con media $\beta$ y matriz de covarianzas $\Sigma$.

Si denotamos a $\lambda_i = \dfrac{\exp(-\hat{\beta}'x_i)}{1+\exp(-\hat{\beta}'x_i)}$, (i= 1, 2, ..., n)

a

$$x' = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \text{ (donde } n = n_1 + n_0\text{)}$$

y

$$A = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

entónces $\Sigma$ estará dada por

$$\Sigma = XAX'$$

Esto permite hacer pruebas estadísticas de significancia para determinar si las $\beta$'s difieren significativamente de un valor preestablecido (generalmente cero).

3.- Consideraciones prácticas sobre los métodos de estimación de los parámetros $\alpha$ y $\beta$ descritos anteriormente.

En muchos casos prácticos el supuesto que $h(x \mid \theta)$ es una distribución normal multivariada es muy fuerte. Además, al subdividir a la muestra total en dos grupos es muy poco probable que ambos grupos tengan la misma matriz de covarianzas. Estos dos supuestos limitan mucho el uso de la función lineal discriminante para estimar $P(E \mid x)$. La ventaja más importante del uso de la función lineal discriminante consiste en que es muy facil computarla. Esta ventaja se vuelve más importante cuando las facilidades de computación son limitadas. Las desventajas teóricas del uso de la función lineal discriminante para estimar $P(E \mid x)$ se pueden resumir en que cuando el supuesto de normalidad no se cumple, los estimadores de $\alpha$ y $\beta$ son sesgados. Además, estos estimadores no son función de una estadística suficiente, y por tanto, no tienen el mínimo error cuadrático medio.

El método de máxima verosimilitud tiene todas las ventajas teóricas que ya se mencionaron. Además, el supuesto de normalidad y desigualdad de la matriz de covarianzas, no es relevante en su aplicación.

El principal problema con este método es que es difícil de computarse y cuando las facilidades de computación son limitadas es casi imposible obtener los estimadores de máxima verosimilitud.

Existen estudios empíricos que comparan, a través de métodos ad-hoc, la "bondad del ajuste" que se obtiene al estimar los parámetros de la función de distribución de la logística cuando se usa como método de estimación la función lineal discriminante o el de máxima verosimilitud.

Ejemplos de estos estudios son los de Halperin, Blackwelder y Verter (1971) y Press y Wilson (1979). En el primer estudio los autores encontraron que en general los estimadores obtenidos a través de la función lineal discriminante son sesgados y que el sesgo no disminuye al aumentar el tamaño de la muestra, esto es, los estimadores son inconsistentes. En el caso de la intercepción del modelo lineal transformado ($\alpha$), su estimador no sólo es inconsistente sino que tiende a afectar mucho las estimaciones de $P(E \mid X)$ cuando el número de casos en la muestra con el atributo E dista mucho del 50%. En el segundo estudio Press y Wilson encontraron que el modelo de regresión logística clasifica mejor si los parámetros del modelo se estiman usando el método de máxima verosimilitud en vez del método de la función lineal discriminante.

A pesar de la evidencia presentada, en este estudio se ha usado la función lineal discriminante para estimar $P(E \mid x)$ con la idea de que una vez que esté disponible un programa para obtener los estimadores de máxima verosimilitud, ambos resultados puedan ser comparados.

## APLICACION DEL MODELO A LA INFORMACION SOBRE DESEMPLEO EN EL AREA METROPOLITANA DE MONTERREY.

La información que se utilizó para estimar la probabilidad de que un individuo esté empleado o desempleado en el Area Metropolitana de Monte