

UN METODO DE DISCRIMINACION EN DOS GRUPOS POR MEDIO DE VARIABLES DICOTOMICAS USANDO DESARROLLO BINARIO*

J. HUMBERTO MAYORGA A.

Profesor Asociado Universidad Nacional de Colombia

RESUMEN. Cuando todas las variables elegidas para fines de discriminación sean de carácter dicotómico, puede adoptarse para llevar a cabo la clasificación de unidades estadísticas, un procedimiento de análisis discriminante en dos grupos, basado en el desarrollo binario de enteros. Este artículo presenta los procesos de *codificación, jerarquización y condensación* requeridos para las variables elegidas en el análisis, previos a la aplicación del procedimiento en consideración; igualmente determina su regla de clasificación y presenta un ejemplo ilustrativo de la misma.

1. INTRODUCCION

Dos grupos de unidades estadísticas, denominados grupo "0" y grupo "1", conforman un determinado universo dicotómico. Las probabilidades que una unidad pertenezca al grupo "1", denotada por δ_1 , y que una unidad pertenezca al grupo "0", denotada por δ_0 , son las denominadas probabilidades a priori.

Las variables aleatorias observables X_1, X_2, \dots, X_p , escogidas adecuadamente para los propósitos de discriminación en los dos grupos, pueden disponerse en un arreglo de tipo vectorial $\mathbb{X} = (X_1, X_2, \dots, X_p)$, cuyo vector de realizaciones particulares x , $x \in \mathbb{R}^p$, genera el espacio de las observaciones, representado por \mathcal{H} , y una regla de clasificación \mathcal{R} , induce una partición de \mathcal{H} en dos regiones \mathcal{H}_0 y \mathcal{H}_1 , regla que permite asignar una unidad particular a un grupo determinado, utilizando los valores específicos que constituyen el vector x para dicha unidad.

Al ubicar incorrectamente una unidad, es decir, al asignarla a un grupo al cual no pertenece, se configura una decisión equivocada que se denomina *error de clasificación incorrecta*. La idea de regla óptima está asociada con valores bajos de la probabilidad

*Artículo derivado de la tesis de Maestría en Estadística, Universidad Nacional de Colombia, dirigida por el Dr. Jimmy A. Corzo S. .

de clasificación incorrecta, y su determinación está sujeta al modelo probabilístico que se asuma para el vector \mathbb{X} . Sin embargo, la escogencia inapropiada de dicho modelo, genera otro tipo de error, como puede ser por ejemplo el ignorar la naturaleza discreta que puedan tener algunas variables y proceder tal y como se procede dentro del análisis discriminante con las variables de tipo continuo, particularmente asumiendo el modelo normal multivariado, Goldstein y Dillon (1978).

Diversas reglas de clasificación se han desarrollado para el caso en el cual \mathbb{X} está conformado por componentes de tipo discreto, como lo presenta Hand (1986); igualmente se han propuesto procedimientos frente a situaciones de mezcla de tipos de variables, Krzanowski (1980). Dentro de este marco, cuando la totalidad de las variables elegidas para propósitos de discriminación sean dicotómicas, el procedimiento que expone este artículo es un procedimiento alternativo de análisis discriminante en dos grupos, evaluado por Mayorga (1992) frente a los métodos multinomial completo, vecino más cercano de orden $r=1$ y de independencia de primer orden.

En la primera parte se describe el sentido de *condensación de variables dicotómicas*, como elemento estructural del procedimiento, en la segunda parte se expone el procedimiento de *discriminación basada en desarrollo binario* y finalmente se ilustra su uso por medio de un ejemplo.

2. CONDENSACIÓN DE VARIABLES DICOTOMICAS.

2.1. Definición de una variable condensadora.

Sea $\mathbb{X} = (X_1, X_2, \dots, X_p, Y)$ un vector aleatorio tal que cada componente X_k , $k = 1, 2, \dots, p$, se asume como una variable aleatoria dicotómica con distribución Binomial de parámetros 1 y π_k , siendo $\pi_k = P[X_k = 1]$, la respectiva probabilidad de éxito. El último de los componentes Y , es una variable aleatoria real, con función de densidad marginal $f_Y(y)$, con valor esperado μ_Y y varianza σ_Y^2 .

La j -ésima observación de una muestra de tamaño n de una población $(p + 1)$ -variada representada por \mathbb{X} , se denota por el vector aleatorio

$$\mathbb{X}_j = (X_{1j}, X_{2j}, \dots, X_{pj}, Y_j) \quad j = 1, 2, \dots, n$$

Como el objetivo inmediato se centra en la definición de una variable que *condense* la información contenida por las variables dicotómicas X_1, X_2, \dots, X_p , en una única variable, para efectos metodológicos en su utilización, resulta conveniente precisar para cada una de ellas, cuál evento debe determinarse como éxito (circunstancia que

puede implicar la recodificación de una o más variables); para ello es preciso establecer la siguiente definición:

Definición 1. Las observaciones muestrales, $X_{1j}, X_{2j}, \dots, X_{pj}$, correspondientes al vector aleatorio $\mathbb{X} = (X_1, X_2, \dots, X_p, Y)$ con $X_k \sim \text{Bin}(1, \pi_k)$, $k = 1, 2, \dots, p$, se dice que están codificadas coherentemente por Y , si y sólo si, para todo valor de k ,

$$\hat{\mathbb{E}}[Y|X_k = 1] \geq \hat{\mathbb{E}}[Y|X_k = 0].$$

donde

$$\hat{\mathbb{E}}[Y|X_k = h] = \bar{Y}_{k/h} = \frac{1}{\#\mathbb{H}} \sum_{\mathbb{H}} Y_j, \quad \mathbb{H} = \{j : X_k = h\}, \quad h = 0, 1, \quad \#(\mathbb{H}) = \text{Card}(\mathbb{H}).$$

La definición 1, además de permitir la identificación para cada variable dicotómica del evento que debe considerarse como éxito, codifica las variables X_k de tal manera que la covarianza muestral entre la variable Y y cada variable X_k , $k = 1, 2, \dots, p$, es un valor no negativo. En efecto, denotando por $\hat{\mathbb{E}}[Y]$ y $\hat{\mathbb{E}}[X_k Y]$ los estimadores de $\mathbb{E}[Y]$ y de $\mathbb{E}[X_k Y]$ respectivamente

$$\begin{aligned} \hat{\mathbb{E}}[Y] &= \hat{\mathbb{E}}[\hat{\mathbb{E}}[Y|X_k]] = (1 - \hat{\pi}_k)\hat{\mathbb{E}}[Y|X_k = 0] + \hat{\pi}_k\hat{\mathbb{E}}[Y|X_k = 1] \\ \hat{\mathbb{E}}[X_k Y] &= \hat{\pi}_k\hat{\mathbb{E}}[Y|X_k = 1] \end{aligned}$$

entonces,

$$\begin{aligned} \text{Cov}(X_k, Y) &= \hat{\mathbb{E}}[X_k Y] - \hat{\mathbb{E}}[Y]\hat{\pi}_k \\ &= \hat{\pi}_k\hat{\mathbb{E}}[Y|X_k = 1] - \left\{ (1 - \hat{\pi}_k)\hat{\mathbb{E}}[Y|X_k = 0] + \hat{\pi}_k\hat{\mathbb{E}}[Y|X_k = 1] \right\} \hat{\pi}_k \\ &= \hat{\pi}_k(1 - \hat{\pi}_k) \left\{ \hat{\mathbb{E}}[Y|X_k = 1] - \hat{\mathbb{E}}[Y|X_k = 0] \right\}. \end{aligned}$$

Como $\hat{\pi}_k(1 - \hat{\pi}_k) > 0$ e igualmente $\left\{ \hat{\mathbb{E}}[Y|X_k = 1] - \hat{\mathbb{E}}[Y|X_k = 0] \right\} \geq 0$, por estar codificadas coherentemente por Y las observaciones muestrales $X_{1j}, X_{2j}, \dots, X_{pj}$, por tanto, $\text{Cov}(X_k, Y) \geq 0$; consecuentemente el coeficiente de correlación muestral es no negativo.

Para definir la variable condensadora se parte del hecho que todo entero positivo a , se puede expresar de manera única, usando la base $b = 2, 3, \dots$, como

$$a = a_m b^m + a_{m-1} b^{m-1} + \dots + a_0 b^0$$

siendo a_r entero, $0 \leq a_r \leq b - 1$, $r = 0, 1, 2, \dots, m$, $a_m \neq 0$ y $0 \leq j_0 < \dots < j_m$.

Al utilizar la base $b = 2$, $j_0 = 0$, $j_k = k$, $k = 1, 2, \dots, p$, con $m \leq p - 1$, los coeficientes $a_r = x_{p-r}$ para $r = 0, 1, 2, \dots, m$, $a_r = 0$ para $r = m + 1, m + 2, \dots, p$, el aprestamiento para la determinación de un valor particular de una variable condensadora está dado, y la definición de variable condensadora se establece a continuación.

Definición 2. Si las observaciones muestrales, $X_{1j}, X_{2j}, \dots, X_{pj}$, correspondientes al vector aleatorio $\mathbb{X} = (X_1, X_2, \dots, X_p, Y)$, están codificadas coherentemente por Y , una variable condensadora de las variables dicotómicas X_1, X_2, \dots, X_p , está determinada por la expresión,

$$I_p = \sum_{k=1}^p 2^{p-k} X_k$$

en consecuencia la variable correspondiente a la condensación de la información de la unidad j en la muestra, se denota

$$I_{p_j} = \sum_{k=1}^p 2^{p-k} X_{k_j}$$

La condensación de variables dicotómicas tal como se describe en este artículo, puede considerarse como un proceso reductor del número de variables, y al mismo tiempo como la construcción de una combinación lineal de las variables dicotómicas, la cual produce una variable adicional, combinación que se elige fundamentalmente con base en la ubicación de las variables. Pero lo más relevante de la condensación, reside en la posibilidad de regresar a la información inicial de las variables, partiendo de un valor dado de la variable adicional; esto es, dado un valor particular l de la variable I_p , existe una única organización de dígitos cero y uno, que constituye la representación binaria de l y que corresponde a un vector particular x de \mathbb{X} .

3. DISCRIMINACION EN DOS GRUPOS BASADA EN DESARROLLO BINARIO.

Dependiendo de la tarea para la cual se pretende utilizar una variable condensadora, se establece una condición o criterio de *jerarquización* de las variables dicotómicas, propicio en la realización de dicha tarea, entendiéndose por *jerarquía*, una organización especial de las variables o la asignación de la posición definitiva a cada una de ellas.

Al adjudicar un sitio a una variable y por lo tanto, su participación en el valor de I_p , el subíndice de cada variable dicotómica X_k , $k = 1, 2, \dots, p$, constituye sin duda alguna, el aspecto de mayor relevancia dentro de la estructura de la variable condensadora. En consecuencia es necesario saber cual de las $p!$ posibles permutaciones de las X_k , debería adoptarse como organización definitiva de las variables, para determinar enteramente la expresión de I_p .

Debido a que en este artículo se presenta sólo una tarea particular de la variable condensadora, la discriminación en dos grupos, el criterio de jerarquización adoptado y la descripción de él será una especie de ejemplo de la forma como éste puede escogerse.

3.1 Determinación de la jerarquía para discriminación en dos grupos.

El procedimiento que se desarrolla, es un procedimiento de discriminación en dos grupos, cuando el conjunto de variables X_1, X_2, \dots, X_p , elegidas para el efecto, son variables dicotómicas. La regla de clasificación asociada con el procedimiento, está fundamentada en la variable condensadora I_p . Por otra parte, la variable aleatoria Y , último componente del vector aleatorio \mathbb{X} , cuyas observaciones muestrales, $X_{1j}, X_{2j}, \dots, X_{pj}$, deben estar codificadas coherentemente por Y , desempeña una función especial dentro del contexto de discriminación en dos grupos. Se asume Y como una variable dicotómica, con distribución $Bin(1, \delta_1)$, precisamente para que realice la labor de representación de pertenencia de una unidad a uno de los dos grupos, y se denomina *variable grupo*.

Debido a que la regla de clasificación del procedimiento, está basada en una variable condensadora, se requiere establecer una permutación de las variables X_1, X_2, \dots, X_p , que haga máxima la covarianza muestral entre las variables I_p e Y , puesto que en la medida que exista alta asociación entre estas variables, los valores bajos de I_p estarán más coligados con el valor "0" de la variable grupo, y los valores altos más coligados con el valor "1". Con esto se persigue que la regla sea más eficiente en la ubicación de unidades en el grupo "0" en presencia de valores inferiores de I_p , y en la ubicación de unidades en el grupo "1" para valores superiores de I_p . El criterio por lo tanto puede anticiparse como el de máxima covarianza muestral entre las variables I_p e Y , en razón a que las variables dicotómicas X_k , $k = 1, 2, \dots, p$, Y que también es dicotómica, e I_p como combinación lineal de las X_k , son variables sin unidades, lo cual hace prescindir de la utilización de coeficientes para cuantificar la asociación entre ellas.

Sea la covarianza muestral entre las variables I_p e Y ,

$$Cov(I_p, Y) = \hat{E}[I_p Y] - \hat{E}[I_p] \hat{E}[Y].$$

Una expresión para esta covarianza, en términos de las covarianzas muestrales entre

la variable Y y cada una de las variables es la siguiente

$$\begin{aligned} Cov(I_p, Y) &= \hat{E}[I_p Y] - \hat{E}[I_p] \hat{E}[Y] \\ &= \hat{E} \left\{ Y \sum_{k=1}^p 2^{p-k} X_k \right\} - \left\{ \sum_{k=1}^p 2^{p-k} \hat{\pi}_k \right\} \hat{E}[Y] \\ &= \sum_{k=1}^p 2^{p-k} (\hat{E}[X_k Y] - \hat{\pi}_k \hat{E}[Y]) \\ &= \sum_{k=1}^p 2^{p-k} Cov(X_k, Y) \end{aligned}$$

De aquí que, la covarianza muestral entre las variables I_p e Y , sea vista ahora en función de las covarianzas citadas, notándose que cada una de las p covarianzas está ponderada acorde al subíndice k , es decir de acuerdo a la ubicación de la variable correspondiente, dentro de la organización vectorial inicial.

Pero si a la máxima covarianza se le pondera por el mayor coeficiente 2^{p-1} , a la segunda covarianza mayor se le pondera por el segundo coeficiente en magnitud 2^{p-2} , y así sucesivamente, hasta ponderar por 2^{p-p} a la covarianza menor, el valor de $Cov(I_p, Y)$ es máximo. Denotando por:

$$Cov_{(1)} = \max\{Cov(X_k, Y), k = 1, 2, \dots, p\}$$

y para $k = 2, 3, \dots, p$

$$Cov_{(k)} = \max\{Cov(X_k, Y), k = 1, 2, \dots, p\} - \{Cov_{(1)}, \dots, Cov_{(k-1)}\}$$

por tanto,

$$\begin{aligned} Cov_{(p)} &= \max\{Cov(X_k, Y), k = 1, 2, \dots, p\} - \{Cov_{(1)}, \dots, Cov_{(p-1)}\} \\ &= \min\{Cov(X_k, Y), k = 1, 2, \dots, p\} \end{aligned}$$

ésto es,

$$Cov_{(1)} \geq Cov_{(2)} \geq \dots \geq Cov_{(p)},$$

entonces el valor máximo de $Cov(I_p, Y)$ es:

$$\sum_{k=1}^p 2^{p-k} Cov_{(k)},$$

lo cual se demuestra usando inducción sobre p .

De esta manera, puede garantizarse una ordenación de variables $\mathbb{X}^* = [X_1^*, X_2^*, \dots, X_p^*]$,

que hace máxima la covarianza muestral entre la variable Y y la variable condensadora basada en dicha ordenación. Esta variable condensadora se denota así:

$$I_k^* = \sum_{k=1}^p 2^{p-k} X_k^*$$

donde X_k^* corresponde a aquella variable dicotómica cuya covarianza con la variable Y es $Cov_{(k)Y}$, $k = 1, 2, \dots, p$. Por tanto, se dice que esta jerarquía ha sido obtenida utilizando el criterio de jerarquización de Covarianza máxima. Para otro tipo de análisis, que exija este criterio de jerarquización, la manera de determinar la jerarquía es la misma aún siendo Y no dicotómica.

3.2. Discriminación basada en desarrollo binario.

El procedimiento de discriminación estadística que se presenta en este artículo puede ser denominado como *discriminación basada en desarrollo binario*, precisamente en razón a la procedencia de la variable condensadora.

Para representar la pertenencia de una unidad a uno de los dos grupos que definen el universo dicotomo, se considera a Y como una variable dicotómica, con distribución Binomial $(1, \delta_1)$, donde δ_1 es la probabilidad a priori que una unidad pertenezca al grupo "1".

Como requisitos primordiales se exige, que las observaciones muestrales, $X_{1j}, X_{2j}, \dots, X_{pj}$, correspondientes al vector aleatorio $\mathbb{X} = (X_1, X_2, \dots, X_p, Y)$, estén codificadas coherentemente por Y , y que la variable I_p^* sea la variable condensadora de las p variables dicotómicas de \mathbb{X} , determinada por la permutación jerárquica $X_1^*, X_2^*, \dots, X_p^*$, basada en el criterio de covarianza máxima.

De conformidad con lo expuesto en el párrafo anterior, los valores inferiores de I_p^* deben estar asociados con el grupo "0" y los valores superiores con el grupo "1". Por tanto la regla de clasificación \mathcal{R} , debe inducir una partición del recorrido de I_p^* en dos subconjuntos, a saber:

$$\mathcal{H}_0 = \{0, 1, \dots, s\} \quad \text{y} \quad \mathcal{H}_1 = \{s + 1, s + 2, \dots, 2^p - 1\}$$

La regla óptima, dependerá igualmente del proceso de minimizar la probabilidad total de clasificación incorrecta,

$$T(\mathcal{R}, f) = \mathbb{P}[I_p^* > s | Y = 0] \mathbb{P}[Y = 0] + \mathbb{P}[I_p^* \leq s | Y = 1] \mathbb{P}[Y = 1].$$

Al denominar por $F_{I_p^* | Y=y}(i) = \mathbb{P}[I_p^* \leq i | Y = y]$ la función de distribución condicional de la variable I_p^* dado $Y = y$, la probabilidad total de clasificación incorrecta, tendrá

la expresión:

$$\begin{aligned} T(\mathcal{R}, f) &= \{1 - F_{I_p^*|Y=0}(s)\}\delta_0 + F_{I_p^*|Y=1}(s)\delta_1 \\ &= \{1 - F_{I_p^*|Y=0}(s)\}\{1 - \delta_1\} + \delta_1 F_{I_p^*|Y=1}(s) \\ &= (1 - \delta_1) + \{\delta_1 F_{I_p^*|Y=1}(s) - (1 - \delta_1)F_{I_p^*|Y=0}(s)\} \end{aligned}$$

Evidentemente $T(\mathcal{R}, f)$ es mínima en la medida que $\{\delta_1 F_{I_p^*|Y=1}(s) - (1 - \delta_1)F_{I_p^*|Y=0}(s)\}$ sea mínimo. El punto $s = s^*$ donde $T(\mathcal{R}, f)$ produce el menor valor, determina finalmente la regla de clasificación,

\mathcal{R} : Asignar la unidad con $I_p^* = l$ al grupo "0" si y sólo si $l \leq s^*$.

Para efectos de notación sencilla, se denota la variable aleatoria que representa al número de observaciones en cada grupo y , $y = 0, 1$, que toman el valor específico $I_p^* = l$, como N_{yl} , $l = 0, 1, \dots, 2^p - 1$, y su valor particular como n_{yl} . Igualmente se denota y define la variable $N_y = \sum_{l=0}^{2^p-1} N_{yl}$ la cual corresponde al tamaño de la muestra referente al grupo y , una vez se fije su valor, en n_y , $y = 1, 0$.

De esta manera las estimaciones insesgadas y consistentes en error cuadrático medio de $F_{I_p^*|Y=0}(s)$, $F_{I_p^*|Y=1}(s)$ y δ_1 , son respectivamente,

$$\hat{F}_0(s) = \frac{1}{n_0} \sum_{l=1}^s n_{0l} \quad \hat{F}_1(s) = \frac{1}{n_1} \sum_{l=1}^s n_{1l} \quad \hat{\delta}_1 = \frac{n_1}{n_1 + n_2}$$

Entonces s^* se estima estableciendo numéricamente la diferencia mínima de

$$\{\hat{\delta}_1 \hat{F}_1(s^*) - (1 - \hat{\delta}_1) \hat{F}_0(s^*)\}$$

El cálculo de la tasa de error aparente para la regla de clasificación de la discriminación basada en desarrollo binario, puede llevarse a cabo por medio de la expresión:

$$\{(1 - \hat{\delta}_1) + \{\hat{\delta}_1 \hat{F}_1(s^*) - (1 - \hat{\delta}_1) \hat{F}_0(s^*)\}\}$$

4. EJEMPLO ILUSTRATIVO DE DISCRIMINACION BASADA EN DESARROLLO BINARIO

Con base en la información que suministra la Tabla 1, se ilustra primeramente la aplicación del procedimiento. Adicionalmente se aplican otros métodos con la misma

información suministrada por Joan C. Martin y Celia Lamper del Centro Médico de Duke University y utilizados en el artículo de Martin y Bradley en 1972, artículo citado por Goldstein y Dillon (1978). Esta información hace referencia a las consecuencias seguidas a un trauma hipóxico (daño sufrido por un recién nacido a falta de oxígeno durante su nacimiento).

El APGAR es un puntaje evaluativo del recién nacido, indicativo de su nivel de funcionamiento fisiológico en sus primeros instantes de vida. Para definir el grupo "0" se consideran los neonatos con puntaje APGAR inferior o igual a siete. El grupo "1" está conformado por los niños normales, es decir aquellos cuyo APGAR supera los siete puntos. Las variables analizadas están denotadas por X'_1 , X'_2 , X'_3 , y referidas a tres aspectos a saber:

X'_1 a la raza del recién nacido,

X'_2 a antecedentes específicos en la historia de la madre,

X'_3 a si el niño respiró antes o después de los cinco segundos de haber nacido.

La definición de los valores de las variables no fue posible, puesto que la codificación de ellas no es explícita en el artículo de Martin y Bradley.

Es necesario realizar una recodificación de variables, puesto que si Y indica la variable grupo, las observaciones muestrales correspondientes al vector $\mathbb{X} = (X'_1, X'_2, X'_3, Y)$ no están codificadas coherentemente por la variable grupo, como se requiere. En efecto, las tres tablas 2 x 2, que se presentan conjuntamente en la Tabla 2, y los cálculos que de ellas se derivan, así lo indican.

TABLA 1. Distribución del número de niños de acuerdo a las variables originales y grupos definidos.

X'_1	X'_2	X'_3	Grupo "0" Daño	Grupo "1" Normal
0	0	0	24	31
0	0	1	0	0
0	1	0	48	36
0	1	1	3	0
1	0	0	8	22
1	0	1	0	0
1	1	0	21	24
1	1	1	2	0
TOTAL			106	113

TABLA 2. Distribuciones del número de recién nacidos por grupo y cada variable original.

GRUPO	X'_1		X'_2		X'_3	
	0	1	0	1	0	1
0	75	31	32	74	101	5
1	67	46	53	60	113	0

$$\begin{aligned} \hat{E}[Y|X'_1 = 1] &= 0.5974 & \hat{E}[Y|X'_1 = 0] &= 0.4710 \\ \hat{E}[Y|X'_2 = 1] &= 0.4778 & \hat{E}[Y|X'_2 = 0] &= 0.6235 \\ \hat{E}[Y|X'_3 = 1] &= 0.0000 & \hat{E}[Y|X'_3 = 0] &= 0.5280 \end{aligned}$$

Lo anterior sugiere que hay la necesidad de recodificar las variables X'_2 y X'_3 puesto que únicamente $\hat{E}[Y|X'_1 = 1] > \hat{E}[Y|X'_1 = 0]$. Una vez realizada la codificación, la información original se reorganiza como en la Tabla 3, teniendo en cuenta que $X_1 = X'_1$, $X_2 = 1 - X'_2$, $X_3 = 1 - X'_3$, y la organización inicial $\mathbb{X} = (X_1, X_2, X_3, Y)$, constituye por lo tanto un vector cuyas observaciones muestrales están codificadas coherentemente por la variable grupo.

TABLA 3. Distribución del número de niños de acuerdo a las variables recodificadas y grupos definidos.

X'_1	X'_2	X'_3	Grupo "0" Daño	Grupo "1" Normal
0	0	0	3	0
0	0	1	48	36
0	1	0	0	0
0	1	1	24	31
1	0	0	2	0
1	0	1	21	24
1	1	0	0	0
1	1	1	8	22
TOTAL			106	113

A partir de las covarianzas muestrales entre cada variable dicotómica y la variable grupo, $Cov(X_1, Y) = 0.0286274$, $Cov(X_2, Y) = 0.0417422$, $Cov(X_3, Y) = 0.0117804$, la jerarquía de las variables dicotómicas queda determinada como X_2, X_1, X_3 es decir $X_1^* = X_2$, $X_2^* = X_1$, $X_3^* = X_3$.

De esta manera la Tabla 4 recoge todos los cambios en cuanto a codificación y jerarquía de variables; igualmente registra todos los cálculos pertinentes para definir la regla de clasificación de la discriminación basada en desarrollo binario. Además se reordenan los valores de (X_1^*, X_2^*, X_3^*) de manera que los valores de la variable condensadora I_3^* puedan ser presentados en orden ascendente.

TABLA 4. Resumen de los cálculos parciales para la deducción de la regla de clasificación basada en desarrollo binario.

X_1^*	X_2^*	X_3^*	I_3^*	n_{0l}	GRUPO "0"		n_{1l}	GRUPO "1"		
					$\hat{F}_0(l)$	$(1-\hat{\delta}_1)\hat{F}_0(l)$		$\hat{F}_1(l)$	$\hat{\delta}_1\hat{F}_1(l)$	$\hat{\delta}_1\hat{F}_1(l)-(1-\hat{\delta}_1)\hat{F}_0(l)$
0	0	0	0	3	0.0283	0.0137	0	0.0000	0.0000	-0.0137
0	0	1	1	48	0.4811	0.2329	36	0.3186	0.1644	-0.0685
0	1	0	2	2	0.5000	0.2420	0	0.3186	0.1644	-0.0776
0	1	1	3	21	0.6981	0.3379	24	0.5310	0.2740	-0.0639
1	0	0	4	0	0.6981	0.3379	0	0.5310	0.2740	-0.0639
1	0	1	5	24	0.9245	0.4475	31	0.8053	0.4155	-0.0320
1	1	0	6	0	0.9245	0.4475	0	0.8053	0.4155	-0.0320
1	1	1	7	8	1.0000	0.4840	22	1.0000	0.5160	-0.0320

La tabla resumen muestra que la diferencia mínima de $\{\hat{\delta}_1\hat{F}_1(l) - (1 - \hat{\delta}_1)\hat{F}_0(l)\}$ se da cuando $l = 2$, esto es $s^* = 2$; por todo lo anterior la regla que se establece en definitiva es:

\mathcal{R} : Clasificar a un neonato como normal si él presenta un valor de la variable I_3^* mayor o igual a tres.

Al usar el método multinomial completo, una unidad con información $(0, 0, 0)$, $(0, 0, 1)$, ó $(0, 1, 0)$ es clasificada en el grupo "0", debido a que sus correspondientes frecuencias son tales que $n_{0l} > n_{1l}$, $l = 0, 1, 2$; las demás unidades son clasificadas en el grupo "1". Esta ubicación coincide con la ubicación que realiza la regla propuesta; la diferencia que puede presentar es en aquellos puntos donde el método multinomial debe hacer la asignación aleatoria.

Para aplicar el método del vecino más cercano de orden $r = 1$, a la información del ejemplo ilustrativo que se ha venido utilizando, es necesario previamente determinar las frecuencias pertinentes, las cuales se presentan en la Tabla 5.

TABLA 5. Organización de frecuencias para aplicación de la regla del vecino más cercano de orden $r = 1$.

X'_1	X'_2	X'_3	Grupo "0" Daño	Grupo "1" Normal
0	0	0	53	36
0	0	1	96	91
0	1	0	26	24
0	1	1	79	82
1	0	0	27	31
1	0	1	80	89
1	1	0	10	22
1	1	1	53	77

Al comparar cada par de frecuencias, se concluye para este caso particular que la clasificación del vecino más cercano de orden $r = 1$, coincide completamente con la clasificación basada en el desarrollo binario.

5. EVALUACION DEL PROCEDIMIENTO

Como se anotó, Mayorga (1992) evaluó el procedimiento basado en desarrollo binario, a través de la tasa de error aparente, simulando 50,400 muestras para cotejarlo con el método multinomial completo tanto en su versión de una muestra como en aquella de muestras independientes, con la variación correspondiente al método del vecino más cercano de orden $r = 1$, bajo las dos versiones muestrales, y con el método de independencia de primer orden estimando probabilidades a priori. De dicha evaluación se destaca lo siguiente:

1. En casi toda muestra, los procedimientos multinomial completo en sus dos versiones muestrales y el basado en desarrollo binario, forman un grupo caracterizado por presentar las más bajas tasas de error aparente, ésto es, el grupo de mayor eficiencia. En el otro grupo con mayores tasas de error aparente, se encuentran los demás procedimientos, del vecino más cercano de orden $r = 1$ en las dos versiones muestrales y el método de independencia de primer orden que estima probabilidades a priori. Las tasas de error aparente correspondientes a la discriminación basada en desarrollo binario, tienden a estar más cerca de las tasas del procedimiento multinomial, que a las tasas de los demás procedimientos.

2. En términos generales, los procedimientos se pueden ordenar de menor a mayor tasa de error aparente, así:
 1. Multinomial completo, sin estimar las probabilidades a priori δ_y .
 2. Multinomial completo, estimando las probabilidades a priori δ_y .
 3. Discriminación basada en desarrollo binario.
 4. Independencia de primer orden, estimando probabilidades a priori δ_y .
 5. Vecino más cercano de orden $r = 1$, sin estimar las probabilidades a priori δ_y .
 6. Vecino más cercano de orden $r = 1$, estimando las probabilidades a priori δ_y .
3. El tercer lugar que ocupa el procedimiento propuesto, superando en algunos casos ampliamente a los procedimientos de los tres últimos lugares, y situándose muy cerca del método Multinomial completo en sus dos versiones muestrales, no corresponde a sus ventajas frente a los demás procedimientos, puesto que:
 - La regla de clasificación del método de Discriminación basado en desarrollo binario, no realiza asignación aleatoria de unidades, frente a empates de frecuencias ó ante igualdad en las expresiones correspondientes; ésto es, la regla es no aleatorizada.
 - Una vez establecida la regla de clasificación del procedimiento propuesto, sólo se requiere de la información particular de la unidad a clasificar, debido a que la regla se basa en una expresión que hace explícita las variables y su jerarquía, y no es necesaria la información original.
 - La jerarquía implícita en la expresión para la regla de clasificación, es un elemento adicional que el procedimiento aporta en el análisis, pues además de cumplir la función clasificatoria, revela la importancia de cada variable dicotómica, en el proceso.

BIBLIOGRAFIA

- GOLDSTEIN, Matthew, DILLON, William. (1978), *Discrete Discriminant Analysis*, John Wiley, New York, pp. 185.
- HAND, D. J. (1986), *Discrimination and Classification*. Chichester, John Wiley, pp. 71-153, y preface.
- KRZANOWSKI, W. J. (1980), *Mixture of continuous and categorical variables in discriminant analysis*, Biometrics Journal of the Biometric Society 36 No. 3 (sep 1980), 493-499.
- MAYORGA, J. Humberto (1992), *Discriminación con variables aleatorias dicotómicas a través de desarrollo binario*, Tesis de Maestría en Estadística, Facultad de Ciencias. Universidad Nacional de Colombia.