Revista Ciencias Estratégicas
ISSN: 1794-8347
revista.cienciasestrategicas@upb.edu.co
Universidad Pontificia Bolivariana
Colombia

Pérez Ramírez, Fredy Ocaris; Támara Ayús, Armando Lenín
ANÁLISIS DISCRIMINANTE COMO SELECCIONADOR DE VARIABLES INFLUYENTES EN EL CÁLCULO DE LA PROBABILIDAD DE INCUMPLIMIENTO
Revista Ciencias Estratégicas, vol. 20, núm. 27, enero-junio, 2012, pp. 103-118
Universidad Pontificia Bolivariana
Medellín, Colombia

ISSN 1794 – 8347 | Vol. 20 | No. 27 | Enero-Junio 2012| pp. 103-118
Revista Ciencias Estratégicas | Medellín - Colombia

ANÁLISIS DISCRIMINANTE COMO SELECCIONADOR DE VARIABLES INFLUYENTES EN EL CÁLCULO DE LA PROBABILIDAD DE INCUMPLIMIENTO

DISCRIMINANT ANALYSIS TO SELECT INFLUENTIAL VARIABLES IN THE CALCULATION OF PROBABILITY OF DEFAULT

Recibido: 08/02/2012
Aprobado: 15/04/2012

Fredy Ocaris Pérez Ramírez
Magister Matemáticas Aplicadas, Universidad Eafit. Profesor-Investigador de la Universidad de Medellín. Departamento de Ingeniería Financiera. Medellín – Colombia.
foperez@udem.edu.co

Armando Lenín Támara Ayús
Magister en Finanzas por la Universidad EAFIT. Profesor-Investigador de la Universidad EAFIT. Escuela de Economía y Finanzas, Medellín-Colombia.
atamaraa@eafit.edu.co

ISSN 1794 – 8347 | Vol. 20 | No. 27 | Enero-Junio 2012| pp. 307-322
Revista Ciencias Estratégicas | Medellín - Colombia

ANÁLISIS DISCRIMINANTE COMO SELECCIONADOR DE VARIABLES INFLUYENTES EN EL CÁLCULO DE LA PROBABILIDAD DE INCUMPLIMIENTO

Resumen

En este artículo se presenta el marco teórico del análisis discriminante y cómo este se aplica como seleccionador de variables influyentes en el cálculo de la probabilidad de incumplimiento. Es así como a través del análisis de una cartera comercial perteneciente a una entidad financiera, se desarrolla un modelo que permita calcular la probabilidad de default a cada uno de los clientes, adicional, se encuentran las ecuaciones que nos lleven a pronosticar el incumplimiento para futuros cliente.

Palabras clave

Análisis discriminante
Probabilidad de incumplimiento
Pérdida esperada

Clasificación JEL: C15, C52, C53, G32

DISCRIMINANT ANALYSIS TO SELECT INFLUENTIAL VARIABLES IN THE CALCULATION OF PROBABILITY OF DEFAULT

Abstract

This article presents the theoretical framework of discriminant analysis as it applies as coach of influential variables in the calculation of the probability of default. Thus, through analysis of business portfolio belonging to a financial institution develops a model to calculate the probability of default to each customer, in addition, is the equations that lead to the failure to predict future customer.

Key Words

Discriminant analysis
Probability of default
Expected loss

Marco teórico

Concepto

El Análisis Discriminante es una técnica multivariada que permite describir algebraicamente las relaciones entre dos o más poblaciones de manera tal que las diferencias entre ellas se maximicen o se hagan más evidentes.

El análisis discriminante se realiza con fines predictivos o descriptivos. El predictivo está relacionado con la clasificación, sean estas nuevas observaciones o algunas sobre las cuales no se conoce a qué grupo pertenecen. En cambio, el análisis discriminante descriptivo se interesa más en las variables empleadas para diferenciar los grupos, y lo que se desea es determinar cuáles de esas variables son las que más diferencian a los grupos y cuáles son importantes (Pamela, 2009).

Clasificación de discriminación

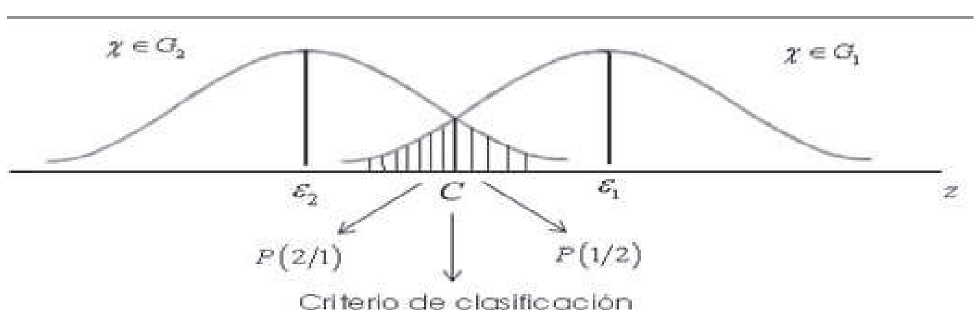
Dado un individuo ω, cuya población de procedencia se desconoce, y sobre el cual se pueden medir variables $X_1.........X_g$, es decir, $\chi=(x_1....x_g)'$, donde $x_1 = X_1(\omega)$ para todo i;$1,2.....,g$, el problema consiste en clasificar este individuo en una de k poblaciones.

Clasificación en una de dos poblaciones normales multivariables con parámetros conocidos

Una manera conceptualmente simple e intuitiva de resolver el problema es abordarlo con el siguiente criterio:

Seleccione α tal que la separación entre ς_1 y ς_2 relativo a σ_z^2. $\Delta^2=\frac{(\varsigma_1-\varsigma_2)^2}{\sigma_z^2}$.
Es decir, hallar α que maximice Δ^2
Además se quiere encontrar los α_i que sean solución al conjunto de ecuaciones $\alpha'\Sigma=(\mu_1-\mu_2)$
Una vez se determine α , la clasificación de un individuo se hará teniendo el score *Z* discrimínate.

Gráfico 1. *Criterio de clasificación bajo normalidad*



*Fuente***.** Elaboración propia.

Si $\chi \in G_2$ pero $z = \alpha' x \geq c$ $\chi \rightarrow G_1$

P(1/2): Probabilidad de la anterior clasificación si $\chi \in G_1$ pero $z < G_2$

Un criterio intuitivo consiste en minimizar *P(1/2) + P(2/1)*, tomando a $C = \frac{\varsigma_1 + \varsigma_2}{2}$

Solución basada en el teorema de Bayes

Sean G_1 y G_2 dos poblaciones donde se tiene definida una variable aleatoria vectorial, χ , p-variante. Suponiendo que X es absolutamente continua y que las funciones de densidad de ambas poblaciones, f_1 y f_2, son conocidas, se estudiará el problema de clasificar un nuevo elemento, X_0 , con valores conocidos de las p variables en una de estas poblaciones. Si se conocen las probabilidades a priori π_1, π_2 con $\pi_1 + \pi_2 = 1$, de que el elemento venga de cada una de las dos poblaciones, su distribución de probabilidad será una distribución mezclada.

$$f(x) = \pi_1 f_1(x) + \pi_2 f_2(x)$$

Y una vez observado χ se pueden calcular las probabilidades a posteriori de que el elemento haya sido generado por cada una de las dos poblaciones, $P(G_i / X)$, con i:1,2. Estas probabilidades se calculan por el teorema de Bayes:

$$P(G_i/X) = \frac{\pi_i P(X/G_i)}{\pi_1 P(X/G_1) + \pi_2 P(X/G_2)}$$, para todo i:1,2 (Probabilidad posteriori)

Cuando $X \sim N_o(\mu_1, \Sigma)$ o $X \sim N_d(\mu_2, \Sigma)$ se usa las densidades $f(X/G_1)$ y $f(X/G_2)$ y, por lo tanto $P(X/G_i)$ puede reexpresarse de la siguiente forma:

$$P(G_i/X) = \frac{\pi_i f_i(x)}{\pi_1 f_1(x) + \pi_2 f_2(x)}$$

Donde:

π_i: Probabilidad a priori de que un individuo provenga de G_i, $i:1,2$,

$P(X/G_i)$ Para todo $i:1,2$: Probabilidad de observar χ en la población G_i.

$P(G_i/X)$ Para todo $i:1,2$: Probabilidad de que la población sea G_i dado que se observa χ.

La aplicación del procedimiento de Bayes lleva al siguiente criterio:

Se clasificará χ en G_2 sí $\pi_1 P(G_1/X) < \pi_2 P(G_2/X)$ y si las probabilidades a priori son iguales, la condición de clasificar χ en G_2 se deduce a $P(G_1/X) < P(G_2/X)$.

Por el contrario se clasificará χ en G_1 sí $\pi_1 P(G_1/X) \geq \pi_2 P(G_2/X)$.

Por otro lado, cuando se tiene poblaciones normales el criterio de clasificación es el siguiente:

Clasificar χ en G_1 si $\frac{\pi_1 f_1(x)}{\pi_2 f_2(x)} \geq 1$ (1) (1)

Clasificar χ en G_2 si $\frac{\pi_1 f_1(x)}{\pi_2 f_2(x)} < 1$

Se puede probar que el anterior criterio de clasificación minimiza la probabilidad de clasificación errónea $\pi_1 P(2/1) + \pi_2 P(1/2)$ (Probabilidad incondicional clasificación errónea).

Incluir costos de clasificación errónea

Al momento de clasificar se pueden cometer errores, los cuales tienen distintas consecuencias que se pueden cuantificar. Por ejemplo, si un cajero automático clasifica equivocadamente un billete de 10.000 COP como de 20.000 COP, y devuelve el cambio equivocado, el costo de clasificación es de 10.000 COP.

Si las consecuencias de un error de clasificación pueden cuantificarse, pueden ser incluidas en la solución del problema formulándolo como un problema bayesiano de decisión.

Suponga que: el error de clasificación se puede cometer cuando un individuo de la población G_1 es clasificado como perteneciente a la población G_2, o un individuo de la población G_2 es clasificado como perteneciente a la población G_1.

Además:
$C(2/1)$: Costo de clasificarlo en G_2 dado que se dio G_1
$C(1/2)$: Costo de clasificarlo en G_1 dado que se dio G_2

El analista siempre quiere maximizar su función de utilidad, lo cual equivale a minimizar el costo esperado.

El costo de clasificar χ en G_2 está dado por:
$C(2/1)P(G_1/\chi)$
Y el costo de clasificar χ en G_1 estará dado por:
$C(1/2)P(G_2/\chi)$
Por lo tanto se asignará χ a G_2 si su costo esperado es menor, es decir, si:

$$\frac{\pi_1 f(G_1/\chi)}{C(1/2)} < \frac{\pi_2 f(G_2/\chi)}{C(2/1)}$$

Con lo anterior, el procedimiento de clasificación bayesiana generalizado para asignar χ a G_1 o G_2, será:

$$z = \sum_{i=1}^{d} \alpha_i \chi_i \geq \frac{\varsigma_1 + \varsigma_2}{2} + \ln\left(\frac{\pi_2 C(1/2)}{\pi_1 C(2/1)}\right)$$

La anterior regla minimiza el costo esperado de la clasificación errónea $\pi_1 C(2/1) + \pi_2 C(1/2) P(1/2)$

Análisis de clasificación y discriminación

En el análisis de clasificación y discriminación, surgen dos problemas:
1) $\chi_0 \to G_i \; i:1,2,3....,g$
2) $\chi_{1i}, \chi_{2i}, ..., \chi_{ni} \in G_i \; i:1,2,3....,g$

Construir una regla que separe los grupos

$f_i(\chi)$: f.d.p condicional de χ dado que $\chi \in G_i$
π_i: Proporción de individuos en la población que pertenecen a G_i
La probabilidad incondicional de χ

$$f_i(\chi) = \sum_{i=1}^{g} \pi_i f_i(\chi)$$

Teorema de Bayes

$$P(\chi_0 \in G_j/\chi = \chi_0) = \frac{f_j(\chi_0)\pi_j}{f(\chi_0)} = \frac{f_j(\chi_0)\pi_j}{\sum_{i=1}^{g} \pi_i f_i(\chi_0)} = g_j(\chi_0)$$

Donde,
$g_j(\chi_0)$ Probabilidad a posteriori de $\chi_0 \in G_j$ dado que $\chi = \chi_0$

La regla de clasificación es asignar χ_0 al grupo G_j si
$g_j(\chi_0) = \max_i q_i(\chi_0)$

Probabilidad total de clasificación errónea (PTM)

Otra forma de clasificación es a través de la PTM, en donde el objetivo primordial es minimizar PTM.

Sea $P(j/i)$ Probabilidad de clasificar un sujeto de G_i en G_j y PTM está dado por $\sum_{i=1}^{g}\left\{\pi_i \sum_{j \neq i} P(j/i)\right\}$, el cual se obtiene analíticamente de la tabla 1.

Tabla Nº 1. *Probabilidad total de clasificación errónea*

	Pertenencia					
Clasificación		1	2	3		g
	1	$P(1/1)$	$P(1/2)$	$P(1/3)$	....	$P(1/g)$
	2	$P(2/1)$	$P(2/2)$	$P(2/3)$	....	
	3	$P(3/1)$	$P(3/2)$	$P(3/3)$	....	
	⋮	⋮	⋮	⋮		⋮
	g	$P(g/1)$	$P(g/21)$	$P(g/3)$	....	$P(g/g)$
		π_1	π_2	π_3		π_g

Fuente. Elaboración propia

Variables normales

Los anteriores procedimientos se realizaron bajo el supuesto que las funciones de densidad o probabilidad $f(\chi/G_j)$ son conocidas, pero en general se tiene que estimar y el modelo más frecuentemente utilizado para estimar $f(\chi/G_j)$ las funciones es el normal multivariante.

Distancia De Mahalanobis (Villardón)

La distancia de Mahalanobis (al cuadrado) entre dos individuos con vectores de observaciones χ y z, es

$$d_m^2 = d_m^2(\chi - z) = (\chi - z)' \sum^{-1} (\chi - z)$$

La distancia de Mahalanobis de un individuo al grupo i es la distancia al centroide del grupo

$$d_m^2 = d_m^2(\chi_0, \mu_j) = (\chi_0 - \mu_j)' \sum^{-1} (\chi_0 - \mu_j)$$

Y la distancia entre dos grupos es la distancia entre sus centroides

$$d_m^2 = d_m^2(\chi_i, \mu_j) = (\chi_i - \mu_j)' \sum^{-1} (\chi_i - \mu_j)$$

Propiedades

- La distancia de Mahalanobis tiene en cuenta las correlaciones entre las variables utilizando sólo la información de cada variable no redundante.
- Es Invariante por transformaciones lineales no singulares, en particular por cambios de escala.

Clasificación de poblaciones normales con matriz de covarianzas Σ común

Si χ pertenece al grupo G_i y a la vez χ sigue una distribución normal con media μ_i y varianza Σ, se puede utilizar

la siguiente regla de decisión para clasificar χ al grupo G_i:
Asigne χ_0 al grupo j que maximice la siguiente ecuación:

$$\log(\pi_j f_j(\chi_0)) = \log(\pi_j) - \frac{d}{2}\log(2\pi) - \frac{1}{2}\log(\Sigma) - \frac{1}{2}(\chi_0 - \mu_j)' \sum^{-1}(\chi_0 - \mu_j)$$

Debe asignarse χ_0 al grupo tal que minimice

$$D_j(\chi_0) = (\chi_0 - \mu_j)' \sum^{-1}(\chi_0 - \mu_j) - 2\log(\pi_j),$$

Donde:

$(\chi_0 - \mu_j)' \sum^{-1}(\chi_0 - \mu_j)$ (Distancia de Mahalanobis)

$D_j(x_0)$ Regla lineal discriminante

En la práctica se trabaja con $\overline{\chi}_0$ en lugar de μ_j

$$\varsigma_p = \frac{(n_1 - 1)\varsigma_1 + (n_2 - 1)\varsigma_2 + \dots + (n_g - 1)\varsigma_g}{n_1 + n_2 + \dots + n_g}$$

- Regla lineal con 2 grupos

Si q=2, asigne χ_0 a G_1 si

$$(\chi_0 - \overline{\chi}_1)' \varsigma_p (\chi_0 - \overline{\chi}_1) - 2\log(\pi_1) < (\chi_0 - \overline{\chi}_2)' \varsigma_p (\chi_0 - \overline{\chi}_2) - 2\log(\pi_2)$$

Así

$$\hat{\lambda}'\left[\chi_0 - \frac{1}{2}(\overline{\chi}_1 + \overline{\chi}_2)\right] > \log\left(\frac{\pi_1}{\pi_2}\right)$$

$$\hat{\lambda}' = \varsigma_p^{-1}\ (\overline{\chi}_1 + \overline{\chi}_2)$$

Clasificación de poblaciones normales con matriz de covarianzas diferentes, $\Sigma_1 \neq \Sigma_2$

La regla óptima de clasificación es:
Para datos normales multivariables con Σ_i´s diferentes asigne χ_0 al grupo j que maximice

$$\log(\pi_j f_j(\chi_0)) = \log(\pi_j) - \frac{d}{2}\log(2\pi) - \frac{1}{2}\log(\Sigma_j) - \frac{1}{2}(\chi_0 - \mu_j)' \sum_j^{-1}(\chi_0 - \mu_j)$$

Clasifique χ_0 en G_i tal que minimice Regla de discriminación cuadrática dada por:

$$D_j(\chi_0) = (\chi_0 - \mu_j)' \sum_j^{-1}(\chi_0 - \mu_j) - 2\log(\pi_j) - \log(\Sigma_j)$$

En la práctica se usa $\overline{\chi}_j$ en lugar de μ_j y δ_j en lugar de Σ_j.

La regla lineal de fisher (bárcena, 2008)

Fisher propuso en 1936 un procedimiento de discriminación lineal que coincide con la regla derivada para dos poblaciones normales con matriz de covarianzas común. En la aproximación de Fisher, la normalidad no es un supuesto.

Dos grupos con matriz de covarianzas Σ común

El razonamiento es el siguiente: se busca una función lineal $\alpha' x$ que separe óptimamente dos grupos. Ello requiere que a $\alpha' x$ tome valores altos en promedio para valores en un grupo, y bajos en otro. Una manera de solucionar esto es buscar maximizar

$$[\alpha'\mu_1 - \alpha'\mu_2]^2 = [\alpha'(\mu_1 - \mu_2)]^2 \quad (1)$$

El cuadrado tiene por objeto eliminar el signo, pues lo que importa la diferencia de a $\alpha' x$ evaluada en μ_1 y μ_2, y no su signo.

Para obtener una única solución se fija la escala de α o fijar $\|\alpha\|^2 = 1$, pero es más práctico hacer $\alpha'\Sigma\alpha = 1$; o, alternativamente, resolver

$$\max_{\alpha}\left(\frac{[\alpha'(\mu_1 - \mu_2)]^2}{\alpha'\Sigma\alpha}\right) (2)$$

Que es de nuevo un problema indeterminado hasta un factor de escala, y normalizar una solución cualquiera de modo que a $\alpha'\Sigma\alpha = 1$.

Derivando (2) respecto de α e igualando el numerador a cero, se obtiene

$$2(\mu_1-\mu_2)\alpha'[\mu_1-\mu_2](\alpha'\Sigma\alpha)-2[\alpha'(\mu_1-\mu_2)]^2\Sigma\alpha=0 \quad (3)$$

Sin tener en cuenta las constantes, se observa que (3) proporciona

$$\Sigma\alpha \propto (\mu_1-\mu_2) \Rightarrow \alpha\Sigma^{-1} \propto (\mu_1-\mu_2)$$

Obsérvese en (2) que el denominador es la varianza de $\alpha'x$. El numerador es el cuadrado de la diferencia entre los valores que toma $\alpha'x$ en μ_1 y μ_2, Por lo tanto, lo que se maximiza es la razón de la diferencia al cuadrado de valores de $\alpha'x$ en términos de su propia varianza, $\mathrm{var}(\alpha'x)$. Se puede interpretar la ecuación (2) como una relación señal/ruido: el numerador es la "señal" y el denominador el "ruido." Por lo tanto se busca una función $\alpha'x$ que maximice la relación señal/ruido (Bárcena, 2008).

Más de dos grupos con matriz de covarianzas Σ común

Si hay K grupos, hay en general no una sino hasta $K-1$ variables discriminantes, combinaciones lineales de las X originales.

Sean pues K grupos, y considerando una muestra con n_i casos $(i=1,....,K)$ en cada grupo. El tamaño total P de la muestra estará dado por $n=\sum_{i=1}^{K} n_i$. Siendo $X_{i(j)}$ la observación i-ésima en el grupo j-ésimo, se define lo siguiente:

$$\bar{X}=n^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n_i}X_{i(j)}$$

$$\bar{X}_i=n^{-1}\sum_{j=1}^{n_i}X_{i(j)}$$

$$T=\sum_{i=1}^{K}\sum_{j=1}^{n_i}(X_{i(j)}-\bar{X})(X_{i(j)}-\bar{X})'$$

$$W_i=\sum_{j=1}^{n_i}(X_{i(j)}-\bar{X})(X_{i(j)}-\bar{X})'$$

$$W=W_1+\ldots+W_K$$

$$B=T-W$$

De lo anterior se puede demostrar que $B=\sum_{i=1}^{K}n_i(\bar{X}_i-\bar{X})(\bar{X}_i-\bar{X})'$ y $\bar{X}=n^{-1}\sum_{i=1}^{K}n_i\bar{X}_i$ y Un razonamiento similar al empleado al obtener el discriminante lineal en el caso de dos grupos, sugeriría ahora maximizar

$$\frac{\sum_{i=1}^{K}\left[\alpha'\sqrt{n_i}(\bar{X}_i-\bar{X})\right]^2}{\sum_{i=1}^{K}\left[\alpha'\sum_{j=1}^{n_i}(\bar{X}_{i(j)}-\bar{X}_i)\right]^2}=\frac{\alpha'B\alpha}{\alpha'W\alpha}=\lambda$$

Derivando respecto a α se obtiene la igualdad matricial $(B-\lambda W)\alpha=0$

Bajo el supuesto de que W tiene inversa, la igualdad anterior es equivalente a $(W^{-1}B-\lambda I)\alpha=0$

Esta tiene solución no trivial para valores λ y vectores a que son respectivamente valores y vectores propios de la matriz cuadrada $W^{-1}B$. Hay a lo sumo $q=\min(p,K-1)$ valores propios no nulos.

Con el anterior resultado si se retiene una sola dirección discriminante, se tomaría la determinada por α_1, siendo (λ,α_1) el par formado por el mayor valor propio y su vector propio asociado. En efecto, tal elección de , maximiza el cociente (Bárcena, 2008).

$$\lambda=\frac{\alpha'B\alpha}{\alpha'W\alpha}$$

Probabilidades posteriores

Otra forma de realizar análisis discriminate es mediante la utilización de probabilidad a posteriori. La ecuación que se utiliza es la siguiente:

$$q_j(\chi_0)=\left(\frac{\pi_j f_j(\chi_0)}{\sum_{i=1}^{g}\pi_i f_i(\chi_0)}\right)=\frac{e^{-\frac{1}{2}D_j(\chi_0)}}{\sum_{i=1}^{g}e^{-\frac{1}{2}D_i(\chi_0)}}$$

En la práctica se sustituye μ_j por $\overline{\chi}_j$, Σ_j por δ_j, en $q_j(\chi_0)$

A menudo es difícil obtener valores de las probabilidades a priori. δ_j Nuestra muestra aleatoria de la población global

$$\hat{\pi}_j = \frac{n_j}{n_1 + n_2 + \ldots + n_g}$$

Cuando se tienen dos poblaciones, a veces es conveniente una prueba F o T^2 de Hotelling como una medida del poder discriminatorio

$$H_o : \mu_1 = \mu_2 \text{ o } H_o : \Delta^2 = 0$$

$$H_o : \mu_1 = \mu_2 = \ldots = \mu_g$$

$$V = \sum n_i (\overline{\chi}_i - \overline{\chi})' \delta^{-1} (\overline{\chi}_i - \overline{\chi})$$

$$\overline{\chi} = \sum_{i=1}^{n} \omega_i \overline{\chi}_i \qquad \omega_i = \frac{n_i}{\sum_{j=1}^{g} n_j}$$

Bajo la hipótesis nula, se trata de minimizar $(n_i) \to \infty$, dado que $V \,\square\, \chi^2(d,(g-1))$

Aplicación

Tomando como referencia las normas emitidas por el Comité de Supervisión Bancaria de Basilea y en especial la dirigida a fomentar la mejora en la gestión de los riesgos en las entidades de crédito, se procede a desarrollar un caso donde es factible la aplicabilidad del análisis discriminante, para eso se introduce el concepto de Riesgo de Crédito planteada por la Superfinanciera como "la probabilidad de que el Banco incurra en pérdidas (no esperadas) y se disminuya el valor de sus activos, como consecuencia de que sus deudores fallen en el cumplimiento oportuno o no cumplan con los términos acordados en los contratos de crédito". Es así como la circular 011 del 2002 plantea cómo a través de un sistema de administración de riesgo crediticio (SARC) se realiza una evaluación del riesgo crediticio considerando un proceso que tenga en cuenta mínimamente los siguientes parámetros:

- Característica de los sujetos de crédito.
- Definición de nicho de mercado.
- Análisis para otorgamiento.
- Otorgamiento.
- Evaluación y seguimiento del portafolio crediticio.

El método basado en calificaciones internas "IRB" planteado en Basilea II se basa en el cubrimiento que deben hacer las entidades crediticia, por un lado las pérdidas inesperadas las cuales son respaldadas con recursos propios, y por otro lado, la pérdidas esperadas las cuales son cubiertas con las provisiones. De lo anterior, la pérdida esperada se considera como el monto de capital que podría perder una institución como resultado de la exposición crediticia en un horizonte de tiempo dado (Wilson & Press, 1978), bajo el esquema de Basilea (Torres Avendaño, 2005) se plantea el siguiente esquema para el cálculo de la pérdida esperada:

Grafico 2. *Cálculo de la Pérdida Esperada*

Fuente. Elaboración propia.

De lo anterior, se puede deducir que la pérdida esperada está en función de la probabilidad de incumplimiento, dado que todo lo demás está dado. Dicha probabilidad de incumplimiento se modela a través del análisis discriminante utilizando factores que determinen la probabilidad de default en una entidad financiera.

La muestra

Para este estudio se utilizó una base de datos correspondiente a 1.000 clientes, los cuales poseían obligaciones financieras con una entidad específica en la ciudad de Medellín, de esta base se extrajo una muestra aleatoria de 400 clientes entre todos aquellos que tenían formalizada una operación de crédito, estuvieran en probabilidad de default ("Circular Externa 011," 2002) o no, adicional, todos se encuentran clasificados dentro del portafolio de cartera comercial[1] y por un período de un año. Para todos ellos la entidad financiera poseía registros que permiten caracterizar a cada uno estos clientes. De toda esta información la más relevante fue la calificación del deudor ("Carta Circular," 2002), la cual estaba acorde con los días de mora que registraba en ese momento, además de esta información se contaba con el saldo del crédito durante todo el año de estudio, lo cual nos permite contabilizar la liberalización de capital que se hubiera podido hacer con un modelo de análisis discriminante.

1 Hace referencia a la colocación que realiza la entidad financiera a personas naturales o jurídicas cuya actividad económica está definida, además, el objeto de los recursos es la producción, transformación o comercialización, y por ningún caso se consideraron los microempresarios.

Selección de variables

En primer lugar se considera una variable que manifieste la probabilidad de default por parte del cliente sobre la entidad financiera, dicha variable toma valores de 0 y 1, este último caracteriza la probabilidad de default.

En cuanto a las variables independientes solo se seleccionaron aquellas que la entidad financiera considerara vital para medir el nivel de riesgo sobre aquellas operaciones de crédito. Las variables son las siguientes:

- **Margen operativo:** margen operativo de la persona i. Es un indicador de rentabilidad y representa el beneficio obtenido en la actividad por cada unidad vendida. Se espera que a mayor margen operativo menor sea la probabilidad de incumplimiento.
- **Nivel de endeudamiento:** endeudamiento de la persona i. Es un indicador a través del cual se ve la participación de los terceros en el valor total de la empresa. Se espera que a mayor endeudamiento mayor probabilidad de incumplimiento tenga el cliente.
- **Activos:** valor de los activos que tiene la persona i. A través de esta variable se puede observar el nivel de respaldo que tiene el cliente sobre la deuda en un caso de incumplimiento.
- **Edad:** es la edad de la persona i. En este caso los clientes con mayor edad responden de manera positiva a los compromisos adquiridos con el sistema financiero.
- **Margen neto:** margen neto de la persona i. Es un indicador de rentabilidad y determina el porcentaje que queda en cada venta después de deducir todos los gastos incluyendo los impuestos.

Resultados

La selección de los 400 clientes se hizo al azar (Palepu, 1986) arrojando 298 clientes cumplidos y 102 clientes incumplidos. Con esta muestra se procede a modelar la probabilidad de incumplimiento a través del análisis discriminante con el objeto de determinar las variables más relevantes, el programa SPSS nos arrojo que las variables a tener en cuenta son las que aparecen en la tabla 2.

Tabla 2. *Variables in the analysis*

Variables in the Analysis				
Step		Tolerance	F to Remove	Wilks' Lambda
1	Margen Neto	1.000	246.447	
2	Margen Neto	.880	325.848	.931
	Endeudamiento	.880	82.558	.618
3	Margen Neto	.795	317.035	.911
	Endeudamiento	.879	78.876	.606
	Margen Operativo	.880	4.320	.511
4	Margen Neto	.765	258.523	.827
	Endeudamiento	.870	71.816	.590
	Margen Operativo	.858	5.680	.507
	Edad	.916	4.835	.506

Fuente. Resultados SPSS 20. Cálculos propios.

Para este estudio se aplicó el criterio de minimización de la Lambda de Wilks y la prueba de menor razón F, con lo cual las dos variables más significativas son en su orden el Margen Neto seguida por el nivel de endeudamiento. Por otra parte, al mirar los valores en la tolerancia, estos al encontrarse en un intervalo entre 0,8 y 1, se puede afirmar que las variables explicativas poseen un alto grado de independencia, es decir, se posee un nivel de tolerancia ideal, con lo cual estas cuatro variables se utilizan en la predicción de la función discriminante.

La tabla 3 mide la fuerza de relación entre la variable predictora y los grupos, en este caso existe una buena relación, ya que nos arroja un valor de 0,707

Tabla 3. *Eigenvalues*

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.001[a]	100.0	100.0	.707

a. First 1 canonical discriminant functions were used in the analysis.

Fuente: Resultados SPSS 20. Cálculos propios

A continuación la tabla 4 muestra la función discriminante estimada para los 400 clientes, los cuales contienen cuatro variables independientes, dichas funciones están sin estandarizar y estandarizadas. Tal como se esperaba, las variables que más contribuyen a la función discriminante son el nivel de endeudamiento y el margen neto, esto se puede ver por los altos coeficientes que muestran estas variables.

Tabla 4. *Función discriminante estimada*

Función discriminante estimada		
	Standardized Canonical Discriminant Function	Canonical Discriminant Function
Margen Operativo	-.770	-.182
Margen Neto	2.684	1.017
Edad	.127	.162
Endeudamiento	-4.019	-.594
(Constant)	.784	

Fuente. Resultados SPSS 20. Cálculos propios

En cuanto a la predicción del modelo, se tiene que de los 298 clientes que se catalogaron como cumplidos el modelo predijo correctamente 296, por lo que su precisión es del 99,3%; mientras que de los 102 clientes que se catalogaron incumplidos, el modelo predijo correctamente el 51%. Por lo tanto el modelo como lo muestra la tabla 5 es preciso en un 87% de los casos, por lo cual el modelo se considera bueno.

Tabla 5. *Classification Results*

Classification Results[a]					
		Default	Predicted Group Membership		Total
			CUMPLIDOS	INCUMPLIDOS	
Original	Count	CUMPLIDOS	296	2	298
		INCUMPLIDOS	50	52	102
	%	CUMPLIDOS	99.3	.7	100.0
		INCUMPLIDOS	49.0	51.0	100.0

a. 87,0% of original grouped cases correctly classified.

Fuente. Resultados SPSS 20. Cálculos propios.

Por otro lado, la función discriminante para hacer pronósticos para nuevos clientes está dada en la tabla 6. Aquí se presentan los coeficientes de la función de clasificación llamados coeficientes de clasificación de Fisher.

Tabla 6. *Classification Results*

Classification Function Coefficients		
	Default	
	CUMPLIDOS	INCUMPLIDOS
Margen Operativo	-.016	1.748
Margen Neto	-.649	-6.795
Edad	.953	.662
Endeudamiento	11.079	20.284
(Constant)	-2.197	-6.349

Fuente. Resultados SPSS 20. Cálculos propios

Por lo tanto, las funciones discriminantes serían en este caso las siguientes:

Función Discriminante para los clientes que son cumplidos.

$$Z = -2{,}197 - 0{,}16\,(Margen\ Operativo) - 0{,}649\,(Margen\ Neto) + 0{,}953\,(Edad) + 11{,}079\,(Endeudamiento)$$

Función Discriminante para los clientes que son incumplidos.

$$Z = -6{,}349 + 1{,}748\,(Margen\ Operativo) - 6{,}795\,(Margen\ Neto) + 0{,}662\,(Edad) + 20{,}284\,(Endeudamiento)$$

La función discriminante que tenga el mayor puntaje será el grupo al que pertenezca el nuevo cliente, es así, como a través de las funciones discriminantes se puede hacer la clasificación de los clientes que se deseen.

Por último, se realiza una comparación entre las probabilidades de default de la institución versus la Superfinanciera y el análisis discriminante, es así, como en la tabla 7.

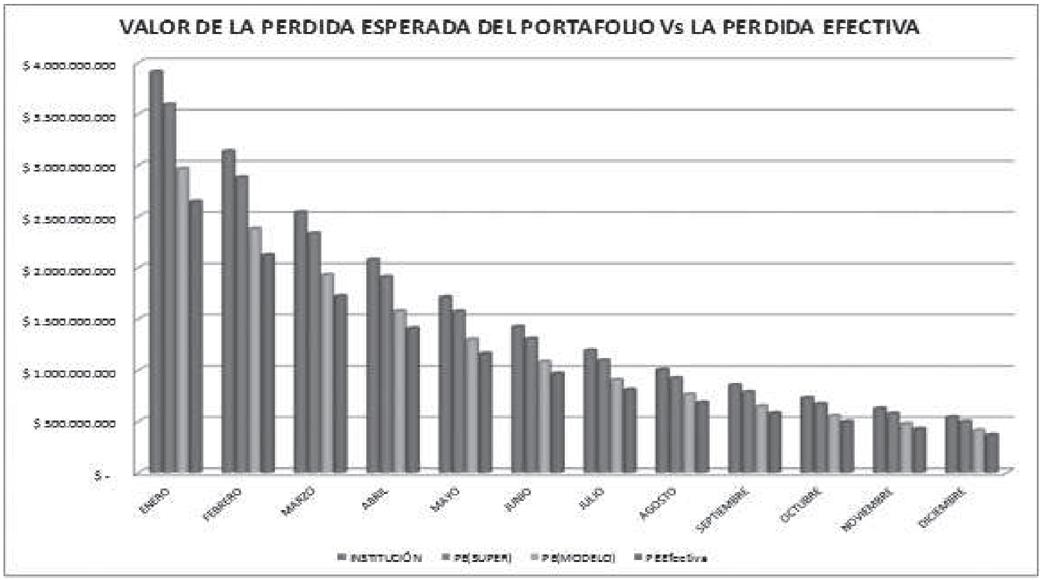
Tabla 7. *Probabilidad de Default*

	SFC	MODELO	INSTITUCION
AA	5,27%	3,26%	6,03%
A	9,41%	7,09%	10,74%
BB	22,36%	21,49%	25,60%
B	25,81%	23,49%	28,70%
CC	37,01%	28,71%	39,20%

Fuente. Cálculos propios

Con base en la tabla anterior se calculó la pérdida esperada para el portafolio de la institución financiera en estudio, como se puede ver en la figura 3, el modelo que arrojó el análisis discriminante permite liberar fondos sin comprometer la estabilidad financiera de la institución, esto debido a que con dicho modelo la provisión que se hace es menor que la estipulada por la institución, pero a la vez mayor que la pérdida efectiva que se dio en el portafolio durante el periodo de estudio.

Gráfico 3.



Fuente. Cálculos propios

Finalmente, se observa que a medida que transcurre el periodo de estudio la diferencia entre el modelo obtenido por análisis discriminante y el desarrollado por la institución financiera disminuye, esto debido al mismo proceso de pago realizado por los deudores, lo que sí queda claro es que el modelo desarrollado en este estudio hubiese ofrecido la oportunidad de liberalización de fondos conduciendo a un aumento en la rentabilidad de la institución.

Conclusiones

El modelo estimado bajo los parámetros del análisis discriminante es significativo, tal como lo muestran los estadísticos calculados para valorar la significación de cada modelo. Además, el nivel de predicción es bueno, ya que posee un nivel de acierto del 87%. En cuanto a las variables que más influyen en la predicción de la probabilidad de default son el margen neto y el nivel de endeudamiento, dichas variables hacen parte del desarrollo operativo de cada uno de los clientes.

El modelo obtenido por análisis discriminante permite mostrar que se produce una asignación eficiente de recursos que son menores a los que hizo la institución financiera y los que prevé el ente regulador, esto es un aspecto relevante porque deja abierta la posibilidad a cada entidad de recurrir a estos modelos con el objeto de poder disponer de mayores recursos sin incurrir en un mayor riesgo que implique detrimentos patrimoniales.

Referencias

Bárcena, M. (2008). Análisis discriminante. Análisis multivariante. In C. González (Ed.), (pp. pp 107 - 122).

Carta Circular. (2002) (SB 059 ed.): Superintendencia Bancaria de Colombia.

Circular Externa 011. (2002). Superintendencia Bancaria de Colombia.

Palepu, K. G. (1986). Predicting takeover targets. *Journal of Accounting & Economics, 8*(1), 3-35.

Pamela, C. S. (2009). Aplicación de métodos estadísticos multivariados en el estudio de calidad de enmiendas orgánicas sólidas y líquidas preparadas en las provincias de Guayas, Los Ríos y El Oro. In R. B. Omar (Ed.).

Torres Avendaño, G. (2005). El acuerdo de Basilea *Estado del arte del SARC en Colombia*. Revista Ad Minister: Universidad EAFIT.

Villardón, J. L. V. (s/f). Análisis discriminante: introducción. Universidad de Salamanca: Departamento de estadistica.

Wilson, S., & Press, S. J. (1978). Choosing Between Logistic Regression and Discriminant Analysis. *Journal of the American Statistical Association, 73*(364), 699-705.