**BANCO CENTRAL DE COSTA RICA**
**DIVISIÓN ECONÓMICA**
**DEPARTAMENTO DE INVESTIGACIONES ECONÓMICAS**
**DIE-NT-03-98**
**AGOSTO, 1998**

# LA TÉCNICA DE ÁNALISIS DISCRIMINANTE: UNA APLICACIÓN PARA EL ÁREA BANCARIA[1]

Evelyn Muñoz Salas

**RESUMEN EJECUTIVO**

*El objetivo principal de esta nota técnica es brindar a la División Económica y al Departamento de Investigaciones Económicas en particular, una herramienta adicional de análisis que permite abordar con bastante flexibilidad temas relacionados sobre todo con el sector financiero, aunque no exclusivamente.*

*El análisis discriminante busca identificar, a partir de una serie de indicadores, si es posible "discriminar" si una observación pertenece a un determinado grupo de entre varios existentes; seleccionar cuál o cuáles de esos indicadores contribuyen más al proceso de discriminación, y adicionalmente permite estimar funciones de clasificación para ubicar nuevos casos.*

*En este trabajo, se realizó una aplicación con la ayuda del paquete estadístico SPSS, para identificar las variables que permiten discriminar entre bancos con altas y bajas utilidades. Se llegó a la conclusión de que los ingresos por intermediación financiera es la variable que más contribuye a la discriminación, dentro de las consideradas para el ejercicio.*



[1] Se agradecen los valiosos comentarios de Msc Ana Cecilia Kikut Valverde.

# I. *INTRODUCCIÓN*

En múltiples ocasiones, un investigador se ve en la necesidad de introducir variables cualitativas dentro de un análisis de regresión. Esta posibilidad permite enfrentar con mayor flexibilidad una serie de problemas.

Concretamente estas variables cualitativas pueden ser del tipo indicadoras o truncadas. En el caso de las variables indicadoras, éstas pueden ser tanto explicativas como dependientes. Las variables truncadas se emplean en ocasiones en que no es posible obtener información para todos los casos sobre la variable dependiente.

Dentro de los usos que se les ha dado como variables explicativas se encuentran, capturar cambios en las pendientes de las series, probar la estabilidad de los coeficientes de una regresión, etc.[2]. Estos casos se han manejado con bastante frecuencia dentro de las formulaciones elaboradas en la División Económica, concretamente en el Departamento de Investigaciones Económicas.

Ahora bien, no ha sido tan frecuente emplear una variable cualitativa como variable dependiente. Este es precisamente el caso de la técnica de análisis discriminante, que busca establecer si una observación pertenece o no a un grupo en particular, con base en uno o más indicadores.

Una razón que motiva elaborar este documento es el hecho de que muchos analistas financieros han señalado las limitaciones que presentan las razones financieras como método para identificar bancos en problemas en forma oportuna. Esta limitación radica, entre otras cosas, en la posibilidad de comportamientos deshonestos e ineficiencia por parte de los administradores bancarios, o de asumir riesgos excesivos. Ante esta situación ellos han propuesto nuevos métodos para detectar bancos atípicos, concretamente aplicando el análisis discriminante a razones financieras y otros criterios adicionales.

[2] A este tipo de variables también se les denomina como artificiales o dummy. Cuando asumen únicamente dos valores (0 y 1) se les llama dicotómicas.

Esta aplicación concreta en el área bancaria y, en general, la financiera, puede enriquecer los trabajos que en esta materia realiza y planea desarrollar la División Económica.

De esta forma, se presenta la siguiente nota técnica que pretende ampliar aspectos referentes a la técnica del análisis discriminante, comentados en forma general en el Anexo 2 del documento "Bancos Privados: análisis discriminante por área de riesgo y su relación con el rendimiento"[3]

Además, se tratará de exponer el procedimiento a seguir en el paquete estadístico SPSS, empleando un ejemplo e interpretando algunos de los principales resultados.

## *II. GENERALIDADES*

El Análisis de Función Discriminante es una técnica estadística de la rama del análisis multivariante, en la cual la variable dependiente es indicadora y no numérica como en el análisis de regresión.

Las variables indicadoras son un tipo especial de variables empleadas en las ecuaciones de regresión múltiple; cuando aparecen como explicativas permiten desde probar la estabilidad de los coeficientes de una regresión, hasta introducir aspectos cualitativos en la misma, tal es el caso de las variables dummy[4].

La variable indicadora puede tomar dos o más valores, no obstante, el caso más común es el que emplea dos valores: cero y uno, en este caso se conoce como variable dicotómica.

Por ejemplo, la variable Y es una variable indicadora, que denota la ocurrencia o no de un evento y se define de esta forma:

$$\boxed{Y = \begin{cases} 0 \rightarrow \textit{hombre} \\ 1 \rightarrow \textit{mujer} \end{cases}}$$

[3] Véase, Alfaro y Muñoz (1998).
[4] Véase, Madala G.S.(1996).

En este contexto, el Análisis de Función Discriminante se emplea para determinar cuál o cuáles variables contribuyen a discriminar entre dos o más grupos que se observan en la práctica.

Por ejemplo, suponga que se tiene una medida de la estatura de un grupo de individuos, 50 hombres y 50 mujeres. En promedio, la estatura de las mujeres es inferior a la de los hombres, de forma que esta diferencia puede reflejarse en la diferencia entre las medias de ambos grupos. Así, la variable "estatura" permite discriminar entre hombres y mujeres de forma más adecuada*: ´si una persona es alta, entonces es probable que sea un hombre; si una persona es baja, es posible que sea una mujer'*.

En resumen, la idea básica que subyace en el Análisis de Función Discriminante es determinar si unos grupos difieren en función de la media de una variable[5], y emplear luego esa variable para predecir la pertenencia de una nueva observación a determinado grupo.

## *III. LA FUNCIÓN DISCRIMINANTE LINEAL*

El método está basado en modelos de probabilidad lineal, y se conoce como función discriminante lineal (FD). Supone que si se tienen n entidades[6] para las que se conocen *k* variables explicativas, y se observa que $n_1$ de ellas pertenece a un grupo (1) y $n_2$ a otro grupo (2), donde: $n_1+n_2=n$; es posible construir una función lineal de las *k* variables que puede usarse para predecir si una nueva observación pertenece a un grupo u otro con una probabilidad determinada. La función lineal general se define como sigue:

$$\boxed{(1): Z = l_0 + \sum_{i=1}^{K} l_i x_i}$$

[5] Se realiza una interpretación similar en el caso de que se trate de más de una variable de discriminación.
[6] Las entidades pueden referirse a individuos, instituciones, etc.

El problema de la función de análisis discriminante desde el punto de vista del análisis de variancia consiste en responder a la pregunta de si dos o más grupos son significativamente diferentes uno de otro respecto a la media de una variable en particular. Debe tenerse presente que si la media de una variable es significativamente diferente en varios grupos, puede decirse que esta variable discrimina entre grupos.

Al igual que en el caso de conglomerados, análisis factorial y de correlación canónica, el análisis discriminante realiza diferentes desgloses de las variancias de un conjunto de datos para someterlos a una serie de pruebas estadísticas y determinar el grado de asociación entre esas variancias y, por tanto, entre las variables. De esta forma, la mejor discriminación se tiene al maximizar:

$$\frac{\textbf{Variancia de Z entre grupos}}{\textbf{Variancia de Z dentro de grupos}}$$

En el caso de una única variable explicativa, la prueba final de significancia de si esta variable discrimina o no entre grupos es una prueba F, que es básicamente una razón de las variancias entre grupos sobre el promedio de la variancia dentro de los grupos. Si la variancia entre grupos es significativamente mayor, deberá haber diferencias significativas entre las medias.

Para el caso de más de una variable, se busca determinar cuál o cuáles de ellas contribuyen a la discriminación entre grupos. En este caso, se tiene una matriz de variancias y covariancias. Puede compararse las matrices con una prueba F multivariable, para determinar si hay o no diferencias significativas en las medias entre grupos.

## *IV. INTERPRETACIÓN DE LA FUNCIÓN DISCRIMINANTE*

En esta sección se estudia la aplicación de la ecuación (1) de la sección anterior a dos casos particulares: para dos grupos y para múltiples grupos. También se incluye la significancia de la función discriminante.

### 4.1. La Función Discriminante para Dos Grupos

En este caso, la función discriminante puede verse como una caso de análisis de regresión lineal múltiple[7]. Si se codifican los dos grupos como 1 y 2, y se emplea tal variable como dependiente en un análisis de regresión lineal múltiple, pueden obtenerse resultados similares a los que se obtendrían de un análisis discriminante. En general, en el caso de dos grupos se ajusta una ecuación lineal del tipo:

$$(2): GRUPO = a + b_1 x_1 + b_2 x_2 + \cdots + b_m x_m$$

Donde a es una constante y $b_1$ a $b_m$ son coeficientes de regresión. La interpretación de estos resultados es similar a la de un modelo de regresión múltiple. Los más significativos son lo que contribuyen más a la predicción de pertenencia a un grupo.

Para efectuar el análisis, es posible emplear diferentes procedimientos, como por ejemplo:

- Análisis discriminante "Stepwise": probablemente la forma más común de aplicación es incluir muchas medidas en el estudio, para determinar las que discriminan entre grupos. Visto de otra forma, se desea construir un modelo de cómo se puede lograr predecir de la mejor forma a cuál grupo pertenece una observación o caso particular.

- Análisis discriminante "Stepwise" hacia adelante: la idea es construir un modelo paso a paso, revisando todas las variables y evaluando cuál puede contribuir más a la discriminación entre grupos. Esta variable podrá ser incluida en el modelo.

- Análisis discriminante "Stepwise" hacia atrás: es posible incluir primero todas las posibles variables en el modelo, y luego en cada paso, eliminar la variable que contribuye menos a la predicción de la pertenencia a un grupo. Como resultado de

[7] Posterior al año 1936, el análisis de función discriminante para dos grupos se conoce también como Análisis Discriminante lineal de Fisher.

un modelo de Función Discriminante exitoso, deben mantenerse en él las variables significativas para discriminar.

### 4.2. La Función Discriminante para Múltiples Grupos

Cuando es posible identificar más de dos grupos, puede estimarse más de una función discriminante similares a la presentada anteriormente. Por ejemplo cuando se tienen tres grupos, puede estimarse 1) una función para discriminar entre grupo 1 y grupos 2 y 3 combinados, y 2) otra función para discriminar entre grupo 2 y grupo 3.

En la práctica, cuando se realiza un análisis discriminante entre varios grupos, no debe especificarse cómo combinar los grupos para formar las diferentes funciones. El paquete computacional que se emplee, por ejemplo SPSS, automáticamente las conforma de manera que la primera es la que ofrece la mayor discriminación como un todo entre grupos, la segunda provee una menor y así sucesivamente. Las funciones son independientes u ortogonales, esto es, su contribución a la discriminación entre grupos no se sobrepone.

### 4.3. Significancia de La Función Discriminante

Puede probarse el número de variables que agregan significancia a la discriminación entre grupos. Solo aquellas que sean estadísticamente significativas deben ser usadas para interpretar, las no significativas deben ignorarse.

En resumen, cuando se interpretan funciones discriminantes múltiples, que surgen del análisis con más de dos grupos y más de una variable, se puede probar primero la significancia estadística de las diferentes funciones, y considerar solo las significativas para las siguientes pruebas. Luego, se observan los coeficientes b estandarizados (véase ecuación 2) para cada variable para cada función significativa. Cuanto mayores sean más alta es la contribución a la discriminación especificada por la respectiva función. Finalmente, pueden verse las medias para las funciones discriminantes significativas para determinar entre cuáles grupos discrimina la respectiva función.

## V. SUPUESTOS IMPLÍCITOS EN EL ANÁLISIS DISCRIMINANTE

- ✓ Distribución normal: se asume que los datos para las variables representan una muestra proveniente de una distribución normal multivariable. No obstante, el no cumplimiento de este supuesto no es problema para el análisis.
- ✓ Homogeneidad de variancias y covariancias: se supone que las matrices de variancias y covariancias son homogéneas entre grupos; de nuevo, si no se cumple tampoco se generan problemas.
- ✓ Correlaciones entre medias y variancias: el principal obstáculo para la validez de las pruebas de significancia se presenta cuando la media de las variables entre grupos están correlacionadas con las variancias. Si hay gran variabilidad en un grupo con alta media en algunas variables, entonces esas medias grandes no son confiables. Sin embargo, la prueba de significancia global está basada en variancias ponderadas, es decir en variancias promedio entre todos los grupos. La prueba de significancia de las medias relativamente grandes (con grandes variancias) estará basada en variancias ponderadas relativamente menores, resultando erróneamente en significancia estadística. Esto ocurre cuando un grupo tiene unos pocos valores extremos que afectan mucho la media y aumentan la variabilidad.
- ✓ Variables no redundantes: se supone que las variables empleadas para discriminar entre grupos no son completamente redundantes, por ejemplo que una variable no sea la suma de otras dos que también están en el modelo.

## VI. FUNCIÓN DE CLASIFICACIÓN

Otro propósito importante de la función discriminante es en cuanto a aspectos de predecir la clasificación de nuevos casos, en el sentido de que una vez que se ha construido el modelo, qué tan bien se puede predecir la pertenencia al grupo de un caso particular.

Por ejemplo, la función discriminante clasifica una entidad como de utilidades bajas o utilidades altas, dos poblaciones diferentes, en función de los niveles de las variables que los identifican. Es decir, definiría las características para pertenecer a uno u otro grupo.

Para ello, el análisis discriminante calcula una función de clasificación, que no debe confundirse con la función discriminante. Hay tantas funciones de clasificación como grupos, cada una permite calcular los puntajes de clasificación para cada caso en cada grupo con la siguiente fórmula:

$$S_i = c_i + w_{i1}x_1 + w_{i2}x_2 + \cdots + w_{im}x_m$$

Donde el subíndice i denota el grupo respectivo y los números 1,...,m las variables; $c_i$ es una constante para el i-ésimo grupo, $w_{ij}$ es el ponderador para la j-ésima variable en el cálculo del puntaje de clasificación para el i-ésimo grupo, $x_j$ es el valor observado para el respectivo caso para la j-ésima variable. $S_i$ es el puntaje de clasificación resultante.

Los siguientes pasos son necesarios para resolver el problema de la discriminación:

1. **1.** Conocer las densidades de probabilidad $P_1$(y)... $P_n$(y) para clasificar las variables y en n grupos o poblaciones diferentes.
2. **2.** Conocer las probabilidades a priori $\pi_1 \ldots \pi_n$ de que una unidad estadística pertenezca a alguna de las poblaciones, las cuales son frecuencias relativas de unidades estadísticas de las n poblaciones.
3. **3.** Especificar valores $r_{ij}$ que representen la pérdida por identificar una variable y en el grupo i cuando en realidad pertenece a la población j.

A las variables y se les asocia un puntaje S que consiste en un promedio ponderado de las probabilidades de que cada variable muestre los atributos que definen a una población en particular. Es decir:

$$\boxed{(3): S_i = \sum_{n=1}^{z} p_n \, p_n(y) r_{ni}}$$

La variable $y_i$ se asigna a la población para la cual su puntaje discriminante es el más alto. En el caso de que las variables sean normales, el puntaje discriminante se puede calcular como:

$$(4): S_i = -\frac{1}{2}\ln\left|\sum_i\right| - \frac{1}{2}(y - y_i)'\sum_i^{-1}(y - y_i) + \ln p_i$$

El cual corresponde al logaritmo de la función de verosimilitud de la variable yi.

Cuando existen solo dos poblaciones, la regla de decisión para la asignación de una unidad estadística en un grupo o en el otro dada por la diferencia de dos puntajes discriminantes: $S_1$-$S_2$. En términos de la verosimilitud normal, la diferencia de los discriminantes sería:

$$(5): (y_1' - y_2')\sum^{-1} y - \frac{1}{2}(y_1'\sum^{-1} y_1 - y_2') + \ln p_1 - \ln p_2$$

Si se denota el primer sumando de la ecuación anterior como L(Y) y los dos últimos como c, la regla de decisión que aplica el análisis discriminante es la siguiente: asigne la i-ésima observación al grupo 1 si L(Y)>c ó, al contrario, al grupo 2 si L(Y)<c.

Si se tiene un caso nuevo, se aplica la fórmula y se obtiene su puntaje de clasificación: se dice que éste pertenece al grupo para el cual presenta el mayor puntaje de clasificación.

## *VII. ALGUNOS INDICADORES*

### 7.1. Lambda de Wilks

Para determinar la significancia de la variables que se introducen, se emplea el estadístico Lambda de Wilks que se obtiene de la razón entre el determinante de la matriz de variancias y covariancias dentro de grupos y el determinante de la matriz de variancias y covariancias total; este se aproxima con una F.

Para saber si una variable es o no significativa para discriminar se realiza una prueba de significancia de esos estadísticos de la forma usual.

### 7.2. Distancia de Mahalanobis

Esta técnica permite el cálculo de un indicador llamado distancia de mahalanobis, que es una medida de la distancia entre dos puntos en el espacio, definido por dos o más variables correlacionadas.

Por ejemplo si hay dos variables no correlacionadas, pueden graficarse los puntos en un espacio de dos dimensiones, la distancia de mahalanobis entre puntos es la distancia euclidiana. Si hay tres variables no correlacionadas puede usarse una regla para medir las distancias en el espacio de tres dimensiones para determinar la distribución entre puntos. Con más de tres variables ya no es posible graficar las distancias.

Por otro lado, cuando las variables están correlacionadas los ejes no están posicionados en ángulo recto en esos casos la distancia euclidiana no es una medida apropiada, por tanto se usa la distancia de mahalanobis.

Para cada grupo se puede determinar la localización de un punto que representa la media para todas las variables en el espacio multivariable definido por las variables en el modelo. Estos puntos se llaman centroides de los grupos. Para cada caso puede calcularse su distancia de mahalanobis respecto del centroide del grupo. De nuevo, puede clasificarse una observación como perteneciente al grupo al que esté más cerca o sea al cual la distancia de mahalanobis es menor.

Al usar la distancia de mahalanobis para clasificar pueden derivarse probabilidades. La probabilidad de que un dato pertenezca a un grupo particular es proporcional a la distancia de mahalanobis al centroide del grupo (no es exactamente proporcional porque se asume una distribución normal multivariada alrededor de cada centroide), se le denomina probabilidad a posteriori dado que se calcula a partir de valores conocidos. Esta probabilidad se basa en el conocimiento que se tiene de los valores de las variables.

Para determinar qué tan bien predice la pertenencia a un grupo, la función de clasificación, es útil revisar la matriz de clasificación que muestra el número de casos que fueron mal clasificados. No obstante, la mejor forma de evaluarla es tomar un dato que no haya sido empleado para estimar la función, debido a que solo la clasificación de casos nuevos permite conocer el valor predictivo de la función de clasificación; la clasificación de casos viejos solo provee una herramienta de diagnóstico útil para identificar casos extremos o áreas donde la función de clasificación parece ser menos adecuada.

## *VIII. EJERCICIO PRÁCTICO*

### 8.1. Generalidades

Con el fin de ilustrar el uso que se puede dar a esta técnica estadística, se presenta un ejemplo aplicado a la banca privada costarricense. En este se busca determinar cuáles variables ayudan a discriminar entre bancos de altas utilidades y de bajas utilidades, (de forma que se constituye en la variable de clasificación) empleando como variables explicativas: obligaciones con el público, cartera de crédito vigente, ingresos financieros e ingresos por comisiones (servicios). La información empleada corresponde al año 1995.

El siguiente cuadro presenta los estadísticos descriptivos de la variable de clasificación[8]:

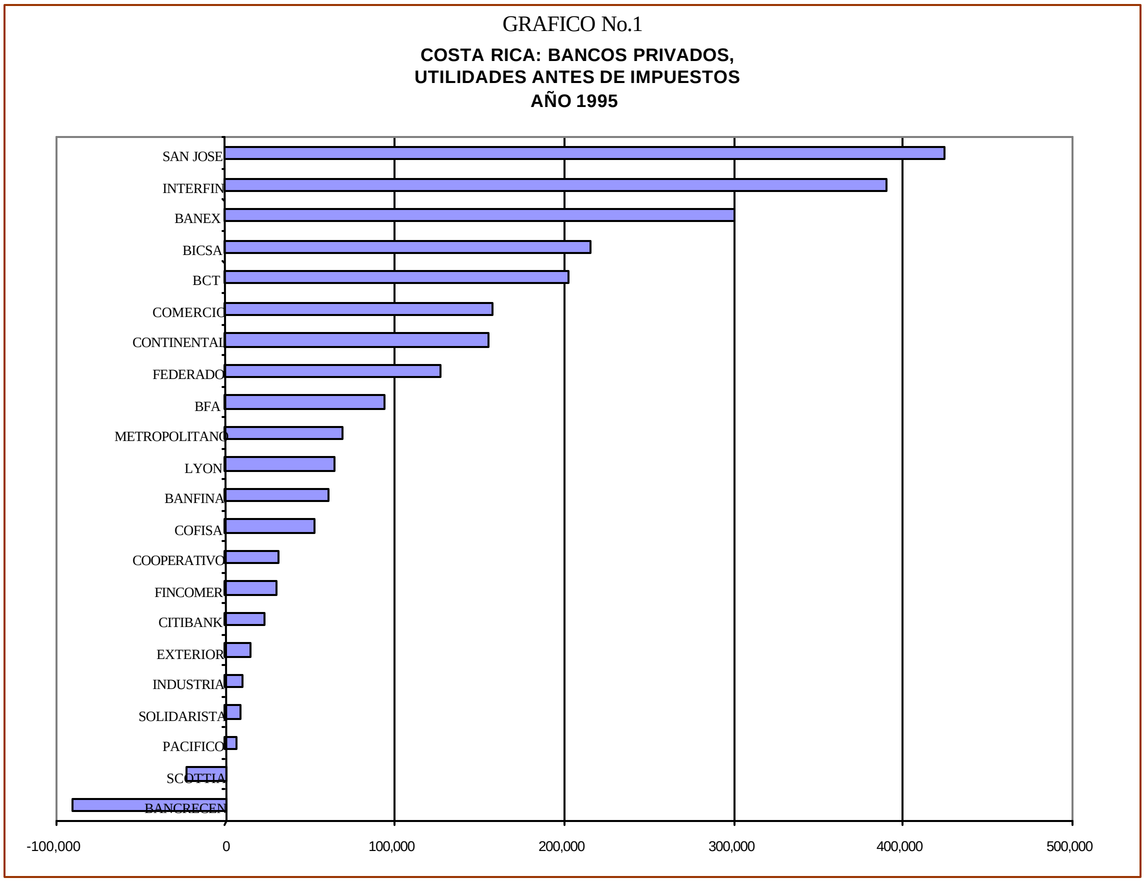
**CUADRO NO.1**
**BANCOS PRIVADOS: UTILIDADES NETAS**
**Año 1995**
**Miles de colones**

**Estadísticos descriptivos**

| | N | Mínimo | Máximo | Media | Desv. típ. | Varianza |
|---|---|---|---|---|---|---|
| | 22 | -90,119 | 424,001 | 105621.0 | 131,811.00 | 1.737E+10 |

[8] Este cuadro se obtiene con la instrucción "Estadísticos; Resumir; Descriptivos" de SPSS.

Gráficamente se observa:




El primer paso consiste en definir con cuántos grupos se va a trabajar, esto depende de aspectos como la orientación del estudio particular que esté realizando un investigador, las características de la muestra que se emplea, etc.

Para simplificar el ejemplo, y dado que se tienen únicamente 22 observaciones, se define como un banco con utilidades bajas, si éstas se ubican por debajo de C100.000.00, valor que es cercano a la media de la muestra.

De esta forma, y para emplear la notación de la primera parte del documento, se tiene que:

$$Y = \begin{cases} 0 \rightarrow bajasU \\ 1 \rightarrow altasU \end{cases}$$

Donde U son utilidades.

El software SPSS puede realizar la agrupación de las observaciones y asignarles 0 y 1 con las instrucciones “Datos; Seleccionar casos; Si se satisface la condición”, indicándole que la variable de clasificación sea mayor al límite establecido. De esta forma construye una variable filtro que se convierte en la variable Y.

Para ejecutar el análisis discriminante se emplean las instrucciones: “Estadísticos; Clasificar; Discriminante”. En este se introducen las diferentes opciones para efectuar el procedimiento:

- Lo primero es indicar cuál es la variable de clasificación y su rango, que para el ejemplo es 0 (mínimo) y 1 (máximo);
- Luego introducir la o las variables independientes;
- Indicar que se calculen algunos estadísticos básicos, tales como medias y variancias de grupos y los coeficientes de la función;
- Indicar en “Clasificación” que presente los cálculos para cada caso;
- Finalmente, en “guardar” debe pedírsele que presente el grupo de pertenencia pronosticado, las puntuaciones discriminantes y las probabilidades de pertenencia al grupo.

Básicamente hay dos opciones: “introducir independientes juntas”, o “usar método de inclusión por pasos”, ambas se verán con el ejemplo.

### 8.2. Presentación de Resultados

Uno de los primeros aspectos que debe revisarse es la media y la desviación que presentan las diferentes variables en los diferentes grupos establecidos, estos son importantes dado que la probabilidad de una mala clasificación es más baja entre mayor sean las diferencias entre los perfiles de los grupos considerados.

Observando la medias de las variables de los grupos conformados, puede concluirse que son relativamente diferentes, es decir los grupos parecen no estar muy cercanos, por tanto se esperaría que la probabilidad de errar en la clasificación a partir de estas variables no sea muy alta.

## CUADRO No.2

**Estadísticos del grupo**

| | | Media | Desv. típ. |
|---|---|---|---|
| .00 | CVIGENTE | 2726500 | 4477498 |
| | INGFIN | 463576.0 | 357666.9 |
| | INGSERV | 86859.61 | 85064.02 |
| | OBLIGP | 3071265 | 4724842 |
| 1.00 | CVIGENTE | 6922625 | 4505698 |
| | INGFIN | 1926744 | 970706.0 |
| | INGSERV | 149309.2 | 87233.91 |
| | OBLIGP | 4641628 | 3840629 |
| Total | CVIGENTE | 4252363 | 4842132 |
| | INGFIN | 995637.0 | 955132.6 |
| | INGSERV | 109568.6 | 89226.61 |
| | OBLIGP | 3642306 | 4397083 |

Donde:

CVIGENTE: Cartera de crédito vigente

INGFIN: Ingresos financieros

INGSERV: Ingresos por servicios

OBLIGP: Obligaciones con el público

Con ayuda de la prueba de igualdad de las medias de los grupos se ve que la variable que es menos importante o significativa es obligaciones con el público, y que la más significativa es ingresos financieros, en tanto que las variables ingresos por comisiones no es significativa al 10%.

## CUADRO No.3

**Pruebas de igualdad de las medias de los grupos**

| | Lambda de Wilks | F | gl1 | gl2 | Sig. |
|---|---|---|---|---|---|
| CVIGENTE | .818 | 4.451 | 1 | 20 | .048 |
| INGFIN | .431 | 26.393 | 1 | 20 | .000 |
| INGSERV | .881 | 2.695 | 1 | 20 | .116 |
| OBLIGP | .969 | .638 | 1 | 20 | .434 |

### *8.2.1. Resultados de la opción: "Introducir independientes juntas"*

Los resultados del análisis se presentan en los siguientes cuadros:

**CUADRO No.4**

**Coeficientes estandarizados de las funciones discriminantes canónicas**

| | Función |
|---|---|
| | 1 |
| CVIGENTE | 2.604 |
| INGFIN | 1.119 |
| INGSERV | -.360 |
| OBLIGP | -2.804 |

Los centroides de los grupos corresponden a la función discriminante canónica[9] que se construye con los coeficientes del cuadro anterior, evaluadas en las medias de los diferentes grupos, como se ve, hay diferencias apreciables, los valores del grupos 0 se agrupan alrededor de –1.355 y los del grupo 1 alrededor de 2.372.

**CUADRO No.5**

**Funciones en los centroides de los grupos**

| | Función |
|---|---|
| VC | 1 |
| .00 | -1.355 |
| 1.00 | 2.372 |

Funciones discriminantes canónicas no tipificadas evaluadas en las medias de los grupos

[9] Se refiere a la función discriminante ya comentada.

Tal como se mencionó anteriormente, resulta útil para un segundo objetivo de clasificar nuevos casos, contar con los coeficientes de la función de clasificación que se presentan en el siguiente cuadro:

**CUADRO No.6**[10]

**Coeficientes de la función de clasificación**

| | VC | |
|---|---|---|
| | .00 | 1.00 |
| CVIGENTE | -1.30E-07 | 2.032E-06 |
| INGFIN | -9.98E-08 | 6.389E-06 |
| INGSERV | 1.053E-05 | -5.08E-06 |
| OBLIGP | 2.202E-07 | -2.14E-06 |
| (Constante) | -1.288 | -8.545 |

Funciones discriminantes lineales de Fisher

Hay tantas funciones de clasificación como grupos, y cada observación se asigna a aquel grupo para el cual obtiene el mayor puntaje de clasificación. Esto se aplica en el siguiente cuadro, donde se compara la clasificación original o real de cada caso, con el grupo pronosticado.

[10] VC: se refiere a valor del coeficiente de la función de clasificación

## CUADRO No.7

**Estadísticos por casos**

| | Número de casos | Grupo real | Grupo mayor: Grupo pronosticado | Grupo mayor: P(D>d \| G=g): p | Grupo mayor: P(D>d \| G=g): gl | Grupo mayor: P(G=g \| D=d) | Grupo mayor: Distancia de Mahalanobis al cuadrado hasta el centroide | Segundo grupo mayor: Grupo | Segundo grupo mayor: P(G=g \| D=d) | Segundo grupo mayor: Distancia de Mahalanobis al cuadrado hasta el centroide | Puntuaciones discriminantes: Función 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Original | 1 | 0 | 0 | .578 | 1 | 1.000 | .309 | 1 | .000 | 18.348 | -1.911 |
| | 2 | 0 | 0 | .741 | 1 | 1.000 | .109 | 1 | .000 | 16.465 | -1.686 |
| | 3 | 0 | 0 | .367 | 1 | 1.000 | .813 | 1 | .000 | 21.430 | -2.257 |
| | 4 | 0 | 0 | .833 | 1 | .998 | .044 | 1 | .002 | 12.369 | -1.145 |
| | 5 | 0 | 0 | .892 | 1 | .998 | .019 | 1 | .002 | 12.895 | -1.219 |
| | 6 | 0 | 0 | .870 | 1 | .998 | .027 | 1 | .002 | 12.699 | -1.192 |
| | 7 | 0 | 0 | .795 | 1 | .997 | .068 | 1 | .003 | 12.022 | -1.095 |
| | 8 | 0 | 0 | .791 | 1 | .997 | .070 | 1 | .003 | 11.989 | -1.091 |
| | 9 | 0 | 0 | .733 | 1 | 1.000 | .116 | 1 | .000 | 16.553 | -1.697 |
| | 10 | 0 | 0 | .854 | 1 | .998 | .034 | 1 | .002 | 12.551 | -1.171 |
| | 11 | 0 | 0 | .872 | 1 | .998 | .026 | 1 | .002 | 12.719 | -1.194 |
| | 12 | 0 | 0 | .415 | 1 | .980 | .665 | 1 | .020 | 8.477 | -.540 |
| | 13 | 0 | 0 | .879 | 1 | .999 | .023 | 1 | .001 | 15.046 | -1.507 |
| | 14 | 0 | 0 | .933 | 1 | .999 | .007 | 1 | .001 | 13.272 | -1.271 |
| | 15 | 1 | 0** | .132 | 1 | .792 | 2.264 | 1 | .208 | 4.940 | .149 |
| | 16 | 1 | 1 | .032 | 1 | 1.000 | 4.585 | 0 | .000 | 34.440 | 4.513 |
| | 17 | 1 | 1 | .076 | 1 | .580 | 3.159 | 0 | .420 | 3.803 | .595 |
| | 18 | 1 | 1 | .197 | 1 | .894 | 1.668 | 0 | .106 | 5.933 | 1.080 |
| | 19 | 1 | 1 | .977 | 1 | .999 | .001 | 0 | .001 | 14.107 | 2.401 |
| | 20 | 1 | 1 | .323 | 1 | 1.000 | .976 | 0 | .000 | 22.235 | 3.360 |
| | 21 | 1 | 1 | .213 | 1 | 1.000 | 1.551 | 0 | .000 | 24.728 | 3.617 |
| | 22 | 1 | 1 | .374 | 1 | 1.000 | .789 | 0 | .000 | 21.304 | 3.260 |

**. Caso mal clasificado

El número de casos se refiere a las observaciones, las cuales se presentan en el orden en el cual fueron introducidas en la base de datos. Para cada una de ellas se observa el grupo al cual pertenece antes de realizar el análisis, luego se presenta el grupo al cual se pronostica que pertenece; como se aprecia en este caso únicamente un banco fue clasificado con la función de clasificación en el grupo al cual no pertenecía originalmente.

Este cuadro tiene dos grandes divisiones, "Grupo mayor" (el más numeroso) y "Segundo grupo mayor", dado que solo se trabajó con dos grupos. Para cada uno de ellos se tiene una función de clasificación tal como ya se explicó, y se evalúan todas las observaciones en ambas funciones, para determinar finalmente a cuál grupo pertenece.

También se muestra para ambas funciones la distancia de mahalanobis al cuadrado hasta el centroide, se observa que por ejemplo los bancos clasificados en el grupo 0 al ser evaluados en la función de clasificación 0, presentan una distancia corta comparada con la que presentan los bancos del grupo 1, un caso particular es el banco que se clasificó mal pues se asignó finalmente al grupo 0 y al ser evaluado en la función de clasificación 0 su distancia al centroide es la mayor de todos los bancos del grupo 0. No obstante, está más lejos del centroide del grupo 1.

El cuadro presenta además, el puntaje discriminante para cada observación, éste permite comprobar lo comentado en el sentido de que el banco No.15 de la muestra no pertenece al grupo 1 sino al 0.

Tal como lo presenta el siguiente cuadro, un 95,5% de los casos agrupados originalmente se clasificaron correctamente con las funciones conformadas.

**CUADRO No.8**

**Resultados de la clasificación**[a]

| | | VC | Grupo de pertenencia pronosticado .00 | 1.00 | Total |
|---|---|---|---|---|---|
| Original | Recuento | .00 | 14 | 0 | 14 |
| | | 1.00 | 1 | 7 | 8 |
| | % | .00 | 100.0 | .0 | 100.0 |
| | | 1.00 | 12.5 | 87.5 | 100.0 |

a. Clasificados correctamente el 95.5% de los casos agrupados originales.

### *8.1.2. Resultados del procedimiento de "inclusión por pasos"*

A diferencia del caso anterior, en esta oportunidad el programa empleó finalmente solo una de las variables para construir el modelo, es decir ingresos financieros aspecto que era de esperar pues ya en el análisis anterior se presentó como la variable más significativa para discriminar entre bancos de altas utilidades y bajas utilidades. Las demás variables fueron excluidas del análisis.

**CUADRO No.9**
**Estadísticos por pasos**

**Variables introducidas/eliminadas**[a,b,c,d]

| | | Lambda de Wilks | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | F exacta | | | |
| Paso | Introducidas | Estadístico | gl1 | gl2 | gl3 | Estadístico | gl1 | gl2 | Sig. |
| 1 | INGFIN | .431 | 1 | 1 | 20.000 | 26.393 | 1 | 20.000 | .000 |

En cada paso se introduce la variable que minimiza la lambda de Wilks global.

a. El número máximo de pasos es 8.

b. La F parcial mínima para entrar es 3.84.

c. Maximum partial F to remove is 2.71.

d. El nivel de F, la tolerancia o el VIN son insuficientes para continuar los cálculos.

## CUADRO No.10

**Funciones en los centroides de los grupos**

| VC | Función 1 |
|---|---|
| .00 | -.828 |
| 1.00 | 1.449 |

Funciones discriminantes canónicas no tipificadas evaluadas en las medias de los grupos

## CUADRO No.11

**Coeficientes de la función de clasificación**

| | VC .00 | VC 1.00 |
|---|---|---|
| INGFIN | 1.123E-06 | 4.666E-06 |
| (Constante) | -.953 | -5.188 |

Funciones discriminantes lineales de Fisher

## CUADRO No.12

**Estadísticos por casos**

| | Número de casos | Grupo real | Grupo mayor: Grupo pronosticado | Grupo mayor: P(D>d \| G=g): p | Grupo mayor: P(D>d \| G=g): gl | Grupo mayor: P(G=g \| D=d) | Grupo mayor: Distancia de Mahalanobis al cuadrado hasta el centroide | Segundo grupo mayor: Grupo | Segundo grupo mayor: P(G=g \| D=d) | Segundo grupo mayor: Distancia de Mahalanobis al cuadrado hasta el centroide | Puntuaciones discriminantes: Función 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Original | 1 | 0 | 0 | .645 | 1 | .974 | .213 | 1 | .026 | 7.498 | -1.289 |
| | 2 | 0 | 0 | .400 | 1 | .663 | .707 | 1 | .337 | 2.062 | .013 |
| | 3 | 0 | 0 | .642 | 1 | .975 | .216 | 1 | .025 | 7.519 | -1.293 |
| | 4 | 0 | 0 | .608 | 1 | .977 | .263 | 1 | .023 | 7.781 | -1.341 |
| | 5 | 0 | 0 | .537 | 1 | .982 | .381 | 1 | .018 | 8.374 | -1.445 |
| | 6 | 0 | 0 | .618 | 1 | .977 | .249 | 1 | .023 | 7.708 | -1.327 |
| | 7 | 0 | 0 | .693 | 1 | .970 | .155 | 1 | .030 | 7.135 | -1.222 |
| | 8 | 0 | 0 | .939 | 1 | .918 | .006 | 1 | .082 | 4.842 | -.752 |
| | 9 | 0 | 0 | .903 | 1 | .946 | .015 | 1 | .054 | 5.755 | -.950 |
| | 10 | 0 | 0 | .816 | 1 | .958 | .054 | 1 | .042 | 6.297 | -1.060 |
| | 11 | 0 | 0 | .907 | 1 | .911 | .014 | 1 | .089 | 4.666 | -.711 |
| | 12 | 0 | 0 | .373 | 1 | .638 | .793 | 1 | .362 | 1.922 | .063 |
| | 13 | 0 | 0 | .537 | 1 | .766 | .380 | 1 | .234 | 2.756 | -.211 |
| | 14 | 0 | 0 | .446 | 1 | .702 | .581 | 1 | .298 | 2.293 | -.065 |
| | 15 | 1 | 0** | .620 | 1 | .812 | .246 | 1 | .188 | 3.173 | -.332 |
| | 16 | 1 | 1 | .816 | 1 | .887 | .054 | 0 | .113 | 4.177 | 1.216 |
| | 17 | 1 | 1 | .702 | 1 | .848 | .146 | 0 | .152 | 3.590 | 1.067 |
| | 18 | 1 | 0** | .620 | 1 | .812 | .247 | 1 | .188 | 3.170 | -.331 |
| | 19 | 1 | 1 | .557 | 1 | .778 | .345 | 0 | .222 | 2.854 | .861 |
| | 20 | 1 | 1 | .560 | 1 | .981 | .340 | 0 | .019 | 8.181 | 2.032 |
| | 21 | 1 | 1 | .036 | 1 | .999 | 4.387 | 0 | .001 | 19.109 | 3.543 |
| | 22 | 1 | 1 | .037 | 1 | .999 | 4.357 | 0 | .001 | 19.046 | 3.536 |

**. Caso mal clasificado

**CUADRO No.13**

**Resultados de la clasificación[a]**

| | | | Grupo de pertenencia pronosticado | | |
|---|---|---|---|---|---|
| | | VC | .00 | 1.00 | Total |
| Original | Recuento | .00 | 14 | 0 | 14 |
| | | 1.00 | 2 | 6 | 8 |
| | % | .00 | 100.0 | .0 | 100.0 |
| | | 1.00 | 25.0 | 75.0 | 100.0 |

a. Clasificados correctamente el 90.9% de los casos agrupados originales.

En este caso, empleando solo una variable explicativa, el 90.9% de los casos resultó bien la clasificación, a diferencia de cuando se emplean todas las variables que se clasificaron correctamente un 95.5%. Para decidir con cuáles funciones se deberá trabajar para efectos de clasificar nuevos casos, debe observarse además el número de grados de libertad que se pierden en comparación con el número de observaciones disponibles al emplear más de una variable explicativa. Como recomendación, podría realizarse un nuevo cálculo con la opción "todas las explicativas juntas", pero eliminando las obligaciones con el público que inicialmente no fue significativa.

A manera de conclusión global de esta sección, puede decirse que la variable que más aporta a la discriminación entre bancos de altas y bajas utilidades es ingresos por concepto de intermediación financiera, seguido de la cartera de créditos vigente, variables que en el fondo están relacionadas una con otra, los ingresos por comisiones por servicios aportan un poco a la discriminación, pero en lo que respecta a las obligaciones que los bancos mantienen con el público no hay aporte importante a la discriminación.

## *IX. CONSIDERACIONES FINALES*

En un contexto en el cual el Banco Central de Costa Rica está mostrando interés por profundizar en el conocimiento y estudio del sector financiero del país, resulta provechoso el conocimiento de nuevas técnicas que permitan abordar diferentes temas relacionados con este sector, en forma más ágil y técnica.

En este sentido, la advertencia que realizan algunos analistas sobre los métodos tradicionales de estudio de temas financieros, concretamente el uso de razones financieras, hacen necesario tomar en consideración la recomendación de emplear el análisis discriminante para identificar instituciones o comportamientos atípicos.

En Costa Rica, se ha empleado escasamente esta metodología, únicamente existe referencia de trabajos desarrollados por Dr. Arnoldo Camacho, en el área bancaria[11].

Finalmente, del ejercicio práctico realizado se tiene que: de las variables consideradas, la que más aporta a la discriminación entre bancos de altas y bajas utilidades es ingresos por concepto de intermediación financiera, seguido de la cartera de créditos vigente, variables que en el fondo están relacionadas una con otra, los ingresos por comisiones por servicios aportan un poco a la discriminación, pero en lo que respecta a las obligaciones que los bancos mantienen con el público no hay aporte importante a la discriminación.

Los resultados anteriores no deben ser considerados como absolutos dado que se trata de un ejercicio para describir la técnica expuesta a lo largo del documento, sin embargo si dan luz acerca de la importancia de estudiar la banca privada costarricense no solo en forma global, sino por grupos afines, además de analizar otros determinantes de las utilidades, además de los ya comentados.

[11] Se dispone de resúmenes de dos trabajos del Dr. Camacho publicados por el INCAE.

# *X. BIBLIOGRAFÍA*

Alfaro V, Gabriela y Muñoz Evelyn. (1988). “Bancos privados: análisis discriminante por área de riesgo y su relación con el rendimientos”. Proyecto de graduación para optar por el título de Magister en Economía. Universidad de Costa Rica, Escuela de Economía.

Camacho, Arnoldo. (1996). “Mercados Financieros en la Encrucijada: la reforma y supervisión prudencial en América Latina”. Alajuela INCAE/Programa de estudios y entrenamiento económico. S.1: programa de finanzas rurales The Ohio State University.

Gujarati, Damodar. (1992). “Econometría”. McGraw-Hill Interamericana de México S.A. de C.V. México.

Manual y Guía de Usuario. Paquete estadístico SPSS. Versión 7.0

Madala G.S.(1996). “Introducción a la Econometría”. Prentice-Hall Hispanoamericana S.A. México.

munozse@bccr.fi.cr

F:\...\NT\NT98\NT0398.DOC