

Cuadernos del CIMBAGE N° 11 (2009) 39-57

UTILIZACIÓN DEL ANÁLISIS DISCRIMINANTE LOGÍSTICO PARA EXPLORAR LAS CAUSAS DE LA EFICIENCIA DEL SISTEMA BANCARIO ARGENTINO

Claudia B. Peretto

Facultad de Ciencias Económicas – Universidad Nacional de Córdoba

Av. Valparaíso s/n - Ciudad Universitaria - Córdoba

cperetto@eco.unc.edu.ar

Recibido 23 de octubre de 2008, aceptado 22 de diciembre de 2008

Resumen

En el presente trabajo se desarrolla uno de los Métodos de Clasificación Supervisada conocido con el nombre de Análisis Discriminante Logístico o Regresión Logística. Se comentan en forma general los Métodos de Clasificación Supervisada y las reglas de clasificación para dos poblaciones. Luego, se hace un breve resumen del Método de Análisis Discriminante Lineal como nexa para el desarrollo del Análisis Discriminante Logístico. Los conceptos expresados sobre este tema son aplicados a una muestra de 66 entidades bancarias de la República Argentina con la finalidad de explorar las causas que expliquen la clasificación de las mismas en eficientes o ineficientes.

Palabras claves: Análisis Discriminante Logístico, entidades bancarias, eficiencia.

**USE OF LOGISTIC DISCRIMINANT ANALYSIS TO EXPLORE
THE CAUSES OF THE ARGENTINE BANKING SYSTEM'S
EFFICIENCY**

Claudia B. Peretto

Facultad de Ciencias Económicas – Universidad Nacional de Córdoba
Av. Valparaíso s/n - Ciudad Universitaria - Córdoba
cperetto@eco.unc.edu.ar

Received 23 October 2008, accepted 22 December 2008

Abstract

In this work it is developed one of the well-known Supervised Classification Methods, the Logistic Discriminant Analysis or Logistic Regression. The Supervised Classification Methods and the classification rules for two populations are commented in general form. Then, a brief summary of the Linear Discriminant Analysis Method is made as a nexus for the Logistic Discriminant Analysis development. All concepts used to express about this topic are applied to a sample of 66 bank entities of the Argentine Republic in order to explore the causes that divide the classification in efficient or inefficient.

Keywords: Logistic Discriminant Analysis, bank entities, efficiency.

1. INTRODUCCIÓN

Discriminación y clasificación son técnicas multivariadas relacionadas con la *separación* de conjuntos de distintos objetos (u observaciones) y con la *asignación* de nuevos objetos (u observaciones) a grupos previamente definidos. Estas técnicas reciben el nombre de *Métodos de clasificación supervisada* para indicar que conocemos una muestra de elementos bien clasificados que sirve como pauta o modelo para la clasificación de las siguientes observaciones.

El planteo estadístico es el siguiente: a partir de una muestra de entrenamiento, es decir, un conjunto de objetos cuya pertenencia a uno de los grupos preestablecidos se conoce, se deriva una regla que permite asignar cada una de las observaciones a uno de los grupos mutuamente excluyentes, minimizando la probabilidad de clasificar incorrectamente a los individuos.

Los objetivos inmediatos de discriminación y clasificación son, respectivamente, los siguientes:

- describir, gráfica o algebraicamente, las características diferenciales de los objetos (observaciones) de varios grupos conocidos (poblaciones). Se trata de encontrar “discriminantes” cuyos valores numéricos sean tales, que los grupos sean separados tanto como sea posible.
- clasificar objetos (observaciones) en dos o más clases identificadas. El énfasis radica en derivar una regla que pueda ser usada para asignar óptimamente objetos nuevos a las clases identificadas.

Estos objetivos frecuentemente se superponen, y la distinción entre separación y asignación se torna confusa.

2. REGLAS DE CLASIFICACIÓN PARA DOS POBLACIONES

Las unidades consideradas pueden ser separadas o clasificadas a partir de mediciones de, por ejemplo, p variables aleatorias $X^t = [x_1, \dots, x_p]$. Los valores observados de X difieren entre una clase y otra¹. Se pueden considerar dos poblaciones, que llamaremos G_1 y G_2 , que pueden ser descritas por sus funciones de densidad de probabilidad $f_1(\mathbf{x})$ y $f_2(\mathbf{x})$. Podemos pensar en la totalidad de los

¹ Si los valores de X no fueran muy diferentes para objetos en clases distintas, no habría problemas; es decir, las clases o poblaciones serían indistinguibles y un nuevo objeto podría asignarse a una u otra indistintamente.

valores de la primera clase como la población de los valores de X para G_1 y los de la segunda clase como la población de los valores de X para G_2 y, en consecuencia hablar, de asignación de observaciones a poblaciones u objetos a clases indistintamente.

Antes de introducirnos en el tema, es preciso definir algunos conceptos:

- **Probabilidad a priori:** probabilidad de que una observación seleccionada en forma aleatoria provenga de G_i :

$$\pi_i (i = 1, 2, \dots, g)$$

- **Función de probabilidad condicional a la clase:** función de probabilidad del vector X de variables en el grupo i -ésimo:

$$f(\mathbf{x}/i) = f_i(\mathbf{x})$$

- **Probabilidad a posteriori:** las reglas de decisión permiten obtener una estimación de la probabilidad condicional de que la clase sea i , fijado el vector de variables que corresponde a un elemento:

$$pr(Y = i/\mathbf{x}) \quad (i = 1, 2, \dots, g)$$

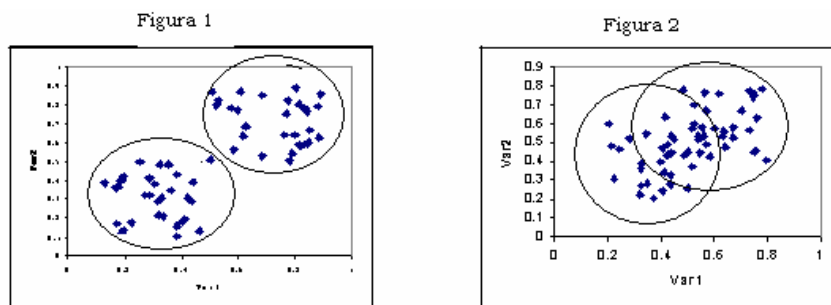
- **Probabilidad de clasificar mal:** probabilidad de asignar al grupo j una observación perteneciente al grupo i -ésimo seleccionada aleatoriamente:

$$pr(Y = j/\mathbf{x} \in G_i) = pr(j/\mathbf{x} \in G_i)$$

Las reglas de asignación o clasificación son generalmente desarrolladas a partir de muestras de estudio. Las características medidas de objetos seleccionados aleatoriamente, conocidas por provenir de cada una de las dos poblaciones, son examinadas para diferenciarlas.

Esencialmente, el conjunto de todas las salidas muestrales posibles es dividido en dos regiones, R_1 y R_2 , tal que si una nueva observación cae en R_1 , ésta es asignada a la población G_1 , y si cae en R_2 , la asignamos a la población G_2 .

Las reglas de clasificación proporcionan métodos de asignación que no están libres de error. Esto se debe a que puede no haber una clara distinción entre las características medidas de las poblaciones, es decir, los grupos pueden superponerse. Entonces, es posible clasificar objetos de la población G_2 como pertenecientes a G_1 u objetos de G_1 como pertenecientes a G_2 .



Figuras 1 y 2. Gráfico de dos variables medidas sobre una muestra de entrenamiento

En la Figura 1 se muestra un ejemplo de dos poblaciones totalmente separadas, mientras que en la Figura 2 se presenta un caso de dos poblaciones bastante superpuestas.

En este último caso, se pueden cometer uno de los dos tipos de errores:

- Asignar \mathbf{x} a G_2 cuando pertenece a G_1 , cuya probabilidad se denota $pr(2/\mathbf{x} \in G_1)$,
- asignar \mathbf{x} a G_1 cuando pertenece a G_2 , cuya probabilidad se denota $pr(1/\mathbf{x} \in G_2)$.

Un buen procedimiento de clasificación daría por resultado pocas malas clasificaciones. En otras palabras, las probabilidades de clasificar mal serían pequeñas.

Otro aspecto de la clasificación es el *costo*. Un procedimiento de clasificación óptimo debería, siempre que sea posible, considerar los costos asociados con a malas clasificaciones. Supongamos que clasificar objetos de G_1 como pertenecientes a G_2 representa un error más serio que clasificar objetos de G_2 como pertenecientes a G_1 . En estos casos, uno debería ser precavido al realizar la asignación. Por ejemplo, no diagnosticar una potencial enfermedad fatal es substancialmente más costoso que concluir que la enfermedad está presente cuando, en realidad, no lo está. Se puede especificar el costo de cometer cada uno de los errores antes mencionados y denotarlos como $C(2/\mathbf{x} \in G_1)$ y $C(1/\mathbf{x} \in G_2)$ respectivamente (Tabla 1).

		Grupo asignado	
		G_1	G_2
Grupo verdadero	G_1	0	$C(2/\mathbf{x} \in G_1)$
	G_2	$C(1/\mathbf{x} \in G_2)$	0

Tabla 1. Costo de cometer errores

El costo esperado de mala clasificación (ECM) es obtenido multiplicando los costos de los errores por sus probabilidades de ocurrencia, calculadas a partir de la probabilidad de cometer cada error, las que se ponderan por la probabilidad *a priori* de ocurrencia en cada uno de los grupos:

$$\begin{aligned} \text{ECM} &= C(2/\mathbf{x} \in G_1)p_1 p(2/\mathbf{x} \in G_1) + C(1/\mathbf{x} \in G_2)p_2 p(1/\mathbf{x} \in G_2) \\ &= C(2/\mathbf{x} \in G_1)p_1 \int_{R_2} f_1(\mathbf{x})dx + C(1/\mathbf{x} \in G_2)p_2 \int_{R_1} f_2(\mathbf{x})dx \end{aligned}$$

El menor valor de ECM se obtiene asignando a R_1 aquellos valores de X para los que se cumple la siguiente desigualdad (ver Anderson, 1984):

$$C(2/\mathbf{x} \in G_1)p_1 f_1(\mathbf{x}) > C(1/\mathbf{x} \in G_2)p_2 f_2(\mathbf{x})$$

De la última expresión se deriva que las regiones óptimas R_1 y R_2 son definidas por los valores de X que satisfacen las siguientes desigualdades:

$$\begin{aligned} R_1: \hat{y}=1 \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} &\geq \frac{p_2 C(1/\mathbf{x} \in G_2)}{p_1 C(2/\mathbf{x} \in G_1)} & R_2: \hat{y}=2 \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} &< \frac{p_2 C(1/\mathbf{x} \in G_2)}{p_1 C(2/\mathbf{x} \in G_1)} \end{aligned}$$

Debido a la naturaleza arbitraria del establecimiento de costos de mala clasificación, en la mayoría de las aplicaciones los mismos se consideran iguales y se define la probabilidad total de clasificar mal (TPM):

$$TPM = p_1 p(2/\mathbf{x} \in G_1) + p_2 p(1/\mathbf{x} \in G_2)$$

El menor valor de TPM se obtiene asignando a R_1 aquellos valores de X para los que se cumple que:

$$p_1 f_1(\mathbf{x}) \geq p_2 f_2(\mathbf{x})$$

$$R_1 : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{p_2}{p_1}$$

$$\hat{y} = 1$$

$$R_2 : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{p_2}{p_1}$$

$$\hat{y} = 2$$

3. ANÁLISIS DISCRIMINANTE LINEAL

Cuando se asume Normalidad Multivariada Homocedástica, se obtiene una combinación lineal de las variables que se conoce como Función Discriminante Lineal.

Si se supone que $\mathbf{X} : N_p(m_i, S)$ en $G_i (i=1,2)$ con $S_1 = S_2 = S$,

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} |S|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-m_i)' S^{-1}(\mathbf{x}-m_i)}$$

La regla de clasificación se obtiene comparando el cociente de verosimilitudes con el punto de corte. Si en caso que las probabilidades *a priori* sean iguales, la regla asignará la observación a G_1 si la distancia a su media es menor que la distancia a la de G_2 . Cuando las probabilidades son distintas, se asignará a G_1 si la distancia a G_2 es mayor que la distancia al primero más dos veces $\ln \frac{p_2}{p_1}$.

$$L(\mathbf{x}) = (\mathbf{m}_1 - \mathbf{m}_2)' S^{-1} \mathbf{x} - \frac{1}{2} (\mathbf{m}_1 - \mathbf{m}_2)' S^{-1} (\mathbf{m}_1 - \mathbf{m}_2) \geq \ln \frac{p_2}{p_1}$$

en forma equivalente:

$$Z(\mathbf{x}) = \ln \frac{p_2}{p_1} + \frac{1}{2} (\mathbf{m}_1 - \mathbf{m}_2)' S^{-1} (\mathbf{m}_1 - \mathbf{m}_2) - (\mathbf{m}_1 - \mathbf{m}_2)' S^{-1} \mathbf{x} \geq 0$$

$$Z(\mathbf{x}) = b_0 + b_1 \mathbf{x}$$

donde $b_0 = \ln \frac{p_2}{p_1} + \frac{1}{2} (\mathbf{m}_1 - \mathbf{m}_2)' S^{-1} (\mathbf{m}_1 - \mathbf{m}_2)$ y $b_1 = (\mathbf{m}_1 - \mathbf{m}_2)' S^{-1}$.

Luego de desarrollar una regla de clasificación, es posible evaluarla calculando su "tasa de error". Si se dispone de información relativa a los costos, se puede analizar el costo esperado de clasificar mal. Pero generalmente esta información no es fácil de obtener, por ello el criterio más empleado es evaluar la probabilidad de clasificar mal.

Si bien esto permite llegar a buenos resultados, existen muchas aplicaciones donde los datos no son normales y algunas o todas las variables son categóricas o cualitativas.

4. ANÁLISIS DISCRIMINANTE LOGÍSTICO

Es un procedimiento semiparamétrico que si bien no requiere asumir una forma específica funcional para $f_i(\mathbf{x})$ $i = 1, 2, \dots, g$, debe cumplir algunos supuestos para su correcta aplicación. Se utiliza en situaciones en las que las variables predictoras no están distribuidas normalmente, y algunas o todas las variables son discretas o categóricas.

Recordando la regla de asignación que surge de minimizar la probabilidad total de clasificar mal (TPM) para una clasificación binaria, una observación se asigna a G_1 si se cumple que:

$$p_1 f_1(\mathbf{x}) \geq p_2 f_2(\mathbf{x})$$

$$R_1 = \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{p_2}{p_1} \qquad R_2 = \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{p_2}{p_1}$$

Aplicando el modelo *logit* y suponiendo que las probabilidades a priori son las mismas:

$$P(y=1/x_i) = \frac{P(x_i/y=1)P(y=1)}{P(x_i)} = C \prod_{j=1}^p p_{1j}^{x_{ij}} (1-p_{1j})^{1-x_{ij}}$$

donde $c = P(y=1)P(y=1)/P(\mathbf{x}_i)$. La transformación logística será:

$$g_i = \log \frac{P(y=1/x_i)}{1-P(y=1/x_i)} = \sum x_{ij} \log \left(\frac{p_{1j}}{p_{2j}} \right) + \sum (1-x_{ij}) \log \left(\frac{1-p_{1j}}{1-p_{2j}} \right)$$

que es una función lineal en las variables que podemos escribir como:

$$g_i = b_0 + b_1' x_i$$

Los coeficientes de la función lineal definida anteriormente son iguales a:

$$b_0 = \sum \log \left(\frac{1-p_{1j}}{1-p_{2j}} \right)$$

$$b_1' = \sum \log \left(\frac{p_{1j}(1-p_{2j})}{p_{2j}(1-p_{1j})} \right)$$

Una ventaja adicional de este modelo es que cuando las variables son normales también verifican el modelo logístico.

4.1. Interpretación de coeficientes

Si bien los coeficientes en una regresión logística son b_0 y b_1' , lo que se interpreta es $\exp(b_0)$ y $\exp(b_1')$, conocidos como cocientes de "Odds ratios" y definidos como la probabilidad de que el evento ocurra.

La probabilidad de que el evento ocurra (*odds ratio*) se define como el cociente entre la probabilidad de que el evento ocurra $p(y=1/x) = p(x)$ y la probabilidad de que no ocurra $p(y=0/x)$:

$$\text{Odds ratio} = \frac{p(x)}{1-p(x)} = e^z$$

$$\ln \frac{p(x)}{1-p(x)} = Z$$

Los coeficientes estimados se interpretan como el cambio en el logaritmo del *odds* por un cambio unitario en la variable independiente asociada, permaneciendo las demás variables constantes.

El cociente de *odds* para una variable independiente x_h dicotómica es igual a:

$$\text{cociente de odds} \quad \Psi = \frac{\frac{\pi(x_h=1)}{1-\pi(x_h=1)}}{\frac{\pi(x_h=0)}{1-\pi(x_h=0)}} = e^{\beta_h} \quad \ln \Psi = \beta_h$$

El valor de Ψ indica cuánto aumenta (o disminuye) la probabilidad de que se presente el evento entre los individuos que pertenecen a la categoría $\mathbf{x}=1$, en relación con la de los que corresponden a la categoría de referencia ($\mathbf{x}=0$). La interpretación de los coeficientes varía según el tipo de variable independiente considerada:

✓ **Variables cuantitativas**

Indica cuánto se modifica el cociente de probabilidad cuando la variable x_h aumenta en una unidad.

✓ **Variables cualitativas**

Indica cuánto se modifican las probabilidades de que se presente el evento si cambia la categoría de la variable x_h en relación a la categoría de referencia.

5. APLICACIÓN

En la aplicación del Análisis discriminante logístico, se deberían seguir los siguientes pasos:

- 1) Definir el problema bajo estudio. Establecer la variable dependiente, la cual debe ser una variable categórica binaria y las variables independientes.
- 2) Dividir la muestra en dos submuestras: una para estimación de los coeficientes y otra para validación del modelo.
- 3) Estimar el modelo de regresión logística y determinar las variables independientes significativas para predicción.
- 4) Obtener los coeficientes de la función discriminante logística.
- 5) Clasificar los individuos de cada grupo de acuerdo con la función obtenida y evaluar la precisión de la clasificación en la muestra de estimación.
- 6) Evaluar la precisión de la clasificación en la muestra de validación.

5.1. Definición del problema y variables a utilizar

En este trabajo se realizará una aplicación del Análisis Discriminante Logístico en los bancos de la República Argentina.

Sobre el particular se puede decir que el sistema financiero de un país puede ser descrito como un conjunto de unidades denominadas

intermediarios financieros (bancos o entidades financieras), que realizan transacciones con cierto tipo de elementos denominados activos financieros, bajo las normas y el control de una autoridad - Banco Central o entidad similar-, que ejerce las funciones de supervisión y de autoridad de aplicación de las normas correspondientes. Los intermediarios financieros cumplen una doble función: por un lado, contribuyen a crear un mecanismo de pago para las transacciones económicas de una sociedad, en adición al constituido por la propia moneda; por el otro, actúan en la intermediación de activos financieros, efectuando un arbitraje de montos y plazos, y asumiendo el riesgo de que sus deudores resulten insolventes.

Las mencionadas entidades financieras deben presentar balances, cuentas de ganancias y pérdidas, y demás documentación referida a su estado económico financiero al Banco Central de la República Argentina. Este publica esta información en forma periódica y detallada para cada entidad. Las variables y los datos a utilizar serán obtenidos de estas publicaciones. Para el año 2006, se cuenta con información relativa a 66 entidades bancarias incluyendo bancos públicos y privados. Las variables disponibles para realizar el estudio son:

✓ **Variables numéricas:**

ACTIVO: total del activo de la entidad bancaria, medido en millones de pesos.

PRÉSTAMOS: volumen de préstamos efectuados por cada entidad, saldo del balance al 31 de diciembre, medido en millones de pesos.

INGRESOS POR SERVICIOS: mide en forma indirecta, por intermedio de los ingresos que genera, los otros servicios proporcionados por la entidad (distintos de los que generan depósitos y préstamos).

EGRESOS POR SERVICIOS: mide las erogaciones de la entidad financiera, pero vinculados a la prestación de otros servicios no financieros tales como cobro de impuestos, corresponsalías de valores, apertura de cuentas de depósitos, seguros, tarjetas de crédito, etc.

GASTOS ADMINISTRATIVOS: mide diferentes ítems relacionados a la administración del negocio (luz, agua, teléfono, impuestos sobre inmuebles propios y fundamentalmente sueldos del personal).

CAPITAL: total de capital, aportes y reservas, medido en millones de pesos. Este rubro junto con Resultados no Asignados forma el patrimonio neto de cada entidad.

PATRIMONIO NETO: total de patrimonio neto, medido en millones de pesos.

INGRESOS FINANCIEROS: monto de ingresos financieros (originados por los préstamos otorgados), medido en millones de pesos.

EGRESOS FINANCIEROS: monto de egresos financieros (originados por los depósitos recibidos), medido en millones de pesos.

CARGO POR INCOBRABILIDAD: comprende las pérdidas netas originadas en la imposibilidad real o potencial de recuperar créditos provenientes de la intermediación habitual entre la oferta y la demanda de recursos financieros.

✓ **Variables categóricas:**

EFICIENCIA DEA: eficiencia técnica relativa de cada unidad con respecto al conjunto de todas las unidades analizadas. Este índice es obtenido a través de la aplicación del método DEA (*Data Envelopment Analysis*). Asume el valor 1 cuando la entidad es eficiente y 2 cuando es ineficiente.

EMPLEADOS: dotación de personal de cada institución, medido en número de personas empleadas. Esta variable tiene cuatro categorías.

DEPÓSITOS: volumen de depósitos recibidos por cada entidad, saldo del balance al 31 de diciembre, medido en millones de pesos. Esta variable tiene tres categorías.

TAMAÑO: clasificación de las entidades bancarias en pequeñas, medianas y grandes de acuerdo con el patrimonio neto, y con el volumen de depósitos y préstamos. Esta variable tiene tres categorías.

ORIGEN DE CAPITAL: clasificación de las entidades según el origen de su capital, que puede ser público, privado nacional y privado extranjero. Esta variable tiene tres categorías.

Sería de interés obtener una regla de clasificación que permita determinar si una nueva entidad tiene más probabilidades de ser eficiente o ineficiente, y además conocer qué variables predicen mejor la eficiencia bancaria.

En cuanto a las variables a utilizar:

✓ **Variable respuesta:**

La variable respuesta debe ser una variable categórica binaria. En este caso, la variable respuesta será *eficiencia dea*.

✓ **Variabes independientes:**

Se estudiaron distintos modelos, incluyendo distintas combinaciones de variables numéricas y categóricas. Se calculó la correlación entre las variables numéricas, lo que permitió observar que todas las variables están altamente correlacionadas.

Para completar el análisis en busca del modelo más adecuado, se procesó en SPSS la regresión logística utilizando el método Backward: Wald, según el cual las variables predictoras, variables independientes que se deberían incluir en el estudio, son: Préstamos (numérica), Ingresos por Servicios (numérica), Depósitos y Origen de Capital (categóricas).

5.2. Submuestras de estimación y validación

La muestra disponible de 66 entidades bancarias ha sido dividida en dos submuestras de estimación y validación, cuidando de mantener en cada una de ellas la misma proporción de unidades eficientes e ineficientes. La submuestra de estimación posee 51 casos y la de validación, 15 (22,7% del total). Ver ANEXO 1, las filas sombreadas corresponden a la muestra de validación.

5.3. Estimación del modelo de regresión logística

Se procesó con el software SPSS v.10.0.1 el modelo de regresión logística con las variables que se indicaron en el apartado 5.1. Es importante destacar que en las variables categóricas Empleados, Depósitos y Tamaño se utilizó como categoría de referencia la última, mientras que en origen de Capital, la categoría de referencia fue la primera. Se obtuvieron los resultados que se muestran a en las tablas 2, 3 y 4.

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	51	77.3
	Missing Cases	0	.0
	Total	51	77.3
Unselected Cases		15	22.7
Total		66	100.0

Tabla 2. Casos procesados

a. If weight is in effect, see classification table for the total number of cases.

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^b		
			EFICIENC		Percentage Correct	EFICIENC		Percentage Correct
Eficiente	Ineficiente	Eficiente	Ineficiente					
Step 1	EFICIENC	Eficiente	35	1	97.2	9	1	90.0
		Ineficiente	4	11	73.3	3	2	40.0
Overall Percentage					90.2			73.3

a. Selected cases SELEC EQ 1

b. Unselected cases SELEC NE 1

c. Tie criterion is .500

Tabla 3. Clasificación

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1						
PRÉSTAMO	.010	.005	4.655	1	.031	1.010
ING_SS	-.432	.189	5.208	1	.022	.649
OR_CAP			3.396	2	.183	
OR_CAP(1)	5.708	3.133	3.320	1	.068	301.187
OR_CAP(2)	5.176	2.948	3.083	1	.079	176.971
DEPOSITO			5.411	2	.067	
DEPOSITO(1)	-8.075	3.878	4.336	1	.037	.000
DEPOSITO(2)	.859	1.839	.218	1	.641	2.360
Constant	2.129	2.083	1.044	1	.307	8.404

a. Variable(s) entered on step 1: PRESTAMO, ING_SS, OR_CAP, DEPOSITO.

Tabla 4. Variables y resultados

5.4. Análisis de resultados y conclusiones

Si bien las tasas totales de clasificación correcta son altas tanto para la muestra de estimación como para la muestra de validación, en la de validación son algo inferior, ascendiendo a un 73,3%. En cuanto a la clasificación en cada grupo, se observa una baja tasa de clasificación correcta en el grupo ineficiente. El modelo diseñado tiende a clasificar las unidades como eficientes, lo que constituye una deficiencia del modelo formulado.

En lo que respecta a las variables de este modelo seleccionado, ellas resultan significativas al 0,05 o al 0,10. Analizando los cocientes de probabilidades (*odds ratios*) que se muestran en la columna $exp(b)$ podemos sacar algunas conclusiones:

✓ Préstamos (variable numérica): por cada peso que aumente esta variable, aumentará la probabilidad de que la unidad considerada sea ineficiente.

✓ Ingresos por Servicios (variable numérica): por cada peso que aumente esta variable, disminuirá la probabilidad de que la unidad considerada sea ineficiente.

✓ Depósitos (variable categórica): hay más probabilidad de ser ineficiente si pertenece a la categoría 2 que si pertenece a la categoría 3. Es decir, si el volumen de sus depósitos está entre 130 y 500 pesos.

✓ Origen de capital (variable categórica): la variable no es significativa, pero en el análisis de los *odds ratios* se puede ver que tiene más probabilidad de ser ineficiente un banco privado nacional que uno privado extranjero.

Según las variables que mejor predicen la ineficiencia bancaria, para el modelo diseñado serían Préstamos, Ingresos por Servicios y Depósitos.

BIBLIOGRAFÍA

Anderson, D.; Sweeney, D.; Williams, T. (1984). *Estadística para Administración y Economía*. Cengage Learning Editores, México D.F.

Banco Central de la República Argentina (2003). *Información de Entidades Financieras*- Diciembre 2002. Publicación oficial. Buenos Aires.

Díaz, M. (2005). "Notas de Cátedra" – *Magister en Estadística Aplicada*. Córdoba.

Fisher, R.A. (1936). "The use of multiple measurement in taxonomic problems". *Annales Eugenics* N° 7, pp. 179-188.

Hosmer, D.; Lemeshow, S. (1990). *Applied Logistic Regression*. John Wiley and Sons, New York.

Johnson, R.A.; Wichern, D.W. (1998). *Applied Multivariate Statistical Analysis*. Fourth Edition. Prentice Hall, Upper Saddle River, New Jersey.

Peña, D. (2002). *Análisis de datos multivariantes*. McGraw – Hill, Madrid.

Peretto, C., Pérez Mackeprang, C. (2005-A). "Reducción del número de variables utilizando un método de Estadística Multivariada". *Revista de*

la Escuela de Perfeccionamiento en Investigación Operativa N° 25, pp. 93-112.

Peretto, C.; Alberto de Azcona, C. (2005-B). "Ordenación de unidades de acuerdo a su eficiencia: estudio comparativo de los métodos DEA y Análisis de Componentes Principales". *Anales 34° JAIIO* (Jornadas Argentinas de Informática e Investigación Operativa) – SIO 2005 (Simposio de Investigación Operativa). Rosario.

Peretto, C. ; Pérez Mackeprang, C. (2005-C). "Análisis Discriminante de Ratios: Evaluación de eficiencia y ordenación de los Bancos de la República Argentina". *Anales del XVIII Encuentro Nacional de Docentes de Investigación Operativa*, Córdoba.

ANEXO I

AÑO 2002	Efi- ciencia	Préstamo	Ing. Serv.	Depó- sito	Orig. Cap.
ABN AMOR Bank N.V.	1	311.50	55.20	2	4
Banca Nazionale del Lavoro S.A.	1	1611.90	145.70	3	2
Banco B.I. Creditanstalt S.A.	2	271.40	0.50	1	4
Banco Bansud S.A.	1	588.40	64.20	3	4
Banco CMF S.A.	2	184.60	3.30	1	4
Banco Comafi Sociedad Anónima	2	105.00	9.60	2	2
Banco Credicoop Coop. Limitado	1	1191.90	71.40	3	2
Banco de Corrientes S.A.	2	112.30	7.00	2	1
Banco de la Ciudad de Buenos Aires	1	3518.60	94.70	3	1
Banco de la Nación Argentina	1	13554.60	458.20	3	1
Banco de La Pampa	2	332.80	13.20	3	1
Banco de la Prov. de Buenos Aires	1	7668.90	428.60	3	1
Banco de la Provincia de Córdoba	1	1412.30	80.70	3	1
Banco de la Rep. Or. del Uruguay	1	4.10	0.80	1	2
Banco de Santa Cruz S.A.	2	122.40	14.20	2	2
Banco de Servicios Financieros S.A.	1	18.40	3.70	1	4
Banco de Valores S.A.	2	49.00	7.30	2	2
Banco del Chubut S.A.	2	164.00	8.40	2	1
Banco del Sol S.A.	1	7.60	0.40	1	2
Banco del Tucumán S.A.	2	68.70	10.70	2	2
Banco do Brasil S.A.	2	10.00	2.20	1	2
Banco Emp. de Tucumán Coop Ltda.	2	122.60	1.90	2	2
Banco Europeo para Am. Latina S.A.	1	113.70	9.70	1	2
Banco Finansur S.A.	1	24.10	1.30	1	2
Banco Itaú Buen Ayre S.A.	1	624.60	54.70	2	4
Banco Julio Sociedad Anónima	1	10.40	0.40	1	2
Banco Macro S.A.	1	278.00	73.70	3	4
Banco Municipal de La Plata	1	64.50	11.70	2	1

Banco Privado de Inversiones S.A.	1	33.80	8.10	1	2
Banco Provincia de Tierra del Fuego	1	53.80	6.50	2	1
Banco Provincia del Neuquén S.A.	1	199.30	47.40	2	1
Banco Regional de Cuyo S.A.	2	69.70	7.90	2	2
Banco Río de la Plata S.A.	1	6213.40	360.50	3	4
Banco Roela S.A.	1	25.30	2.20	1	2
Banco Saenz S.A.	2	57.50	7.20	1	2
Banco San Luis S.A.	2	135.90	22.80	2	2
Banco Soci��t�� G��n��rale S.A.	1	540.50	71.50	3	4
Bank of America	1	0.00	2.60	1	4
BankBoston	1	2674.10	241.40	3	4
BBVA Banco Franc��s S.A.	1	6874.70	325.00	3	4
Deutsche Bank S.A.	1	53.20	26.50	2	4
HSBC Bank Argentina S.A.	2	2674.90	68.90	3	4
LLOYDS TSB Bank PLC	1	261.00	50.10	3	4
MBA Banco de Inversiones S.A.	1	4.60	2.70	1	2
Nuevo Banco Bisel S. A.	1	105.30	30.50	3	4
Nuevo Banco de Entre R��os S.A.	1	56.90	19.20	3	4
Nuevo Banco de La Rioja S.A.	1	10.00	11.00	1	2
Nuevo Banco de Santa Fe S.A.	1	666.80	73.60	3	2
Nuevo Banco del Chaco S.A.	1	92.00	17.30	2	1
Nuevo Banco Suqu��a S.A.	1	109.00	40.90	3	4
Providian Bank S.A.	1	6.20	9.60	1	2
Banco Bradesco Argentina S.A.	1	40.10	0.30	1	4
Banco de Formosa S.A.	1	17.50	12.00	1	2
Banco de Galicia y Buenos Aires S.A.	1	9186.90	340.20	3	2
Banco de San Juan S.A.	2	42.30	24.20	2	2
Banco de Santiago del Estero S.A.	1	100.10	15.20	2	2
Banco Hipotecario S.A.	1	2918.00	94.30	2	2
Banco Mariva S.A.	1	35.90	4.40	1	2
Banco Municipal de Rosario	1	29.40	13.20	1	1

Banco Patagonia S.A.	2	96.70	13.70	2	2
Banco Piano S.A.	2	56.70	4.40	2	2
Banco Sudameris Argentina S.A.	1	1211.90	127.60	3	4
BNP Paribas	2	176.70	26.90	2	4
CITIBANK N.A.	1	3586.30	297.60	3	4
ING BANK	1	141.30	10.70	1	4
<u>Nuevo Banco Industrial de Azul S.A.</u>	<u>2</u>	<u>52.90</u>	<u>5.60</u>	<u>1</u>	<u>2</u>

Las filas sombreadas corresponden a la submuestra de validación.