

# El análisis factorial y el análisis discriminante en la estimación de la pérdida esperada para una institución financiera

**Factor Analysis and Discriminant Analysis in the Estimation of Expected Loss in a Financial Institution**

**A análise fatorial e a análise discriminante na estimação da perda esperada para uma instituição financeira**

DOI rces.v25n38.a10

Recibido: 25/08/2017

Aceptado: 12/10/2017

## **Armando Lenin Támara Ayús**

Docente e investigador Asociado al Departamento de Finanzas de la Universidad EAFIT. Economista Universidad de Antioquia. Esp. En Diseño y Evaluación de Proyectos Universidad del Norte. Msc. En Finanzas EAFIT. Estudiante del Doctorado en Administración de la Universidad de Medellín. Tel 2619500 Ext. 9937 atamaraa@eafit.edu.co

## **Angélica María Villarraga Peña**

Contador Público U. del Valle de Cali. Maestría en Administración financiera. Universidad EAFIT. Angelica.villarraga.p@gmail.com

## **Yelitza Carolina Vera Álvarez**

Contador Público U. Libre de Cali. Maestría en Administración financiera. Universidad EAFIT. Carolina0756@gmail.com  
Correspondencia: Universidad EAFIT. Bloque 26 Oficina 508. Carrera 49 N° 7 Sur-50. Medellín-Colombia.

- 1 Este artículo es resultado de un proyecto de investigación realizado en la Maestría en Administración Financiera de la Universidad EAFIT.

## **El análisis factorial y el análisis discriminante en la estimación de la pérdida esperada para una institución financiera**

### **Palabras clave**

Probabilidad de incumplimiento, análisis factorial, análisis discriminante, pérdida esperada.

### **Resumen**

Este trabajo presenta un análisis de los diferentes estudios que se han hecho alrededor del riesgo de crédito, y, con base en la información de una cartera comercial perteneciente a una institución financiera, se ofrece un modelo para pronosticar la probabilidad de incumplimiento de un deudor, mediante la utilización del análisis factorial como seleccionador de las variables en el modelo y, del análisis discriminante, como herramienta para encontrar las funciones discriminantes que permitan determinar al deudor como incumplido y cumplido.

**Clasificación JEL: C15, C52, C53, G32**

## **Factor Analysis and Discriminant Analysis in the Estimation of Expected Loss in a Financial Institution**

### **Keywords**

Probability of default, Discriminant analysis, Factor Analysis, Expected loss.

### **Abstract**

This project presents an analysis of the different studies that have been conducted on the topic of credit risk. Based on the information from the commercial file of a financial institution, a model has been developed to predict the probability of default using factor analysis as a selector of model variables and discriminant analysis as a tool to find the discriminating functions which make it possible to determine whether a customer is in compliance or not.

## **A análise fatorial e a análise discriminante na estimação da perda esperada para uma instituição financeira**

### **Palavras-chave**

Probabilidade de incumprimento, Análise Fatorial, Análise Discriminante, Perda Esperada.

### **Resumo**

Este trabalho apresenta uma análise dos diferentes estudos que se há feito ao redor do risco de crédito, e, com base na informação de uma carteira comercial pertencente a uma instituição financeira, se oferece um modelo para prognosticar a probabilidade de incumprimento de um devedor mediante a utilização da análise fatorial como selecionador das variáveis no modelo e da análise discriminante como ferramenta para encontrar as funções discriminantes que permitam determinar ao devedor como descumprido e cumprido.

## Introducción

El sector bancario tiene como principal actividad la intermediación financiera, enfocada hacia el otorgamiento de créditos; y sus mayores esfuerzos están orientados a la gestión del riesgo de crédito para garantizar una rentabilidad que compense la inversión realizada. El riesgo de crédito, en lugar de evitarse, debe gestionarse de manera apropiada. Razón por la cual, a lo largo del tiempo, los reguladores han estructurado normas con el objeto de que las instituciones financieras cuenten con el capital necesario para respaldar posibles pérdidas. Como consecuencia del impago del deudor en las operaciones de intermediación crediticia, una de las regulaciones más aplicadas en el mundo es el Acuerdo de Capital de Basilea, de obligatorio cumplimiento por medio de los bancos centrales de los países firmantes del acuerdo.

En la actualidad, el riesgo de crédito tiene como objetivo administrar y minimizar el riesgo para lograr una rentabilidad acorde con el nivel de riesgo asumido, mediante la aplicación, al mismo tiempo, de un capital adecuado y la aplicación de la normatividad vigente. Se han desarrollado diversos modelos como el logit, probit, matrices de probabilidad de transición, cadenas de Markov, logit multinomial, entre otros, con el fin de estudiar y aplicar la forma de reducir el riesgo en las entidades financieras, permitir la cuantificación anticipada de las posibles pérdidas en que pueden llegar a incurrir por la generación de provisiones, y de entregar elementos de juicio para la toma de decisiones en el manejo del riesgo crediticio.

Este estudio, a través de la aplicación de los modelos discriminante y factorial, busca calcular la probabilidad de incumplimiento de una entidad financiera, permitiendo comparar lo anterior con la pérdida registrada por la entidad y la solicitada por

los entes reguladores. Según Pérez (2007), el análisis factorial tiene como objeto simplificar las múltiples y complejas relaciones que pueden existir entre un conjunto de variables observadas  $X_1, X_2, \dots, X_p$ . Para ello, se trata de encontrar dimensiones comunes o factores que ligan a las variables no relacionadas en apariencia. Los mismos autores plantean que el análisis discriminante es una técnica que resulta útil para las situaciones en las que se desea construir un modelo predictivo para pronosticar el grupo al que pertenece una observación, a partir de determinadas características observadas que delimitan su perfil. Se trata de una técnica estadística que permite asignar o clasificar nuevos individuos u observaciones dentro de grupos definidos con anticipación, razón por la cual es una técnica de clasificación ad hoc.

## 1. Antecedentes

Debido a la importancia del tema, diferentes autores han estudiado metodologías que al aplicarlas tienen como propósito estimar la probabilidad de incumplimiento. Uno de los primeros investigadores fue Altman (1968), quien aplicó el análisis discriminante para estudiar el riesgo del crédito y la probabilidad de que las empresas lleguen a un estado de liquidación, tomando como insumos el balance general y el estado de resultados; concluyó que la terminación de una empresa se ve afectada por tres razones financieras entre las que están: la rentabilidad del activo, el apalancamiento y el flujo de efectivo. Lennox (1999) estudio la probabilidad de default aplicando los modelos logit y probit para posteriormente compararlos con el análisis discriminante elaborado por Altman (1968), concluyendo que una entidad tiene mayor probabilidad de default cuando presenta una baja rentabilidad, altos niveles de apalancamiento y problemas con el flujo de caja. Finalmente, se deter-

minó que estos modelos tienen una mejor explicación y capacidad de predicción que el modelo de análisis discriminante.

Alves (2004) estimó, para el sector corporativo en la Unión Europea, un modelo de cointegración VEC, con el fin de establecer la relación que existe entre algunas variables macroeconómicas y la probabilidad de incumplimiento esperada. Los resultados señalan que las variables macroeconómicas influyen en la determinación de los valores del estado estacionario de la probabilidad de incumplimiento esperada de cada sector y ayudan a identificar el perfil de riesgo de los deudores. Wong J. (2005) hizo un análisis de sensibilidad para el portafolio de créditos de los bancos de Hong Kong, en el que se afirma que la probabilidad de incumplimiento de un deudor se encuentra influenciada por los cambios en los factores macroeconómicos. Sin embargo, esta estimación tenía una debilidad al no permitir involucrar cambios en diversas variables. Las falencias que tenía el análisis propuesto por Wong J. (2005), las soluciona Hoggarth (2005) a través de las funciones de impulso de un modelo VAR, con el que se clarificó que existe una relación negativa entre el PIB y el indicador de mora; adicional, analizó el impacto del incremento de la tasa de interés e inflación; sin embargo, estas no tienen un alto impacto como el PIB.

Koopman, Lucas, y Klaassen (2005) realizaron un estudio en empresas estadounidenses buscando la relación entre la probabilidad de incumplimiento de los deudores, el PIB real y el margen de tasas de interés, deduciendo que los ciclos de riesgo del crédito se presentan simultáneamente con el ciclo económico; sin embargo, el riesgo de crédito fluctúa a la inversa de la actividad económica. Por su parte, Ruano (2006) empleó un modelo de selección de Heckman en compañías españolas para analizar la probabilidad de incumplimiento y sus determinantes. Los resultados evidenciaron que una empresa

puede disminuir la probabilidad de incumplimiento en la medida que incrementa el nivel de cobertura, la rentabilidad y la liquidez.

A nivel nacional, Zapata (2003) utilizó las matrices de probabilidades de transición, teniendo como objetivo hallar las probabilidades de transición para la cartera comercial. Este análisis arrojó que las probabilidades de transición y de default son regularmente distintas en cada estado del ciclo cuando se segmenta con base en este. Posteriormente, Amaya (2005) estimó la relación de largo plazo que existe entre el PIB, las tasas de interés, el precio de la vivienda nueva y los indicadores de mora de la cartera hipotecaria y de consumo. Los resultados del estudio evidenciaron que la actividad económica y las tasas de interés influyen altamente en el comportamiento de los indicadores de mora.

Zamudio (2007) utilizó el modelo logit multinomial para estimar la probabilidad de incumplimiento de las empresas, soportando que el incumplimiento de las mismas propende más cuando tienen bajos niveles de liquidez y de rentabilidad. Gutiérrez Rueda (2008) incluye la variable tasa de desempleo al modelo de Amaya (2005) como una variable de análisis para las modalidades de crédito de la cartera comercial. Se evidenció la relación negativa entre el indicador de mora de cada cartera y la actividad económica, siendo el desempleo una variable influyente a largo plazo en el indicador de mora.

Siguiendo los estudios, tenemos a Gómez González (2009) quien aplica un modelo de duración, con el objeto de encontrar la probabilidad de incumplimiento en la cartera comercial del sistema financiero colombiano, arrojando que, la migración de los créditos hacia bajas calificaciones, se ven afectados por el comportamiento de la liquidez, la composición de la deuda, el tamaño y la eficiencia. Más adelante, el mismo Gómez González (2009), con el objeto de

probar los supuestos Markovianos, utiliza matrices de transición involucrando variables como edad, PIB, tasa de interés real y composición de la deuda, concluyendo que para el SARC es importante la herramienta de matrices de transición. Gómez González y Kiefer (2009) evalúan con matrices de transición en tiempo homogéneo, el comportamiento del crédito modelando variables como la liquidez, tamaño, costo de oportunidad, composición de la deuda y el PIB, evidenciando que la condición económica del país tiene influencia sobre la probabilidad de incumplimiento.

Estudios más recientes muestran la búsqueda de nuevas metodologías para pronosticar la probabilidad de incumplimiento, este es el caso de Calabrese y Osmetti (2014) quienes desarrollan un modelo basado en la teoría de eventos binarios, para hacer la pronosticación en las PYMES italianas de la probabilidad de incumplimiento, obteniendo mayor precisión frente al modelo logístico.

Un estudio innovador es el de Jang y Yoon (2016), desarrollando un modelo estructural que incorpora el riesgo macroeconómico para pronosticar la probabilidad de incumplimiento para empresas con altas calificaciones de crédito, concluyen que los factores macroeconómicos son importantes en el modelo y aportan al pronóstico.

Luo y Nie (2016) desarrollan un modelo de supervivencia discreto, basado en splines de regresión, adaptado a la banca minorista; dado que estos modelos poseen capacidad para analizar datos censurados, y a la vez, complementan el modelo Cox existente. Los resultados indican que el modelo proporciona una estadística similar y mejora la predicción para el modelo de desgaste que tiene baja tasa de eventos.

Sousa, Gama y Brandao (2016) proponen un nuevo modelo dinámico para la evaluación del riesgo de cré-

dito, que amplía los modelos de calificación de crédito predominantes, basados en parámetros estáticos de datos históricos para tarjetas de crédito de una institución financiera en Brasil. La contribución del estudio es proporcionar la idea de que diferentes cantidades de memoria se pueden explorar simultáneamente, las cuales se utilizan para la estimación.

Costin, Gordy, Huang y Szerszen (2016) desarrollan dos enfoques novedosos para resolver la transformación de Laplace de un proceso estocástico, cambiado en el tiempo. Las aplicaciones de los métodos permiten introducir el cambio de tiempo estocástico a la clase estándar de modelos de riesgo de crédito por defecto, y demuestran que el cambio de tiempo estocástico tiene un efecto muy grande en la fijación de precios de opciones profundas fuera del dinero en swaps de incumplimiento crediticio.

Con el análisis de la literatura, se puede deducir que, cada vez más, los estudios relacionados a la modelación de riesgo de crédito aportan al desarrollo del pronóstico de la probabilidad de incumplimiento, lo que permite disminuir la provisión de la pérdida esperada sin que esto conlleve al incremento del riesgo de la entidad y, por lo tanto, acrecentado la rentabilidad a través de la disponibilidad de recursos para la inversión.

## 2. La pérdida esperada

El proceso de crédito llevado a cabo por una entidad financiera se encuentra soportado en el conocimiento que esta tiene, o espera tener, del cliente a través de la información que aporta el solicitante, mediante el diligenciamiento de la solicitud de crédito y la entrega de los estados financieros que soportan la actividad generadora de ingresos, su nivel de inversiones, la

capacidad de pago y la capacidad de endeudamiento. Con esta información, el banco procede de acuerdo a las políticas que se tienen para su análisis de crédito, identificando variables cualitativas y cuantitativas que permitan tomar una decisión favorable o desfavorable a la solicitud de crédito hecha por el cliente; teniendo en cuenta, además de lo mencionado anteriormente, su historial de pagos ya sea con la entidad, el sector financiero, el sector real o con otras que considere pertinente la entidad.

La pérdida esperada la define Wilson y Press (1978) como el monto de capital que podría perder una institución como resultado de la exposición crediticia en un horizonte de tiempo dado. Por lo tanto, la probabilidad de incumplimiento de un cliente está dada cuando este alcance una altura de mora  $n$ , en la cual la institución financiera asume la pérdida del capital. De tal forma que la pérdida esperada se calcula como:

$$PE = PI * S * E$$

Donde:

**PI:** Probabilidad de Incumplimiento para el período dado. Mide la probabilidad que el prestatario incumpla el pago de su crédito durante un horizonte temporal determinado.

**S:** Severidad o pérdida en caso de incumplimiento. Calcula la proporción de la exposición que se perdería si se produjera el incumplimiento. Tiene en cuenta las garantías y la variabilidad tanto de la exposición como de la garantía.

**E:** Exposición al incumplimiento. Se calcula como el monto legalmente adecuado al banco (bruto de provisiones y amortizaciones parciales).

Al aplicar esta definición, las pérdidas esperadas aumentan en función del monto del crédito y de la

probabilidad de deterioro asignada a cada activo, por lo tanto, las pérdidas serán menores cuanto más baja sea la severidad. Por otro lado, es de anotar que las provisiones de capital se basan en el resultado obtenido de esta ecuación y son precisamente las reglamentadas por la Superintendencia Financiera de Colombia.

### 3. Serie de datos

La información utilizada para el caso de estudio corresponde a una muestra de 1.500 clientes con obligaciones en una institución financiera, a los cuales se les hizo seguimiento durante un período de 12 meses (agosto del 2009 a julio del 2010); para cada uno de ellos la entidad posee registros de las características que los identifica. Entre la información disponible aparece: número de obligación, valor desembolsado, fecha de desembolso, inversión, tipo de cartera, activos, pasivos, ingresos, edad, actividad económica, garantías y calificación del deudor de acuerdo con los días de mora que registraba en ese momento.

#### 3.1 Definición de variables

Las variables seleccionadas en este estudio fueron:

1. **Margen Operativo;** margen operativo de la persona  $i$ . Indicador que permite evaluar lo que queda de la actividad económica del cliente luego de cubrir los costos y gastos para atender parte del servicio a la deuda, provisión de impuestos y generar utilidades; está dado por la relación de Utilidad Operativa/ Ingresos. Se espera que a mayor margen operativo disminuya la probabilidad de incumplimiento.
2. **Margen Neto;** margen neto de la persona  $i$ . Indicador que permite evaluar lo que le queda al cliente luego de haber cubierto su costo, gasto,

carga financiera y haber realizado su respectiva provisión para pago de impuesto. Se espera que a mayor margen neto disminuya la probabilidad de incumplimiento.

3. **Activos<sub>i</sub>**; valor estimado de los activos que tiene la persona *i*. A través de esta variable se puede identificar cuál es la base real de respaldo que tienen los clientes del banco, ya que son los activos los que reflejan el nivel de inversión y a su vez miden el tamaño o estructura, frente al nivel de obligaciones contraídas. Por lo tanto, se debe presentar una relación inversa entre esta variable y la probabilidad de incumplimiento.
4. **Edad<sub>i</sub>**; 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.
5. **Endeudamiento<sub>i</sub>**; endeudamiento de la persona *i*. Indicador que refleja la capacidad de maniobra para llevar a cabo inversiones apalancadas, que le permitan crecer en su nivel de activos. Está dado por la relación pasivo total sobre activo total. Se espera que el nivel de endeudamiento esté ligado a las inversiones, las cuales generarán un crecimiento en los ingresos, por lo que la relación esperada entre esta variable y la probabilidad de incumpliendo debe ser negativa.

## 4. Análisis factorial

Como punto de partida, se aplica el análisis factorial al conjunto de variables de estudio (margen operativo, margen neto, activos, edad, endeudamiento) y se analizan las correlaciones entre estas variables. La Tabla 1 muestra la matriz de correlación donde se puede observar que no existen correlaciones significativas. La correlación más alta se presenta entre las variables *activos* y *margen operativo* (0,463); sin embargo,

sigue siendo baja. Lo anterior muestra un indicio de que no será clara la identificación de las variables que conforman cada factor, recurriendo, por lo tanto, a técnicas de rotación. Ver Tabla 1, siguiente página.

La Tabla 2 nos muestra el valor del determinante de la matriz de correlaciones de 0.487, el cual no es cercano a cero, confirmando que el grado de intercorrelación entre las variables es bajo. En cuanto al valor del estadístico KMO (0.676) no es tan alto, lo que permite determinar que la aplicación del análisis factorial, al conjunto de datos observados, es adecuada.

La Tabla 3 (siguiente página) muestra que la varianza explicada por el primer y segundo factor es 41.54% y 22.13% respectivamente, para una variabilidad total explicada del 63.67%; un valor considerable para el estudio.

Los coeficientes de correlación entre cada dos variables, después de estar en función de los factores y conocidos como coeficientes de correlación reproducidos, se muestran en la Tabla 4. Según Pérez (2007), estos coeficientes no tienen por qué coincidir con los de la matriz de correlación inicial, pero no deben diferenciarse en más de 0.05. Es así como ninguno de los errores supera el umbral de 0.05, indicando que la bondad de ajuste de las variables es buena. Ver tabla 4 siguiente página.

Para identificar qué variables son representadas por cada factor se aplica una rotación Varimax, cuya técnica permite conservar incorrelacionados los factores. La Tabla 5 muestra que las variables *margen operativo*, *margen neto*, *activos* y *endeudamiento* se encuentran representadas en el primer factor y la variable *edad*, en el segundo factor. De esta manera, se puede decir que el primer factor agrupa variables de tipo financiero y el segundo factor agrupa variables de tipo demográfico.

**Tabla 1.** Matriz de correlaciones

		Margen Operativo	Margen Neto	Activos	Edad	Endeudamiento
Margen Operativo	Correlación de Pearson	1	0,316	,463	,240**	0,114*
	Sig. (bilateral)		0,000	0,000	0,000	0,012
	N	1500	1500	1500	1500	1500
Margen Neto	Correlación de Pearson	0,316**	1	,415	,381**	,100*
	Sig. (bilateral)	0,000	0,000	0,000	0,000	0,051
	N	1500	1500	1500	1500	1500
Activos	Correlación de Pearson	,463**	,415	1	,221**	,185**
	Sig. (bilateral)	0,000	0,000	0,000	0,000	0,000
	N	1500	1500	1500	1500	1500
Edad	Correlación de Pearson	0,24**	0,381	,221	1	-0,092
	Sig. (bilateral)	0,000	0,000	0,000	0,000	0,067
	N	1500	1500	1500	1500	1500
Endeudamiento	Correlación de Pearson	,114**	0,100	,185	-0,092	1
	Sig. (bilateral)	0,012	0,051	0,000	0,067	
	N	1500	1500	1500	1500	1500

\*\*La correlación es significativa al nivel 0,01 (bilateral)  
+ La correlación es significativa al nivel 0,05 (bilateral)

Fuente: Elaboración propia. Resultados SPSS 21.

**Tabla 2.** Determinante de la matriz de correlación y prueba de esfericidad de Bartlett

Matriz de correlaciones	Medida de adecuación muestral de Kaiser-Meyer-Olkin		0,676	
Determinante	0,487	Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	285,155
			gl	10
			Sig.	0,000

Fuente: Elaboración propia. Resultados SPSS 21.



**Tabla 3.** Varianza total explicada

Factor	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	2,077	41,543	41,543	1,655	33,108	33,108
2	1,107	22,130	63,673	0,759	15,188	48,295
3	0,746	14,914	78,587			
4	0,584	11,674	90,261			
5	0,487	9,739	100,000			

Método de extracción: Mínimos cuadrados no ponderados.

Fuente: Elaboración propia. Resultados SPSS 21.

**Tabla 4.** Coeficientes de correlación reproducidos

		Margen Operativo	Margen Neto	Activos	Edad	Endeudamiento
Correlación reproducida	Margen Operativo	,371	0,338	0,477	0,241	0,131
	Margen Neto	0,388	,339*	0,418	0,379	0,087
	Activos	0,477	0,418	,623(a)	0,221	0,187
	Edad	0,241	0,379	0,221	1,000(a)	-0,089
	Endeudamiento	0,131	0,087	0,187	-0,089	,083(a)
Residual	Margen Operativo		-0,002	0,004	-0,001	-0,008
	Margen Neto	-0,002		-0,004	0,002	0,013
	Activos	-0,004	-0,004		0,000	-0,002
	Edad	-0,001	-0,002	0,000		-0,003
	Endeudamiento	-0,008	-0,013	-0,002	-0,003	

Método de extracción: Mínimos cuadrados no ponderados.

a. Comunalidades reproducidas

b. Los residuos se calculan entre las correlaciones observadas y reproducidas. Hay 0 (0%) residuales no redundantes con valores absolutos mayores que 0,05.

Fuente: Elaboración propia. Resultados SPSS 21.

**Tabla 5.** Cargas de las variables en cada factor rotado

Matriz de factores rotados (a)		
	Factor	
	1	2
Margen Operativo	0,577	0,196
Margen Neto	0,47	0,343
Activos	0,773	0,161
Edad	0,079	0,997
Endeudamiento	0,265	-0,11

Método de extracción: Mínimos cuadrados no ponderados.  
Método de rotación: Normalización Varimax con Kaiser.  
a. La rotación ha convergido en 3 iteraciones.

Fuente: Elaboración propia. Resultados SPSS 21.

## 5. Análisis discriminante

Con base en los resultados del análisis factorial se procede a determinar qué variables pueden generar una mayor discriminación en el análisis, para esto se realiza una prueba de igualdad de medias, que permite determinar si el valor medio de una variable, en cada uno de los grupos, es estadísticamente diferente; donde:

$H_0$ : El valor medio de una variable es igual tanto en el grupo  $i$  como en el grupo  $j$

$H_1$ : El valor medio de una variable difiere en el grupo  $i$  y el grupo  $j$ .  $ij=1,2$

La Tabla 6 (siguiente página) arroja los resultados donde el p-valor de las pruebas realizadas es menor que 0,05, lo que indica que se rechaza la hipótesis nula. Esto quiere decir, que todas las variables, consideradas en el análisis, pueden ayudar a clasificar correctamente a una persona como cumplida o incumplida.

Finalmente, la Tabla 7 (siguiente página) muestra el resultado de la técnica de selección de variable, por lo tanto, las variables que se deben incluir en el modelo son: *margen neto, endeudamiento, margen operativo y edad.*

Para verificar que las variables sean estadísticamente significativas, y que realmente permiten efectuar una adecuada discriminación, se analiza el Lambda de Wilks. En este caso, la Tabla 8 (siguiente página) muestra que el p-valor es inferior a un nivel de significancia de 0.05, por lo tanto, se valida el supuesto.

Las funciones discriminantes están dadas en la Tabla 9 y permiten hacer pronósticos para nuevos clientes. Por lo tanto, las funciones discriminantes serían:

Para cumplidos (C):

$$F_c = -1,039 - 0,015(MO) - 0,538(MN) + 0,847(E) + 10,162(ENDE)$$

Para incumplidos (INC)

$$F_i = -5,291 + 1,421(MO) - 5,472(MN) + 0,587(E) + 18,346(ENDE)$$

De esta manera, el individuo que obtenga el mayor puntaje en una de las dos funciones pertenece a dicho grupo.

Tabla 9 siguiente página

La Tabla 10 muestra que la variabilidad total, explicada por el modelo, es del 85%. Lo cual era de esperarse ya que, como se mencionó anteriormente, el estadístico Lambda de Wilks arrojó una adecuada discriminación.

**Tabla 6.** Prueba de igualdad

Pruebas de igualdad de las medias de los grupos					
	Wilks' Lambda	F	gl 1	gl 2	Sig.
Margen Operativo	0,988	4,756	1	1498	0,030
Margen Neto	0,618	146,447	1	1498	0,000
Activos	0,949	21,406	1	1498	0,000
Edad	0,885	51,769	1	1498	0,000
Endeudamiento	0,931	29,545	1	1498	0,000

Fuente: Elaboración propia. Resultados SPSS 21.

**Tabla 7.** Variables en el análisis

Paso		Tolerancia	F para salir	Lambda de Wilks
1	Margen Neto	1,000	246,447	
2	Margen Neto	0,88	325,848	9,931
	Endeudamiento	0,88	82,558	0,618
3	Margen Neto	0,795	317,035	0,911
	Endeudamiento	0,879	78,876	0,606
	Margen Operativo	0,88	4,32	0,511
4	Margen Neto	0,765	258,523	0,827
	Endeudamiento	0,87	71,816	0,59
	Margen Operativo	0,858	5,68	0,507
	Edad	0,916	4,835	0,506

Fuente: Elaboración propia. Resultados SPSS 21.

**Tabla 8.** Lambda de Wilks

Paso	Número de variables	Lambda	gl 1	gl 2	gl 3	F exacta			
						Estadístico	gl 1	gl 2	Sig.
1	1	0,618	1	1	1498	246,447	1	1498	0,000
2	2	0,511	2	1	1498	189,754	2	1498	0,000
3	3	0,506	3	1	1498	129,000	3	1498	0,000
4	4	0,500	4	1	1498	98,896	4	1498	0,000

Fuente: Elaboración propia. Resultados SPSS 21.

**Tabla 9.** Coeficientes de las funciones canónicas discriminantes

	Default	
	Cumplidos	Incumplidos
Margen Operativo	-0,15	1421
Margen Neto	-0,538	-5472
Edad	0,847	0,587
Endeudamiento	10162	18346
Constante	-1039	-5291

Fuente: Cálculos propios. Resultados SPSS 21.

Es importante aclarar que los individuos se clasifican en cada grupo según las probabilidades que tienen, a priori, de pertenecer a los mismos. Una vez conocidas las puntuaciones discriminantes, cada individuo se clasificará en el grupo que tenga mayor probabilidad, a posteriori, de pertenecer según sus puntuaciones discriminantes (Pérez, 2007).

En la Tabla 11 (siguiente página) se puede observar que, en general, la clasificación correcta es del 85%; por su parte, el modelo está clasificando mejor los verdaderos cumplidos (99,28%), en comparación a los verdaderos incumplidos (50,79%). Sin embargo, el porcentaje de clasificación es significativamente alto, por lo que se puede concluir que el resultado es bueno y la clasificación será adecuada.

Tomando como referencia los resultados del análisis discriminante, se procede a realizar una comparación entre las probabilidades de default encontradas y las que otorga la SFC, al igual que la aplicada por la institución. Por lo tanto, dados los resultados que se muestran en la Tabla 12, se puede concluir que 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.

**Tabla 10.** Autovalores

Función	Autovalor	Autovalores		Correlación canónica
		% varianza	% acumulado	
1	1,001(a)	100	100	0,85

a. Se han empleado las 1 primeras funciones discriminantes canónicas en el análisis.

Fuente: Elaboración Propia. Resultados SPSS 21.

**Tabla 11.** Resultados de clasificación

		Default Cumplidos	Grupo de pertenencia pronosticado		Total
			Cumplidos	Incumplidos	
Original	Recuento	Cumplidos	1110	8	1118
		Incumplidos	188	194	382
	%	Cumplidos	99,28	0,72	100
		Incumplidos	49,21	50,79	100

a. Clasificados correctamente el 85% de los casos agrupados originales.

Fuente: Elaboración Propia. Resultados SPSS 21.

**Tabla 12.** Probabilidad de *Default*

MES	INSTITUCIÓN	PE(SUPER)	PE(MODELO)	PE Efectiva
ENERO	\$ 3.908.331.270	\$ 3.591.727.551	\$ 1.306.477.908	\$ 1.126.354.862
FEBRERO	\$ 3.136.733.876	\$ 2.882.828.475	\$ 1.407.163.215	\$ 956.012.365
MARZO	\$ 2.541.016.078	\$ 2.335.460.095	\$ 847.324.148	\$ 789.654.123
ABRIL	\$ 2.077.075.845	\$ 1.909.132.526	\$ 691.961.079	\$ 594.160.837
MAYO	\$ 1.712.555.191	\$ 1.574.134.833	\$ 570.087.728	\$ 479.632.158
JUNIO	\$ 1.423.589.698	\$ 1.308.552.361	\$ 473.616.257	\$ 379.628.513
JULIO	\$ 1.192.481.087	\$ 1.906.130.930	\$ 396.559.840	\$ 302.165.946
AGOSTO	\$ 1.006.026.102	\$ 924.743.177	\$ 334.460.780	\$ 284.397.052
SEPTIEMBRE	\$ 854.312.976	\$ 7785.283.940	\$ 283.979.171	\$ 210.930.628
OCTUBRE	\$ 729.851.755	\$ 670.871.975	\$ 242.596.080	\$ 198.364.752
NOVIEMBRE	\$ 626.943.584	\$ 576.270.977	\$ 208.398.608	\$ 159.884.573
DICIEMBRE	\$ 541.221.489	\$ 497.467.978	\$ 179.923.650	\$ 100.926.415

Fuente: Cálculos propios.

## Conclusiones

Por medio de la técnica de análisis factorial se logró determinar que a partir de dos factores es posible reducir la dimensionalidad de las variables observadas. En el inicio no fue fácil identificar cuáles variables estaban representadas en cada factor; sin embargo, por medio del procedimiento de rotación Varimax fue posible determinar que las variables margen operativo, margen neto, activos y endeudamiento se encontraban representadas en el primer factor y la variable *edad*, en el segundo. Así, el primero agrupa variables de tipo financiero y el segundo otras de tipo demográfico. Con este resultado es posible realizar otra clase de análisis multivariante, como el de conglomerados, a partir, únicamente, de los factores obtenidos considerados como nuevas variables.

El desarrollo del modelo por medio del análisis discriminante arrojó que las variables son significativas desde el punto de vista estadístico, lo que permite realizar una adecuada significación que hace relación a los resultados de la investigación teórica; de manera adicional, lleva a cabo una clasificación correcta de los datos en el 85%.

Es importante precisar que en el análisis discriminante, al ajustar el modelo por medio de la selección de variables, se utilizaron las variables incluidas. Se destacan el endeudamiento, los ingresos y los activos como las variables que influyen en mayor medida en el incumplimiento de un deudor.

Los resultados del modelo arrojaron una menor provisión por pérdida esperada para la entidad, comparado con los registros actuales y la exigencia de los entes reguladores. Dicha herramienta se suministra a la entidad para análisis de aplicación, lo cual le permitirá

mejorar sus resultados financieros, sin exponer la entidad a un riesgo mayor.

## Referencias

- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4), 589-609.
- Alves, I. (2004). Sectoral fragility: factors and dynamics (Vol. 22, pp. 450-480): Bank for International Settlements.
- Amaya, C. A. (2005). Evaluación del Riesgo de Crédito en el sistema financiero Colombiano (pp. 67 - 78). Reporte de Estabilidad Financiera: Banco de la República.
- Calabrese, R. y Osmetti, S. A. (2014). A Generalized Additive Model for Binary Rare Events Data: An Application to Credit Defaults. *Analysis & Modeling of Complex Data In Behavioral & Social Sciences*, 73. doi:10.1007/978-3-319-06692-9\_9
- Costin, O., Gordy, M. B., Huang, M. y Szerszen, P. J. (2016). Expectations of functions of stochastic time with application to credit risk modeling. *Mathematical Finance*, 26(4), 748-784. doi:10.1111/mafi.12082
- Gomez Gonzalez, J. E., y Kiefer, N. M. (2009). Evidence of a Non-Markovian Behavior in the Process of Bank Rating Migrations. *Cuadernos de economía*(133), 33-50.
- Gutiérrez Rueda, J. (2008). Un análisis de cointegración para el riesgo de crédito. In D. Vásquez (Ed.). Bogotá: Reporte de Estabilidad Financiera.
- Gómez Gonzalez, J. E. (2009). Un Modelo de Alerta Temprana para el sistema Financiero Colombiano. In P. I. Orozco Hinojosa (Ed.): Borradores de Economía.
- Gómez González, J. E. (2009). Estimation of conditional time homogeneous credit quality transition matrices for commercial banks in Colombia. In I. P. Orozco Hinojosa (Ed.). Borradores de Economía: Bogotá.
- Hoggarth, G., Sorensen, S., y Zicchino L. (2005). Stress tes of UK banks using VAR approach (Working Paper no. 282 ed.): Bank of England.

- Jang, B., Rhee, Y. y Yoon, J. H. (2016). Business cycle and credit risk modeling with jump risks. *Journal of Empirical Finance*, 39(Part A), 15-36. doi:10.1016/j.jempfin.2016.08.001
- Koopman, S. J., Lucas, A., y Klaassen, P. (2005). Empirical credit cycles and capital buffer formation. *Journal of Banking and Finance*, 29, 3159-3179. doi:10.1016/j.jbankfin.2005.01.003
- Lennox, C. S. (1999). The Accuracy and Incremental Information Content of Audit Reports in Predicting Bankruptcy. *Journal of Business Finance & Accounting*, 26(5/6), 757-778.
- Luo, S., Kong, X. y Nie, T. (2016). Interfaces with Other Disciplines: Spline based survival model for credit risk modeling. *European Journal of Operational Research*, 253869-879. doi:10.1016/j.ejor.2016.02.050
- Sousa, M. R., Gama, J. y Brandão, E. (2016). A new dynamic modeling framework for credit risk assessment. *Expert Systems with Applications*, 45341-351. doi:10.1016/j.eswa.2015.09.055
- Peréz, C. B., D. (2007). *Minería de datos técnicas y herramientas*: Magallanes: Editorial International Thomson Editores Spain Paraninfo.
- Ruano, S. a. S.-F., Vicente. (2006). Dilatoriness of the Banking Enterprise Debt in Spain, 1992-2003 (Morosidad De La Deuda Empresarial Bancaria En España, 1992-2003): Banco de Espana Research Paper No. WP-0622.
- Wilson, S., y Press, S. J. (1978). Choosing Between Logistic Regression and Discriminant Analysis. *Journal of the American Statistical Association*, 73(364), 699-705.
- Wong J., C., K., y Fong, T. (2005). A framework for macro stress testing the creditrisk of banks in Hong Kong: Hong Kong Monetary Authority Quarterly Bulletin.
- Zamudio, N. (2007). Determinantes de la probabilidad de incumplimiento de las empresas Colombianas (Vol. 466). Borradores de Economía: Banco de la República.
- Zapata, A. (2003). Modelando el riesgo de crédito en Colombia *Matrices de transición para la cartera comercial*. Apuntes de banca y finanzas: Asobancaria.