



# Modelación de la probabilidad de incumplimiento y cálculo de la pérdida catastrófica en una institución financiera en Colombia

## Modeling the probability of default and calculation of the catastrophic loss in a financial institution in Colombia

DOI: <https://doi.org/10.17981/econcuc.42.2.2021.Econ.2>

Artículo de estudio de caso.  
 Fecha de recepción: 20/09/2020  
 Fecha de devolución: 22/02/2021  
 Fecha de aceptación: 03/03/2021  
 Fecha de publicación: 15/03/2021

**Armando Lenin Támara Ayús**   
 Universidad EAFIT  
 Medellín, Antioquia (Colombia)  
[atamaraa@eafit.edu.co](mailto:atamaraa@eafit.edu.co)

**José Eduardo Segura Ramos**   
 Politécnico de Colombia  
 Medellín, Antioquia (Colombia)  
[jesebruar@eafit.edu.co](mailto:jesebruar@eafit.edu.co)

**Ignacio Emilio Chica Arrieta**   
 Universidad de Córdoba  
 Montería, Córdoba (Colombia)  
[ignaciochica@correo.unicordoba.edu.co](mailto:ignaciochica@correo.unicordoba.edu.co)

Para citar este artículo:  
 Támara-Ayús, L., Segura-Rámos, J. & Chica-Arrieta, I. (2021). Modelación de la probabilidad de incumplimiento y cálculo de la pérdida catastrófica en una institución financiera en Colombia. *Económicas CUC*, 42(2), 173–186. DOI: <https://doi.org/10.17981/econcuc.42.2.2021.Econ.2>

JEL: D31, E51, L81.

### Resumen

La pérdida esperada en una institución financiera es el monto de capital que se perdería producto de la exposición que tienen la deuda en el tiempo. Este trabajo se enfoca en modelar la probabilidad de incumplimiento para una cartera de crédito bajo dos escenarios de estudio, uno con nivel de mora normal y otro con un nivel de mora restringido. Se toma como referencia una base de datos perteneciente a una cartera de crédito de consumo, con una muestra de 5.000 obligaciones analizadas en el período de enero a diciembre del 2019. El método estadístico utilizado es la regresión logística basado en las variables financieras de liquidez y endeudamiento, más una variable no financiera como es la edad. Los resultados muestran un modelo con un nivel de ajuste global mayor al 85% en los dos escenarios de estudio, donde, la variable ingreso es la que más influencia posee sobre el modelo de regresión logística. Finalmente, se ratifica la aplicabilidad de la regresión logística como una herramienta estadística en la búsqueda de modelos de pronóstico, con lo cual se logra reducir la pérdida esperada en una cartera de crédito sin aumentar la exposición al riesgo.

**Palabras clave:** Probabilidad de incumplimiento; regresión logística; riesgo de crédito; liquidez; endeudamiento; ajuste global

### Abstract

The expected loss in a financial institution is the amount of capital that would be lost as a result of the exposure that the debt has over time. This work focuses on modeling the probability of default for a loan portfolio under two study scenarios, one with a normal default level and the other with a restricted default level. A database belonging to a consumer loan portfolio is taken as a reference, with a sample of 5,000 obligations analyzed in the period from January to December 2019. The statistical method used is the logistic regression based on the financial variables of liquidity and indebtedness, plus a non-financial variable such as age. The results show a model with a global adjustment level greater than 85% in the two study scenarios, where the income variable is the one that has the most influence on the logistic regression model. Finally, the applicability of logistic regression as a statistical tool in the search for forecasting models is ratified, with which it is possible to reduce the expected loss in a loan portfolio without increasing risk exposure.

**Keywords:** Probability of default; logistic regression; credit risk; liquidity; indebtedness; global adjustment

## INTRODUCCIÓN

Las entidades financieras fundamentan su operación en el crédito, siendo este su mayor fuente de generación de ingresos, por consiguiente, esto implica que es preciso que tengan un manejo adecuado de su riesgo crediticio. Asimismo, deben garantizar la maximización en la colocación del efectivo, sea en personas jurídicas o naturales, por lo tanto, se hace necesario que en el contexto de esta actividad se trabaje en la minimización del riesgo en que se incurre al realizar la recuperación de la cartera.

Chen, Ribeiro y Chen (2016), Imtiaz y Brimicombe (2017), Cappon, Gorenstein, Mignot y Manuel (2018) y Assef y Steiner (2020) plantean que las diferentes metodologías utilizadas en el otorgamiento y seguimiento del crédito en el sistema financiero han venido en constante desarrollo. No obstante, el objetivo final no ha cambiado y sigue siendo la cuantificación anticipada de las posibles pérdidas en las que pueda incurrir una institución financiera a través de la colocación de fondos, lo que es lo mismo, en el otorgamiento de créditos. Este estudio toma como referencia la definición otorgada por la Superfinanciera (2002), el cual define el riesgo crediticio como la probabilidad de que una institución financiera tenga pérdidas no esperadas, lo que implica una reducción en el valor de sus activos como consecuencia del incumplimiento por parte de sus clientes en los contratos de créditos adquiridos.

El estudio, análisis y modelación del riesgo de crédito poseen importancia académica y práctica relacionada con la aplicabilidad en la búsqueda de un modelo de pronóstico de la probabilidad de incumplimiento que logre reducir el nivel de pérdida esperada en cualquier tipo de cartera correspondiente a una institución financiera. Chen et al. (2016), Durovic (2017), Támara, Villegas y De Andrés (2019) y Assef y Steiner (2020) han encontrado a través de sus trabajos, que la técnica estadística más utilizada en relación al riesgo de crédito entorno a la probabilidad de incumplimiento es la regresión logística, razón por la cual en este trabajo se utiliza dicha técnica.

Este trabajo se presenta por secciones, así: La segunda sección hace referencia a los materiales y métodos estableciendo con total claridad la base de datos utilizada en el trabajo, discriminando la variable dependiente de las variables independientes y, especifica la técnica estadística utilizada, en este caso, la regresión logística. En la tercera sección se muestran los resultados con relación al modelo encontrado y las implicaciones de este sobre la pérdida esperada. La sección cuatro es una discusión en torno a los resultados obtenidos, y, se finaliza con la sección cinco donde se hacen conclusiones en relación con el efecto que el modelo tiene sobre la cartera de crédito.

## CONTEXTUALIZACIÓN TEÓRICA

Según Superfinanciera (2002), el riesgo crediticio es la probabilidad de que una institución financiera tenga pérdidas no esperadas que trae como consecuencia la disminución en el valor de sus activos, dado que sus clientes incumplen con los

pagos derivados del crédito. Así mismo, el sistema financiero establece la calificación crediticia que utilizan las instituciones para hacerse una primera idea de qué tan cumplido es el cliente, dichas categorías son A, B, C y D, catalogando esta última como default.

En torno al análisis del riesgo de crédito, el incumplimiento se considera uno de los elementos más importantes a la hora de realizar el análisis en torno de la pérdida esperada. **Castro y Soto (2017)** definen la Pérdida Esperada (PE) como el monto de capital que una institución financiera puede perder por la exposición que tiene de la deuda en un período de tiempo determinado. Es así como la pérdida por incumplimiento se dará cuando un cliente alcance una altura de mora  $n$ , con lo que se asume la pérdida de capital para el período analizado y su cálculo estará dado por la ecuación 1:

$$PE = PI * S * E \quad (1)$$

Donde  $S$  representa la severidad y corresponde al porcentaje del monto de la pérdida al que una institución financiera se expone si el cliente no paga sus obligaciones;  $E$  es la exposición y representa la cantidad de los activos que se comprometen frente al no pago durante el período establecido. Esto implica que la pérdida esperada de la institución se incrementa con el monto del crédito y con la probabilidad de deterioro de cada activo, por tanto, un bajo nivel de  $S$  implica menores pérdidas.

### MÉTODOLOGÍA

En el ámbito del sistema financiero todo proceso de crédito está soportado en el comportamiento histórico de los clientes y se complementan con la información que se obtiene a través del diligenciamiento de las solicitudes de crédito, al igual que los estados financieros que los solicitantes aportan a sus solicitudes. Con base en toda esta información, las instituciones financieras logran ubicar el origen de los ingresos, el grado de inversión, el poder de pago y el nivel de endeudamiento que posean al momento de la petición. Asimismo, las instituciones establecen internamente políticas y protocolos para el análisis de las solicitudes, y para ello clasifican la información en dos tipos de variables: cualitativas y cuantitativas, a ello se le suman algunos elementos de apoyo como el comportamiento de pago, la edad y el sector socioeconómico.

Los estudios de **Cappon et al. (2018)**, **Támara et al. (2019)** y **Assef y Steiner (2020)** ratifican que la técnica estadística más utilizada en los estudios del riesgo de crédito ha sido la regresión logística. Adicionalmente, los trabajos de **Affes y Hentati (2017)**, **Lohmann y Ohliger (2018)**, **Támara, Villegas, Leones y Salazar (2018)** y **Do Prado, Carvalho, Benedicto y Lima (2019)** han mostrado que esta técnica posee un nivel alto de predicción en torno a la probabilidad de incumplimiento.

Como lo indica su nombre, la regresión logística es una técnica estadística basada en un tipo de análisis que permite predecir los diferentes resultados sobre una variable categórica en función de unas variables predictoras. Esta técnica está enmarcada en un conjunto de modelos lineales generalizados (General linear models-

GLM) soportados en el uso de una función logística, por tal razón, toman el nombre de modelo Logit.

Siguiendo a **Do Prado et al. (2019)**, el modelo Logit binario tiene las siguientes condiciones establecidas en la ecuación 2:

$$\begin{aligned}
 P_i &= P(y_i = 1 | X_{i1}, X_{i2}, X_{i3}, \dots, X_{in}) \\
 P_i &= \frac{e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}}} \quad (2) \\
 P_i &= \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik})}}
 \end{aligned}$$

Donde la función logística es dada por la ecuación 3:

$$F(z) = \frac{1}{1 + e^{-z}} \quad (3)$$

A diferencia de los parámetros de regresión lineal, los de regresión logística se interpretan como efectos aditivos en el logaritmo de la razón de momios, es decir, como una unidad de cambio en la  $j$ -ésima variable explicada. En este trabajo, esta se considera como una variable dicotómica, por lo que los  $\beta_k$  son razones de momios, conocidas como odds ratios.

### *Muestra*

Para esta investigación se toma una muestra de 5 000 clientes que conforman una cartera de crédito de consumo, perteneciente a una entidad financiera ubicada en el departamento de Córdoba (Colombia), cuyo período de estudio va de enero a diciembre de 2019. Para cada uno de los clientes se cuenta con los estados financieros que permiten conocer sus indicadores de liquidez y endeudamiento, adicionalmente, se considera una variable no financiera como es la edad y su calificación crediticia.

Para la elaboración del modelo se procede a segmentar la base en dos muestras, de entrenamiento y de prueba. Para la segmentación se utiliza el paquete estadístico **SPSS 24** (Statistical Package for the Social Sciences) implementando el método de validación cruzada de  $n$ -folds, de tal manera que, el 60% es para el entrenamiento y el 40% restante se toma para la prueba.

### *Variable dependiente*

En el modelo esta variable será el incumplimiento y se denotada como  $Y$ , donde, se le dará valores de 1 si la persona incumple en el pago de sus obligaciones, es decir, se encuentra en mora, y, 0 en caso contrario. La institución financiera de referencia en el trabajo establece que una persona ha tenido incumplimiento en su pago cuando se ha rezagado en un término de tiempo mayor a 45 días.

### *Variable independiente*

Dada la información suministrada por la institución financiera, las variables explicativas del modelo serán los ingresos, el nivel de endeudamiento y la edad. La primera hace referencia al salario mensual registrado en su solicitud de crédito y respaldado por su declaración de renta o por un certificado no menor a 30 días de la empresa donde trabaja, la segunda, se calcula como el total de su pasivo sobre el total de su activo. En relación con la variable edad, esta se toma como los años cumplidos a la fecha de la solicitud del crédito. La selección de las variables se realiza con base en los estudios de [Trejo, Martínez y Venegas \(2017\)](#) y [Fantazzini y Zimin \(2020\)](#).

## RESULTADOS

La [Tabla 1](#) muestra los resultados de los parámetros que arroja el modelo de regresión logística con relación a cada una de las variables explicativas del modelo. Los coeficientes estimados poseen un nivel de significación del 5%, dado que su P-value es menor a 0.05, lo que quiere decir que son estadísticamente significativos. Por otra parte, de las tres variables explicativas, solo la variable ingreso dio signo positivo, lo que implica que, individuos que tengan ingresos altos tenderán a tener una menor probabilidad de incumplimiento. Con respecto a las variables endeudamiento y edad, estas arrojaron signo negativo, lo que implica que, a mayor nivel de endeudamiento mayor será la probabilidad de incumplimiento, esta misma implicación ocurre con los individuos que tengan mayor edad.

**TABLA 1.**  
*Estimadores del modelo logístico.*

VARIABLES	$\beta$	Error estándar	Estadístico Z	P-value
Ingreso	1.8488	0.3716	4.9755	0.0000
Edad	-1.2814	0.2265	-5.6565	0.0000
Endeudamiento	-0.5881	0.1294	-4.5463	0.0000
Constante	34.3290	8.0267	4.2769	0.0000

Fuente: Elaborado con la salida de [RiskSimulator \(2020\)](#).

Teniendo como referencia los coeficientes estimados que muestra la [Tabla 1](#), el modelo correspondiente a la regresión logística queda de la siguiente manera (formula 4):

$$p_i = \frac{1}{1 + e^{-(1,8488 \cdot \text{ingresos} - 1,2814 \cdot \text{edad} - 0,5881 \cdot \text{endeudamiento} + 34,3290)}} \quad (4)$$

### *Cálculo de la pérdida esperada*

Seguidamente, se procede a calcular la probabilidad de la pérdida esperada para cada uno de los clientes de la base de datos y a reorganizarlos en las cuatro carteras que se plantean según el porcentaje de la probabilidad de incumplimiento misma.

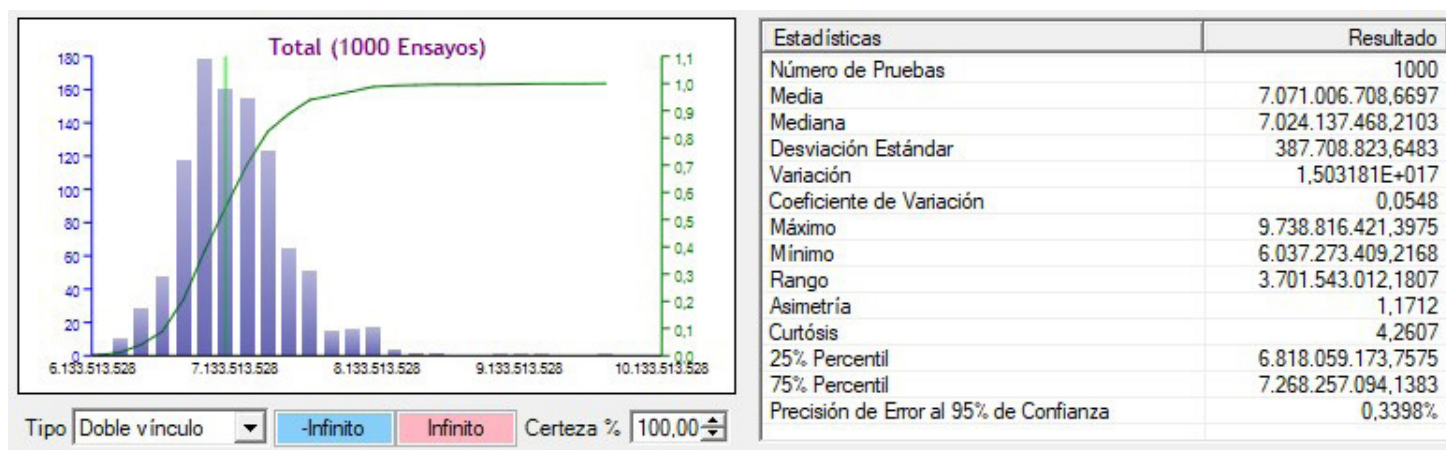
Así, en la cartera 1 se ubican los que poseen una probabilidad de entre el 0% y el 25%, en la cartera 2 los que están entre el 25% y el 50%, en la cartera 3 los que están entre el 50% y el 75%, y, por último, en la cartera 4 los que están entre el 75% y el 100%. La anterior asignación permite realizar un ajuste de distribución (prueba de Kolmogorov-Smirnov) a cada una de las carteras, con el objeto de realizar las simulaciones necesarias para obtener las probabilidades de incumplimiento ajustadas, en este caso, se realizaron 10 000 iteraciones a través de **RiskSimulator** (2020). Adicionalmente, como se muestra en la **Tabla 2**, las calificaciones crediticias A, B, C y D, las cuales son otorgadas a cada cliente por el sistema financiero, sirvieron para establecer su participación en el portafolio.

**TABLA 2.**  
*Cálculo de la pérdida esperada primer escenario.*

Calificación	Participación	Exposición (en pesos)	Probabilidad de incumplimiento	Severidad	Pérdida esperada (en pesos)
A	24%	6 087 010 065	0.23%	55%	7 635 885
B	26%	6 600 575 229	38.26%	55%	1 388 917 576
C	25%	6 230 134 783	62.41%	55%	2 138 652 396
D	25%	6 339 583 096	99.77%	55%	3 478 637 696

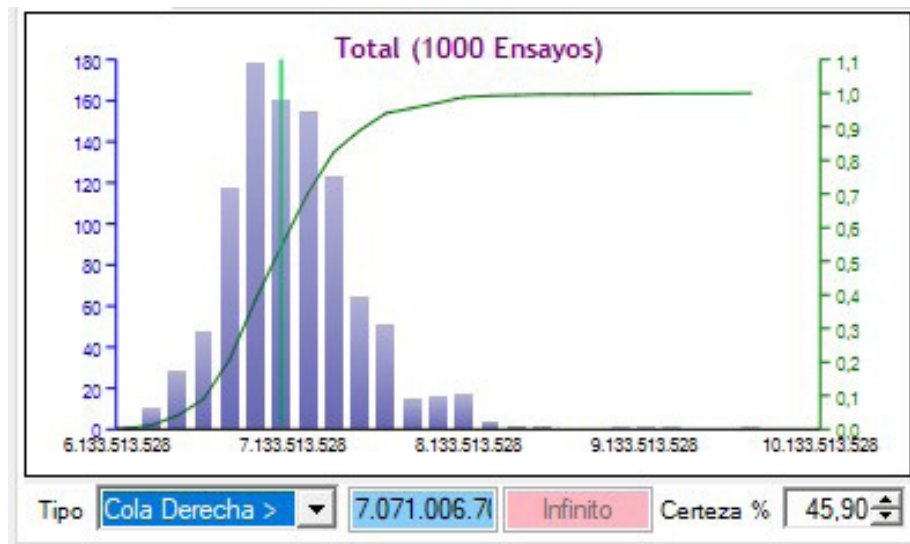
Fuente: Elaborado con la salida de **RiskSimulator** (2020).

Para el cálculo de la exposición se tuvo en cuenta la participación de cada una de las carteras multiplicada por el valor total del portafolio, que, para el caso de estudio, es de \$25 257 303 173. La institución financiera tiene establecida una tasa de recuperación (gestión de cartera) del 45%, lo que implica tener un índice de severidad (S) del 55% que sería para el caso, la pérdida producida por el incumplimiento en la cartera. Lo anterior implica tener una pérdida esperada total del portafolio de \$ 7 013 843 553, sin embargo, para mayor exactitud se procede a realizar una simulación de la pérdida esperada, con lo cual se espera una pérdida esperada media del portafolio de \$ 7 071 006 708, tal y como se muestra en la **Figura 1**.



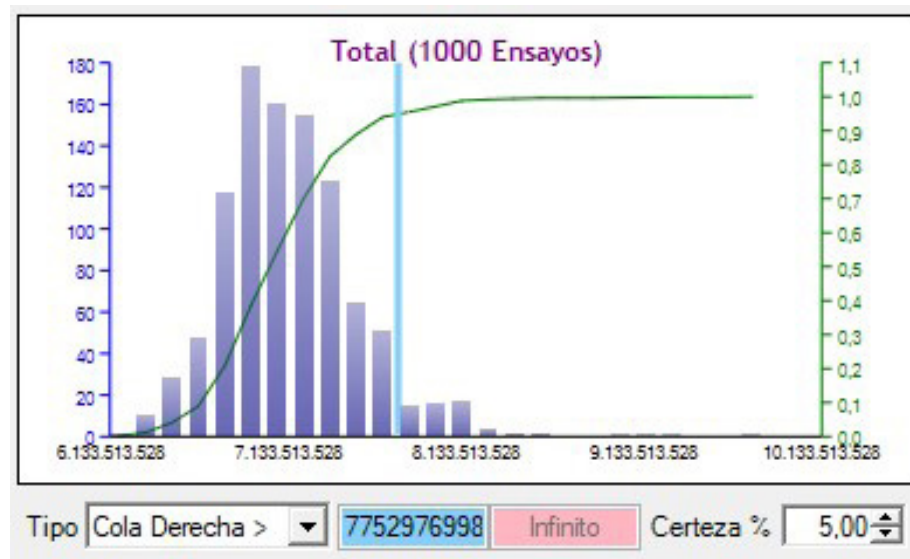
**Figura 1.** Simulación de la pérdida esperada.  
Fuente: Salida de **RiskSimulator** (2020).

Complementando el análisis anterior, los resultados indican que la pérdida esperada del portafolio oscilara entre un mínimo de \$6 037 273 409 y un máximo de \$9 738 816 421. Por otra parte, al tomar la pérdida esperada media del portafolio de \$7 071 006 708 y analizarla con base en la cola derecha de la función de distribución, encontramos que la probabilidad de la pérdida esperada media del portafolio es del 45.9% (Figura 2), la cual será cubierta con las reservas de la institución financiera.



**Figura 2.** Probabilidad de la pérdida esperada.  
Fuente: Salida de RiskSimulator (2020).

Un dato importante en esta investigación es el cálculo de la pérdida catastrófica, para ello se trabaja en RiskSimulator (2020) con la opción de cola derecha y una certeza del 5%, arrojando un valor de \$7 752 976 998 (Figura 3). Lo anterior implica que, se tendrá una pérdida no esperada equivalente a \$681 970 289, resultado de restarle a la pérdida catastrófica la pérdida esperada. Cabe resaltar que el monto de la pérdida no esperada deberá ser cubierta con capital por parte de la institución financiera.



**Figura 3.** Pérdida catastrófica.  
Fuente: Salida de RiskSimulator (2020).

### *Evaluación del poder predictivo del modelo*

La institución financiera caso de estudio maneja un nivel de tolerancia del 30%, por tanto, si la probabilidad de incumplimiento calculada a través del modelo de regresión logística supera dicho nivel, se le asigna a cada cliente el valor de 1 y, en caso contrario, 0. Los resultados obtenidos de este proceso se muestran en la **Tabla 3**.

**TABLA 3.**  
*Evaluación del poder predictivo del modelo.*

Predicción	0	1	Total
0	1.494	1	1.495
1	6	1.499	1.505
Total	1.500	1.500	3.000

Fuente: Elaborado con la salida de **RiskSimulator** (2020).

Para evaluar el poder predictivo del modelo se procede en primera instancia a calcular la sensibilidad, que muestra la tasa de verdaderos positivos (formula 5).

$$\text{Sensibilidad} = \frac{\text{Verdaderos positivos}}{(\text{Verdaderos positivos} + \text{Falsos negativos})} \quad (5)$$

Tomando los datos de la **Tabla 3** y reemplazando, se tiene (formula 6):

$$\text{Sensibilidad} = \frac{1.499}{(1 + 1.499)} = 99,93 \% \quad (6)$$

El resultado indica que una sensibilidad del 99.93% es señal de que el modelo de regresión logística está pronosticando de manera correcta a los clientes cumplidos. Aunque esto es bueno, en el ámbito del riesgo crediticio se le da mayor importancia a la predicción de los clientes incumplidos y para ello debe calcularse la especificidad (formula 7).

$$\text{Especificidad} = \frac{\text{Verdaderos negativos}}{(\text{Verdaderos negativos} + \text{Falsos positivos})} \quad (7)$$

De manera similar a la sensibilidad, se toman los datos de la **Tabla 3** y se reemplazan (formula 8).

$$\text{Especificidad} = \frac{1.494}{(1.494 + 6)} = 99,60 \% \quad (8)$$

Este nivel del 99.60% indica que el modelo de regresión logística propuesto pronostica de manera eficiente a los clientes incumplidos, lo cual, en esencia, es lo que le interesa a la institución financiera.



Finalmente, para que el modelo tenga mayor credibilidad se procede calcular el nivel de ajuste global, que es una razón de la especificidad y la sensibilidad (formula 9).

$$\begin{aligned} \text{Nivel de ajuste global} &= \frac{\text{Verdaderos negativos} + \text{Verdaderos positivos}}{\text{Total general}} \\ \text{Nivel de ajuste global} &= \frac{1.494 + 1.499}{3.000} = 99,77 \% \end{aligned} \quad (9)$$

Este nivel del 99.77 % indica que el modelo en general está enfocado tanto en seleccionar a los clientes incumplidos como a los cumplidos. El modelo también es utilizado en la base de prueba, que representa el 40% de la base inicial de los 5 000 clientes. Para efectos de la comparación se hizo el mismo proceso descrito anteriormente. La **Tabla 4** muestra un comparativo de las pruebas de sensibilidad, especificidad y nivel de ajuste global.

**TABLA 4.**  
*Evaluación del modelo en las bases de entrenamiento y prueba.*

	Base de entrenamiento	Base de prueba
Sensibilidad	99.93 %	86.02 %
Especificidad	99.60 %	84.26 %
Nivel global de ajuste	99.77 %	85.14 %

Fuente: Elaborado con la salida de **RiskSimulator** (2020).

Los resultados de la **Tabla 4** señalan que, aunque las pruebas de sensibilidad, especificidad y nivel global de ajuste en la base de prueba son menores que en la de entrenamiento, estas se encuentran por encima del 80%, denotando así que el modelo encontrado cumple de manera acertada con las predicciones de clientes, tanto incumplidos como cumplidos.

#### *Evaluación del poder predictivo del modelo con restricción*

A continuación, se procede a realizar una restricción en la calificación de cada uno de los clientes basados en una reducción en los días de mora, es decir, en el primer escenario de estudio los clientes se consideraban en mora cuando sobrepasaban el incumplimiento en un rango mayor a 45 días. En este segundo escenario se considera que el cliente es incumplido si supera un rango de 30 días, es así como, se pone a prueba el modelo encontrado en torno a que tan eficaz es en la pronosticación de los clientes incumplidos.

Manteniendo el nivel de tolerancia del 30% y el índice de severidad del 55% por parte de la institución financiera, el modelo muestra una disminución en las pruebas de sensibilidad, especificidad y nivel global de ajuste para el segundo escenario. Sin embargo, a pesar de que las tres pruebas disminuyen, estas se mantienen por

encima del 85% en la base de entrenamiento y del 80% en la base de prueba, los resultados se muestran en la **Tabla 5**.

**TABLA 5.**  
*Evaluación del modelo con diferente nivel de mora.*

	Nivel de mora entre 30 y 45 días		Nivel de mora superior a 45 días	
	Entrenamiento	Prueba	Entrenamiento	Prueba
Sensibilidad	85.2 %	80.7 %	99.93 %	86.02 %
Especificidad	85.6 %	80.0 %	99.60 %	84.26 %
Nivel global de ajuste	85.4 %	80.3 %	99.77 %	85.14 %

Fuente: Elaborado con la salida de **RiskSimulator** (2020).

En relación con el cálculo de la pérdida esperada en el segundo escenario, esta se incrementa de \$7 013 843 553 a \$7 546 246 967, es decir, la institución financiera tendría que aumentar la protección del portafolio en \$529 403 414, lo que implica un aumento del 7.59%. La **Tabla 6** muestra la pérdida esperada por subcartera, al igual que la exposición de estas.

**TABLA 6.**  
*Cálculo de la pérdida esperada segundo escenario.*

Calificación	Participación	Exposición (en pesos)	Probabilidad de incumplimiento	Severidad	Pérdida esperada (en pesos)
A	20%	5 051 460 635	0.19%	55%	5 278 776
B	23%	5 809 179 730	32.15%	55%	1 027 208 206
C	28%	7 072 044 888	64.08%	55%	2 492 471 500
D	29%	7 324 617 920	99.82%	55%	4 021 288 484

Fuente: Elaborado con la salida de **RiskSimulator** (2020).

## DISCUSIÓN

El riesgo de crédito ha sido un tema alrededor del cual se han venido desarrollando investigaciones, dada la importancia que tiene sobre el sistema financiero y más exactamente sobre las instituciones que la componen. Los primeros trabajos inician con **Myers y Forgy (1963)**, **Lane (1972)** y **Apilado, Warner y Dauten (1974)**, donde se muestra la necesidad de abordar el tema desde el punto de vista de la clasificación del riesgo crediticio y la forma de medirlo.

Con el avance de la estadística y de programas de modelación se desarrollan trabajos basados en la técnica de la regresión logística, como es el caso de **Wiginton (1980)** y **Srinivasan y Kim (1987)** que muestran un aumento en el nivel de pronóstico frente a los encontrados bajo la técnica del análisis discriminante. Por otra parte, están los trabajos de **Davis, Edelman y Gammerman (1992)** y **Tsai y Hung**

(2014) quienes aplican la técnica de la inteligencia artificial, que, aunque mejora levemente los resultados de la pronosticación, el procedimiento y entendimiento de esta técnica deja dudas en torno al camino utilizado para llegar a los resultados obtenidos.

Las revisiones de literaturas realizadas en los trabajos de Alaka et al. (2018) y Támara et al. (2019) encontraron que la técnica más implementada en la búsqueda de modelos de pronosticación, tanto para la probabilidad de incumplimiento como para la pronosticación de la quiebra empresarial, ha sido la técnica de la regresión logística. No obstante, cabe anotar que a esta técnica le siguen en utilización las redes neuronales y los árboles de decisión. Prueba de lo anterior, se encuentran los trabajos de Dendramis, Tzavalis y Adraktas (2018), Van Thiel y Van Raaij (2019) y Song, Wang y Xu (2020).

Finalmente, los resultados obtenidos en este trabajo ratifican lo encontrado en los estudios de Do Prado et al. (2019), Grundke, Pliszka y Tuchscherer (2019), Russo, Lagasio, Brogi y Fabozzi (2020) y Sariev y Germano (2020).

### CONCLUSIONES

Esta investigación ratifica la aplicabilidad de la regresión logística como una herramienta estadística en la búsqueda de modelos de pronosticación, que, para el caso del riesgo crediticio aporta a los diferentes estudios realizados. El modelo plantea la importancia de las variables financieras ingreso y endeudamiento de los clientes que conforman una cartera de consumo en una institución financiera, a la vez, ratifica el papel que juega la variable edad con relación en el incumplimiento de los deudores frente a las obligaciones crediticias. De esta forma, se concluye con base en los parámetros estimados que la variable ingreso es la que más influencia tiene en el modelo y, por ende, en el cálculo de la probabilidad de incumplimiento.

A la hora de analizar la estructuración de una cartera la calificación crediticia de los clientes es importante, dado que permite tener un nivel de participación equitativa en cada una de las calificaciones (A, B, C y D), determinando un rango de exposición del portafolio entre \$6 087 010 065 y \$6 339 583 096. Lo anterior, facilita conocer la probabilidad de incumplimiento de las sub-carteras del portafolio y ratifica que los clientes con calificación crediticia A tienen la probabilidad de incumplimiento más baja (0.23%), mientras que, aquellos con calificación crediticia D poseen la más alta (99.77%).

La evaluación del poder predictivo del modelo queda confirmada con los análisis de sensibilidad y especificidad realizados, que están por encima del 90% en la base de datos de entrenamiento y por encima del 80% en la base de prueba. En relación con el nivel de ajuste global, este se encuentra por encima del 85% en ambas bases, lo que garantiza que el modelo no solo enfatiza en la predicción de los clientes malos, sino también en la predicción de los clientes buenos. Finalmente, el modelo se prueba con una restricción en el nivel de mora, al pasar de 45 a 30 días el retraso en el pago de las obligaciones crediticias, arrojando un aumento leve de solo el 7.59% sobre la pérdida esperada.

REFERENCIAS

- Affes, Z. & Hentati, R. (2017). Predicting US Banks Bankruptcy: Logit Versus Canonical Discriminant Analysis. *Computational Economics*, (54), 199–244. <https://doi.org/10.1007/s10614-017-9698-0>
- Alaka, H., Oyedele, L., Owolabi, H., Kumar, V., Ajayi, S., Akinade, O. & Bilal, M. (2018). Systematic Review of Bankruptcy Prediction Models: Towards A Framework for Tool Selection. *Expert Systems With Applications*, (94), 164–184. <https://doi.org/10.1016/j.eswa.2017.10.040>
- Apilado, V., Warner, D. & Dauten, J. (1974). Evaluative Techniques in Consumer Finance--Experimental Results and Policy Implications for Financial Institutions. *The Journal Of Financial And Quantitative Analysis*, 9(2), 275–283. <https://doi.org/10.2307/2330105>
- Assef, F. & Steiner, M. (2020). Ten-year evolution on credit risk research: A systematic literature review approach and discussion. *Ingenieria e Investigacion*, 40(2), 1–21. <https://doi.org/10.15446/ing.investig.v40n2.78649>
- Cappon, A., Gorenstein, A., Mignot, S. & Manuel, G. (2018). Credit ratings, default probabilities, and logarithms. *Journal of Structured Finance*, 24(1), 39–49. <https://doi.org/10.3905/jsf.2018.24.1.039>
- Castro, D. y Soto, A. (2017). Estimación de la probabilidad de incumplimiento para las firmas del sector económico industrial y comercial en una entidad financiera colombiana entre los años 2009 y 2014. *Cuadernos de Economía*, 36(71), 293–319. <https://doi.org/10.15446/cuad.econ.v36n71.55273>
- Chen, N., Ribeiro, B. & Chen, A. (2016). Financial credit risk assessment: a recent review. *Artificial Intelligence Review*, 45(1), 1–23. <https://doi.org/10.1007/s10462-015-9434-x>
- Davis, R., Edelman, D. & Gamberman, A. (1992). Machine learning algorithms for credit-card applications. *IMA Journal of Management Mathematics*, 4(1), 43–51. <https://doi.org/10.1093/imaman/4.1.43>
- Dendramis, Y., Tzavalis, E. & Adraktas, G. (2018). Credit risk modelling under recessionary and financially distressed conditions. *Journal of Banking and Finance*, 91, 160–175. <https://doi.org/10.1016/j.jbankfin.2017.03.020>
- Do Prado, J., Carvalho, F., Benedicto, G. y Lima, A. (2019). Análisis del riesgo de crédito que enfrentan las empresas de capital abierto en Brasil: un enfoque utilizando análisis discriminante regresión logística y redes neuronales artificiales. *Estudios Gerenciales*, 35(153), 347–360. <https://doi.org/10.18046/j.estger.2019.153.3151>
- Durovic, A. (2017). Estimating probability of default on peer to peer market - Survival analysis approach. *Journal of Central Banking Theory and Practice*, 62(2), 149–167. <https://doi.org/10.1515/jcbtp-2017-0017>
- Fantazzini, D. & Zimin, S. (2020). A multivariate approach for the simultaneous modelling of market risk and credit risk for cryptocurrencies. *Journal of Industrial and Business Economics*, 47(1), 19–69. <https://doi.org/10.1007/s40812-019-00136-8>

- Grundke, P., Pliszka, K. & Tuchscherer, M. (2019). Model and estimation risk in credit risk stress tests. *Review of Quantitative Finance and Accounting*, 55, 163–199. <https://doi.org/10.1007/s11156-019-00840-5>
- Imtiaz, S. & Brimicombe, A. (2017). A Better Comparison Summary of Credit Scoring Classification. *International Journal of Advanced Computer Science and Applications*, 8(7), 1–4. <https://doi.org/10.14569/IJACSA.2017.080701>
- Lane, S. (1972). Submarginal Credit Risk Classification. *The Journal Of Financial And Quantitative Analysis*, 7(1), 1379–1385. <https://doi.org/10.2307/2330069>
- Lohmann, C. & Ohliger, T. (2018). Nonlinear relationships in a logistic model of default for a high-default installment portfolio. *Journal of Credit Risk*, 14(1), 45–68. <https://doi.org/10.21314/JCR.2017.232>
- Myers, J. & Forgy, E. (1963). The Development of Numerical Credit Evaluation Systems. *Journal Of The American Statistical Association*, 58(303), 799–806. <https://doi.org/10.2307/2282727>
- RiskSimulator. (2020). [Software]. Available from [www.software-shop.com/producto/risk-simulator](http://www.software-shop.com/producto/risk-simulator)
- Russo, V., Lagasio, V., Brogi, M. & Fabozzi, F. (2020). Application of the Merton model to estimate the probability of breaching the capital requirements under Basel III rules. *Annals of Finance*, 16(1), 141–157. <https://doi.org/10.1007/s10436-020-00358-0>
- Sariev, E. & Germano, G. (2020). Bayesian regularized artificial neural networks for the estimation of the probability of default. *Quantitative Finance*, 20(2), 311–328. <https://doi.org/10.1080/14697688.2019.1633014>
- Song, S., Wang, Y. & Xu, G. (2020). On the probability of default in a market with price clustering and jump risk. *Mathematics and Financial Economics*, 14(2), 225–247. <https://doi.org/10.1007/s11579-019-00253-x>
- Srinivasan, V. & Kim, Y. (1987). Credit Granting: A Comparative Analysis of Classification Procedures. *The Journal Of Finance*, 42(3), 665–681. <https://doi.org/10.2307/2328378>
- SPSS. (versión 24). Statistical Package for the Social Sciences. [Software]. Armonk: IBM. Disponible en <https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-24>
- Superfinanciera. (2002). . *Circular externa 011*. Bogotá, D.C.: Superfinanciera. Disponible en [http://legal.legis.com.co/document/Index?obra=legcol&document=legcol\\_75992041c75992048a75992043f75992034e75990430a75010151f75992034](http://legal.legis.com.co/document/Index?obra=legcol&document=legcol_75992041c75992048a75992043f75992034e75990430a75010151f75992034)
- Támara, A., Villegas, G. y De Andrés, J. (2019). Una revisión sistemática de la literatura en torno a la quiebra empresarial para el período 2012-2017. *Espacios*, 40(4). Disponible en <http://www.revistaespacios.com/a19v40n04/a19v40n04p25.pdf>
- Támara, A., Villegas, G., Leones, M. y Salazar, J. (2018). Modelación del riesgo de insolvencia en empresas del sector salud empleando modelos logit. *Revista de métodos cuantitativos para la economía y la empresa*, 26, 128–145. Disponible en <https://www.upo.es/revistas/index.php/RevMetCuant/article/view/2757>

- Trejo, J., Martínez, M. & Venegas, F. (2017). Credit risk management at retail in Mexico: An econometric improvement in the selection of variables and changes in their characteristics. *Contaduría y Administración*, 62(2), 399–418. <https://doi.org/10.1016/j.cya.2017.02.006>
- Tsai, C. & Hung, C. (2014). Modeling credit scoring using neural network ensembles. *Kybernetes*, 43(7), 1114–1123. <https://doi.org/10.1108/k-01-2014-0016>
- Van Thiel, D. & Van Raaij, W. (2019). Artificial intelligence credit risk prediction: An empirical study of analytical artificial intelligence tools for credit risk prediction in a digital era. *Journal of Risk Management in Financial Institutions*, 12(3), 268–286. <https://doi.org/10.33423/jaf.v19i8.2622>
- Wiginton, J. (1980). A note on the comparison of logit and discriminant models of consumer credit behaviour. *Journal of Financial and Quantitative Analysis*, 15(3), 757–770. <https://doi.org/10.2307/2330408>

#### BIODATA

**Armando Lenin Támara Ayús** es Economista de la Universidad de Antioquia (Colombia). Especialista en diseño y evaluación de proyectos de la Universidad del Norte (Colombia). Magister en ciencias de las finanzas de la Universidad EAFIT y PhD de la Universidad de Medellín (Colombia). Sus intereses de investigación incluyen temas relacionados con el riesgo de crédito, quiebra empresarial y opciones reales. ORCID: <https://orcid.org/0000-0003-3619-5164>

**José Eduardo Segura Ramos** es Economista de CECAR (Colombia). Especialista en finanzas y magister en administración financiera de la Universidad EAFIT (Colombia). Sus intereses de investigación incluyen temas relacionados con el riesgo de crédito y la probabilidad de incumplimiento. ORCID: <https://orcid.org/0000-0001-8672-8841>

**Ignacio Emilio Chica Arrieta** es Ingeniero Químico de la Universidad de Antioquia (Colombia). Magister en administración de la Universidad EAFIT (Colombia). Sus intereses de investigación incluyen temas relacionados con la administración. ORCID: <https://orcid.org/0000-0001-6192-4491>