

# FACTORES DETERMINANTES DEL INCUMPLIMIENTO EN CRÉDITOS DE CONSUMO DE UNA COOPERATIVA DE AHORRO Y CRÉDITO

Diana López Avilés \* ~ Andrés Tobar Vidal \*\* ~ René Vergara Toro \*\*\*

## RESUMEN

Determinar el riesgo de crédito es una de las principales dificultades para cualquier entidad financiera, puesto que si no son cuantificados de una manera adecuada pueden generar graves problemas, como la incapacidad para responder ante pérdidas en las operaciones crediticias otorgadas.

Las instituciones utilizan diversas metodologías para cuantificar este tipo de riesgo, las cuales van desde el juicio experto de un analista hasta metodologías más complejas, como modelos de predicción de comportamiento.

Este documento se estructura de la siguiente forma: en la primera sección se explican los sustentos teóricos para entender el problema de incumplimiento en una cartera de créditos de consumo. La segunda sección contempla el análisis de los factores influyentes en el incumplimiento o cese de pagos en una cartera de créditos de consumo. En la tercera, y última etapa, se realiza la construcción de un modelo logit, con el cual se logra cuantificar el impacto y significancia de las variables que explican el incumplimiento (*default*), y además se establece una estrategia para aceptar o rechazar una solicitud de crédito.

**PALABRAS CLAVE:** Modelo *logit*, probabilidad de incumplimiento, *default*, créditos de consumo, *scoring*, Cooperativa de Ahorro y Crédito

## ABSTRACT

Determining the credit risk is one of the major challenges for any financial entity, since if these are not properly quantified, this may create serious issues, such as the inability to answer for losses in the credit operations granted.

The institutions use various methodologies to quantify this type of risk, from the expert opinion of an analyst to more complex methodologies, such as models for predicting behaviors.

The structure of this document is as follows: in the first section, the theoretical support for understanding the issue of consumer credit portfolio default is explained. The second section deals with the analysis of factors that influence default or cessation of payments on a consumer credit portfolio. The third and last section creates a logit model that allows to quantify the impact and significance of the variables explaining the default, in addition to establishing a strategy to accept or deny a credit application.

**KEYWORDS:** Logit model, probability of default, default, consumer credits, scoring, Credit Union

**Códigos JEL:** C58

Fecha de Recepción 30 septiembre 2018

Fecha de Aceptación 01 de octubre de 2018

\*Ingeniero Comercial, Mención Economía, Universidad de Chile. Posee especialización en Métodos Estadísticos y Big Data, con diplomados en Pontificia Universidad Católica de Chile para las respectivas áreas, además de un Magíster en Finanzas (c), en la Universidad de Chile. Ostenta 5 años de experiencia en recopilación de información, compilación de estadísticas, investigación y análisis de mercados internacionales, primero en la Facultad de Economía y Negocios de la Universidad de Chile y desde junio de 2012 en el Banco Central de Chile.

\*\*Egresado de Ingeniería Comercial, Universidad Tecnológica Metropolitana. Analista de datos y Analista de proyectos en la Cooperativa de Ahorro y Crédito Talagante LTDA. Correo electrónico: atobarvidal@gmail.com / atobar@coocretal.cl.

\*\*\* Egresado de Ingeniería Comercial, Universidad Tecnológica Metropolitana. Correo electrónico: renevergaratoro@gmail.com.

## INTRODUCCIÓN

En esta investigación se muestra la construcción de un modelo *logit*, con el objetivo de determinar cuáles son los factores más influyentes para que un individuo cometa incumplimiento<sup>1</sup> de sus obligaciones en un crédito de consumo. Los resultados que se obtienen están sujetos al contexto de la entidad financiera<sup>2</sup>, que para este caso es una Cooperativa de Ahorro y Crédito del mercado financiero chileno, y serán utilizados para modelar un proceso de *scoring*.

La medición del *riesgo crédito* es relevante para cualquier tipo de entidad financiera, puesto que su mitigación puede conllevar un gran beneficio económico y financiero para la entidad respectiva, lo que a su vez puede generar una mayor estabilidad y desarrollo del sistema financiero.

Existen distintos tipos de metodologías que actualmente se usan para cuantificar el riesgo de crédito, siendo la probabilidad de incumplimiento la base para medir este fenómeno.

## 1. OBJETIVOS

### 1.1. Objetivo general

Determinar los factores más influyentes que explican el incumplimiento de un socio en una cartera de consumo dentro de una Cooperativa de Ahorro y Crédito.

### 1.2. Objetivos específicos

- Estimar la probabilidad de incumplimiento (default) mediante la construcción de un modelo *logit*.

- Definir la estrategia para aceptar o rechazar las solicitudes de créditos.
- Crear un sistema de *Scoring* para contabilizar el riesgo de solicitudes de créditos.

## 2. MARCO TEÓRICO

### 2.1. Estado del arte

Moreno (2013) utilizó modelos *probit*, *logit* y *logit mixto* para comparar la calidad de los ajustes de cada uno de ellos y así determinar cuál de las tres metodologías es la más adecuada para ser utilizada posteriormente en la elaboración de un *scoring* de crédito. Para esto utilizó una muestra total de 10,841 registros, con una proporción de 96.49% clientes no default y 3.51% de default. Se incorporaron 24 variables y fueron divididas en los siguientes factores: información referente a la operación de crédito, situación financiera, comportamiento de crédito histórico, situación laboral, características sociodemográficas y factor económico. Para evaluar los factores más influyentes en el default, realizó la modelación de la variable objetivo contra cada uno de los factores. Una vez construidos los tres modelos, se llega a la conclusión de que en general las tres metodologías empleadas presentan en promedio una tasa de especificidad sobre el 90% y una tasa de sensibilidad promedio de 50%. Además, se destaca el modelo *logit mixto* como el más robusto y sensible para la explicación del default, teniendo la menor tasa de falsos 0; sin embargo, presenta la mayor cantidad de falsos 1.

Velandia (2013) utilizó un modelo *logit* para medir la probabilidad de incumplimiento en una institución financiera de una cartera de créditos para 201 pymes, a las cuales se les concedió algún crédito. Finalmente, el modelo construido concluye que las principales variables que explican el default son las referidas a las de calificación interna<sup>3</sup>, y también el género, donde

1. La Superintendencia de Bancos e Instituciones Financieras define el incumplimiento como una mora igual o superior a 90 días.

2. Por motivos de privacidad, y a pedido de la entidad que facilitó la información para esta investigación, cada vez que se requiera nombrar a la institución en cuestión, se le denominará como *entidad financiera* o *Cooperativa de Ahorro y Crédito*.

3. Las calificaciones son: AAA, AA, A y A-, que son principalmente el resultado de los análisis financieros y de los niveles de mora que han presentado en la entidad.

destaca que los hombres tienen mayores probabilidades de entrar en default.

Medina y Ulfe (2015) realizaron un trabajo que compara las metodologías de árboles de decisión, regresión logística y redes neuronales, con el objetivo de determinar cuál de los tres mecanismos era más efectivo para predecir la probabilidad de default en una cartera de créditos, y así luego construir un credit scoring. Una vez construidas las tres metodologías se concluyó que la mejor estructura es la *regresión logística*, la cual arrojó una curva ROC de 79.2%.

Covarrubias (2012) utilizó un modelo de regresión logit para calcular la probabilidad de incumplimiento y posteriormente construir una metodología de seguimiento para este tipo de modelos. Para efectos prácticos con la presente investigación, solo se analizará el modelo de regresión logit construido y las variables utilizadas para pronosticar el default. La base de datos utilizada contaba con 15.000 registros, de los cuales el 94% corresponde a no default y el 6% restante corresponde a default. Las variables explicativas del modelo son: de comportamiento en la entidad financiera, demográficas, características particulares de los clientes, como la edad y el número de dependientes. Las variables que lograron ser significativas en explicar el default fueron la edad, la cantidad de moras entre 30 y 60 días dentro de la institución, la presencia o no de mora mayor a 90 días, el número de hipotecas y la mora entre 60 y 90 días. Finalmente se destaca que el modelo construido posee un poder discriminante del 80.4% (curva ROC), una tasa de sensibilidad de un 72.6% y una especificidad del 79.4%.

## 2.2. Contextualización

Una Cooperativa de Ahorro y Crédito es una entidad financiera no bancaria, cuya función es la intermediación financiera. Su mercado objetivo, generalmente, son los estratos socioeconómicos medio-bajos de nuestro país y su principal producto ofrecido son los créditos de consumo.

## 2.3. Marco normativo y legal

Una de las principales tareas de una institución financiera es la mitigación de riesgos, por esto es vital establecer procedimientos y políticas para lograrlo. Gran parte de las normativas y legislación vigente tienen su origen en los principales acuerdos internacionales y los establecidos por el ente regulador, que en el caso chileno es la Superintendencia de Bancos e Instituciones Financieras.

## 2.4. Problema de default

Dentro de toda entidad financiera existen fenómenos y procesos críticos, como el default y el proceso de otorgamiento de crédito, por lo cual se busca la mitigación de riesgo de incurrir en estos mediante la predicción de comportamiento futuro del solicitante.

La Superintendencia de Bancos e Instituciones Financieras, en el compendio de normas contables de Cooperativas en su capítulo G-2 hoja 1, define dicho evento como: “La cartera en incumplimiento comprende todas las colocaciones y el 100% del monto de los créditos contingentes, de los deudores que al cierre de un mes presenten un atraso igual o superior a 90 días en el pago de intereses o capital de algún crédito” (Superintendencia de Bancos e Instituciones Financieras, 2015).

Es por lo anterior, que en esta investigación se toma como parámetro esta definición y se establece como default aquel socio que presente *una mora igual o mayor a 90 días*, y un socio no default es aquel que presente una mora menor a 90 días.

## 2.5. Metodología WOE (Weight of Evidence)

Naeem Siddiqi (2016) define a la metodología WOE como un indicador de riesgo relativo de ocurrencia del evento (en nuestro caso el incumplimiento o default) en un rango determinado de las variables independientes que son del tipo dicotómicas o continuas que fueron

discretizadas. Esta metodología se puede calcular realizando la siguiente operación:

$$WOE: \left[ \ln \left( \frac{\text{Distribución No default}}{\text{Distribución Default}} \right) \right] \times 100$$

Donde la interpretación de su resultado será basada en la distribución de los valores obtenidos, los que serán asociados a un nivel de riesgo. En el caso de los WOE será lo siguiente:

- Si es positivo, disminuye la probabilidad de incumplimiento. Mientras mayor sea el valor, menos riesgoso será el segmento.
- Si es negativo, aumenta la probabilidad de incumplimiento. Mientras menor sea el valor, más riesgoso será el segmento.

### 3. METODOLOGÍA

#### 3.1. Regresión logística binaria

La regresión logística binaria es parte de los modelos lineales generalizados (GLM, es su acrónimo en inglés). Es binaria debido a que la variable respuesta (dependiente) plantea dos posibles escenarios: (0) cuando no hay presencia de default y (1) cuando hay presencia de default.

La regresión logística se puede expresar como:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k \quad (1)$$

Donde  $p$  es la probabilidad de ocurrencia del evento estudiado. En este caso es la probabilidad de incumplimiento o default, la cual se puede expresar como:

$$p(x_i) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}} \quad (2)$$

Donde  $z$  queda expresado como:

$$Z = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k \quad (3)$$

Si realizamos la siguiente ecuación:

$$\log\left(\frac{p}{1-p}\right) = y_i \quad (4)$$

Donde:

$$Y_i = \begin{cases} 1, & \text{si el socio presenta default} \\ 0, & \text{si el socio no presenta default} \end{cases}$$

A diferencia de una regresión lineal, la estimación de los parámetros en este tipo de modelos se realiza mediante el método de máxima verosimilitud.

#### 3.2. Curva ROC

La curva ROC<sup>4</sup> reconoce el poder predictivo del modelo logit, donde se ilustra la distribución de probabilidad de los socios que alcanzan el estado de default, en contraste a los que no. La distribución de los falsos positivos (1 – especificidad) se encuentra en el eje X y la distribución de verdaderos positivos (sensibilidad) se encuentra en el eje Y (Moreno, 2013).

Para la interpretación de la curva ROC, debemos observar los valores debajo de dicha curva (AUC<sup>5</sup>), los cuales fluctúan entre los valores 0,5 y 1. La interpretación de los resultados de la curva ROC indica que mientras más cercano a 1 sea el valor de área bajo la curva, mejor será el poder discriminante del modelo.

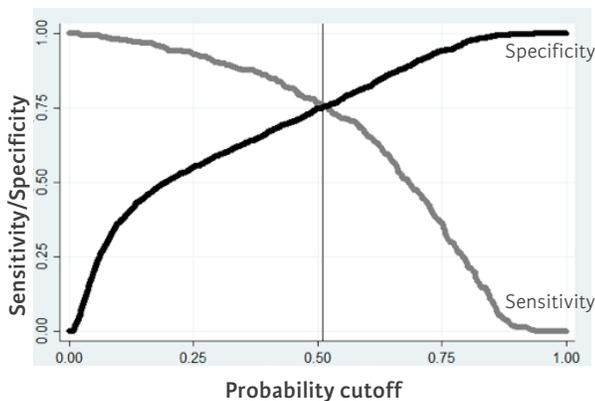
4. Receiver operating characteristic.

5. Área bajo la curva.

### 3.3. Curva de sensibilidad y especificidad

La curva de sensibilidad y especificidad permite observar aquel punto en el cual se maximiza la clasificación de los individuos en las clases respectivas (default o no default) para posteriormente construir la matriz de confusión. Este punto de corte se establece al maximizar la capacidad que posee el modelo para identificar aquellos individuos que poseen el fenómeno de estudio (default) junto con la capacidad que posee el modelo para identificar a los individuos que no poseen el fenómeno de estudio (no default). Cabe destacar que cuando el punto corte se acerca al valor cero aumenta la sensibilidad y, en caso contrario, si a uno aumenta la tasa de especificidad del modelo. La ilustración 1 muestra donde se maximizan estas curvas:

**Ilustración 1.** Ejemplo curva de sensibilidad y especificidad



Fuente: elaboración propia en Stata 14.

### 3.4. Tabla de clasificación (matriz de confusión)

La matriz de confusión es una medida que permite observar el poder de predicción del modelo construido. La matriz de confusión se puede expresar como:

**Tabla 1.** Matriz de confusión

		Observado	
		Default	No default
Predicción	Default	Verdaderos positivos (VP)	Falsos positivos (FP)
	No default	Falsos negativos (FN)	Verdaderos negativos (VN)

Fuente: elaboración propia.

Desde la Tabla 1 se desprende una serie de parámetros utilizados para medir la capacidad de clasificar correctamente, algunas de las cuales son:

#### Tasa de error tipo I

Corresponde a los casos en que el modelo clasifica mal a un individuo que no posee el fenómeno de estudio<sup>6</sup>; es decir, se rechaza la solicitud de crédito de un buen socio, lo que se traduce en una pérdida de costo-oportunidad. Se calcula como:

$$\frac{FP}{VP + FP}$$

#### Tasa de error tipo II

Corresponde a los casos cuando el modelo clasifica bien a un individuo que posee el default; es decir, se acepta la solicitud de crédito de un mal socio, lo que se traduce en una pérdida monetaria directa por el no retorno del crédito. Se calcula como:

$$\frac{FN}{FN + VN}$$

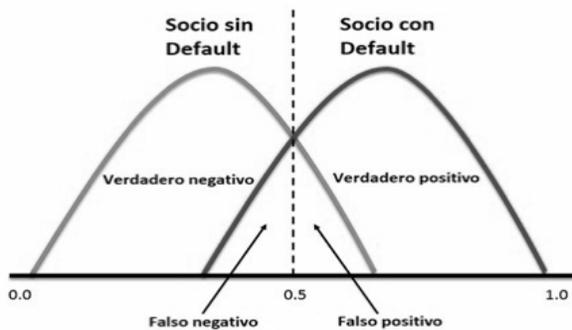
Cuando un modelo predice perfectamente los resultados, permite identificar y distinguir claramente entre socios default y no-default; es decir, la tasa de aciertos es la

6. En esta investigación un individuo sin el fenómeno de estudio es homólogo a un socio sin default, y un individuo con el fenómeno de estudio es homólogo a un socio con default.

máxima y no existen las tasas de error. Esto no es posible en la realidad, ya que siempre existe un punto donde las curvas de sensibilidad y especificidad se traslapan, lo cual no permite distinguir ambos grupos.

Como se observa en la Ilustración 2, existen dos puntos donde estas curvas se traslapan y se pueden observar los errores tipo I y tipo II, conceptos explicados anteriormente.

**Ilustración 2.** Curvas de sensibilidad y especificidad traslapadas



Fuente: elaboración propia

### 3.5. Oferta crediticia del sistema financiero

El mercado financiero oferta diferentes tipos de créditos. Entre estos se distinguen los créditos comerciales, créditos de consumo y créditos hipotecarios.

La investigación se limitará al análisis de créditos de consumo. Este tipo de crédito puede canalizarse a través de dos productos: créditos en cuotas o tarjetas de créditos (que contempla también la línea de crédito). El primer producto será el eje central de la investigación, quedando las tarjetas de crédito y línea de crédito fuera del alcance de este estudio.

## 4. APLICACIÓN: ESTIMACIÓN Y VALIDACIÓN DE MODELOS

Después de procesar la base de datos y antes de estructurar el modelo, es necesario balancear la data disponible. Se observa una desproporción entre socios buenos (sin default) y socios malos (con default) (ver Tabla 2); lo que, de no ser corregido, provocaría que el modelo clasificaría de mejor manera a la clase predominante de la muestra, en desmedro de la menos numerosa.

**Tabla 2.** Frecuencia por clase muestra original

Estado	Frecuencia	Porcentaje
No default	3,865	91%
Default	380	9%
<b>Total</b>	<b>4,245</b>	<b>100%</b>

Fuente: elaboración propia.

Para resolver este problema existen diversas técnicas, dentro de las cuales se distinguen dos enfoques: en el primero se opta por asignar un costo diferencial a las instancias de entrenamiento según las frecuencias de las clases, mientras que en el segundo se remuestrea el conjunto de datos originales, ya sea agregando casos repetidos de la clase minoritaria o submuestreando las clases mayoritarias (Drozdowicz, 2007).

En esta investigación se optó por el segundo enfoque, donde se realizó un remuestreo de la clase de menor frecuencia de la muestra original; es decir, un *oversampling*<sup>7</sup> con la proporción de 70% de socios sin default y 30% de socios con default, para posteriormente dar inicio al proceso de modelamiento final.

### 4.1. Datos y variables

Luego de realizado el proceso de balanceo de la muestra, esta cuenta con 5,516 (Tabla 3) datos de socios

7. Procedimiento mediante el cual se replica la clase de menos frecuencia de una base de datos.

pertencientes a la cartera de créditos de consumo concedidos entre agosto de 2016 y mayo de 2018.

**Tabla 3.** Frecuencia por clase muestra balanceada

Estado	Frecuencia	Porcentaje
No default	3,861	70%
Default	1,655	30%
<b>Total</b>	<b>5,516</b>	<b>100%</b>

Fuente: elaboración propia.

La muestra contiene la variable dependiente “default” y 40 variables sociodemográficas y financieras del cliente, además de variables netamente del crédito, las cuales permiten explicar el fenómeno de default. La variable dependiente es del tipo binaria y fue codificada bajo el valor 0 cuando el cliente no presenta el fenómeno de default, y 1 cuando sí lo presenta, la que equivale a una mora igual o mayor a 90 días, acorde con lo establecido en el compendio de normas contables. En cuanto a las variables independientes, presentan información que es almacenada al momento de analizar el otorgamiento del crédito. Las variables serán agrupadas en 6 factores, los cuales son: crédito, financiero, comportamiento crediticio, laboral, sociodemográficos y factores económicos.

## 4.2. Análisis descriptivo de la muestra

A continuación, se presenta el análisis y las principales características de las variables utilizadas:

1. La población se encuentra concentrada en las edades mayores, ya que la media de edad es de 51.8 años. Además se presenta una proporción similar entre hombres y mujeres.
2. El plazo cuenta con un mínimo de 12 meses, lo cual está acorde con el promedio del mercado financiero, donde los plazos mínimos de 6 y 12 meses son los más recurrentes.
3. El nivel educacional de la muestra está concentrado en los niveles básico y secundario.

4. Existe una alta concentración de individuos sin problemas de moralidad externa, lo que indica una cartera crediticia sana. Esto es confirmado, porque el 9% de la distribución inicial de la muestra presenta default.
5. Los socios, por lo general, poseen una tenencia propia o familiar.
6. La renta promedio de la población corresponde a \$375.226, lo cual indica que el mercado objetivo de la entidad financiera son los estratos socioeconómicos que no son atendidos por la banca tradicional.

## 4.3. Estimación de modelos

Previa a la estimación de los modelos se realizó un proceso de discretización a las variables continuas, bajo el criterio *Weight of Evidence* (WOE), para encontrar patrones de comportamiento dentro de las categorías generadas e identificando los grupos de riesgo correspondientes. Con esto, se logra obtener con mayor precisión en qué categorías o intervalos existe mayor evidencia de default. El modelo inicial cuenta con todas las variables que fueron seleccionadas debido a su importancia e interés en el fenómeno por evaluar.

### 4.3.1. Modelo I

En la Tabla 4 se muestran los resultados del modelo inicial, donde las variables que no son significativas se detallan a continuación:

#### Aporte de acciones

Esta variable fue incorporada por el hecho de que, si se le da condición de socio, presenta un aporte de acciones de manera metódica, esto reflejaría que fuese más seguro otorgarle un crédito, debido a la periodicidad del aporte a la Cooperativa; sin embargo, se concluye que, al no ser significativa, no existe evidencia que confirme la explicación del default.

#### Nivel escolar

Esta variable resulta no significativa; sin embargo, se debe precisar que la entidad financiera en estudio es

una Cooperativa de Ahorro y Crédito, esto indica que el mercado objetivo de estas instituciones difiere de los de la banca tradicional, ya que está sesgado hacia un sector de la población donde la educación básica y media son los niveles educativos más frecuentes.

### Estado civil

Esta variable mostró que no existe evidencia que aporte a explicar el default en un crédito de consumo.

### N° cargas familiares

Esta variable buscaba relacionar que al tener mayor número de cargas familiares y la probabilidad de entrar en default sea positiva; sin embargo, al no ser significativa, se descarta que exista evidencia.

### Monto en UF

Esta variable en las investigaciones revisadas es de suma relevancia, ya que se relaciona con la idea de que a un mayor monto otorgado, el default sea más probable; sin embargo, las características del mercado objetivo de la entidad financiera no tienen mucha variabilidad.

### Renta líquida en UF

Esta variable sigue la misma lógica que *monto* y *nivel escolar*, debido a que el mercado objetivo de la Cooperativa, al poseer características similares de ingreso, no permite construir categorías de la variable que indiquen en qué rangos se presentan mayores evidencias de default.

Tabla 4. Modelo I

Variabes	Coficiente	P> z
<b>Plazo</b>		
12 a 17	0	
18 a 23	0.271	0.048
24 a 35	0.592	0.000
36 o más	0.713	0.000
<b>Moralidad externa</b>		
Sin problemas de moralidad	0	
Con problemas de moralidad	1.038	0.000
<b>Comportamiento interno</b>		
0-30	0	
31-60	0.459	0.022

61-90	0.966	0.013
Más de 90	1.202	0.006
<b>Género</b>		
Femenino	0	
Masculino	0.418	0.000
<b>Antigüedad laboral</b>		
0 a 4	0	
5 a 14	-0.317	0.002
15 o más	-1.036	0.000
<b>Antigüedad socio</b>		
Menor a 4	0	
Mayor a 4	-1.242	0.000
<b>Edad</b>		
19 a 30	0	
31 a 49	0.295	0.011
50 o más	-0.670	0.000
<b>Situación laboral</b>		
Independiente	0	
Dependiente	-0.592	0.000
Jubilados, pensionados y montepiados	-0.856	0.000
<b>Tenencia</b>		
Propia	0	
Allegado	0.432	0.000
Arrendatario	0.575	0.000
Familiar	0.445	0.000
<b>Aporte acciones</b>		
No Metódico	0	
Metódico	-0.060	0.732
<b>Nivel escolar</b>		
Básica	0	
Secundario	-0.153	0.108
Técnica	-0.146	0.446
Universitaria	0.022	0.946
<b>Estado civil</b>		
Divorciado	0	
Casado separación de bien	-0.245	0.256
Casado sociedad conyugal	-0.204	0.198
Soltero	-0.238	0.138
Viudo	-0.169	0.469
<b>N° cargas familiares</b>	0.013	0.887
<b>Monto UF</b>	0.000	0.918
<b>Renta líquida UF</b>	0.002	0.762
<b>Tasa</b>	2.343	0.000
<b>N° acreedores financieros</b>	0.121	0.011
<b>Constante</b>	-6.128	0.000
<b>Pseudo R2</b>	0.2730	
<b>Observaciones</b>	4,413	

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Fuente: elaboración propia en Stata 14.

4.3.2. Modelo II

Tabla 5. Modelo II

VARIABLES	COEFICIENTE	P> z
<b>Plazo</b>		
12 a 17	0	
18 a 23	0.277**	0.042
24 a 35	0.587***	0.000
36 o más	0.698***	0.000
<b>Moralidad externa</b>		
Sin problemas de moralidad	0	
Con problemas de moralidad	1.021***	0.000
<b>Comportamiento interno</b>		
0-30	0	
31-60	0.474**	0.017
61-90	0.956**	0.014
Más de 90	1.228***	0.005
<b>Género</b>		
Femenino	0	
Masculino	0.418***	0.000
<b>Antigüedad laboral</b>		
0 a 4	0	
5 a 14	-0.318***	0.002
15 o más	-1.057***	0.000
<b>Antigüedad socio</b>		
Menor a 4	0	
Mayor a 4	-1.257***	0.000
<b>Edad</b>		
19 a 30	0	
31 a 49	0.333***	0.003
50 o más	-0.609***	0.000
<b>Situación laboral</b>		
Independiente	0	
Dependiente	-0.592***	0.000
Jubilados, pensionados y montepiados	-0.829***	0.000
<b>Tenencia</b>		
Propia	0	
Allegado	0.416***	0.000
Arrendatario	0.566***	0.001
Familiar	0.421***	0.000
<b>Tasa</b>		
	2.348***	0.000
<b>N° acreedores financieros</b>		
	0.126***	0.005
<b>Constante</b>		
	-6.461***	0.000
<b>Pseudo R2</b>		0.2720
<b>Observaciones</b>		4,413

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Fuente: elaboración propia en Stata 14

4.3.3. Efectos marginales: Modelo II

Tabla 6. Efectos marginales Modelo II

VARIABLES	dy/dx	Error estándar	P> z
<b>Plazo</b>			
12 a 17	0	(base)	
18 a 23	0.038**	0.019	0.041
24 a 35	0.083***	0.016	0.000
36 o más	0.100***	0.018	0.000
<b>Moralidad externa</b>			
Sin problemas de moralidad	0	(base)	
Con problemas de moralidad	0.156***	0.028	0.000
<b>Comportamiento interno</b>			
0-30	0	(base)	
31-60	0.071**	0.030	0.019
61-90	0.146**	0.060	0.015
Más de 90	0.188***	0.067	0.005
<b>Género</b>			
Femenino	0	(base)	
Masculino	0.061***	0.012	0.000
<b>Antigüedad laboral</b>			
0 a 4	0	(base)	
5 a 14	-0.050***	0.016	0.002
15 o más	-0.151***	0.022	0.000
<b>Antigüedad socio</b>			
Menor a 4	0	(base)	
Mayor a 4	-0.189***	0.017	0.000
<b>Edad</b>			
19 a 30	0	(base)	
31 a 49	0.055***	0.018	0.002
50 o más	-0.092***	0.020	0.000
<b>Situación laboral</b>			
Independiente	0	(base)	
Dependiente	-0.089***	0.014	0.000
Jubilados, pensionados y montepiados	-0.122***	0.019	0.000
<b>Tenencia</b>			
Propia	0	(base)	
Allegado	0.062***	0.018	0.000
Arrendatario	0.085***	0.025	0.001
Familiar	0.063***	0.015	0.000
<b>Tasa</b>			
	0.344***	0.027	0.000
<b>N° acreedores financieros</b>			
	0.018***	0.006	0.004

Note: dy/dx for factor levels is the discrete change from the base level.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Fuente: elaboración propia en Stata 14.

#### 4.3.4. Interpretación de las variables

##### **Plazo**

Esta variable presenta categorías con plazos que presenten patrones similares de comportamiento, distinguiendo a su vez cuál de estos es el más riesgoso. Los signos presentan coherencia, puesto que a mayor plazo se genera un aumento en la probabilidad de incumplimiento. El pasar de un crédito con un plazo de 12 a 17 meses, a uno de 18 a 23, 24 a 35 y más de 36 meses, provoca que la probabilidad de entrar en default aumente en un 3.8%, 8.3% y 10% respectivamente.

##### **Moralidad externa**

Esta variable indica la presencia o ausencia de mora en el sistema financiero en los últimos 2 años previos a la solicitud del crédito. Los signos indican que, ante la presencia de moralidad externa, la probabilidad de incumplimiento aumenta en un 15.6%, lo cual es coherente, ya que es normal concluir que, si el socio ya presenta un fenómeno de mora en el sistema financiero en los últimos dos años, podría reincidir en este comportamiento.

##### **Comportamiento interno**

Esta variable indica el historial de comportamiento que presenta el socio en la entidad financiera dos años previos a la solicitud del crédito. Esta variable posee 4 categorías agrupadas en un aumento de 30 días de mora. Los signos reflejan que ante un mayor nivel de mora que presente en su historial de comportamiento interno, mayor es su probabilidad de default; es decir, el pasar de un nivel de 0 a 30 días de mora, hacia un nivel de 31 a 60, 61 a 90 y más de 90, aumenta la probabilidad de incumplimiento en 7.1%, 14.6% y 18.8% respectivamente, lo cual es coherente, ya que es lógico concluir que dependiendo del grado de mora que ha presentado el socio con la entidad financiera durante los últimos dos años, podría tener más posibilidades de reincidir en este comportamiento.

##### **Género**

En este caso los signos reflejan que el ser hombre aumenta la probabilidad de default en 6.1%. Esto

concuera con estudios que confirman estos resultados (Ormazábal, 2014) sin embargo, no es materia de esta investigación buscar los motivos del por qué los hombres presentan una mayor probabilidad de default

##### **Antigüedad laboral**

La variable presenta 3 categorías: 0 a 4, 5 a 14 y 15 o más años, donde los signos de estas variables indican que, al pasar del tramo de 0 a 4 años de antigüedad laboral, hacia los tramos 5 a 14 y 15 o más años, disminuyen en 5% y 15.1% la probabilidad de incumplimiento respectivamente.

##### **Antigüedad como socio**

La variable presenta dos categorías: antigüedad menor a 4 años y antigüedad mayor a 4 años. Los signos indican que ante una mayor permanencia del socio en la entidad, su probabilidad de default disminuye; es decir, el tener más de cuatro años de antigüedad como socio disminuye la probabilidad de default en 18.9%.

##### **Edad**

Esta variable presenta tres categorías: de 19 a 30 años, 31 a 49 y 50 o más años. Los signos indican que al pasar de estar en el segmento de 19 a 30 años al segmento de 31 a 49 años, aumenta la probabilidad de incumplimiento en 5.5%, y pasar de 19 a 30 años hacia el segmento de 50 o más años, disminuye la probabilidad de incumplimiento en 9.2%. Las interpretaciones anteriores indican que los segmentos más riesgosos son el de 19 a 30 y 31 a 49 años, y el segmento de menor riesgo es el de 50 años o más, justamente el segmento que presenta una mayor concentración de socios, con un 56.27%.

##### **Situación Laboral**

La variable presenta tres categorías, trabajadores dependientes, independientes y por último a los jubilados, pensionados y montepiados. Los signos indican que el pasar de ser un trabajador independiente a dependiente o jubilados, pensionados y montepiados es menos riesgoso, disminuyendo su probabilidad de default en 8.9% y 12.2% respectivamente. Lo anterior se podría explicar debido a que los trabajadores dependientes y los jubilados, pensionados y montepiados, presentan

ingresos y/o sueldos de forma segura por medio de contratos o pensiones.

### Tenencia

La variable presenta cuatro categorías: tenencia propia, familiar, allegado y arrendatario. El signo refleja que al pasar de tenencia propia a una del tipo allegado, familiar o arrendatario, aumenta la probabilidad de incumplimiento en 6.2%, 8.5% y 6.3% respectivamente. Lo anterior es coherente, debido a que poseer una tenencia propia puede reflejar una mayor solidez y estabilidad financiera.

### Tasa

La variable fue ingresada de manera continua y su distribución va desde 1.8% a 2.7%. El signo indica que, ante el aumento de un punto porcentual de la tasa, aumenta en 34.4% la probabilidad de default; es decir, si aumenta en 0.1% la tasa de interés, la probabilidad de default aumenta en 3.44%. Lo anterior es lógico y coherente desde el punto de vista financiero, debido a que, si aumenta la tasa de interés, la carga financiera aplicada al monto solicitado es mayor, aumentando su costo y amortización mensual.

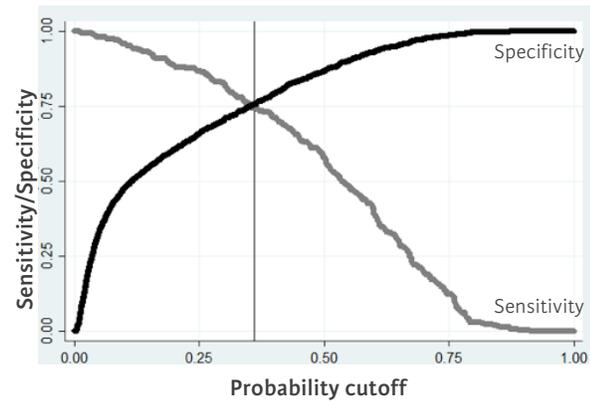
### Nº acreedores financieros

La variable indica que si el socio presenta mayor cantidad de acreedores financieros, su probabilidad de default aumenta en 1.8% por cada acreedor adicional.

#### 4.3.5. Curva de sensibilidad y especificidad

La curva de sensibilidad y especificidad muestra el punto de corte (C) exacto donde el modelo discrimina con mayor exactitud el default y no-default, donde muestra desde que probabilidad (sobre o bajo C) clasifica a los socios en default o no default.

**Ilustración 3.** Punto de corte curvas sensibilidad y especificidad



Fuente: elaboración propia en Stata 14.

Como se observa en la Ilustración 3, en este caso el punto de corte (C) ilustrado con la línea vertical roja posee un valor de 0.36; es decir, el modelo realiza la clasificación bajo el siguiente criterio: si un socio presenta una probabilidad de incumplimiento mayor o igual a C (0.36), este es clasificado como default; y en caso de que el socio presente una probabilidad de incumplimiento menor a C (0.36) este es clasificado como no-default.

#### 4.3.6. Punto de corte: aceptación o rechazo crediticio

El siguiente punto de corte define las condiciones de aceptación y rechazo en una hipotética solicitud de crédito, donde deben considerarse dos factores: el costo incurrido por clasificar de manera errónea un socio y la tasa error global del modelo estimado.

La Tabla 7 muestra las tasas de error global, donde para determinar el punto de rechazo a la solicitud de crédito se considera la mínima. El punto de corte que minimiza la tasa de error total es el 0.49 con valor de 21.69% de error total.

El segundo factor por considerar son los costos asociados a los errores tipo I y II; es decir, el costo de rechazar a un socio bueno y aceptar a uno malo. Para determinar estos valores, se asume que el error tipo

II es tres veces más costoso que el error tipo I, esto debido a que incurrir en el error tipo I es una pérdida monetaria directa. En la Tabla 8 se determina el punto de corte de aceptación directa a una hipotética solicitud de crédito. El punto donde se minimiza el costo total es el 0.3 con un valor de 1,557.

toda solicitud crediticia que se encuentre en él debe ser analizada por un experto.

Como consecuencia de lo anterior, la estrategia de aceptación o rechazo de es la siguiente:

**Tabla 7.** Punto de corte según error total.

Punto de corte rechazo			
Corte	Error tipo I	Error tipo II	Error total
0.05	61.63%	2.2%	47.54%
0.10	56.28%	3.53%	38,00%
0.15	53.34%	5.18%	33.76%
0.20	51.15%	7.35%	31.04%
0.25	48,00%	7.8%	27.8%
0.30	45.5%	9.32%	25.72%
0.35	43.84%	12,00%	24.84%
0.40	40.68%	13.37%	23.14%
0.45	37.52%	15.21%	22.09%
<b>0.49</b>	<b>35.3%</b>	<b>16.49%</b>	<b>21.69%</b>
0.50	35.09%	17.07%	21.87%
0.55	32.16%	19.52%	22.21%
0.60	29,00%	21.62%	22.84%
0.65	26.35%	23.7%	24.02%
0.70	23.15%	25.91%	25.7%
0.75	19.9%	27.39%	27.05%
0.80	22,00%	29.25%	29.16%
0.85	18.52%	29.48%	29.41%
0.90	16.67%	29.73%	29.71%
0.95	0,00%	29.8%	29.8%

Fuente: elaboración propia.

Al combinar los dos enfoques anteriores se define la estrategia para aceptar o rechazar un crédito por parte de la institución financiera. Basándose en la tasa de error total se puede definir que cualquier crédito con una probabilidad de incumplimiento mayor o igual a 0.49, debe ser rechazado inmediatamente. Sin embargo, basándose en el segundo enfoque, se puede definir que cualquier crédito con una probabilidad igual o menor a 0.3 debe ser aceptado. Los puntos de aceptación y rechazo generan un intervalo de indecisión, donde

**Tabla 8.** Punto de corte según costo

Punto de corte aceptación			
Corte	Error tipo I	Error tipo II	Error total
0.05	1993	78	2071
0.10	1577	141	1718
0.15	1363	255	1618
0.20	1189	411	1600
0.25	1027	534	1561
<b>0.30</b>	<b>882</b>	<b>675</b>	<b>1557</b>
0.35	763	957	1720
0.40	635	1128	1763
0.45	495	1332	1827
0.50	410	1656	2066
0.55	310	1959	2269
0.60	214	2325	2539
0.65	150	2676	2826
0.70	91	3054	3145
0.75	48	3306	3354
0.80	21	3729	3750
0.85	6	3858	3864
0.90	1	3927	3928
0.95	0	3942	3942

Fuente: elaboración propia.

**Tabla 9.** Regla de decisión

Intervalo de probabilidad	Regla
[0 – 0.3)	Aceptar
[0.3 – 0.49)	Evaluar
[0.49 - 1]	Rechazar

Fuente: elaboración propia.

Ahora bien, luego de realizar las interpretaciones necesarias e identificadas en el punto de corte que determinan la estrategia de la entidad, se procede a validar estadísticamente los resultados del modelo propuesto.

#### 4.4. Validación del modelo

##### 4.4.1. Prueba de Hosmer-Lemeshow

Esta prueba permite evaluar el ajuste global del modelo; es decir, qué tanto coincide lo observado con lo estimado. Realiza el análisis agrupando los casos en deciles de riesgo y compara las frecuencias observadas con las esperadas dentro de cada subconjunto utilizando el test  $\chi^2$ .

La Tabla 10 muestra que las desviaciones de lo estimado respecto de lo observado son similares, lo cual queda confirmado por el resultado de la prueba de Hosmer-Lemeshow, al obtener una probabilidad del 45.71% siendo esta mayor a 5%, por lo tanto, no rechazamos la hipótesis nula y confirmamos el correcto ajuste global del modelo estimado en la base de entrenamiento (80% de la muestra).

**Tabla 10.** Prueba bondad de ajuste de Hosmer-Lemeshow

<b>(Table collapsed on quantiles of estimated probabilities)</b>	
number of observations =	1103
number of groups =	10
Hosmer-Lemeshow $\chi^2(8)$ =	7.76
Prob > $\chi^2$ =	0.4571

Group	Prob	Obs_1	Exp_1	Obs_o	Exp_o	Total
1	0.0196	1	1.4	110	109.6	111
2	0.0407	3	3.2	107	106.8	110
3	0.0792	2	6.5	109	104.5	111
4	0.1407	13	11.7	97	98.3	110

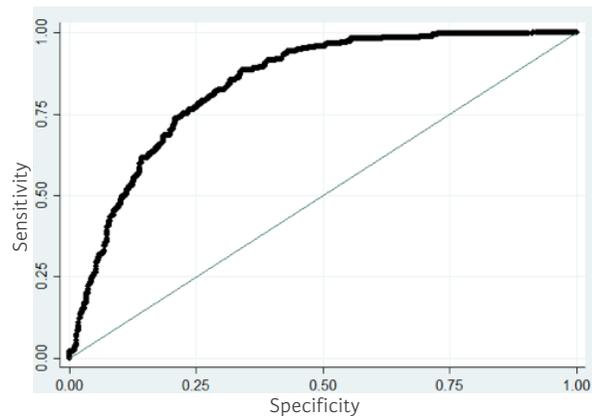
5	0.2447	25	20.7	85	89.3	110
6	0.3490	36	33	74	77	110
7	0.4684	49	45.8	62	65.2	111
8	0.5771	59	58.1	52	52.9	111
9	0.6661	75	68.4	35	41.6	110
10	0.9112	77	80.5	32	28.5	109

Fuente: elaboración propia en Stata 14.

##### 4.4.2. Curva de ROC

En la Ilustración 8 el área bajo la curva establece que la capacidad discriminante del modelo es de un 84.02%, lo cual da un indicio de que el modelo posee una gran capacidad de clasificación.

**Ilustración 8.** Curva ROC



Area under ROC curve = 0.8402

Fuente: elaboración propia en Stata 14.

##### 4.4.3. Matriz de confusión

La Tabla 11 presenta la matriz de confusión del 20% de la muestra (base de validación), donde se pueden observar las predicciones del modelo:

**Tabla 11.** Matriz de confusión de la base de validación

		Observado		
		D	~D	Total
Predicción	+	256	172	428
	-	84	591	675
Total		340	763	1103

Classified + if predicted $Pr(D) \geq 0.36$			
Sensitivity	$Pr(+ D)$		75.29%
Specificity	$Pr(- \sim D)$		77.46%
Positive predictive value	$Pr(D +)$		59.81%
Negative predictive value	$Pr(\sim D -)$		87.56%
False + rate for true ~D	$Pr(+ \sim D)$		22.54%
False - rate for true D	$Pr(- D)$		24.71%
False + rate for classified +	$Pr(\sim D +)$		40.19%
False - rate for classified -	$Pr(D -)$		12.44%
<b>Correctly classified</b>			<b>76.79%</b>

Fuente: elaboración propia en Stata 14.

Según la matriz de confusión se obtienen las siguientes conclusiones respecto de la sensibilidad, especificidad, tasa de aciertos y tasa de error tipo I y tipo II:

- Sensibilidad: dado que el individuo posee el fenómeno de default, la probabilidad de que el modelo lo clasifique como default es del 75.29%.
- Especificidad: dado que el individuo no posee el fenómeno de default, la probabilidad de que el modelo lo clasifique como no default es del 77.46%.
- Tasa de aciertos totales: los individuos correctamente clasificados corresponden al 76.79%.
- Tasa de errores: dado que el modelo clasificó al individuo como default, la probabilidad de que sea no default es del 40.19% (error tipo I), y dado que el modelo clasificó al individuo como no default, la

probabilidad de que sea default es del 12.44% (error tipo II). Además, el modelo cuenta con una tasa de error global del 23.21%.

#### 4.5. Simulación de evaluación crediticia

En esta sección se simulará la evaluación de dos socios pertenecientes a la entidad financiera, demostrando así los tres posibles escenarios establecidos por la estrategia construida en el apartado 4.3.6. para aceptar o rechazar una solicitud.

- El socio 1 presenta las siguientes características:

**Tabla 12.** Primer caso de evaluación crediticia

Variable	Característica socio o crédito
Plazo	18 a 23
Moralidad externa	Con problemas de moralidad
Comportamiento interno	0-30
Género	Masculino
Antigüedad laboral	0 a 4
Antigüedad socio	Menor a 4
Edad	19 a 30
Situación laboral	Dependiente
Tenencia	Arrendatario
Tasa	2.1
N° acreedores financieros	1

Fuente: elaboración propia.

Un socio con las características anteriores posee una probabilidad de incumplimiento correspondiente a un 57.08%; es decir, según la estrategia definida en la Tabla 12, la solicitud de este crédito debe ser rechazada.

- El socio 2, a diferencia del socio 1, solo tendrá el cambio referente a la variable moralidad externa, donde se le asignará sin problemas de moralidad, por lo cual presenta las siguientes características:

**Tabla 13.** Segundo caso de evaluación crediticia

Variable	Característica socio o crédito
Plazo	18 a 23
Moralidad externa	Sin problemas de moralidad
Comportamiento interno	0-30
Género	Masculino
Antigüedad laboral	0 a 4
Antigüedad socio	Menor a 4
Edad	19 a 30
Situación laboral	Dependiente
Tenencia	Arrendatario
Tasa	2.1
Nº acreedores financieros	1

Fuente: elaboración propia.

Un socio con las características anteriores posee una probabilidad de incumplimiento correspondiente a un 32.39%; es decir, según la estrategia definida en la Tabla 13, la solicitud de este crédito debe ser enviada a riesgo para su evaluación.

- El socio 3 tendrá al igual que el socio 2 se le asignará la característica de “sin problemas de moralidad” y se le modificarán las variables antigüedad laboral y antigüedad socio asignándole a una mejor categoría.

**Tabla 14.** Tercer caso de evaluación crediticia

Variable	Característica socio o crédito
Plazo	18 a 23
Moralidad externa	Sin problemas de moralidad
Comportamiento interno	0-30
Género	Masculino
Antigüedad laboral	5 a 14
Antigüedad socio	Mayor a 4
Edad	31 a 49
Situación laboral	Dependiente

Tenencia	Familiar
Tasa	2.1
Nº acreedores financieros	1

Fuente: elaboración propia

Un socio con las características anteriores posee una probabilidad de incumplimiento correspondiente a un 10.69%, es decir, según la estrategia definida en la tabla 11, la solicitud de este crédito debe ser aceptada

## 5. ESCALAMIENTO A SCORE

Para que la Regresión Logística estimada tenga un mejor entendimiento se realiza un *escalamiento a score* o puntaje, el cual será asignará bajo las siguientes premisas:

1. El Score debe ser construido bajo una escala comprensible (ejemplo: 1 – 1000).
2. Cada atributo o apertura de cada variable tendrá asociado un Score negativo o positivo.
3. El Score final debe ser la sumatoria de todos los scores particulares de cada apertura.

Generalmente, la forma de cálculo del Score es el siguiente:

$$Score = Escala + Factor * Ln (Odds)$$

Donde:

Odds: chance de ocurrencia del evento. En este caso, cada 10 individuos clasificados como buenos hay 1 clasificado como malo.

Posteriormente, para la creación del Score debe establecerse lo siguiente:

Pdo: cantidad de puntos necesarios para duplicar las chances de ocurrencia del evento o la cantidad de puntos necesarios para duplicar los Odds. Además esta variable permite establecer la amplitud del Score que queremos generar.

Además, el Score debe cumplir las siguientes restricciones:

$$\begin{aligned} \text{Score} &= \text{Escala} + \text{Factor} * \text{Ln}(\text{Odds}) \\ \text{Score} + \text{pdo} &= \text{Escala} + \text{Factor} * \text{Ln}(2 * \text{Odds}) \\ \text{Ln}(\text{Odds}) &= -(\sum \beta_i x_i + \beta_0) \end{aligned}$$

Al resolver el sistema de ecuaciones:

$$\begin{aligned} \text{Factor} &= \frac{\text{pdo}}{\text{Ln}(2)} \\ \text{Escala} &= \text{Score} - \text{Factor} * \text{Ln}(\text{Odds}) \\ \text{Ln}(\text{Odds}) &= -(\sum \beta_i x_i + \beta_0) \end{aligned}$$

Al realizar una evaluación crediticia, y una vez entregada la resolución al cliente, es importante explicar por qué le fue rechazada. Es importante contar con un Score individual por variable y score por apertura de la variable; por lo tanto, si se tienen V variables, con aperturas o categorías, entonces:

$$\begin{aligned} \text{Score} &= \text{Escala} + \text{Factor} * \left( \sum_{V, A_v} \beta_{k_v} * X_{k_v} + \beta_0 \right) \\ \text{Score} &= \text{Escala} + \sum_V \left( \text{Factor} * \left( \sum_{V, K_v} \beta_{k_v} * X_{k_v} + \frac{\beta_0}{V} \right) \right) \\ \text{Score} &= \sum_V \left( \frac{\text{Escala}}{V} - \text{Factor} * \left( \sum_{V, K_v} \beta_{k_v} * X_{k_v} + \frac{\beta_0}{V} \right) \right) \end{aligned}$$

El Score neutro corresponde a la eliminación de todas las variables y su forma de cálculo es la siguiente:

$$\text{Score neutro} = \frac{\text{Escala}}{V} - \text{Factor} * \frac{\beta_0}{V}$$

Toda variable que presente un puntaje menor al Score neutro quiere decir que tiene un riesgo asociado alto. Para la regresión logística estimada y su posterior escalamiento, se establecieron los siguientes parámetros:

**Tabla 15.** Escalamiento según variables

Escalamiento	
Score	300
pdo	56
Factor	80,791
Odds	10
Offset	113,972
Constante	-6,461
n	12
Neutral	53

Donde:

$$Factor = \frac{56}{Ln(2)} = 80,791$$

$$Offset = Score - Factor * Ln(Odds) = 300 - 80,791 * Ln(10) = 113,972$$

$$Score\ Neutro = \frac{Offset}{V} - Factor * \frac{Constante}{V} = \frac{113,972}{12} - 80,791 * \frac{-6,461}{12} = 53$$

$$Score_{v,A_v} = \frac{Offset}{V} - Factor * (\beta_{k_v} * X_{k_v} + \left(\frac{Constante}{V}\right))$$

Ejemplo de variable plazo apertura 12 a 17:

$$Score_{v,A_v} = \frac{113,972}{12} - 80,791 * (0,277 * 1 + \left(\frac{-6,461}{12}\right)) = 75$$

El Score debe asignarse a cada variable y a su respectiva apertura o categoría (Tabla 16).

**Tabla 16.** Scoring según aperturas de variables

Variable	Clase	Detalle	Coef.	Xi	Score sin/R	Score x apertura	Característica socio	Score socio	MIN	MAX
Plazo	1	12 a 17	-0,277	1	75	75	12 a 17	75	-3	75
	2	18 a 23	0,277	1	31	31				
	3	24 a 35	0,587	1	6	6				
	4	36 o más	0,698	1	-3	-3				
Moralidad externa	1	Sin problemas de moralidad	-1,021	1	135	136	Con problemas de moralidad	-30	-30	136
	2	Con problemas de moralidad	1,021	1	-29	-30				
Comportamiento Interno	1	0 a 30	-0,474	1	91	92	Más de 90	-47	-47	92
	2	31 a 60	0,474	1	15	15				
	3	61 a 90	0,956	1	-24	-25				
	4	Más de 90	1,228	1	-46	-47				
Género	1	Femenino	-0,418	1	87	87	Femenino	87	20	87
	2	Masculino	0,418	1	19	20				
Antigüedad laboral	1	0 a 4	0,318	1	27	28	15 o más	139	28	139
	2	5 a 14	-0,318	1	79	79				
	3	15 o más	-1,057	1	138	139				

Antigüedad Socio	1	Menor a 4	1,257	1	-49	-49	Mayor a 4	155	-49	155
	2	Mayor a 4	-1,257	1	155	155				
Edad	1	19 a 30	-0,333	1	80	80	50 o más	103	27	103
	2	31 a 49	0,333	1	26	27				
	3	50 o más	-0,609	1	102	103				
Situación Laboral	1	Independiente	0,592	1	5	6	Dependiente	101	6	120
	2	Dependiente	-0,592	1	101	101				
	3	Jubilados, pensionados y montepiados	-0,829	1	120	120				
Tenencia	1	Propia	-0,416	1	87	87	Propia	87	8	87
	2	Allegado	0,416	1	19	20				
	3	Arrendatario	0,566	1	7	8				
	4	Familiar	0,421	1	19	19				
Tasa	1	Continua	2,348	0,027	48	48	0,027	48	48	
Nº acreedores financieros	1	Continua	0,126	0	53	53	0	53	53	
							PUNTAJE FINAL	771	56	994

Fuente: Elaboración propia. S/N: Sin problemas de moralidad. C/M: Con problemas de moralidad. J/P/M: Jubilados, pensionados y montepiados.

Luego de realizar todas las iteraciones y combinaciones posibles de las diferentes variables y sus aperturas, se aprecia que la Regresión Logística entrega un Score que no permite mitigar la inclusión de individuos con problemas de moralidad (interna y externa), presentando los siguientes problemas:

1. No envía a los individuos con problemas de moralidad externa e interna (por sí solas) a una instancia previa de evaluación de un analista de riesgo, que permita realizar una evaluación exhaustiva de sus parámetros.
2. No rechaza, de manera inmediata sin pasar a una instancia de evaluación de un analista, a los individuos que presenten al mismo tiempo problemas de moralidad externa e interna.

Debido a lo anterior se reasignaron los puntajes a las variables y sus respectivas aperturas. Para lo anterior, se tiene que considerar lo siguiente:

1. Se disminuyó el valor máximo y mínimo del score total.
2. Se reasignaron los puntajes a todas las variables y sus aperturas, salvo las variables Tasa y Número de Acreedores Financieros. Dicha reasignación se realizó siguiendo las tendencias de los puntajes originales; es decir, se reasignaron analizando la ponderación que entregaba su coeficiente en la Regresión Logística.
3. En el caso de las variables de comportamiento crediticio (interno y externo) se les asignó una mayor relevancia, penalizando con resta de puntaje a la presencia de moralidad interna y/o externa.

Luego de la reasignación de puntajes, el Score final quedó forma explícita en la Tabla 17, que además presenta la evaluación de un individuo aleatorio.

**Tabla 17.** Scoring según aperturas de variables

Variable	Clase	Detalle	Coef.	Xi	Score sin/R	Score x apertura	Característica socio	Score socio	MIN	MAX
Plazo	1	12 a 17	-0,277	1	69	60	12 a 17	60	-10	60
	2	18 a 23	0,277	1	45	40				
	3	24 a 35	0,587	1	31	20				
	4	36 o más	0,698	1	26	-10				
Moralidad externa	1	Sin problemas de moralidad	-1,021	1	101	20	Sin problemas de moralidad	20	-150	20
	2	Con problemas de moralidad	1,021	1	12	-150				
Comportamiento Interno	1	0 a 30	-0,474	1	77	50	0 a 30	50	-150	50
	2	31 a 60	0,474	1	36	-80				
	3	61 a 90	0,956	1	15	-130				
	4	Más de 90	1,228	1	4	-150				
Género	1	Femenino	-0,418	1	75	75	Femenino	75	39	75
	2	Masculino	0,418	1	39	39				
Antigüedad laboral	1	0 a 4	0,318	1	43	20	15 o más	60	20	60
	2	5 a 14	-0,318	1	70	40				
	3	15 o más	-1,057	1	102	60				
Antigüedad Socio	1	Menor a 4	1,257	1	2	3	Mayor a 4	30	3	30
	2	Mayor a 4	-1,257	1	111	30				
Edad	1	19 a 30	-0,333	1	71	40	50 o más	30	20	40
	2	31 a 49	0,333	1	42	20				
	3	50 o más	-0,609	1	83	30				
Situación Laboral	1	Independiente	0,592	1	31	32	Dependiente	70	32	90
	2	Dependiente	-0,592	1	82	70				
	3	Jubilados, pensionados y montepiados	-0,829	1	93	90				
Tenencia	1	Propia	-0,416	1	75	75	Propia	75	33	75
	2	Allegado	0,416	1	39	39				
	3	Arrendatario	0,566	1	32	33				
	4	Familiar	0,421	1	38	39				
Tasa	1	Continua	2,348	0,022	54	55	0,022	55	55	
N° acreedores financieros	1	Continua	0,126	5	29	30	5	30	55	
							<b>PUNTAJE FINAL</b>	<b>555</b>	<b>-53</b>	<b>610</b>

Fuente: Elaboración propia. S/N: Sin problemas de moralidad. C/M: Con problemas de moralidad. J/P/M: Jubilados, pensionados y montepiados.

El puntaje entregado en el Score bajo la regla de decisión establecida, que fue llevada a puntajes, se acepta la solicitud crediticia del individuo (Tabla 18).

**Tabla 18.** Reglas de decisión

Regla de Decisión	
Aceptado	(427 a 610]
Evalúa Riesgo	(311 a 427]
Rechazado	(-53 a 311)

## 6. COMENTARIOS FINALES Y TRABAJO FUTURO

En esta investigación se analizaron los factores determinantes del fenómeno de default en una cartera de consumo. Los resultados fueron satisfactorios en cuanto a la tasa de aciertos, errores y la capacidad predictiva del modelo establecido. Si bien los resultados fueron los esperados, es necesario que para una futura recalibración del modelo se incorporen nuevas variables de comportamiento interno por tramos de mora, analizando diferentes escenarios de tiempo previos al momento del otorgamiento.

Respecto del modelo planteado, se debe hacer un seguimiento que sea capaz de detectar cambios significativos en la distribución de las variables con las que fue construido. Se entiende como un cambio significativo aquel que modifica la capacidad discriminante del modelo, por lo cual necesita ser corregido mediante una recalibración que aplique las modificaciones de la distribución de las variables, incorporando nuevas o eliminando las ya existentes.

## REFERENCIAS

**Banco Central** (2003). *Normas financieras para Cooperativas de Ahorro y Crédito*.

**Covarrubias, G.** (2012). *Construcción y validación de una metodología de seguimiento para modelos de regresión logística*. Santiago: Universidad de Chile.

**Fayyad, U.** (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3).

**Gujarati, D. y Porter, D.** (2010). *Econometría* (quinta edición). McGraw-Hill.

**Hadad, A., Evin, D. y Drozdowicz, B.** (2007). *Modelo para el Tratamiento de Datos Desbalanceados basado en Redes Neuronales Autoorganizadas*. Concepción del Uruguay, provincia de Entre Ríos, Argentina: Universidad Nacional de Entre Ríos.

**Hernández, J., Ramírez, M. y Ferrari, C.** (2004). *Introducción a la Minería de Datos*. Madrid, España: Pearson Prentice Hall.

**Hosmer, D. y Lemeshow, S.** (1989). *Applied Logistic Regression*. Estados Unidos: Wiley Series in Probability and Statistics.

**Medina, M. y Ulfe, H.** (2015). *Modelo de credit scoring para predecir el otorgamiento de crédito en una Cooperativa de Ahorro y Crédito*. Lambayeque, Perú: Universidad Nacional Pedro Ruiz Gallo.

**Moreno, S.** (2013). *El Modelo Logit Mixto para la construcción de un Scoring de Crédito*. Bogotá, Colombia: Universidad Nacional de Colombia.

**Ormazábal, F.** (2014). Variables que afectan la tasa de incumplimiento de créditos de los chilenos. *Revista de Análisis Económico*; Vol. 29, N°1, pp. 3-16.

**Siddiqi, N.** (2016). *Intelligent Credit Scoring: Building and implementing better credit risk scorecards*. Estados Unidos: Wiley.

**Tolvett, C.** (2011). *Estudio de metodologías para el seguimiento de modelos de credit scoring utilizando regresión logística*. Santiago: Universidad de Chile.

**Velandia, N.** (2013). *Establecimiento de un Modelo Logit para la Medición del Riesgo de Incumplimiento en Créditos para una Entidad Financiera del Municipio de Arauca, Departamento de Arauca*. Bogotá, Colombia: Universidad Nacional de Colombia.