

Diseño de un modelo predictivo de fuga de asociados y sus causas en una cooperativa de ahorro y crédito de Santander usando análisis de sobrevivencia

Omar Gabriel Oviedo Uribe

Trabajo de Grado para Optar el título de Especialista en Estadística

Director

Gabriel Yáñez Canal

Doctor en Matemática Educativa

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Matemática

Especialización en Estadística

Bucaramanga

2020

### **Dedicatoria**

A mi amada esposa por su apoyo en todo momento, aliento y comprensión.

A mi madre por siempre impulsarme a caminar la vida por el sendero académico y retarme a superar los obstáculos que se presentan.

### **Agradecimientos**

A la Cooperativa que con cada una de las acciones y apoyos que me ha brindado en mi paso por ahí ha transformado mi vida para bien radicalmente.

Al profesor Gabriel Yañez por su paciencia y su particular pero muy efectiva forma de educar siempre de forma exigente y basada en el respeto.

**Contenido**

	<b>Pág.</b>
Introducción.....	11
1. Objetivos.....	17
1.1 Objetivo General.....	17
1.2 Objetivos Específicos.....	17
2. Cuerpo del Trabajo.....	18
2.1 Marco Referencial.....	18
2.1.1 Método.....	24
2.1.2 Resultados.....	27
2.1.2.1 Análisis descriptivo.....	27
2.1.2.1.1 Variables sociodemográficas.....	28
2.1.2.1.2 Variables transaccionales.....	38
2.1.2.2 Tablas cruzadas (Análisis bivariado).....	47
2.1.2.3 Conclusiones de la exploración univariada y bivariada de los datos.....	59
2.1.2.4 Aplicación del análisis de supervivencia.....	60
2.1.2.4.1 Curvas de supervivencia de Kaplan-Meier.....	60
2.1.2.4.2 Modelo de regresión de Cox.....	75
3. Conclusiones.....	85
Referencias Bibliográficas.....	88

**Lista de Tablas**

	<b>Pág.</b>
Tabla 1. Variables para utilizar en el proyecto .....	25
Tabla 2. Correlación entre las variables ingresos y egresos .....	38
Tabla 3. Rangos por percentiles.....	40
Tabla 4. Correlación de las variables continuas transaccionales .....	47
Tabla 5. Tabla cruzada de Fugado vs Sexo .....	48
Tabla 6. Tabla cruzada de Fugado vs Rango de edad.....	48
Tabla 7. Tabla cruzada de Fugado vs Estado civil.....	49
Tabla 8. Tabla cruzada de Fugado vs Tipo de vivienda.....	49
Tabla 9. Tabla cruzada de Fugado vs Cargo .....	50
Tabla 10. Tabla cruzada de Fugado vs Estrato .....	50
Tabla 11. Tabla cruzada de Fugado vs Nivel educativo .....	51
Tabla 12. Tabla cruzada de Fugado vs Rango de antigüedad .....	51
Tabla 13. Tabla cruzada de Fugado vs Rango de ingresos .....	52
Tabla 14. Tabla cruzada de Fugado vs Rango de egresos.....	52
Tabla 15. Tabla cruzada de Fugado vs Personas a cargo .....	53
Tabla 16. Tabla cruzada de Fugado vs Tenencia de crédito.....	53
Tabla 17. Tabla cruzada de Fugado vs Tenencia de ahorros.....	54
Tabla 18. Tabla cruzada de Fugado vs Tenencia de PAP .....	54

Tabla 19. Tabla cruzada de Fugado vs Tenencia de CDAT .....	55
Tabla 20. Tabla cruzada de Fugado vs Inactividad en la cuenta de ahorros .....	55
Tabla 21. Tabla cruzada de Fugado vs Rango de número de productos.....	56
Tabla 22. Tabla cruzada de Fugado vs Rango de saldo promedio de ahorros .....	56
Tabla 23. Tabla cruzada de Fugado vs Rango de saldo promedio de crédito .....	57
Tabla 24. Tabla cruzada de Fugado vs Rango de saldo promedio de PAP.....	57
Tabla 25. Tabla cruzada de Fugado vs Rango de saldo promedio de CDAT .....	58
Tabla 26. Tabla cruzada de Fugado vs Rango de número de transacciones voluntarias .....	58
Tabla 27. Resultado general regresión de Cox .....	76
Tabla 28. Prueba Ómnibus del modelo .....	77
Tabla 29. Codificación de variables categóricas .....	79
Tabla 30. Variables en la ecuación modelo de regresión de Cox (Parte 1).....	80
Tabla 31. Variables en la ecuación modelo de regresión de Cox (Parte 2).....	82
Tabla 32. Predicción del modelo vs Datos reales en la muestra.....	84

**Lista de Figuras**

	<b>Pág.</b>
Figura 1. Función de Supervivencia .....	19
Figura 2. Función de riesgo .....	20
Figura 3. Curva de sobrevivencia de Kaplan-Meier. ....	22
Figura 4. Distribución por género.....	28
Figura 5. Distribución por rangos de edad .....	29
Figura 6. Distribución por estado civil.....	30
Figura 7. Distribución por tipo de vivienda.....	30
Figura 8. Distribución por ocupación.....	31
Figura 9. Distribución por estrato .....	32
Figura 10. Distribución por nivel educativo .....	33
Figura 11. Distribución por rango de antigüedad .....	34
Figura 12. Distribución por tenencia de personas a cargo .....	34
Figura 13. Histograma de ingresos .....	35
Figura 14. Histograma de egresos.....	36
Figura 15. Distribución por rango ingresos .....	37
Figura 16. Distribución por rango egresos .....	37
Figura 17. Distribución por tenencia de productos.....	39
Figura 18. Histograma de número de productos a corte .....	41

Figura 19. Histograma de número de transacciones en 2014 .....41

Figura 20. Histograma de promedio de saldos de ahorros en 2014 .....42

Figura 21. Histograma de promedio de saldos de CDAT en 2014 .....42

Figura 22. Histograma de promedio de saldos de crédito en 2014 .....43

Figura 23. Histograma de promedio de saldos de PAP en 2014 .....43

Figura 24. Distribución por cantidad de productos y transacciones .....44

Figura 25. Distribución por rangos de promedio de saldos de los productos en 2014 .....46

Figura 26. Curva de supervivencia general .....61

Figura 27. Curva de supervivencia por sexo .....62

Figura 28. Curva de supervivencia por rango de edad .....63

Figura 29. Curva de supervivencia por tipo de vivienda.....64

Figura 30. Curva de supervivencia por ocupación.....65

Figura 31. Curva de supervivencia por estrato .....66

Figura 32. Curva de supervivencia por nivel educativo.....67

Figura 33. Curva de supervivencia por rango de antigüedad .....68

Figura 34. Curva de supervivencia por rango de ingresos .....69

Figura 35. Curva de supervivencia por rango de egresos .....70

Figura 36. Curva de supervivencia por tenencia de productos a dic. de 2014 .....71

Figura 37. Curva de supervivencia por cantidad de productos y transacciones .....73

Figura 38. Curva de supervivencia por nivel rango de promedio de saldos de productos en  
2014.....74

## Resumen

**Título:** Diseño de un modelo predictivo de fuga de asociados y sus causas en una cooperativa de ahorro y crédito de Santander usando análisis de sobrevivencia \*

**Autor:** Omar Gabriel Oviedo Uribe \*\*

**Palabras Clave:** Deserción, Análisis de sobrevivencia, Regresión de Cox, Probabilidad.

### Descripción:

En el contexto económico actual es bien sabida la importancia de dirigir esfuerzos a la retención de clientes. Esta es una actividad imprescindible dada la gran cantidad de oferta de diferentes productos y servicios en entidades financieras que buscan arrebatar a los mejores perfiles que se encuentran en cooperativas y bancos distintos al propio. En procura de evitar que sus asociados deserten, la Cooperativa define que es necesaria una herramienta estadística que permita anticiparse al deseo de las personas de cancelar sus cuentas en la entidad, para el desarrollo de este modelo predictivo se tienen en cuenta diferentes variables para cada persona propias de su comportamiento financiero o características sociodemográficas que se almacenan a lo largo del tiempo. A partir de los anterior, con un seguimiento de 5 años a poco más de seis mil personas de la oficina principal de la empresa, se realiza un análisis de sobrevivencia, en el que mediante las curvas de supervivencia de Kaplan-Meier y la regresión de Cox se obtiene el perfil de los asociados que son más propensos a retirarse de la Cooperativa. Con los resultados del modelo, el área de retención y fidelización de asociados está en capacidad de lanzar campañas comerciales que permitan reducir los niveles de deserción.

---

\* Trabajo de Grado

\*\* Facultad de Ciencias. Escuela de Matemáticas. Director: Gabriel Yañez Canal, Doctor en matemática Educativa.

### Abstract

**Title:** Design of a predictive model of member's churn and its causes in a Santander credit union using survival analysis \*

**Author:** Omar Gabriel Oviedo Uribe \*\*

**Key Words:** Attrition, Survival analysis, Cox regression, Probability

### Description:

In the current economic context, the importance of directing efforts to customer retention is well known. This is an essential activity given the large number of different products and services offered by financial institutions that seek to capture the best profiles found in cooperatives and banks other than their own. In an attempt to prevent its associates from deserting, the Cooperative defines that a statistical tool is necessary to anticipate people's desire to cancel their accounts in the entity, for the development of this predictive model different variables are taken into account for each person specific to their financial behavior or sociodemographic characteristics that are stored over time. From the above, with a 5-year follow-up of just over six thousand people from the company's main office, a survival analysis is carried out, using the Kaplan-Meier survival curves and the regression of Cox profile is obtained from the associates who are most likely to leave the Cooperative. With the results of the model, the area of retention and loyalty of associates is able to launch commercial campaigns that reduce the levels of attrition.

---

\* Degree work

\*\* Science Faculty. School of Mathematics. Director: Gabriel Yañez Canal, Doctor in Educational Mathematics

## Introducción

La cartera de clientes es, sin duda alguna, el activo más importante para todas las empresas, especialmente para aquellas que tienen como actividad económica la disposición de productos de ahorro y crédito, entre ellas, las cooperativas. El presente proyecto se desarrollará en una cooperativa de ahorro y crédito, que en adelante por razones de reserva de información se denominará simplemente la Cooperativa. En el marco del cooperativismo profesado por la entidad no se maneja la figura de “cliente”, la cual es reemplazada por la de “asociado”, quien para ser catalogado como tal debe abrir una cuenta denominada en la Cooperativa como ‘Aportes’. El estudio que aquí se llevará a cabo se centrará en desarrollar un modelo que permita predecir qué asociados cancelarán su cuenta de aportes y dejarán de ser asociados mediante la utilización de variables sociodemográficas, movimientos de los diferentes productos de crédito y de ahorros con el objetivo de identificar los factores que inciden en su fuga.

En el caso concreto de la Cooperativa, se realizan una serie de campañas enfocadas a lograr un grado de fidelización superior incentivando la transaccionalidad, es decir, el aumento de las transacciones y/o movimientos que realicen las personas ya sean pagos o trámites administrativos (Banco de Bogotá, s.f), y ofreciendo beneficios propios del sector solidario. El indicador de deserción mensual (Asociados retirados del mes / Asociados que ingresan en el mes) es muy alto: en el transcurso del 2019 ha estado por encima del 40%, y el índice de deserción anual (Asociados retirados en el año/Asociados al inicio del año) ha venido creciendo desde el 5.5% en el año 2013

a cifras cercanas al 7% en el año inmediatamente anterior, incrementando así las pérdidas de la entidad, ya que es mucho más costoso captar un nuevo cliente que conservar uno que ya se tiene.

Actualmente, la Cooperativa realiza una clasificación destinada a identificar los asociados que son más rentables para la misma, estos se denominan “asociados preferenciales”, quienes se clasifican de esta manera, en función de la trayectoria que hayan manejado con la entidad, antigüedad y saldos en los diferentes productos y servicios de la cooperativa. Esta clasificación se realiza básicamente utilizando únicamente sus movimientos transaccionales, sin embargo, no se tienen en cuenta las variables sociodemográficas de las personas ni las causas que llevan al retiro de los asociados. En consecuencia, la Cooperativa no posee una herramienta estadística que le permita determinar la probabilidad de que sus asociados abandonen la entidad en donde se integren su movimiento transaccional, características sociodemográficas y razones para retirarse. A partir de los resultados del modelo a diseñar se espera que se puedan emprender esfuerzos para retener aquellos asociados que vale la pena retener teniendo en cuenta su clasificación como preferenciales y/o sean rentables para la organización.

El problema de investigación se centra en responder la siguiente pregunta:

- ¿Qué variables sociodemográficas y transaccionales tienen relevancia en el retiro de los asociados de la Cooperativa?

En el contexto económico actual es bien sabido los esfuerzos dirigidos a la retención de asociados. Esta es una actividad imprescindible y generadora de competitividad en el sector, dada la gran cantidad de oferta de diferentes productos y servicios en entidades financieras que buscan arrebatarse a los mejores perfiles que se encuentran en cooperativas y bancos distintos al propio. En productos de colocación, es decir créditos, constantemente surgen diversas promociones,

principalmente de tasas y exención de impuestos, que desencadenan un enorme movimiento de personas entre entidades financieras que buscan mejores condiciones para sus préstamos.

Dadas así las cosas, es primordial para las entidades financieras invertir en la retención de sus asociados, so riesgo de ver disminuidas sus utilidades. Ahora bien, esta inversión debe estar enfocada en mejorar los aspectos de alto riesgo que con alta probabilidad conduzcan a la deserción de los asociados.

Sin embargo, en este momento la Cooperativa no cuenta con un modelo que le permita predecir con confiabilidad cuáles pueden ser aquellos asociados que con alta probabilidad podrían desertar, porque, como ya se ha mencionado previamente, solo se basa en el comportamiento transaccional dejando a un lado las características sociodemográficas e históricas asociadas a su comportamiento como cliente, sin ahondar en las causas que pudieran provocar la insatisfacción y posterior retiro de los asociados.

Es en este contexto en el que se plantea el presente trabajo de grado, cuyo objetivo es obtener la probabilidad de riesgo de deserción de los asociados con ciertas características. Este conocimiento hará posible generar campañas focalizadas a las personas con mayor probabilidad de deserción lo que permitirá optimizar los recursos y esfuerzos enfocados a segmentos definidos. Igualmente, se estará en capacidad de corregir los principales defectos operacionales y de servicio que serán identificados gracias al conocimiento de las razones que motivaron a los asociados a retirarse de la entidad.

Teniendo en cuenta la gran importancia que tienen los clientes para todas las organizaciones, incluyendo las del sector financiero, son diversos los esfuerzos que se han realizado en procura de predecir cuáles son las personas que son más proclives a dar por terminada la relación que tienen con la entidad.

Es importante recalcar que la deserción de clientes se puede clasificar dentro de dos grandes tipos, la deserción voluntaria y la deserción involuntaria. La primera de estas es aquella en la que el cliente por iniciativa propia decide dar por terminado el vínculo que posee con la entidad; mientras que la deserción involuntaria se da por decisiones y políticas de la organización, situaciones que se dan en el sector financiero principalmente por causas como la mora en créditos o algún tipo de fraude y/o delito. Es en la deserción voluntaria donde autores como (Miranda, Rey y Weber, 2005) realizan un estudio para un banco chileno en el que utilizando Máquinas de Soporte Vectorial (Carmona,2014)<sup>1</sup>, predicen si un cliente del banco con cartera vigente se fugará. Para ello utilizan 4 agrupaciones de variables: (i) Variables de comportamiento bancario (ii) Variables sociodemográficas (iii) Variables de percepción de la calidad del servicio (iv) variables de entorno. Los resultados obtenidos fueron mejores que los obtenidos utilizando redes neuronales multicapas.

Las variables que mostraron su fuerza predictiva para la deserción de los clientes fueron: Saldos promedios mensuales, número de transacciones, tenencia de productos, edad, estado civil, nivel de estudios, número de reclamos y antigüedad.

Otro estudio relacionado con la deserción se llevó a cabo en un banco colombiano dirigido al producto de tarjeta de crédito (Herrera, 2016). En este trabajo se desarrollaron varios modelos de regresión logística en el que la variable respuesta era binaria: cancelada o no cancelada, usando diferentes técnicas de selección como backward, forward, stepwise.

---

<sup>1</sup> Técnica que consiste principalmente en un algoritmo de aprendizaje supervisado en el que se separa un hiperplano en dos o más regiones que corresponden a cada clase definidas para la variable objetivo. Una SVM define el hiperplano óptimo para la superficie de decisión, de manera que la separación entre las dos clases en los datos se amplía al máximo.

Después de repetidas pruebas con diferentes modelos, se concluyó que las personas que recibían su pago de nómina en el banco eran menos propensas a desertar, así como las más propensas a desertar eran aquellas con antigüedad entre 2 y 5 años.

Otro estudio realizado en busca de desarrollar un modelo estadístico para predecir el riesgo de abandono de los clientes en un banco español se basó en características sociodemográficas, económicas y se le dio bastante relevancia al nivel de satisfacción y confianza de las personas para con la entidad financiera (Fiestas, Liébana y Rubio, 2013).

En este proyecto se empleó la técnica de regresión logística sobre una muestra determinada del banco español, donde la variable respuesta era dicotómica y un tiempo máximo de 18 meses posteriores a la encuesta de satisfacción ya que de esta manera no se supera el tiempo de diferencia entre la realización de este tipo de encuestas en la organización que es de dos años. El porcentaje de la muestra utilizada para la estimación del modelo fue del 75% y el restante 25% para su validación. Los resultados del modelo, empleando la base de estimación indican que el acierto para la desertación fue de casi 9 de 10 personas, mientras que, al realizar la validación del modelo con la base reservada para tal fin, se aprecia que sus porcentajes de efectividad disminuyen en 2 puntos porcentuales con respecto a la estimación previamente generada, por lo que se deduce que la precisión del modelo puede verse ligeramente afectada cuando se trabaje con datos nuevos, aunque sigue siendo consistente.

Se concluye en este trabajo que se puede definir la lealtad de los clientes de las entidades financieras atendiendo a la satisfacción y la confianza, además la tenencia de tarjetas débito y/o crédito también parecer ser factores determinantes.

Un trabajo más que se desarrolló en el ámbito local (Ruiz, 2015), consistió en la realización de un modelo de predicción de fuga de clientes de una entidad bancaria colombiana. En este

estudio se tuvieron en cuenta variables no solo las alusivas al cliente sino también se tomaron variables económicas y financieras, como la tasa de inflación o los indicadores económicos de la empresa, para que a partir de estas se estime mediante regresión logística la probabilidad de deserción en periodos futuros. Aquí, se definió como desertor aquella persona que dejaba de realizar operaciones en el periodo que correspondía al que estaba siendo analizado.

Se realizó el cálculo para cada uno de los segmentos con los que cuenta el banco y se validó realizando predicciones para un determinado periodo de tiempo obteniendo una precisión general del 64%, sin embargo, el autor indica que el propósito del proyecto era únicamente la ejecución de un modelo teórico, por lo que se utilizaron pocas variables, aunque advierte que es importante considerar variables de servicio, mercados y los clientes.

## 1. Objetivos

### 1.1 Objetivo General

Diseñar un modelo que permita calcular el riesgo de que un asociado deserte de la Cooperativa dados sus valores en un conjunto de variables sociodemográficas y transaccionales usando análisis de supervivencia.

### 1.2 Objetivos Específicos

- Identificar mediante análisis estadísticos las variables que cumplen un papel importante a la hora de diseñar un modelo de deserción de asociados de la Cooperativa.
- Desarrollar un modelo utilizando análisis de supervivencia que permita predecir los clientes que se darán de baja en los próximos meses.
- Utilizar el modelo desarrollado mediante análisis de supervivencia como indicador de alerta para el área de mercadeo con el fin de fortalecer la relación comercial con los asociados antes de que la fuga se haga efectiva.

## 2. Cuerpo del Trabajo

### 2.1 Marco Referencial

Como lo indica el título de este proyecto, el trabajo se realizará usando la técnica de análisis de supervivencia, la cual consiste en una serie de procedimientos estadísticos que permiten estimar el riesgo de que se presente un evento, así como cuales variables influyen para que el mismo suceda (Kleinbaum & Klein, 2012). El análisis de supervivencia fue inicialmente utilizado en el campo de la medicina, en donde se pretendía estimar la probabilidad del tiempo de supervivencia de pacientes sometidos a cierto tratamiento. Sin embargo, sus aplicaciones también se dan para los tiempos de falla de algún producto, el tiempo de deserción de estudiantes en la educación superior, de clientes en una entidad bancaria, etc.

Los datos del análisis de supervivencia se pueden estudiar por medio de dos tipos de probabilidades distintas, que son las siguientes:

- **Función de supervivencia:** Se denota  $S(t)$  y es la probabilidad de que el individuo sobreviva desde la entrada al estudio hasta un tiempo  $t$  determinado para el análisis; en otras palabras,  $S(t)$  da la probabilidad de que la variable exceda el tiempo definido  $t$ . Se convierte en una función clave ya que al obtener la probabilidad de supervivencia en diferentes tiempos  $t$  se puede obtener información crucial en diferentes puntos de tiempo.

La función de supervivencia tiene la forma:

$$S(t) = 1 - F(t) = P [T > t]$$

$$S(t) = \int_t^{\infty} f(t) dt$$

Donde  $F(t)$  es la función de probabilidad acumulada

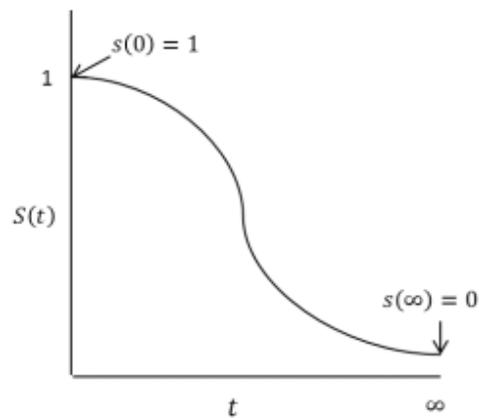
Sus propiedades son:

- $0 \leq S(t) \leq 1$
- Es una función decreciente
- $S(0) = 1 \quad \lim_{t \rightarrow \infty} S(t) = 0$

Y su grafica teóricamente es de la siguiente manera:

**Figura 1.**

*Función de Supervivencia*



Nota. Tomado de Kleinbaum & Klein (2012) pág. 10

- **Función de riesgo:** se define como la tasa instantánea de fallo para que a un individuo que ha sobrevivido hasta un tiempo t le suceda el evento. Se denota h(t) y es:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t < T \leq t + \Delta t | T \geq t)}{\Delta t}$$

Al no ser una probabilidad, sino una tasa instantánea de falla, los valores que puede llegar a tomar se encuentran entre cero e infinito, dependiendo de la escala de tiempo que se utilice (días, meses, años, etc.).

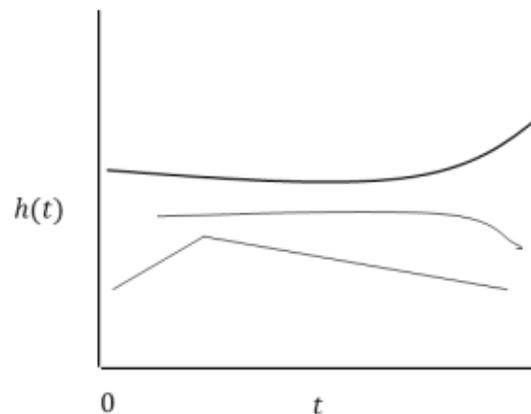
Sus propiedades son:

- No es necesariamente decreciente
- No posee límite superior
- Es una función no negativa

La función de riesgo tiene más opciones tal como se muestra en la Figura 2:

**Figura 2.**

*Función de riesgo*



Nota. Tomado de Kleinbaum & Klein (2012) pág. 13

Por lo tanto, la evidente diferencia con la función de supervivencia es que esta se centra en la no ocurrencia del evento mientras que la función de riesgo se basa en que el evento si suceda.

A continuación, se aprecia la relación entre la función de supervivencia y la función de riesgo, ya que es factible escribir la fórmula de cada una de ellas en términos de una integral utilizando la otra, es decir, la fórmula de  $S(t)$  en función de  $h(t)$  utilizando integrales y viceversa:

$$S(t) = \exp \left[ - \int_0^t h(u) du \right]$$

$$h(t) = - \left[ \frac{dS(t)/dt}{S(t)} \right]$$

A la hora de elegir el modelo adecuado para nuestro conjunto de datos y que los resultados obtenidos sean los que se esperan, se cuenta con 3 opciones diferentes: modelos paramétricos, modelo no paramétricos y modelos semiparamétricos.

- Modelos paramétricos: Este tipo de modelos se utiliza cuando se tiene certeza acerca de la selección de la distribución de los datos. Los métodos de este tipo que habitualmente se manejan son: Weibull, exponencial, Log-logístico y otros.
- Modelos no paramétricos: Este tipo de modelos se utiliza cuando no hay certeza acerca de la selección de la distribución de los datos. El método comúnmente más utilizado es el de Kaplan-Meier (KM). El modelo de Kaplan-Meier permite calcular la probabilidad de que un individuo sobreviva a un evento de interés en un determinado tiempo  $t$ , se denota de la siguiente forma:

$$S(t_j) = S(t_{j-1}) \left( 1 - \frac{d_j}{n_j} \right)$$

Donde:

$d_j$ : Número de eventos ocurrido en  $t_j$

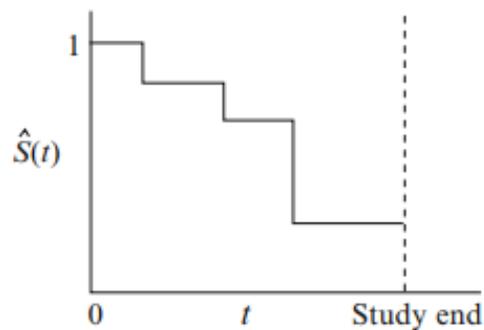
$n_j$ : Número de individuos que no han presentado el evento antes de  $t_j$

$S(t)$  es constante entre los tiempos de los eventos, es por eso que la probabilidad varía solo cuando se presenta el evento dando así una apariencia de escalera a la Figura tal y como se aprecia a continuación:

**Figura 3.**

*Curva de sobrevivencia de Kaplan-Meier.*

$\hat{S}(t)$  in practice:



Nota. Tomado de Kleinbaum & Klein (2012) pág. 10

- Modelos semiparamétricos: Un modelo es semiparamétrico cuando contiene una parte especificada y definida, pero otra no. El principal modelo trabajado aquí es el de Riesgo proporcional de Cox. El modelo de riesgo proporcional de Cox consiste en una regresión por medio de la cual pretende explicarse el riesgo de la ocurrencia de un evento con una serie de covariables. Se define de la siguiente manera:

$$h(t, \mathbf{X}) = h_0(t) \exp \left( \sum_{i=1}^p \beta_i X_i \right)$$

Donde:

- $h_0(t)$  hace referencia a la función de riesgo base e implica el riesgo si todas las variables  $X$  son cero. Este término varía con el tiempo.

- $X_i$  es la covariable  $i$  con  $i = 1, \dots, p$ .
- $\beta_i$  es el parámetro asociado a la variable  $X_i$

Por lo tanto, se puede afirmar que la función de riesgo se basa en una serie de covariables que son o no relevantes dependiendo de sus respectivos coeficientes. Como se puede observar, la variable tiempo no se encuentra en el exponencial, esto indica que se asumen constantes en el tiempo de observación. Estos exponentes, que son los coeficientes del riesgo, pueden tomar valores mayores o menores de uno, un valor superior a 1 indica que una covariable se asocia positivamente con el riesgo de ocurrencia del evento, es decir, aumenta el riesgo y disminuye la supervivencia; un valor inferior a uno implica lo contrario.

Un aspecto a tener en cuenta al ajustar el modelo de Cox es el supuesto de riesgos proporcionales (HR) que no es otra cosa que el cociente del riesgo de dos individuos.

$$\widehat{HR} = \frac{\widehat{h}(t, X^*)}{\widehat{h}(t, X)}$$

Es decir:

$$\widehat{HR} = \exp \left[ \sum \beta_1 (X_i^* - X_i) \right]$$

En esta expresión se observa que la razón de riesgo entre dos individuos solo depende de los valores de las covariables y es constante en el tiempo.

Equivalentemente a  $h(t)$ , la función de supervivencia de Cox es:

$$S(t) = S_0(t) e^{\sum_{i=1}^p \beta_i X_i}$$

Donde  $S_0(t)$  es la función de supervivencia común.

La fórmula del modelo de Cox tiene la propiedad de que, si todas las  $X$  son iguales a cero, la fórmula se reduce a la función de riesgo de referencia. Esta propiedad del modelo de Cox es la razón por la cual  $h(t)$  se denomina función de supervivencia común o de base.

### **2.1.1 Método.**

Teniendo en cuenta las particularidades de este estudio en el que el evento de interés es el retiro de los asociados de la Cooperativa se registra diariamente y se calcula mensualmente al interior de la organización, al contar con periodos de tiempos continuos se utilizara el modelo de riesgo proporcional de Cox para el desarrollo de este.

La población de estudio son los asociados a corte de diciembre de 2014 de la oficina principal de la Cooperativa los cuales conforman 40.374 personas. Sin embargo, con el fin de segmentar aún más la información se excluyen de esta población inicial las personas jurídicas y los que tienen fecha de actualización de sus datos superior a 5 años, es decir, fechas de actualización de la información inferiores al año 2010 ya que serían poco fiables los datos; además sólo se tienen en cuenta los asociados entre 18 y 65 años de edad ya que por política de la entidad son aquellos a los que se le puede ofrecer el producto de crédito, lo cual los convierte en las personas más rentables para la Cooperativa.

La recolección de información para este modelo se realizó extrayendo la información de las bases de datos propias de la Cooperativa a las cuales el autor del proyecto tiene acceso, aquí se

incluyeron variables sociodemográficas y transaccionales, además de las propias de los análisis de supervivencia, como se aprecia en la siguiente tabla:

**Tabla 1.**

*Variables para utilizar en el proyecto*

FACTOR	VARIABLE	DESCRIPCIÓN
MODELO	ID	Identificación del individuo
	TIEMPO	Tiempo de supervivencia
	CENSURA	Presencia o no del evento de interés
SOCIODEMOGRÁFICAS	GENERO	Genero del individuo
	EDAD	Edad del individuo
	ESTADO_CIVIL	Estado civil del individuo
	TIPO_VIVIENDA	Tipo de vivienda del individuo
	OCUPACION	Ocupación del individuo
	ESTRATO	Estrato del individuo
	NIVEL_EDUCATIVO	Nivel educativo del individuo
	ANTIGÜEDAD	Antigüedad como asociado del individuo
	TOTAL_INGRESOS	Total en pesos de los ingresos del individuo
	TOTAL_EGRESOS	Total en pesos de los egresos del individuo
	CIUDAD	Ciudad de residencia del individuo
	DEPARTAMENTO	Departamento de residencia del individuo
	PERSONAS_A_CARGO	Personas a cargo del individuo
TRANSACCIONALES	CREDITO	Tenencia o no de crédito a corte
	AHORROS	Tenencia o no de ahorros a corte
	PAP	Tenencia o no de PAP a corte
	CDAT	Tenencia o no de CDAT a corte
	MORA_CRED	Tenencia de mora de crédito a corte
	CRED_PREPAGO_AÑO	Tenencia de créditos prepagados en el último año
	CRED_REESTRUCTURA_AÑO	Tenencia de créditos reestructurados en el último año
	CRED_NORMALIZADO_AÑO	Tenencia de créditos normalizados en el último año
	CRED_NEGADO_AÑO	Tenencia de créditos negados en el último año
	INACT_AHORRO	Indicador de inactividad de la cuenta de ahorros a corte
	MORA_PAP	Tenencia de mora de PAP a corte
	NUM_PRODUCTOS	Número de productos del portafolio a corte
	PROM_AHORRO_AÑO	Promedio de saldos de ahorro en pesos del último año
	PROM_CREDITO_AÑO	Promedio de saldos de crédito en pesos del último año
	PROM_PAP_AÑO	Promedio de saldos de PAP en pesos del último año
	PROM_CDAT_AÑO	Promedio de saldos de CDAT en pesos del último año
NUM_TRANS_AÑO	Número de transacciones del último año	

Los criterios que se establecieron para el análisis de supervivencia son los siguientes:

- Evento de interés: Para este proyecto se definió el evento de interés como la deserción voluntaria de los asociados personas naturales entre 18 y 65 años y con fecha actualización de sus datos no mayor a 5 años, de la oficina principal de la cooperativa. Debido a que el objetivo es retiro voluntario, además de los filtros previamente mencionados, también se eliminan del análisis aquellas personas que la Cooperativa en uso de sus facultades decidió excluir por razones de incumplimiento o actividades ilegales. Por lo tanto, el número de individuos finales para llevar a cabo el proyecto es de 16.218 personas. No obstante, podría aplicarse alguna otra segmentación adicional que pudiera permitir análisis comparativos entre asociados con ciertas características.

- Tiempo de seguimiento: El tiempo de seguimiento establecido fue de 60 meses, que corresponden al periodo de enero 2015 hasta diciembre de 2019.

- Herramientas: Para llevar a cabo el modelo de análisis de supervivencia utilizaremos el software SPSS Statistics ya que se dispone de su licenciamiento.

Las fases para llevar a cabo el desarrollo exitoso del proyecto son las siguientes:

- La exploración inicial de datos en donde se recopila la información con la que se cuenta, se identifican problemas de calidad de los datos, y se realiza un análisis descriptivo de los datos con la ayuda de tablas y gráficos con el ánimo de identificar en primera instancia las variables que son significativas para el modelo.

- La fase de preparación de datos en la que se transforman todos los datos con los que se cuentan inicialmente en el conjunto de datos final a partir del cual se dé respuesta a la problemática. Aquí se realizan actividades como la limpieza de los datos, transformación de variables y creación de nuevas variables según se requiera de acuerdo con lo observado en la exploración inicial.

- Una vez está cubierta la exploración y preparación de los datos se procede a realizar el modelado. El modelado puede ejecutarse múltiples veces a medida que se descubre información, se ajustan datos o se vuelve a la fase de preparación de datos para llevar a cabo las transformaciones necesarias para garantizar la efectividad del modelo.

Finalmente se ejecuta la evaluación del modelo desarrollado, verificando si se cumplen los objetivos planteados. De igual forma, se revisa el proceso, en función de los resultados obtenidos para identificar si en el transcurso de este se cometieron errores.

### **2.1.2 Resultados.**

A continuación se presenta el análisis descriptivo de los datos que se manejaron en el proyecto, así como la aplicación de la técnica estadística y los resultados de esta aplicación.

**2.1.2.1 Análisis descriptivo.** La población para este proyecto son los 40.374 asociados a corte de diciembre de 2014 de la oficina principal de la Cooperativa. Sin embargo, con el fin de segmentar aún más la información se incluyeron en esta población inicial sólo las personas naturales y los que tienen fecha de actualización de sus datos durante el 2014 ya que así se garantiza la fiabilidad y completitud de los datos; además sólo se tienen en cuenta los asociados entre 18 y 65 años de edad ya que por política de la entidad son aquellos a los que se le puede ofrecer el producto de crédito. De igual forma, se eliminaron del análisis aquellas personas que la Cooperativa en uso de sus facultades decidió excluir por razones de incumplimiento o actividades ilegales. Adicionalmente se eliminaron aquellas personas que en el transcurso del periodo de observación se retiraron y volvieron a asociarse. En un principio se consideraron variables

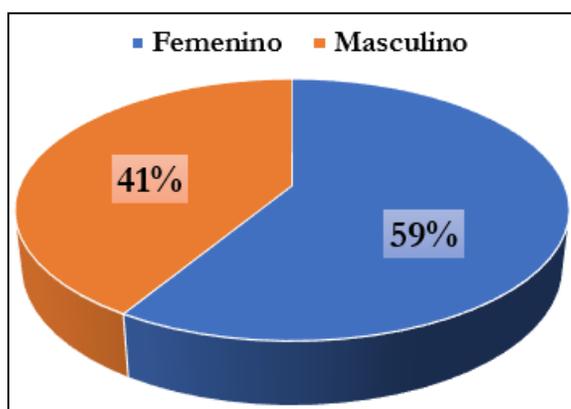
relacionadas con el servicio, como las PQRs radicadas por las personas, y variables de comportamiento de crédito de los asociados, sin embargo, debido a la poca cantidad de personas que cumplían con estas características fueron excluidas aquellas que no se encuentran en el perfil mayoritario. Es así como después de todo este proceso de filtración y limpieza de los datos, el número de individuos que serán considerados en este análisis es de 6.029 personas.

A continuación, se realiza el análisis descriptivo invariado de las variables descritas en la metodología que se encuentran agrupadas en sociodemográficas y transaccionales:

**2.1.2.1.1 Variables sociodemográficas.** Los datos del estudio muestran que la mayoría de las personas son de sexo femenino, las cuales representan el 59% de las observaciones tal y como se aprecia en la Figura 4.

**Figura 4.**

*Distribución por género*

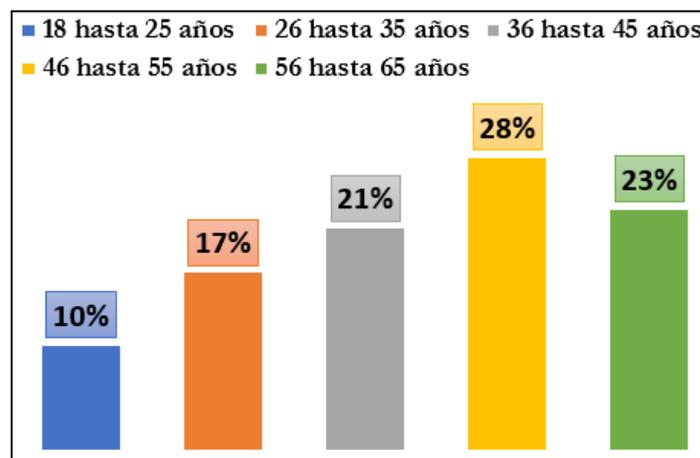


En cuanto a la edad, se decidió agrupar las edades por los rangos que maneja la Cooperativa en la realización de sus diferentes informes. Excepto los jóvenes que están en franca minoría, los

demás grupos etarios participan al menos con aproximadamente la quinta parte de la muestra, siendo la de mayor participación los asociados entre 46 y 55 años. Se debe recordar que uno de los filtros fue analizar personas únicamente entre los 18 y 65 años. Para efectos de la ejecución del modelo se unirán los dos primeros rangos que se muestran en la Figura 5 debido a su baja participación en el total individualmente.

**Figura 5.**

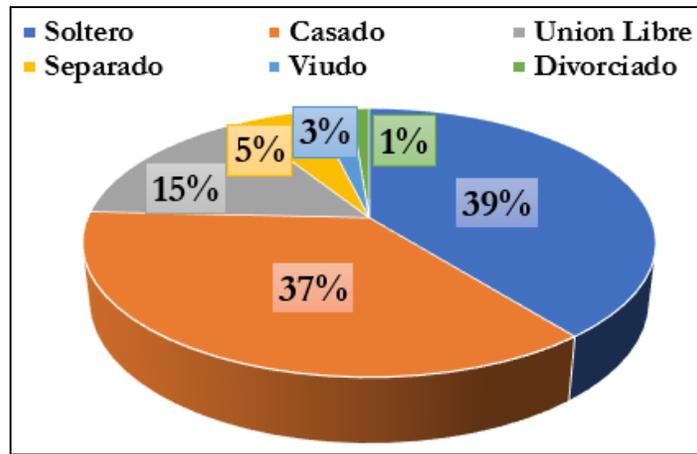
*Distribución por rangos de edad*



Las personas solteras representan casi el 40% de la población en tanto que los casados responden por el 37%; si bien la unión libre tiene alguna representatividad importante (15%) no es el caso de las demás categorías del estado civil de los asociados en consideración. Para efectos de la ejecución del modelo se unirán las categorías de “Unión Libre”, “Separado”, “Viudo” y “Divorciado” debido a su baja participación en el total individualmente. Ver la Figura 6.

**Figura 6.**

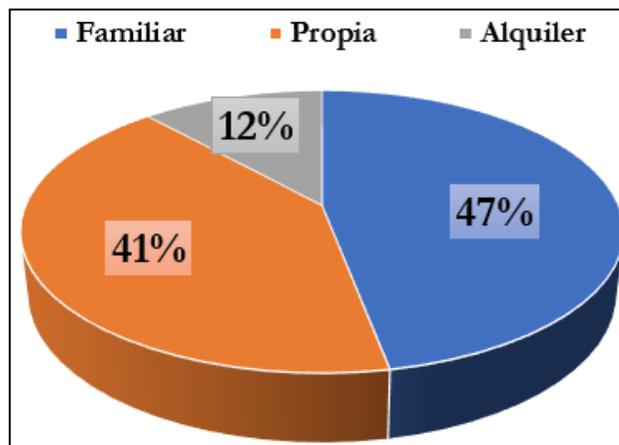
*Distribución por estado civil*



Casi la mitad de las personas habita en una vivienda familiar, mientras que los asociados con vivienda propia representan más de la tercera parte de la muestra (41%) y el menor porcentaje, equivalente al 12%, son las personas que viven en arriendo.

**Figura 7.**

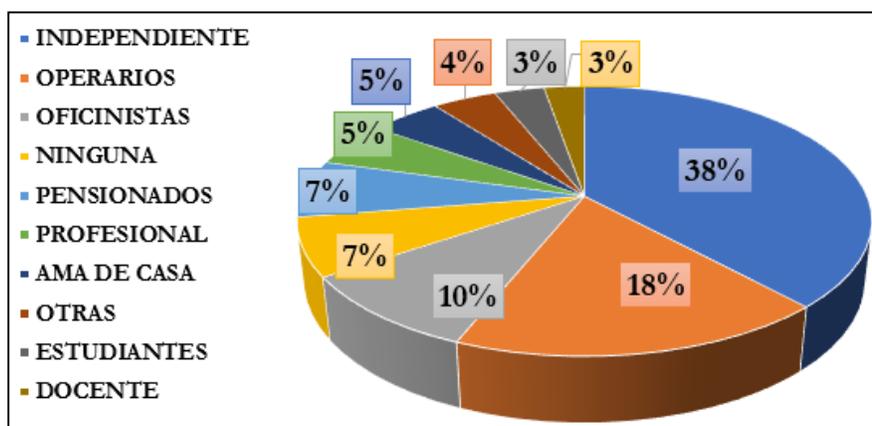
*Distribución por tipo de vivienda*



Las personas cuya actividad económica es independiente representan la gran mayoría de la muestra: 38%; con un porcentaje representativo del 18% siguen los denominados operarios (empleados con labores operativas); las demás participaciones de las ocupaciones de los asociados de la muestra no superan el 10%, tal como se aprecia en la Figura 8. Las categorías que no llegan al 10% para efectos de la ejecución del modelo se unirán en una sola debido a su baja participación en el total individualmente.

**Figura 8.**

*Distribución por ocupación*

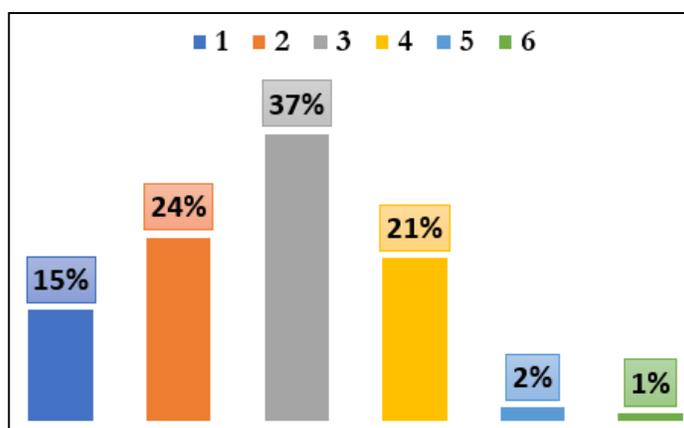


La mayor participación en lo referente al estrato socioeconómico de los asociados es el estrato 3 con un 37% de la población, le sigue el estrato 2 con el 24% y el estrato 4 con el 21%. Esta distribución permite afirmar que la oficina principal de la Cooperativa presta sus servicios financieros básicamente a la clase media y a la clase baja alta de la población que se sabe no tienen acceso fácil a las entidades bancarias tradicionales; la participación de los demás estratos es baja ya que no superan el 15%. Inicialmente al explorar esta variable se apreció un error en el almacenamiento de la información ya que según el DANE el estrato 0 no existe, las nomenclaturas

aceptadas van de la 1 a la 6, por lo tanto, se tomó la decisión de asignarles a estos casos el estrato 1 ya que socioeconómicamente son similares. Para efectos de la ejecución del modelo se unirán los estratos 1 y 2 en una sola categoría y los estratos 4, 5 y 6 conformarán una sola categoría debido a la baja participación de los estratos 1, 5 y 6 en el total individualmente.

**Figura 9.**

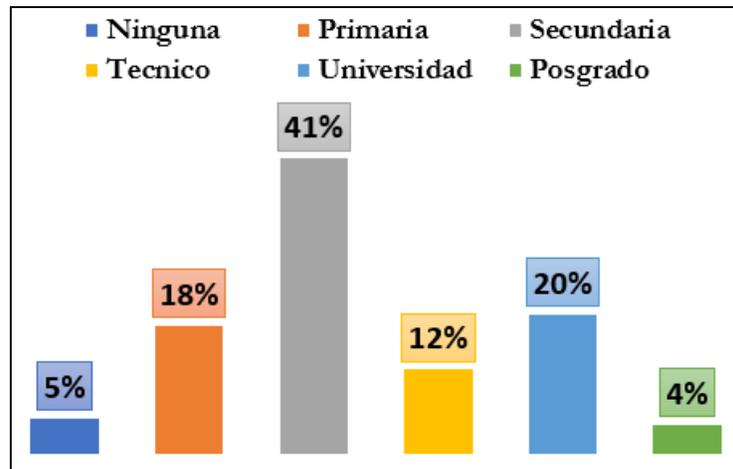
*Distribución por estrato*



En cuanto al nivel educativo de los asociados de la muestra, se observa en la Figura 10 que las personas con nivel educativo de secundaria representan el 40% de la población, con un poco menos de la mitad de este porcentaje se encuentran los niveles educativos de universidad y primaria. Estos tres niveles educativos, constituyen el 79% del total de los asociados de la muestra. Las otras categorías consideradas no son porcentualmente significativas. Para efectos de la ejecución del modelo se unirán las categorías “Ninguna” y “Primaria” en una sola y las categorías de personas con nivel educativo superior a la básica secundaria también conformarán una única categoría debido a su baja participación en el total individualmente.

**Figura 10.**

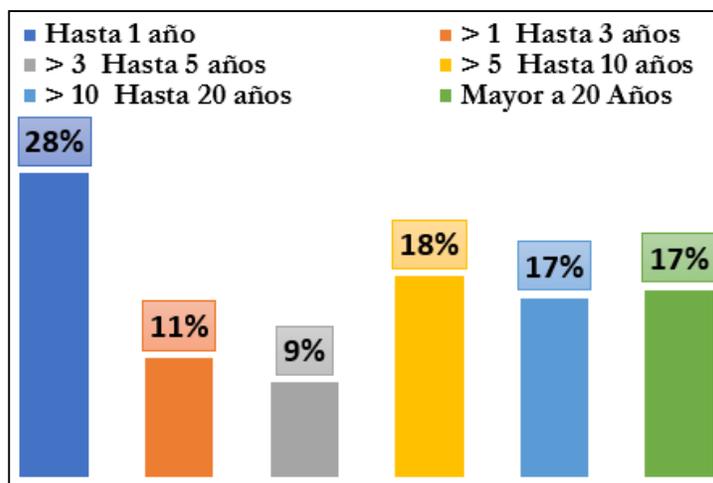
*Distribución por nivel educativo*



Para la variable de antigüedad como asociado, se realizó un procedimiento similar al realizado con la variable edad, se segmentó de acuerdo con los rangos utilizados por la Cooperativa para sus diferentes informes: la muestra de asociados prácticamente está dividida en dos grupos iguales cuyo punto de división son 5 años de antigüedad: la mitad con más de 5 años de antigüedad y la otra mitad con menos de 5 años de antigüedad, tal como se desprende de la Figura 11. Para efectos de la ejecución del modelo se unirán los rangos de antigüedad de entre 1 y 10 años en una sola categoría y lo propio se realizará con los rangos que contienen las personas con más de 10 años de antigüedad en la Cooperativa debido a su baja participación en el total individualmente.

**Figura 11.**

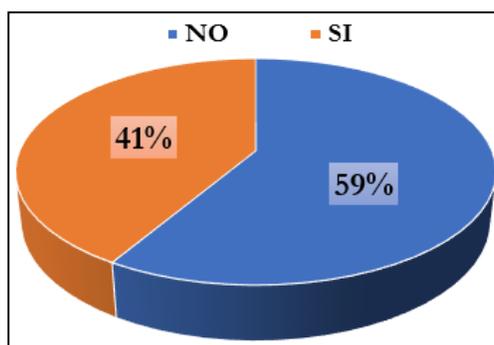
*Distribución por rango de antigüedad*



Según la información de la base de datos, los asociados con personas a cargo constituyen el 42% de la muestra, en tanto que el 58% restante no reportan ser acudientes de otra persona. Ver la Figura 12.

**Figura 12.**

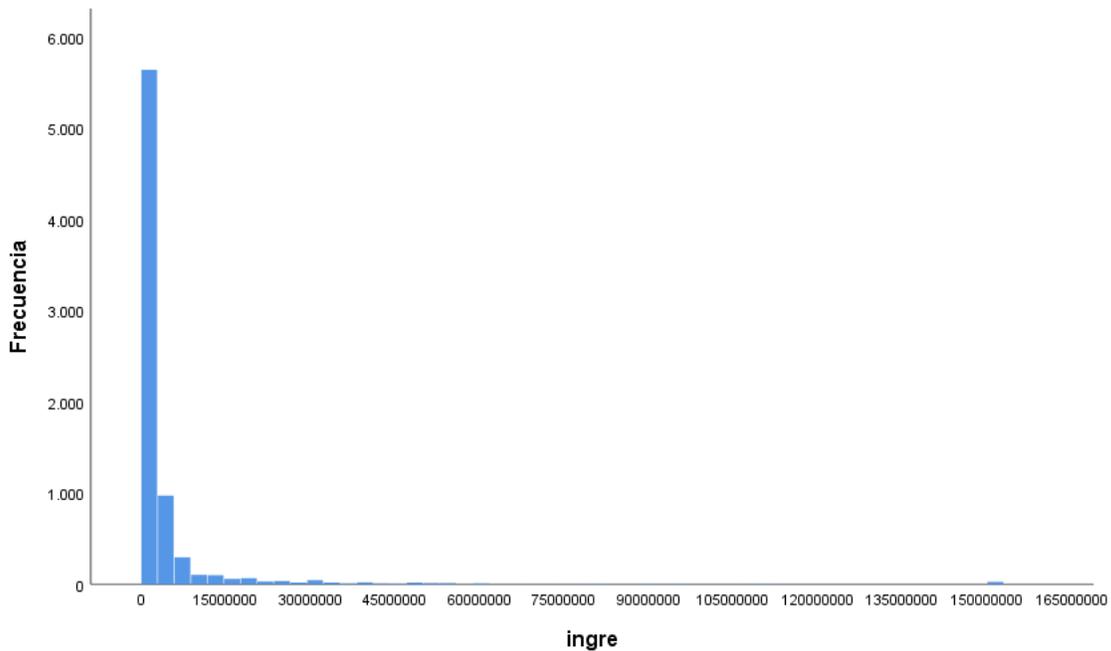
*Distribución por tenencia de personas a cargo*



Los campos de ingresos y egresos de los asociados presentan datos muy dispersos, por lo anterior y en procura de que los datos extremos no afecten la modelación se decidió reemplazar los valores atípicos y dejar como máximo el monto apreciado en la tercera desviación estándar tanto de los ingresos como de los egresos; en cuanto a los egresos, su valor inferior es negativo, lo cual denota un error en los datos que se produce por confusión de las personas que registran los datos en el sistema al asumir que por tratarse de egresos, se debe adicionar el signo menos(-) al valor, esto se corrige multiplicando el valor por menos uno(-1). Para estas 2 variables, se realiza segmentación de los datos con rangos basados en el SMMLV del 2014 en Colombia que es el año de corte de la información. El salario mínimo en ese año era de \$616.000.

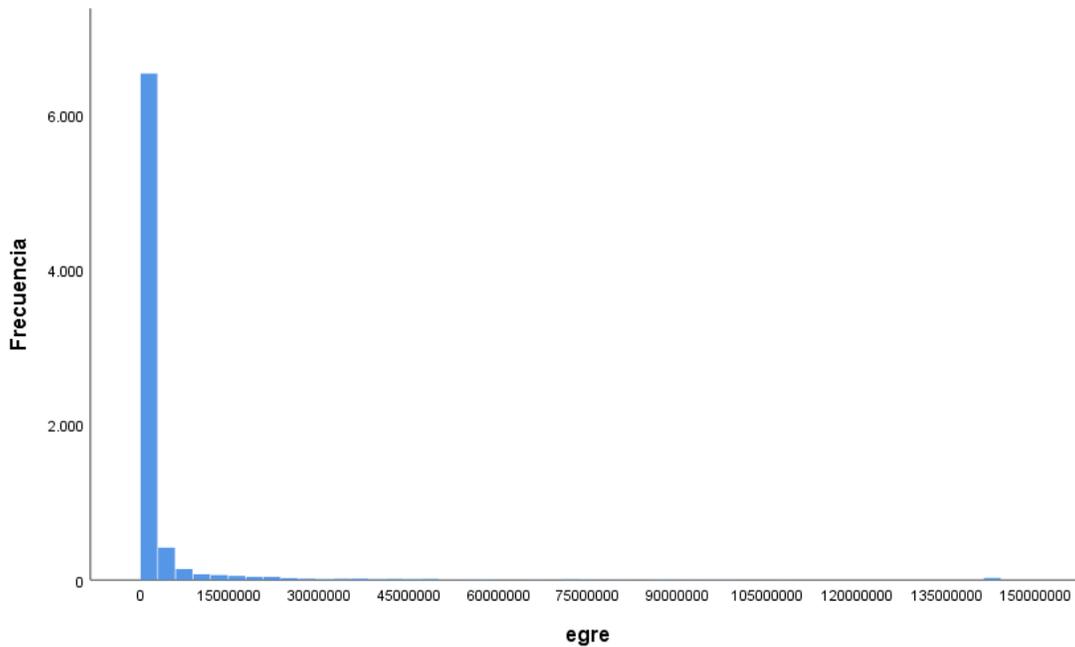
**Figura 13.**

*Histograma de ingresos*



**Figura 14.**

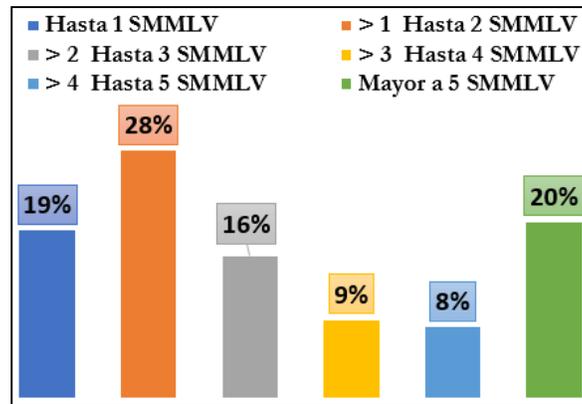
*Histograma de egresos*



En el caso de los ingresos, el 47% de las personas tienen ingresos hasta los 2 SMMLV, lo cual coincide con el hecho de que la Cooperativa presta sus servicios principalmente en las clases baja y media de la población, los asociados con ingresos entre los 2 y los 3 SMMLV representan el 16% de la población que se aproxima a la cantidad de personas con estrato socioeconómico 4; si bien la representatividad de las personas con ingresos superiores a los 5 SMMLV es relevante con un 20%, similar porcentaje de las personas con nivel educativo universitario, no es el caso de las demás categorías de ingresos de los asociados en mención. Para efectos de la ejecución del modelo se unirán los rangos que contienen los valores entre 2 y 5 SMMLV debido a su baja participación en el total individualmente.

**Figura 15.**

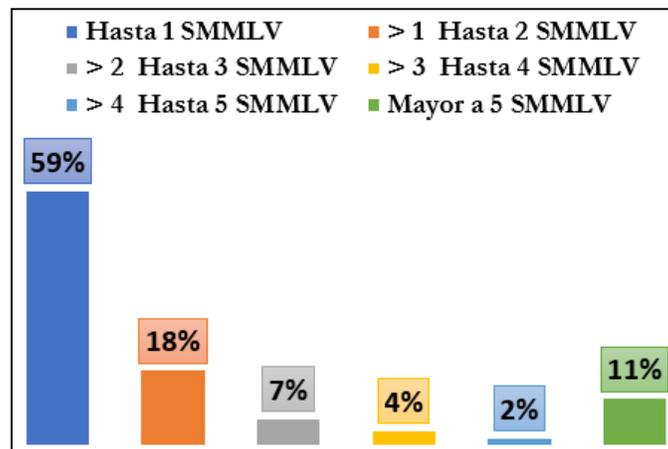
*Distribución por rango ingresos*



Para los egresos, se aprecia que las personas que reportan que sus egresos no superan 1 SMMLV representan el 59% de la población, el 18% reportan egresos entre 1 y 2 SMMLV y las personas con egresos superiores a los 5 SMMLV participan con un 11%, las demás categorías no resultan significativas. Para efectos de la ejecución del modelo se unirán los rangos que contienen los valores superiores a 2 SMMLV debido a su baja participación en el total individualmente.

**Figura 16.**

*Distribución por rango egresos*



Al analizar los datos de ingresos y egresos se observa que a medida que aumentan los ingresos, también lo hacen los egresos, esta situación se confirma en la Tabla 2 donde se aprecia un coeficiente de correlación de Pearson cercano a 1.

**Tabla 2.**

*Correlación entre las variables ingresos y egresos*

Correlaciones			
		TOTAL_INGR ESOS	TOTAL_EGR ESOS
TOTAL_INGRESOS	Correlación de Pearson	1	,904**
	Sig. (unilateral)		,000
	N	6029	6029
TOTAL_EGRESOS	Correlación de Pearson	,904**	1
	Sig. (unilateral)	,000	
	N	6029	6029

\*\* La correlación es significativa en el nivel 0,01 (unilateral).

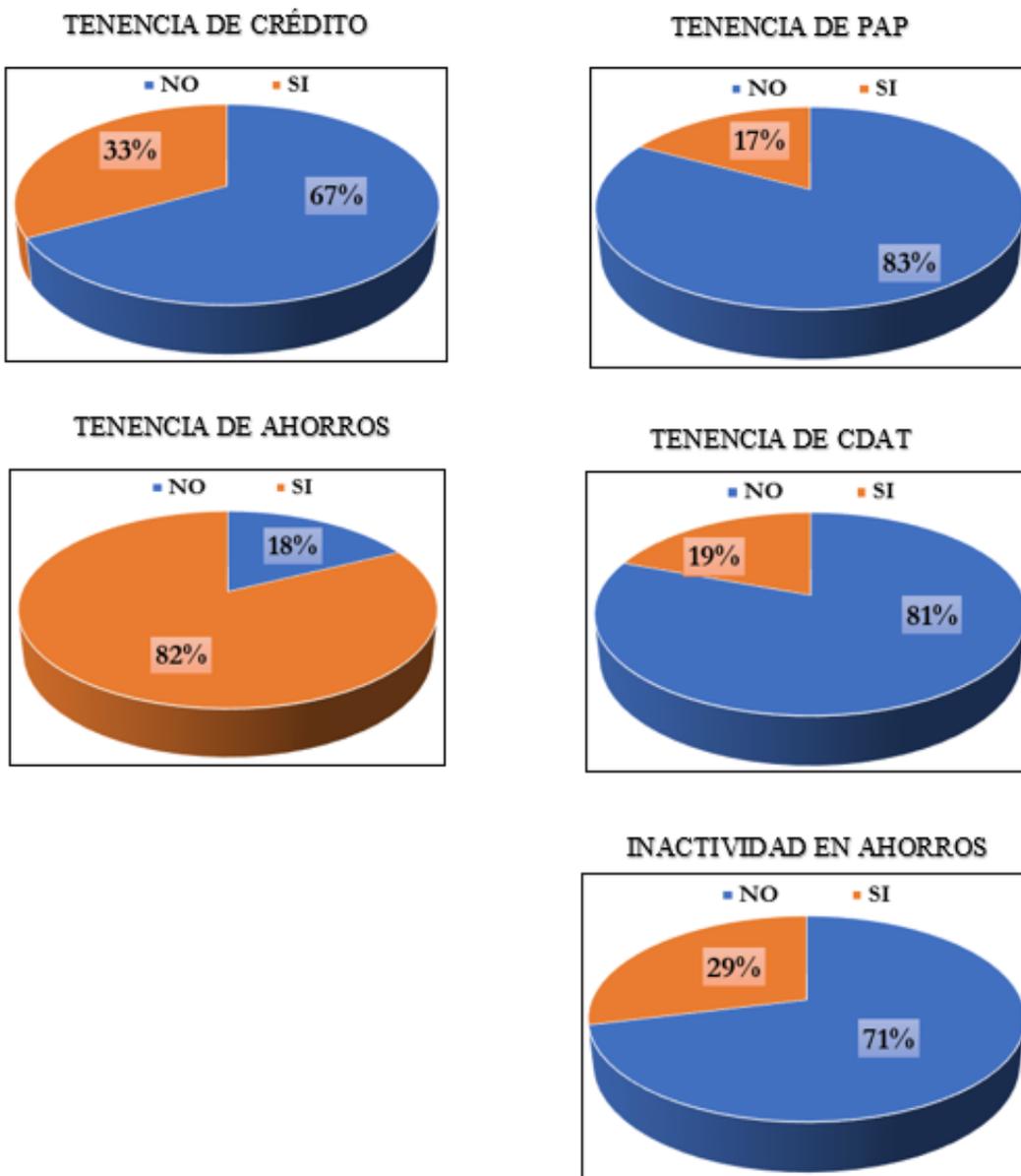
El perfil sociodemográfico mayoritario son personas en mayor porcentaje de mujeres, entre los 46 y 55 años, mayoritariamente asociados solteros, con vivienda familiar y actividad económica independiente, prevalecen la clase media y con estudios básicos; más de 5 años de antigüedad con la Cooperativa. Casi 1 de cada 2 personas ganan menos de 2 SMMLV y aproximadamente el 60% de los asociados reportan egresos inferiores a 1 SMMLV.

**2.1.2.1.2 Variables transaccionales.** Según la información arrojada por la muestra poco más de 2/3 partes de los asociados no poseían crédito al momento de iniciar el seguimiento (diciembre de 2014), además el 82 % de las personas tenían al menos una cuenta de ahorros, aproximadamente 8 de cada 10 personas no contaban con un CDAT y solo el 17% sostenían un PAP. Sin embargo, es importante destacar, para el caso de las cuentas de ahorro, que casi el 30%

de las personas no habían realizado movimiento en sus cuentas de ahorros en los últimos 6 meses, por lo que estas permanecían inactivas.

**Figura 17.**

*Distribución por tenencia de productos*



Las variables continuas referentes a la transaccionalidad de los asociados muestran rangos muy amplios principalmente en las que contienen saldos en pesos. Por ello, al igual que en variables de ingresos y egresos, se reemplazaron los valores atípicos por el observado en la tercera desviación estándar de cada una de estas variables. De igual forma, se procede a realizar la segmentación de estas variables, en el caso de las relacionada con la cantidad de productos se tomaron los rangos de la segmentación comercial de la Cooperativa, en cuanto a la variable de cantidad de transacciones se realizaron rangos en función de percentiles de la opción 2, y para las variables de saldos, se mantienen los rangos en SMMLV, ya que se analizó la posibilidad de crear rangos por percentiles, pero debido a la gran cantidad de personas sin el producto y por ende sin saldo, los rangos que arrojaba esa distribución no se consideran apropiados para el estudio. Ver tabla 3.

**Tabla 3.**

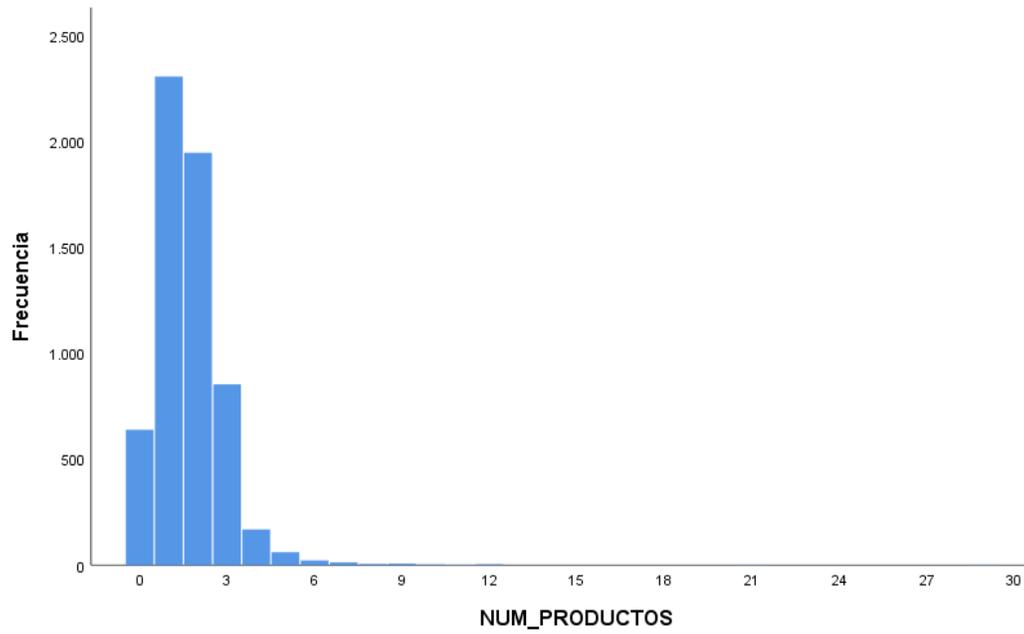
*Rangos por percentiles*

OPCION 1					
	AHORROS	CRÉDITO	PAP	CDAT	TRANSACCIONES
Percentil 25	139	-	-	-	3
Percentil 50	14.407	-	-	-	14
Percentil 75	500.324	2.330.231	-	6.133.333	37

OPCION 2					
	AHORROS	CRÉDITO	PAP	CDAT	TRANSACCIONES
Percentil 20	-	-	-	-	2
Percentil 40	2.499	-	-	-	9
Percentil 60	81.867	-	-	-	22
Percentil 80	824.968	3.955.467	13.565	14.325.129	43

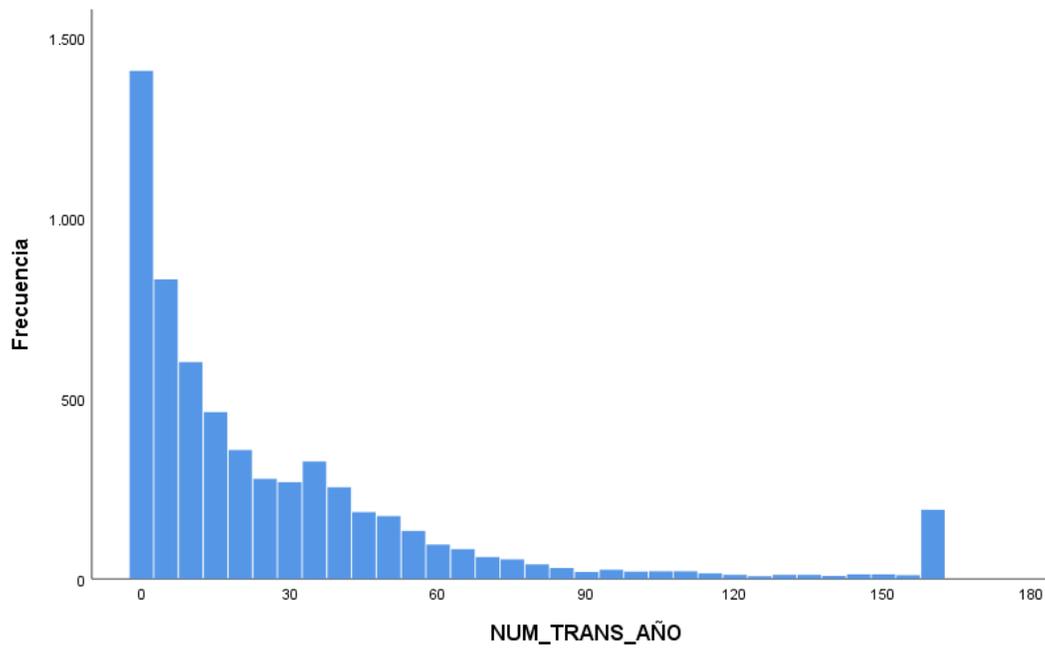
**Figura 18.**

*Histograma de número de productos a corte*



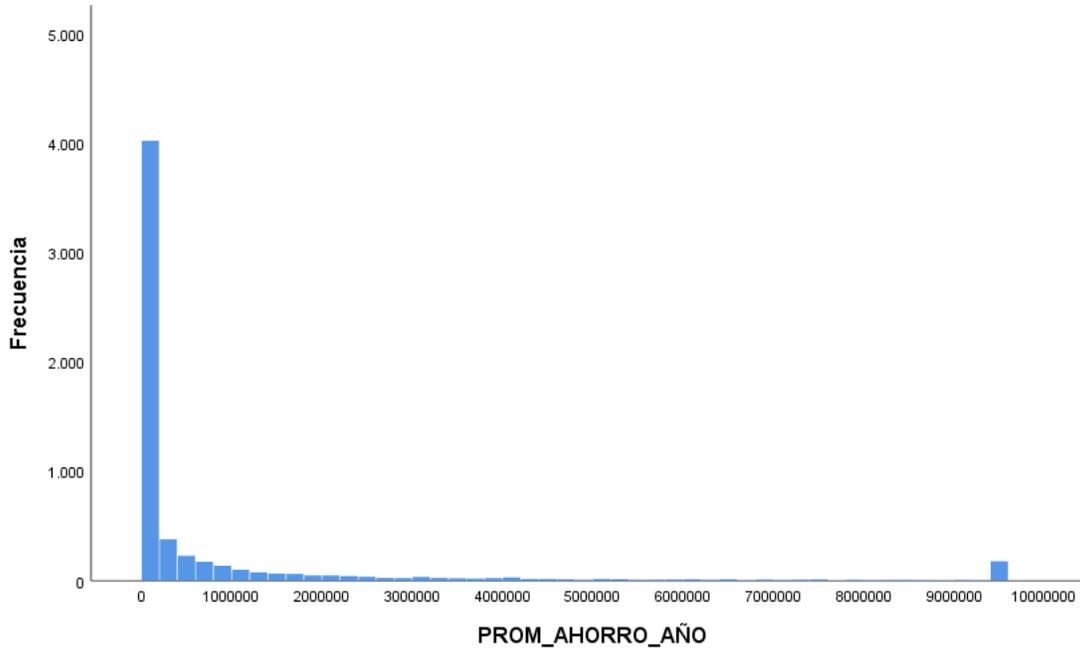
**Figura 19.**

*Histograma de número de transacciones en 2014*



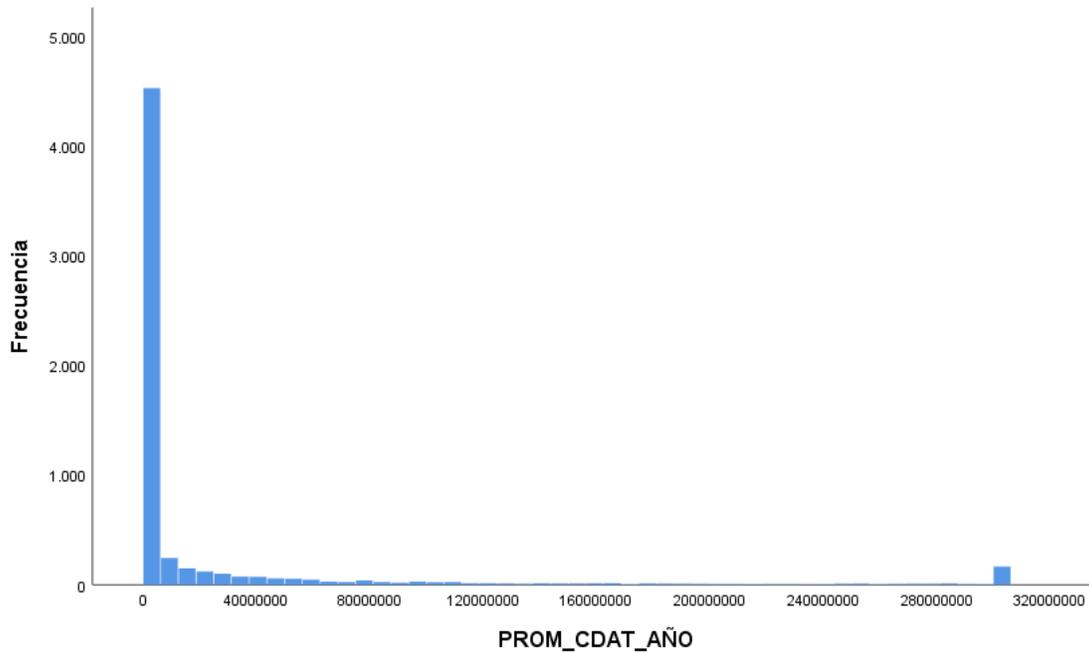
**Figura 20.**

*Histograma de promedio de saldos de ahorros en 2014*



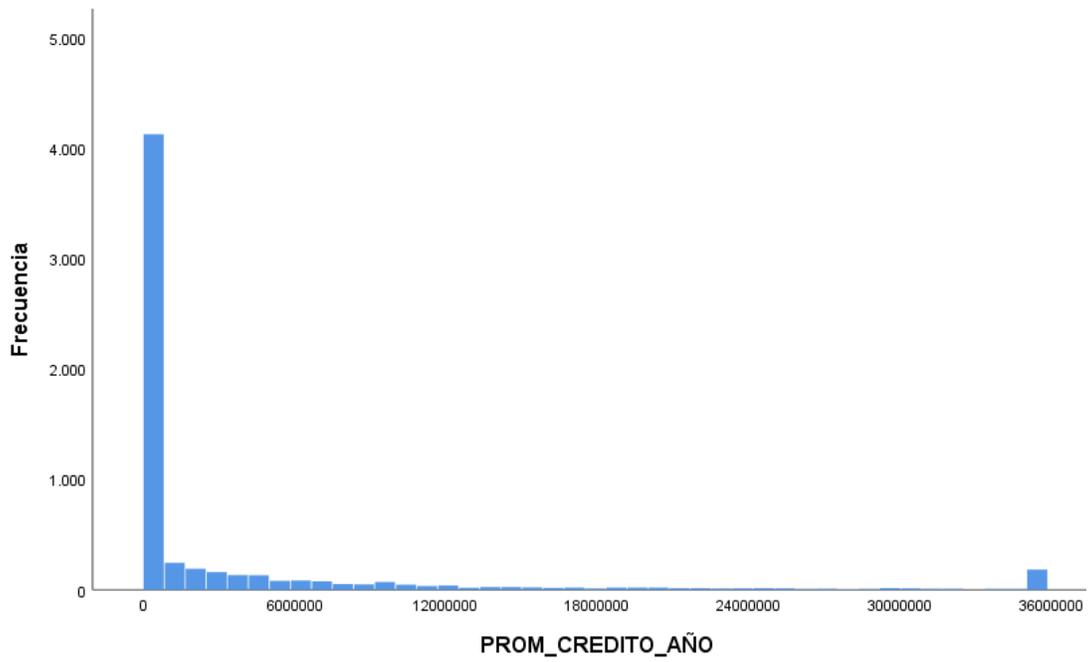
**Figura 21.**

*Histograma de promedio de saldos de CDAT en 2014*



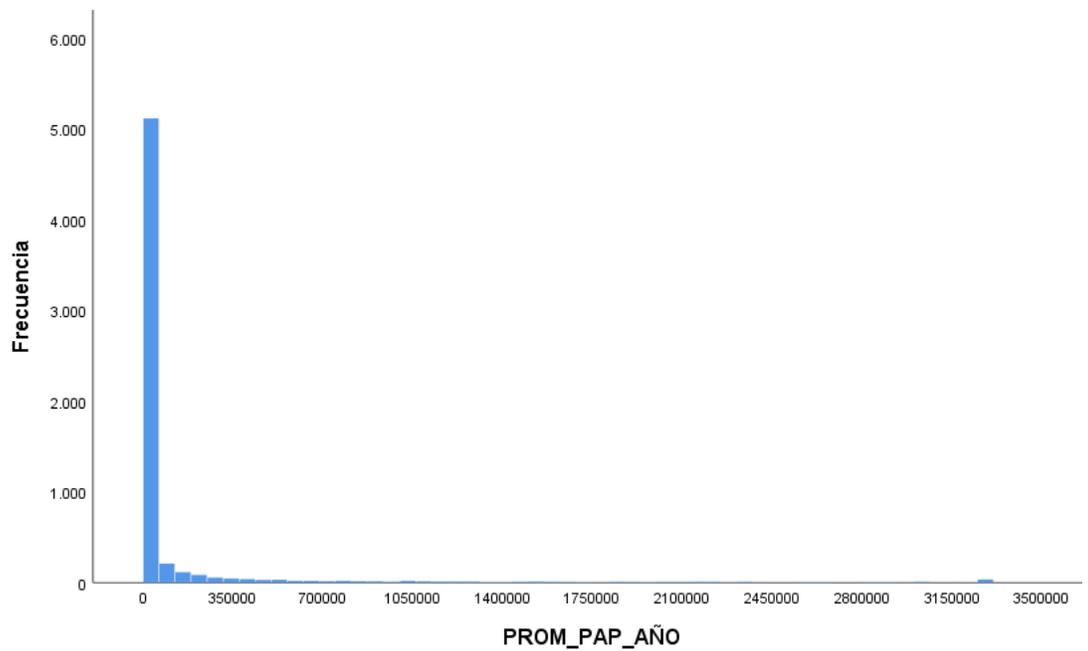
**Figura 22.**

*Histograma de promedio de saldos de crédito en 2014*



**Figura 23.**

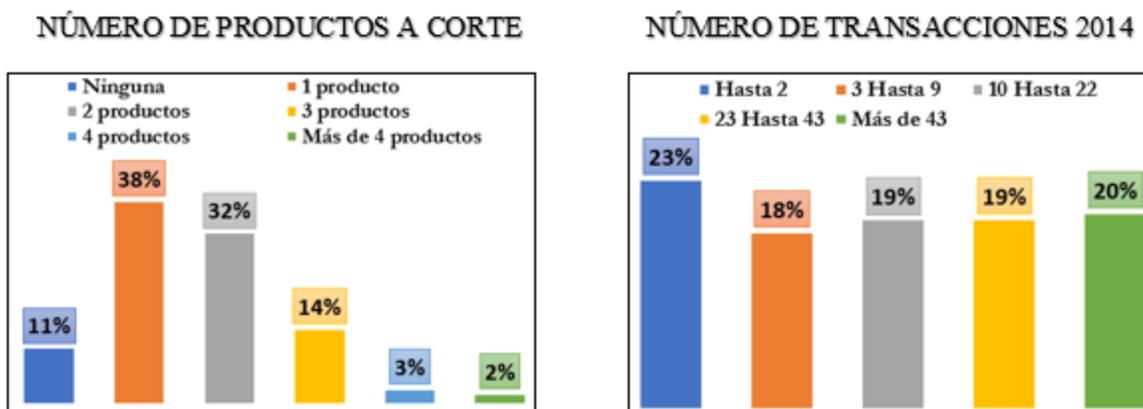
*Histograma de promedio de saldos de PAP en 2014*



Observando la cantidad de productos, preocupa que poco más del 80% de los asociados cuentan con máximo 2 productos, ya que esto indica la poca profundización de la Cooperativa en estas personas, siendo aquellas sin productos o con solo 1 vigente a las que debe prestárseles más atención, todo lo contrario denota el casi 20% de los asociados que tienen 3 productos o más con la Cooperativa. Al seguir analizando el nivel de utilización que los asociados de la muestra hacen de los productos de la cooperativa se detallan la cantidad de movimientos transaccionales voluntarios que realizaron a lo largo del año 2014, donde el 41% de los asociados realizaron menos de 10 transacciones en este periodo de tiempo; aproximadamente el 40% transaron más de 23 veces en la cooperativa en el 2014. Se hace énfasis en la voluntariedad de las transacciones ya que existen muchas transacciones que la Cooperativa realiza automáticamente, como por ejemplo los impuestos (4x1000), comisiones (retiros en cajeros) entre otras, éstas transacciones fueron excluidas ya que no aportaban al propósito de este proyecto. Ver la Figura 24. Para efectos de la ejecución del modelo se unirán los rangos que contienen a partir de 3 productos en una categoría debido a su baja participación en el total individualmente.

**Figura 24.**

*Distribución por cantidad de productos y transacciones*

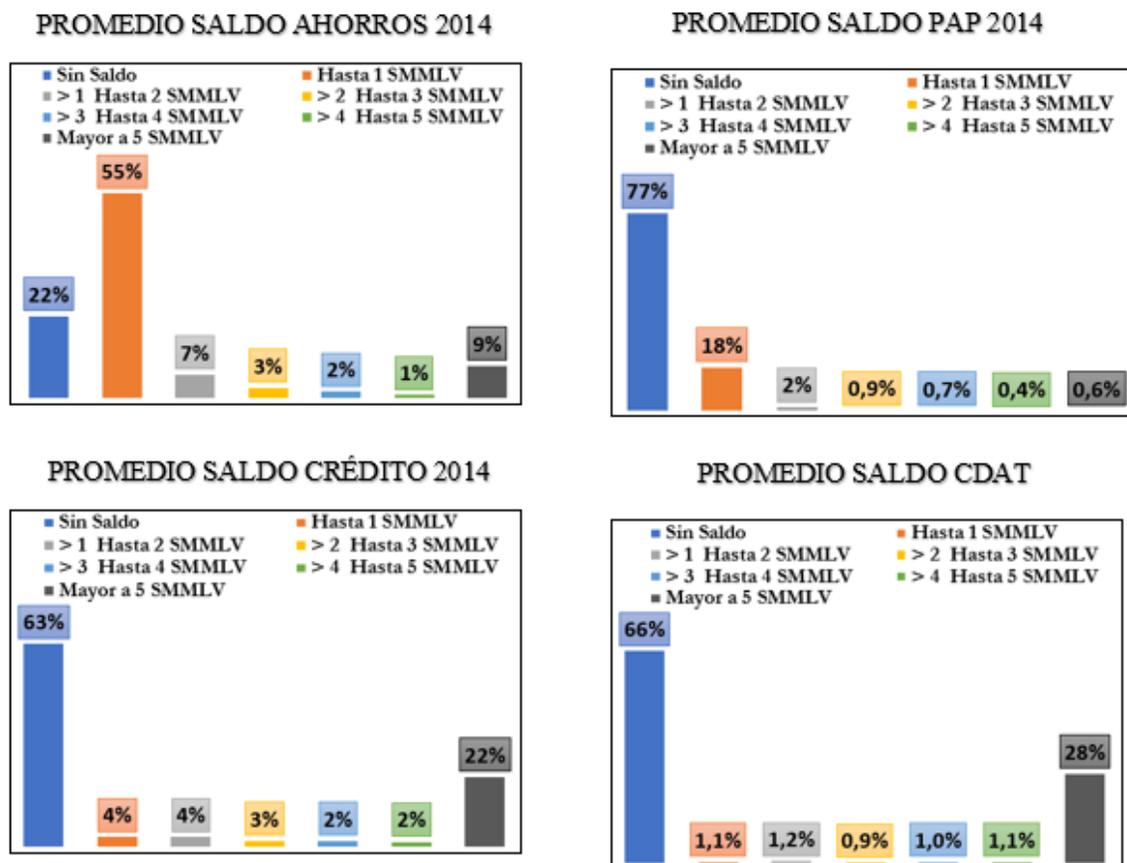


Debido a que hay una participación mayoritaria de las personas que no tienen productos con la cooperativa, se observa que los rangos que contienen los saldos “Hasta 1 SMMLV” son los que abarcan más asociados ya que allí se encuentran tanto las personas sin producto como aquellas con saldos bajos, apreciándose una participación de 77% en ahorros, 67% en crédito, más del 95% en PAP y 67% en CDAT; por otra parte, sólo los productos de CDAT y Crédito cuentan con otras categorías relativamente significativas con un 28% y 22% respectivamente en la categoría de saldos superiores a 5 SMMLV.

Para efectos de la ejecución del modelo y debido a su baja participación en el total individualmente, en el caso de los promedios de ahorros se unirán los rangos que contienen los valores mayores a 1 SMMLV en una sola categoría; para los promedios de crédito se agrupan en una sola categoría las personas con saldo entre \$1 y 5 SMMLV. En lo referente a los saldos promedio de PAP y CDAT se unirán en una sola categoría las personas que tengan saldo teniendo en cuenta el gran porcentaje que no cuenta con saldos en esta variable.

Figura 25.

Distribución por rangos de promedio de saldos de los productos en 2014



Finalmente se procedió a calcular el coeficiente de correlación de Pearson entre las variables continuas transaccionales. En la Tabla 4 se observa que no existen correlaciones significativas con un coeficiente superior a 0,5 y que la máxima correlación se da entre las variables del número de productos vigentes al inicio del periodo de observación y el número de transacciones voluntarias durante el 2014 con un coeficiente del 0,447.

**Tabla 4.**

*Correlación de las variables continuas transaccionales*

		<b>Correlaciones</b>					
		NUM_PROD UCTOS	PROM_AHOR RO_AÑO	PROM_CRED ITO_AÑO	PROM_PAP_ AÑO	PROM_CDAT _AÑO	NUM_TRANS _AÑO
NUM_PRODUCTOS	Correlación de Pearson	1	,122**	,296**	,159**	,284**	,447**
	Sig. (unilateral)		,000	,000	,000	,000	,000
	N	6029	6029	6029	6029	6029	6029
PROM_AHORRO_AÑO	Correlación de Pearson	,122**	1	,087**	,072**	,160**	,140**
	Sig. (unilateral)	,000		,000	,000	,000	,000
	N	6029	6029	6029	6029	6029	6029
PROM_CREDITO_AÑO	Correlación de Pearson	,296**	,087**	1	,014	-,044**	,392**
	Sig. (unilateral)	,000	,000		,132	,000	,000
	N	6029	6029	6029	6029	6029	6029
PROM_PAP_AÑO	Correlación de Pearson	,159**	,072**	,014	1	,030**	,133**
	Sig. (unilateral)	,000	,000	,132		,010	,000
	N	6029	6029	6029	6029	6029	6029
PROM_CDAT_AÑO	Correlación de Pearson	,284**	,160**	-,044**	,030**	1	,057**
	Sig. (unilateral)	,000	,000	,000	,010		,000
	N	6029	6029	6029	6029	6029	6029
NUM_TRANS_AÑO	Correlación de Pearson	,447**	,140**	,392**	,133**	,057**	1
	Sig. (unilateral)	,000	,000	,000	,000	,000	
	N	6029	6029	6029	6029	6029	6029

**2.1.2.2 Tablas cruzadas (Análisis bivariado).** Con el propósito de conocer el perfil de las personas que se retiraron de la Cooperativa en el transcurso de los 5 años de seguimiento del estudio, se muestran las tablas cruzadas en las que se compara cada una de las variables categóricas en función de la variable “FUGADO”, la cual indica si la persona se retiró o no en el periodo de tiempo anteriormente descrito.

**Tabla 5.**

*Tabla cruzada de Fugado vs Sexo*

		Tabla cruzada FUGADO*GENERO		
		GENERO		
FUGADO	NO	Recuento	Femenino	Masculino
		% dentro de FUGADO	59,9%	40,1%
	SI	Recuento	Femenino	Masculino
			689	580
		% dentro de FUGADO	54,3%	45,7%

Según la Tabla 5 las mujeres se retiraron en mayor proporción de la Cooperativa que los hombres durante el periodo de observación.

**Tabla 6.**

*Tabla cruzada de Fugado vs Rango de edad*

		Tabla cruzada FUGADO*RANGO_EDAD					
		RANGO_EDAD					
FUGADO	NO	Recuento	18 hasta 25 años	26 hasta 35 años	36 hasta 45 años	46 hasta 55 años	56 hasta 65 años
				429	429	741	996
		% dentro de FUGADO	9,0%	15,6%	20,9%	29,4%	25,1%
	SI	Recuento	18 hasta 25 años	26 hasta 35 años	36 hasta 45 años	46 hasta 55 años	56 hasta 65 años
			186	292	291	297	203
		% dentro de FUGADO	14,7%	23,0%	22,9%	23,4%	16,0%

Según la Tabla 6 las personas entre los 46 y 55 años fueron los que más desertaron de la Cooperativa con poco más del 23% de participación entre quienes.

**Tabla 7.**

*Tabla cruzada de Fugado vs Estado civil*

			ESTADO_CIVIL					
			Casado	Divorciado	Separado	Soltero	Union Libre	Viudo
FUGADO	NO	Recuento	1749	59	239	1870	716	127
		% dentro de FUGADO	36,7%	1,2%	5,0%	39,3%	15,0%	2,7%
	SI	Recuento	448	15	74	499	206	27
		% dentro de FUGADO	35,3%	1,2%	5,8%	39,3%	16,2%	2,1%

La Tabla 7 muestra que los asociados con estado civil de soltero fueron junto con los casados quienes desertaron en mayor proporción de la Cooperativa en el transcurso del periodo de observación.

**Tabla 8.**

*Tabla cruzada de Fugado vs Tipo de vivienda*

			TIPO_VIVIENDA		
			Alquiler	Familiar	Propia
FUGADO	NO	Recuento	552	2125	2083
		% dentro de FUGADO	11,6%	44,6%	43,8%
	SI	Recuento	151	709	409
		% dentro de FUGADO	11,9%	55,9%	32,2%

En función de la Tabla 8 podemos concluir que las personas que habitan en vivienda familiar fueron quienes desertaron en mayor proporción de la Cooperativa en el transcurso del periodo de observación con un porcentaje del 55,9%.

**Tabla 9.**

*Tabla cruzada de Fugado vs Cargo*

		Tabla cruzada FUGADO*CARGO										
		CARGO										
		AMA DE CASA	DOCENTE	ESTUDIANTES	INDEPENDIENTE	NINGUNA	OFICINISTAS	OPERARIOS	OTRAS	PENSIONADOS	PROFESIONAL	
FUGADO	NO	Recuento	253	122	157	1862	386	440	733	192	353	262
	% dentro de FUGADO	5,3%	2,6%	3,3%	39,1%	8,1%	9,2%	15,4%	4,0%	7,4%	5,5%	
SI	NO	Recuento	37	42	45	438	33	129	350	68	62	65
	% dentro de FUGADO	2,9%	3,3%	3,5%	34,5%	2,6%	10,2%	27,6%	5,4%	4,9%	5,1%	

Según la Tabla 9 los asociados con ocupación de trabajadores independientes fueron junto con los operarios quienes desertaron en mayor proporción de la Cooperativa en el transcurso del periodo de observación con un 34,5% y 27,6% respectivamente.

**Tabla 10.**

*Tabla cruzada de Fugado vs Estrato*

		Tabla cruzada FUGADO*ESTRATO						
		ESTRATO						
		1	2	3	4	5	6	
FUGADO	NO	Recuento	756	1112	1728	1001	97	66
	% dentro de FUGADO	15,9%	23,4%	36,3%	21,0%	2,0%	1,4%	
SI	NO	Recuento	120	324	503	276	30	16
	% dentro de FUGADO	9,5%	25,5%	39,6%	21,7%	2,4%	1,3%	

La Tabla 10 muestra que las personas que residían en estrato 3 fueron con un 39,6% quienes desertaron en mayor proporción de la Cooperativa durante el seguimiento.

**Tabla 11.**

*Tabla cruzada de Fugado vs Nivel educativo*

Tabla cruzada FUGADO*NIVEL_EDU								
			NIVEL_EDU					
			Ninguna	Posgrado	Primaria	Secundaria	Tecnico	Universidad
FUGADO	NO	Recuento	304	214	916	1896	508	922
		% dentro de FUGADO	6,4%	4,5%	19,2%	39,8%	10,7%	19,4%
	SI	Recuento	14	52	167	567	215	254
		% dentro de FUGADO	1,1%	4,1%	13,2%	44,7%	16,9%	20,0%

Gracias a la Tabla 11 se aprecia que los asociados con nivel educativo máximo alcanzado de secundaria fueron quienes desertaron en mayor proporción de la Cooperativa en el transcurso del periodo de observación con una participación del 44,7%.

**Tabla 12.**

*Tabla cruzada de Fugado vs Rango de antigüedad*

Tabla cruzada FUGADO*RANGO_ANTIGUEDAD								
			RANGO_ANTIGUEDAD					
			> 1 Hasta 3 años	> 10 Hasta 20 años	> 3 Hasta 5 años	> 5 Hasta 10 años	Hasta 1 año	Mayor a 20 Años
FUGADO	NO	Recuento	477	855	411	947	1126	944
		% dentro de FUGADO	10,0%	18,0%	8,6%	19,9%	23,7%	19,8%
	SI	Recuento	191	141	122	168	553	94
		% dentro de FUGADO	15,1%	11,1%	9,6%	13,2%	43,6%	7,4%

En función de la Tabla 12 se puede observar que los asociados con menos de 1 año de antigüedad en la Cooperativa fueron con un 43,6% quienes más desertaron durante el seguimiento.

**Tabla 13.**

*Tabla cruzada de Fugado vs Rango de ingresos*

			Tabla cruzada FUGADO*RANGO_INGRESOS					
			RANGO_INGRESOS					
			> 1 Hasta 2 SMMLV	> 2 Hasta 3 SMMLV	> 3 Hasta 4 SMMLV	> 4 Hasta 5 SMMLV	Hasta 1 SMMLV	Mayor a 5 SMMLV
FUGADO	NO	Recuento	1302	734	423	383	982	936
		% dentro de FUGADO	27,4%	15,4%	8,9%	8,0%	20,6%	19,7%
	SI	Recuento	378	233	120	112	163	263
		% dentro de FUGADO	29,8%	18,4%	9,5%	8,8%	12,8%	20,7%

La Tabla 13 muestra que los asociados con ingresos inferiores a 1 smmlv fueron junto con aquellos que devengan más de 5 smmlv quienes desertaron en mayor proporción de la Cooperativa en el transcurso del periodo de observación con una participación el 29,8% y 20,7% respectivamente.

**Tabla 14.**

*Tabla cruzada de Fugado vs Rango de egresos*

			Tabla cruzada FUGADO*RANGO_EGRESOS					
			RANGO_EGRESOS					
			> 1 Hasta 2 SMMLV	> 2 Hasta 3 SMMLV	> 3 Hasta 4 SMMLV	> 4 Hasta 5 SMMLV	Hasta 1 SMMLV	Mayor a 5 SMMLV
FUGADO	NO	Recuento	873	313	183	100	2769	522
		% dentro de FUGADO	18,3%	6,6%	3,8%	2,1%	58,2%	11,0%
	SI	Recuento	192	82	52	29	761	153
		% dentro de FUGADO	15,1%	6,5%	4,1%	2,3%	60,0%	12,1%

La Tabla 14 muestra que los asociados con estado civil de soltero fueron junto con los casados quienes desertaron en mayor proporción de la Cooperativa en el transcurso del periodo de observación.

**Tabla 15.**

*Tabla cruzada de Fugado vs Personas a cargo*

			PERSONAS_A_CARGO	
			NO	SI
FUGADO	NO	Recuento	2819	1941
		% dentro de FUGADO	59,2%	40,8%
	SI	Recuento	710	559
		% dentro de FUGADO	55,9%	44,1%

Según la Tabla 15 los asociados que no reportaban personas a cargo fueron quienes más desertaron de la entidad con casi el 56% durante el seguimiento.

**Tabla 16.**

*Tabla cruzada de Fugado vs Tenencia de crédito*

			CREDITO	
			NO	SI
FUGADO	NO	Recuento	3459	1301
		% dentro de FUGADO	72,7%	27,3%
	SI	Recuento	570	699
		% dentro de FUGADO	44,9%	55,1%

En función de la Tabla 16 se aprecia que los asociados con crédito vigente al momento de iniciar el periodo de observación fueron quienes más se retiraron de la Cooperativa con una participación de poco más del 55%.

**Tabla 17.**

*Tabla cruzada de Fugado vs Tenencia de ahorros*

			AHORROS	
			NO	SI
FUGADO	NO	Recuento	948	3812
		% dentro de FUGADO	19,9%	80,1%
	SI	Recuento	116	1153
		% dentro de FUGADO	9,1%	90,9%

Gracias a la Tabla 17 se aprecia que las personas con cuenta de ahorros a corte de diciembre de 2014, es decir, el inicio del seguimiento fue quienes más desertaron con un porcentaje del 90,9%.

**Tabla 18.**

*Tabla cruzada de Fugado vs Tenencia de PAP*

			PAP	
			NO	SI
FUGADO	NO	Recuento	4013	747
		% dentro de FUGADO	84,3%	15,7%
	SI	Recuento	984	285
		% dentro de FUGADO	77,5%	22,5%

La Tabla 18 muestra que las personas con PAP en el inicio del periodo de observación fueron quienes más participaron entre los retirados con más de  $\frac{3}{4}$  partes del total.

**Tabla 19.**

*Tabla cruzada de Fugado vs Tenencia de CDAT*

		<b>Tabla cruzada FUGADO*CDAT</b>		
		CDAT		
FUGADO	NO	Recuento	3712	1048
		% dentro de FUGADO	78,0%	22,0%
	SI	Recuento	1150	119
		% dentro de FUGADO	90,6%	9,4%

En función de la Tabla 19 se aprecia que los asociados que no tenían un CDAT vigente en diciembre de 2014 son quienes mayor participación tienen en los retirados en el transcurso del periodo de observación con poco más del 90%.

**Tabla 20.**

*Tabla cruzada de Fugado vs Inactividad en la cuenta de ahorros*

		<b>Tabla cruzada FUGADO*INACT_AHORRO</b>		
		INACT_AHORRO		
FUGADO	NO	Recuento	3399	1361
		% dentro de FUGADO	71,4%	28,6%
	SI	Recuento	888	381
		% dentro de FUGADO	70,0%	30,0%

Según la Tabla 20 se observa que 7 de cada 10 personas retiradas no tenían la cuenta de ahorros inactiva al momento de iniciar el seguimiento lo hicieron en el transcurso de este lapso de 5 años.

**Tabla 21.**

*Tabla cruzada de Fugado vs Rango de número de productos*

			RANGO_PRODUCTOS					Más de 4 productos	Ninguna
			1 producto	2 productos	3 productos	4 productos			
FUGADO	NO	Recuento	1889	1412	614	144	110	591	
		% dentro de FUGADO	39,7%	29,7%	12,9%	3,0%	2,3%	12,4%	
	SI	Recuento	417	534	239	24	8	47	
		% dentro de FUGADO	32,9%	42,1%	18,8%	1,9%	0,6%	3,7%	

La Tabla 21 muestra que los asociados con 1 y 2 productos vigentes al momento de empezar el periodo de observación conforman el 75% de las personas que terminaron abandonando la Cooperativa.

**Tabla 22.**

*Tabla cruzada de Fugado vs Rango de saldo promedio de ahorros*

			RANGO_AHORRO						
			> 1 Hasta 2 SMMLV	> 2 Hasta 3 SMMLV	> 3 Hasta 4 SMMLV	> 4 Hasta 5 SMMLV	Hasta 1 SMMLV	Mayor a 5 SMMLV	Sin Saldo
FUGADO	NO	Recuento	322	158	112	76	2427	473	1192
		% dentro de FUGADO	6,8%	3,3%	2,4%	1,6%	51,0%	9,9%	25,0%
	SI	Recuento	89	40	27	11	878	75	149
		% dentro de FUGADO	7,0%	3,2%	2,1%	0,9%	69,2%	5,9%	11,7%

Gracias a la Tabla 22 se evidencia que los asociados con promedio de saldo de ahorros de hasta 1 smmlv participan con casi el 70% de los retirados que se produjeron en el periodo de observación.

**Tabla 23.**

*Tabla cruzada de Fugado vs Rango de saldo promedio de crédito*

		Tabla cruzada FUGADO*RANGO_CREDITO							
		RANGO_CREDITO							
		> 1 Hasta 2 SMMLV	> 2 Hasta 3 SMMLV	> 3 Hasta 4 SMMLV	> 4 Hasta 5 SMMLV	Hasta 1 SMMLV	Mayor a 5 SMMLV	Sin Saldo	
FUGADO	NO	Recuento	156	101	92	83	162	876	3290
		% dentro de FUGADO	3,3%	2,1%	1,9%	1,7%	3,4%	18,4%	69,1%
	SI	Recuento	68	51	55	50	55	467	523
		% dentro de FUGADO	5,4%	4,0%	4,3%	3,9%	4,3%	36,8%	41,2%

En función de la Tabla 23 podemos observar que muestra que los asociados sin saldo promedio de crédito o con más de 5 smmlv en este ítem participan con aproximadamente el 80% de las personas desertoras durante el seguimiento.

**Tabla 24.**

*Tabla cruzada de Fugado vs Rango de saldo promedio de PAP*

		Tabla cruzada FUGADO*RANGO_PAP							
		RANGO_PAP							
		> 1 Hasta 2 SMMLV	> 2 Hasta 3 SMMLV	> 3 Hasta 4 SMMLV	> 4 Hasta 5 SMMLV	Hasta 1 SMMLV	Mayor a 5 SMMLV	Sin Saldo	
FUGADO	NO	Recuento	118	46	36	22	754	35	3749
		% dentro de FUGADO	2,5%	1,0%	0,8%	0,5%	15,8%	0,7%	78,8%
	SI	Recuento	16	7	7	2	319	1	917
		% dentro de FUGADO	1,3%	0,6%	0,6%	0,2%	25,1%	0,1%	72,3%

La Tabla 24 muestra que los asociados sin saldo sin saldo promedio de PAP durante el 2014 tienen la gran mayoría de los desertores al acumular más del 70% de los retiros de la muestra durante los 5 años del periodo de observación.

**Tabla 25.**

*Tabla cruzada de Fugado vs Rango de saldo promedio de CDAT*

		Tabla cruzada FUGADO*RANGO_CDAT							
		RANGO_CDAT							
		> 1 Hasta 2 SMMLV	> 2 Hasta 3 SMMLV	> 3 Hasta 4 SMMLV	> 4 Hasta 5 SMMLV	Hasta 1 SMMLV	Mayor a 5 SMMLV	Sin Saldo	
FUGADO	NO	Recuento	59	43	49	58	53	1510	2988
		% dentro de FUGADO	1,2%	0,9%	1,0%	1,2%	1,1%	31,7%	62,8%
SI	NO	Recuento	15	11	13	8	12	202	1008
		% dentro de FUGADO	1,2%	0,9%	1,0%	0,6%	0,9%	15,9%	79,4%

Según la Tabla 25 y al igual que lo apreciado en la tabla cruzada anterior las personas sin saldo promedio en el transcurso del 2014 pero esta vez en el producto de CDAT son quienes mayor participación tienen en los retirados con un porcentaje cercano al 80%.

**Tabla 26.**

*Tabla cruzada de Fugado vs Rango de número de transacciones voluntarias*

		Tabla cruzada FUGADO*RANGO_TRANSACCIONES					
		RANGO_TRANSACCIONES					
		10 Hasta 22	23 Hasta 43	3 Hasta 9	Hasta 2	Más de 43	
FUGADO	NO	Recuento	909	790	901	1251	909
		% dentro de FUGADO	19,1%	16,6%	18,9%	26,3%	19,1%
SI	NO	Recuento	256	378	184	157	294
		% dentro de FUGADO	20,2%	29,8%	14,5%	12,4%	23,2%

En función de la Tabla 26 se aprecia que los asociados con más de 23 transacciones voluntarias durante el 2014 agrupadas en los rangos de entre 23 y 43 transacciones y más de 43 transacciones participan en conjunto con un porcentaje superior al 50% de los desertores registrados durante el seguimiento.

### 2.1.2.3 Conclusiones de la exploración univariada y bivariada de los datos

- El perfil que predomina en las personas de la muestra es: Mayor porcentaje de mujeres, minoría de jóvenes hasta los 25 años, mayoritariamente asociados solteros y casados con vivienda familiar o propia y actividad económica independiente u operarios, prevalecen la clase media y baja de la población con estudios básicos; más de la mitad de estas personas llevan más de 5 años de antigüedad con la Cooperativa sin personas a cargo. Casi 1 de cada 2 personas ganan menos de 2 SMMLV y aproximadamente el 60% de los asociados reportan egresos inferiores a 1 SMMLV. Sólo 1 de cada 3 personas tenían crédito vigente a corte de diciembre de 2014, el 18% poseían PAP y únicamente el 19% CDAT, mientras que el 82% contaba con al menos una cuenta de ahorros. Casi 8 de cada 10 asociados tenían máximo 2 productos y cerca del 60% había realizado por lo menos 10 transacciones voluntarias en el año 2014; finalmente, su promedio de saldo en los diferentes productos de la Cooperativa se agrupaba significativamente en el rango de hasta 1 SMMLV.

- Las variables continuas, tanto transaccionales como sociodemográficas, poseen una gran dispersión de los datos, se evidencian datos atípicos y extremos. En cuanto a las variables continuas sociodemográficas se aprecia que los ingresos y egresos cuentan con una correlación que se acerca a 1 según el coeficiente de correlación de Pearson, mientras que en las variables continuas transaccionales no se aprecian correlaciones significativas según el mismo coeficiente. Para la aplicación del modelo se distribuyeron los datos en rangos utilizados en los informes de la Cooperativa.

- Teniendo en cuenta que algunas categorías cuentan con una participación baja en el total, se decidió unir las con la más cercana que tenga la menor participación en el total, buscando

así evitar el desbalanceo de clases y sesgar los resultados de la aplicación del modelo de regresión de Cox.

- Si bien las tablas cruzadas pueden dar una idea acerca de que perfil de asociados fueron los que más se retiraron durante el periodo de observación, es importante destacar que en su mayoría coincide con la participación general de la categoría en la muestra, por lo tanto la aplicación del modelo de análisis de supervivencia permitirá encontrar las variables que estadísticamente son significativas a la hora de influir en la deserción de las personas de la Cooperativa.

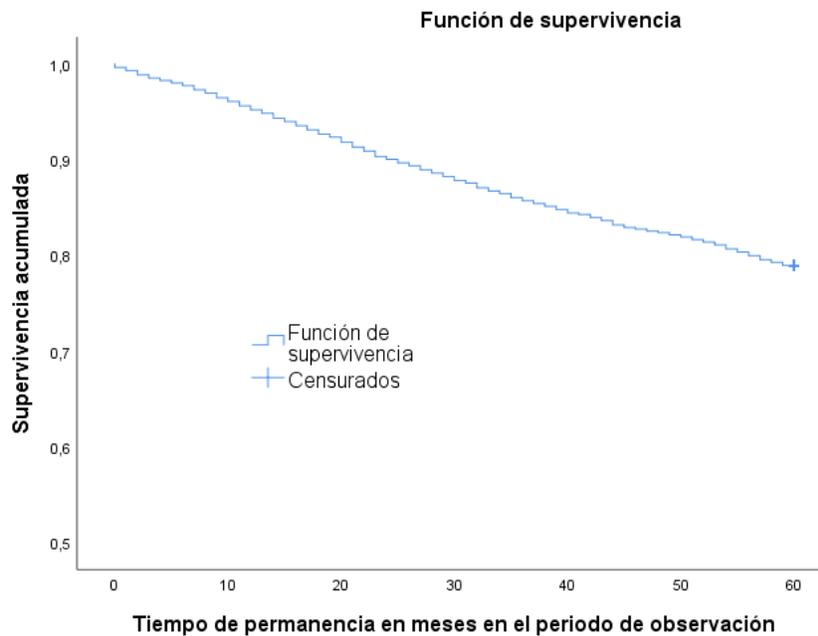
**2.1.2.4 Aplicación del análisis de supervivencia.** El análisis de supervivencia de los datos de esta investigación se realizó en el software SPSS STATISTICS versión 25. Como ya se mencionó, una vez realizados todos los filtros, se cuenta con un total de 6.029 observaciones. A continuación, se presentan las curvas de supervivencia tanto general como para las variables categóricas del estudio.

**2.1.2.4.1 Curvas de supervivencia de Kaplan-Meier.** La Figura 26 muestra la curva de supervivencia general de los datos del estudio, allí se aprecia un descenso constante de la probabilidad de permanecer como asociado a través de los meses situándose en el mes 60 en un porcentaje cercano al 80%, esto se ajusta a los datos de la entidad que indican que mensualmente se retiran de la organización voluntariamente una cantidad similar de asociados. Teniendo en cuenta las características de la base, es un buen indicador que 4 de cada 5 personas permanezcan por lo menos 5 años en la Cooperativa después del inicio del seguimiento, sin embargo, debe

aclararse que no todo el universo de asociados cumple con estas características, por lo que es factible que la curva cambie una vez se generalice el modelo a todas las personas.

**Figura 26.**

*Curva de supervivencia general*



Ahora, se muestran las curvas de supervivencia por cada una de las variables tenidas en cuenta en este estudio que presentaron diferencias en el comportamiento de sus categorías con un nivel de significancia del 5%.

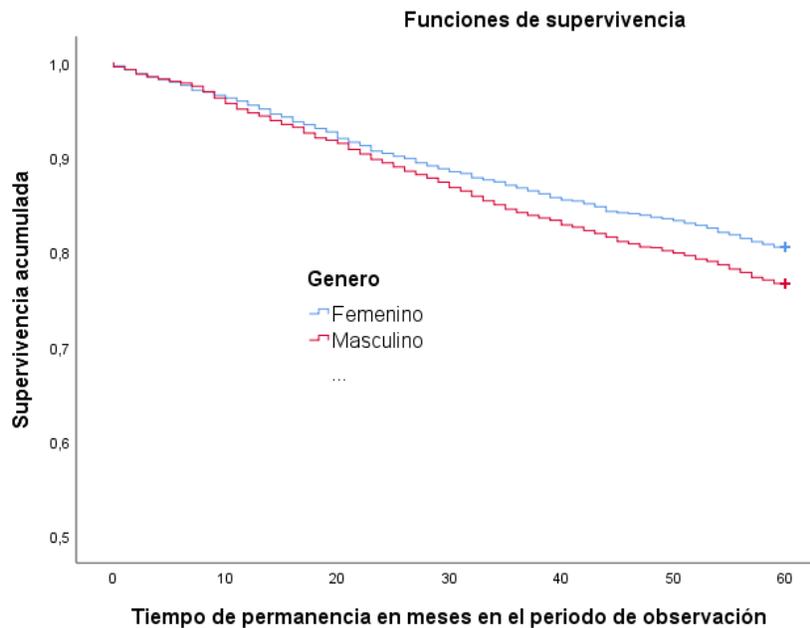
**Sexo**

En cuanto al sexo, se aprecia en la Figura 27 que las curvas de hombres y mujeres presentan una diferencia estadísticamente significativa que se acentúa más con el paso del tiempo: a partir del mes 10 los hombres tienen menos probabilidad de permanecer como asociados, hacia el mes

40 se hace más notoria la disparidad y finaliza el periodo de observación con un riesgo 5% mayor de desertar que las mujeres.

**Figura 27.**

*Curva de supervivencia por sexo*

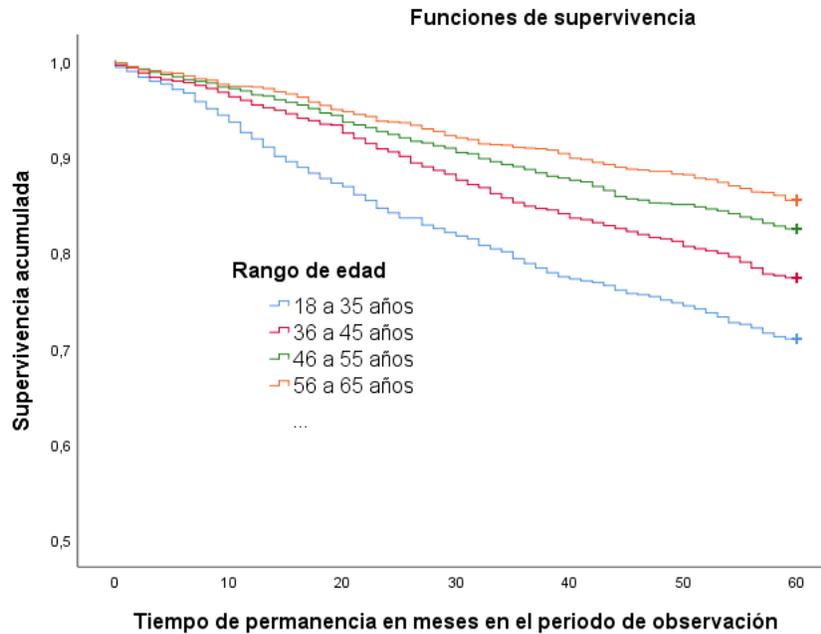


**Edad**

En la Figura 28 se observa que a medida que la edad aumenta, la probabilidad de permanecer como asociado también lo hace, a partir del mes 10 se comienza a notar una diferencia significativa entre las personas hasta 35 años y los mayores a está, llegando por lo menos a una diferencia porcentual de casi 10 puntos y en su pico máximo aproximadamente poco más del 15% de diferencia al final del periodo de observación, es decir, la fidelidad de los clientes es directamente proporcional a su edad.

**Figura 28.**

*Curva de supervivencia por rango de edad*

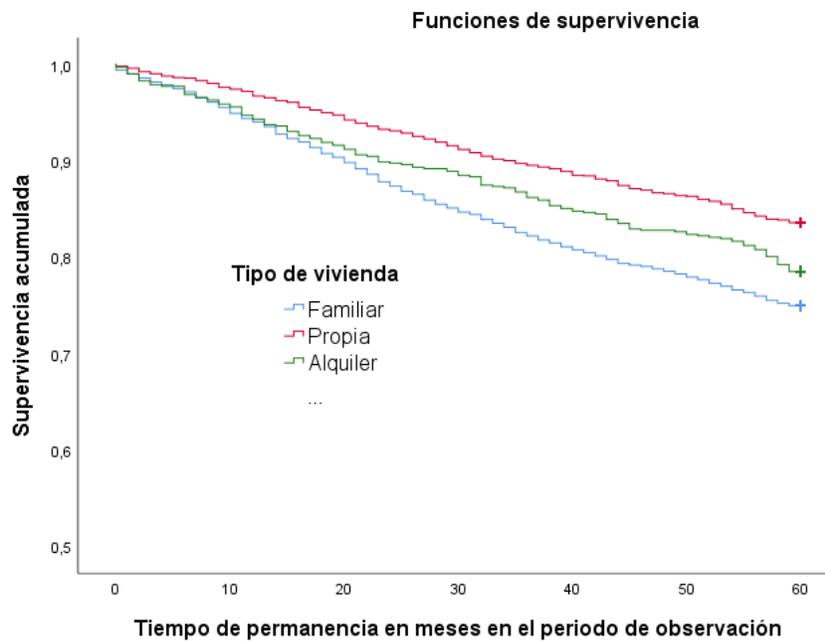


**Tipo de vivienda**

La Figura 29 evidencia que las personas con vivienda propia tienen mayor probabilidad de permanecer en la Cooperativa en comparación con aquellas con domicilio en arriendo o familiar. Con estas últimas la diferencia es aún más significativa, casi de un 10% al final del periodo de observación del estudio. De igual forma, las personas que viven en arriendo tienen alrededor de 3% más posibilidad de permanecer como asociado al finalizar el seguimiento en comparación con las que habitan en vivienda familiar.

**Figura 29.**

*Curva de supervivencia por tipo de vivienda*

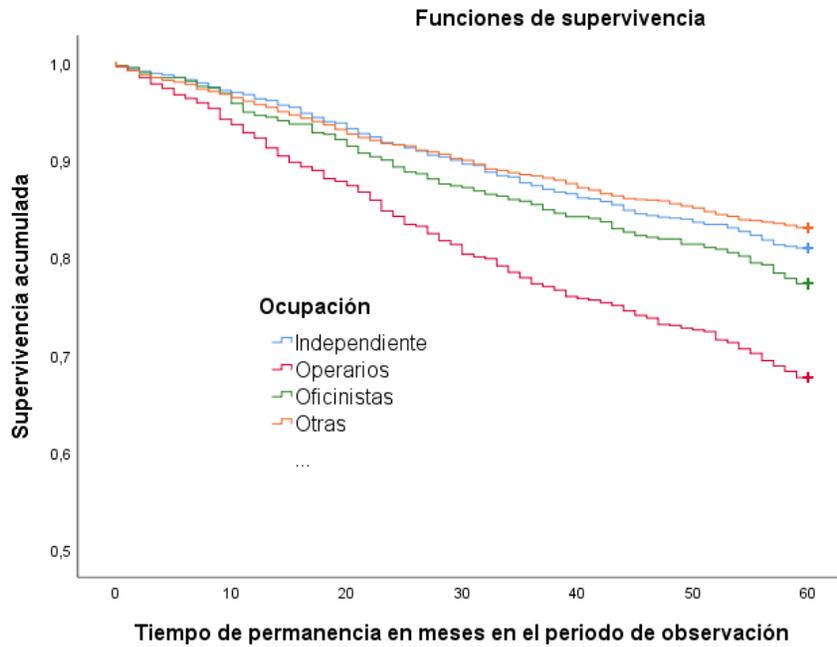


### Ocupación

La Figura 30 muestra que aquellas personas que reportaron no tener una actividad económica vigente como los son las amas de casa o los pensionados que se agrupan En la categoría “OTRAS” muestran el transcurso del tiempo una menor probabilidad de desertar de la Cooperativa. En el extremo opuesto, es decir, las categorías más propensas a desertar, se encuentran los operarios, quienes finalizan con cerca del 35% de probabilidad de abandonar la organización. En alrededor del 80% de probabilidad de persistir como asociados se encuentran las categorías “OTRAS”, ”OFICINISTA”, e “INDEPENDIENTES”.

**Figura 30.**

*Curva de supervivencia por ocupación*

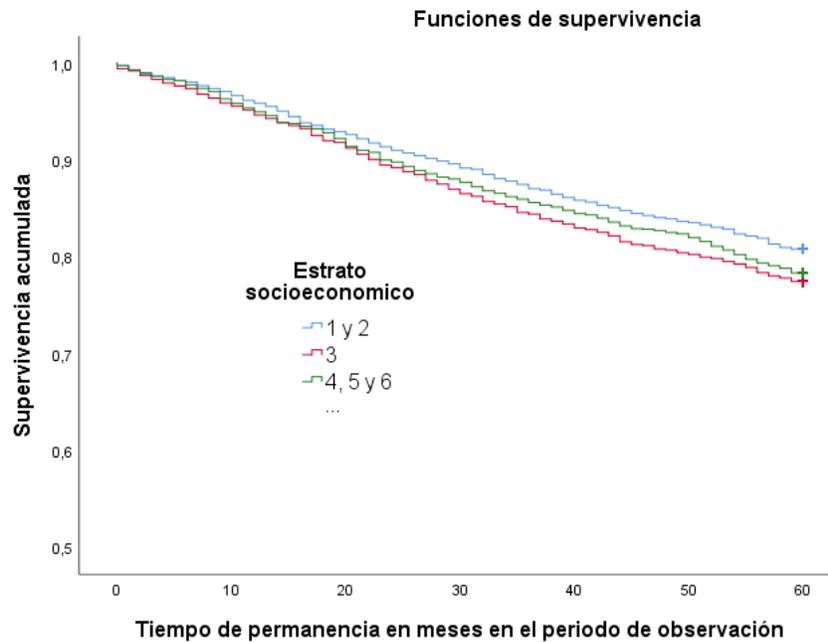


**Estrato**

Como uno de los factores sociodemográficos de este estudio se incluyó el estrato de los asociados; la Figura 31 denota que las personas registradas en el estrato 1 y 2 cuentan con más probabilidad de permanecer en el tiempo como asociados de la Cooperativa. Las categorías que agrupan los estratos superiores a 2 terminan el seguimiento con aproximadamente un 20% de probabilidad de desertar.

**Figura 31.**

*Curva de supervivencia por estrato*

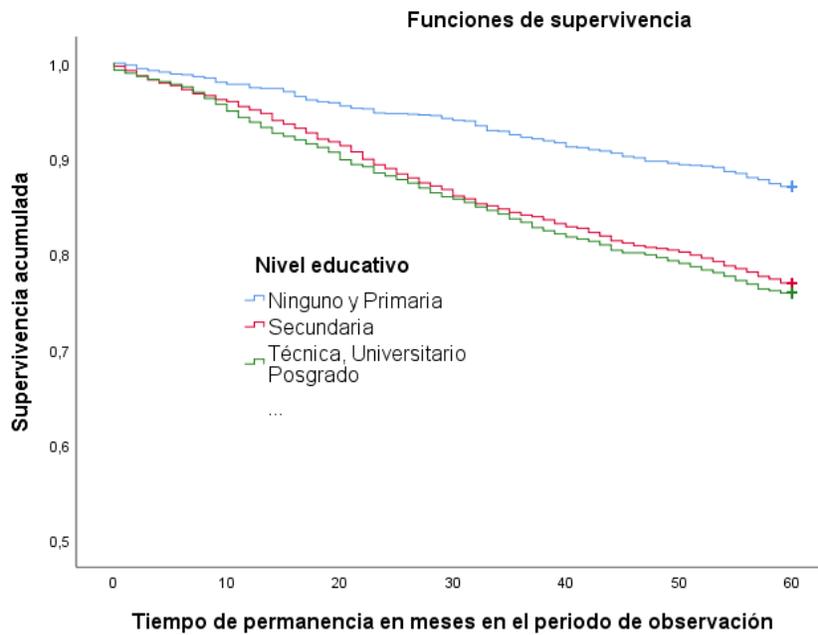


**Nivel educativo**

En cuanto al nivel educativo, puede observarse que las personas sin ningún nivel educativo alcanzado y con primaria, tienen la mayor probabilidad de permanecer como asociados con aproximadamente un 90% en el mes 60, mientras que las personas con nivel educativo técnico o superior tienen menos probabilidad de permanecer en la entidad llegando a menos del 80% al final del periodo de observación. La curva de la categoría “Secundaria” se mantiene cerca a la categoría antes mencionada a través del tiempo y termina sobre el 20% de probabilidad de desertar de la compañía, tal y como se desprende de la Figura 32.

**Figura 32.**

*Curva de supervivencia por nivel educativo*

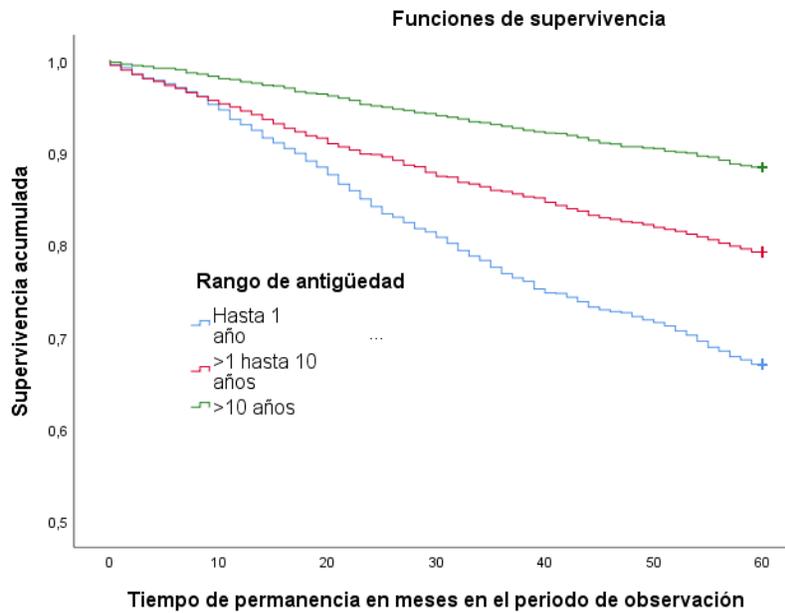


**Antigüedad**

La Figura 33 expone que mientras mayor es la antigüedad como asociado, rasgo que va de la mano con la edad de la persona, más grande es la probabilidad de permanecer en la Cooperativa. Al final del periodo de observación la diferencia en la probabilidad de desertar entre las personas con menos de un año de antigüedad al inicio del estudio y las personas con más de 10 años es de casi un 25%. Cerca de un 80% de probabilidad de quedarse en la Cooperativa tienen aquellos con entre 1 y 10 años de antigüedad.

**Figura 33.**

*Curva de supervivencia por rango de antigüedad*

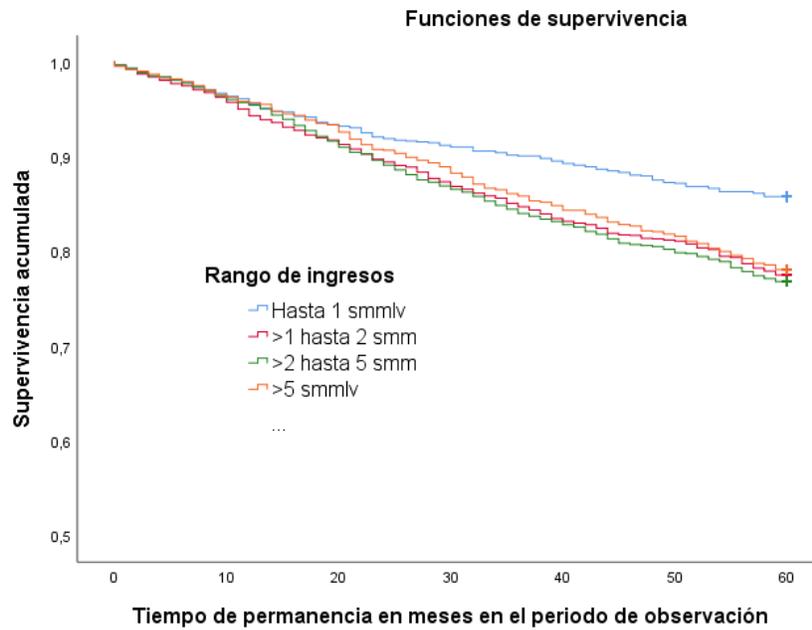


**Ingresos**

Se evidencia que las personas con ingresos hasta 1 smmlv tienen mayor probabilidad de permanecer en la Cooperativa en comparación con las demás categorías de la variable, al finalizar el seguimiento existe casi un 10% de diferencia con las personas de mayores ingresos. Mientras que las personas con ingresos superiores a 1 smmlv tienen una probabilidad de deserción de más del 20% una vez finalizado los 60 meses de observación.

**Figura 34.**

*Curva de supervivencia por rango de ingresos*

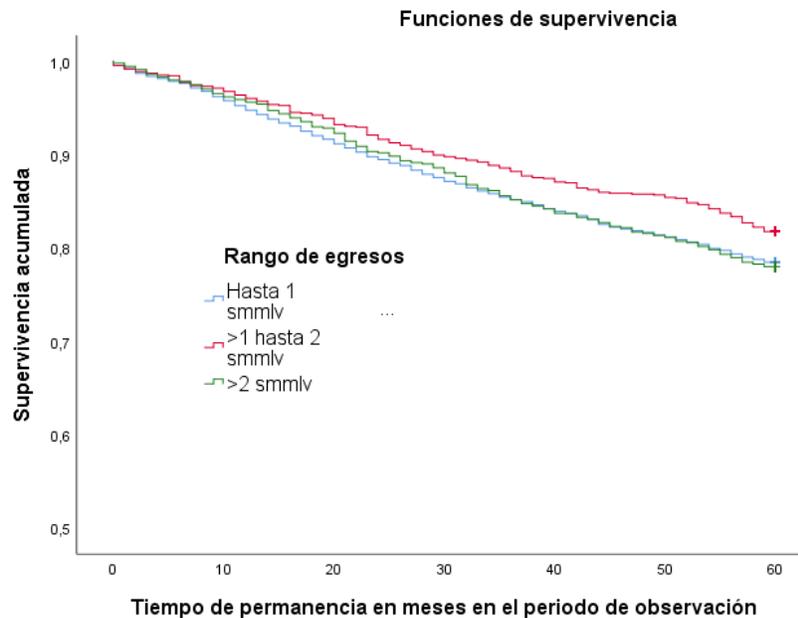


**Egresos**

Se evidencia que las personas con egresos entre 1 y 2 smmlv muestran menos probabilidad de abandonar la Cooperativa. Por otra parte los asociados que registran egresos hasta 1 smmlv o superior a 2 smmlv finalizando el periodo de estudio con una probabilidad de permanecer en la entidad cercana al 80%. La situación anteriormente descrita se aprecia en la Figura 35.

**Figura 35.**

*Curva de supervivencia por rango de egresos*



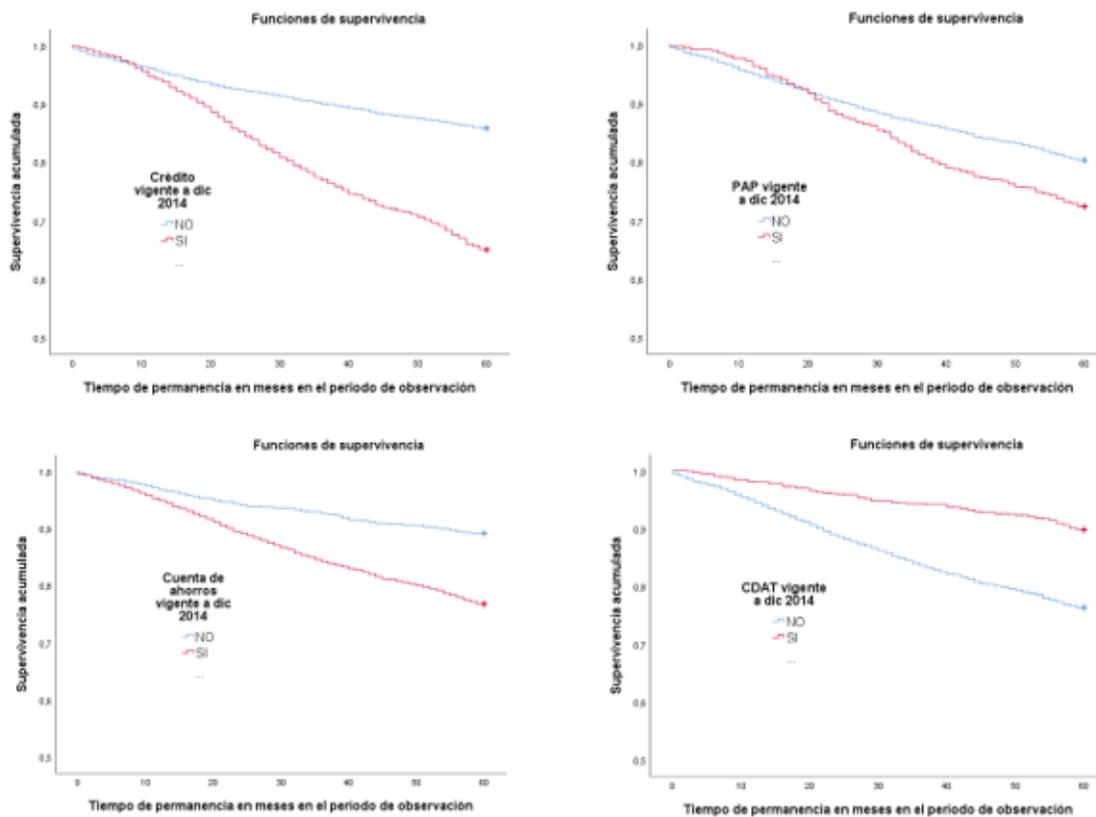
**Tenencia de crédito-ahorros-PAP-CDAT**

Tal y como puede verse en la Figura 36, tanto en crédito, como en ahorro y PAP las personas con presencia del producto vigente cuentan con mayor probabilidad de desertar. En el caso de cartera la diferencia es notable pasados los 10 primeros meses y terminando con alrededor de un 30% de diferencia, para PAP la brecha se amplía desde el mes 20 y concluye con más de 5 puntos porcentuales de disimilitud y en lo que se refiere a ahorros las curvas difieren desde el mes 5 finalizando con un 20% de diferencia entre los que cumplen la condición y los que no. Esto se debe a una mala práctica de la Cooperativa en la que: una vez se desembolsa un crédito, a la persona también se le crean los productos de ahorro y PAP, por tanto, es común que una vez la persona finaliza el pago de su crédito también cancele la cuenta de ahorros, el PAP y finalmente se retire de la entidad.

Por otra parte se observa que: contrario a la situación anteriormente descrita, las personas con CDAT tienen más probabilidad de permanecer en la Cooperativa como asociados en el transcurso del tiempo. De esta manera se demuestra que las personas con perfil inversor tienen menos riesgo de abandonar la organización ya que suelen buscar estabilidad en las inversiones que realizan. La diferencia es visible en cada uno de los meses y aumenta a lo largo del tiempo, a mediados del periodo de observación ya se observa una disparidad de aproximadamente 10 puntos porcentuales y finalizando esté llega aproximadamente al 20%.

**Figura 36.**

*Curva de supervivencia por tenencia de productos a dic. de 2014*



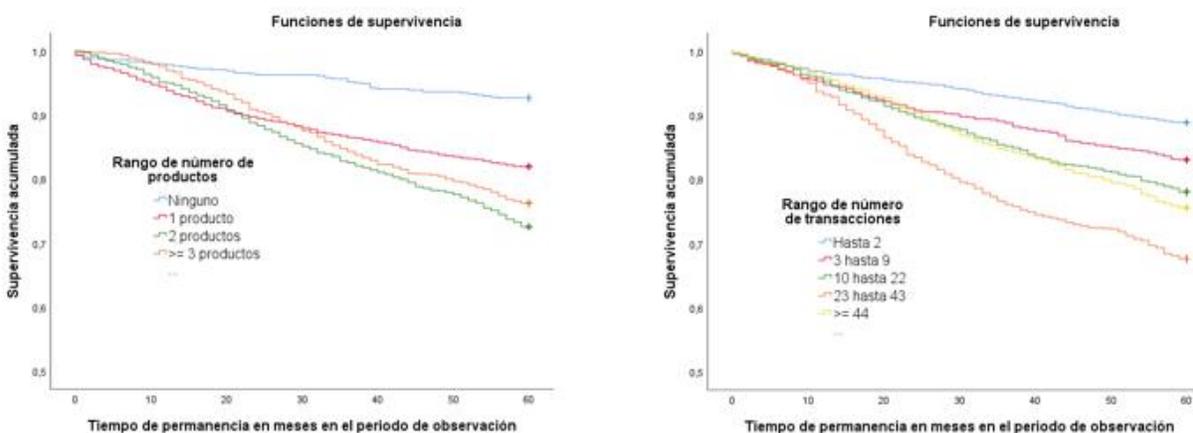
### **Cantidad de productos en diciembre de 2014 y cantidad de transacciones voluntarias durante el 2014**

La Figura 37 enseña que las personas sin productos vigentes a diciembre de 2014 tienen una mayor probabilidad de seguir como asociados, ya que finalizan el periodo de observación con solo un 5% de probabilidad de desertar, también se confirma que aquellas personas que terminan de cancelar su crédito, a su vez retiran el PAP y la cuenta de ahorros para abandonar la organización, abarcando así las categorías de 1, 2, 3 o más productos. Las personas con 1 producto finalizan el seguimiento con poco menos del 20% de probabilidad de desertar; los descensos más acelerados se aprecian en las curvas de los asociados con 2, 3 o más productos especialmente a partir del mes 40 y culminando el seguimiento con probabilidades cercana al 70% de permanecer como asociados de la organización.

Contrario a lo que se pensaba entre más transacciones realizan las personas, mayor probabilidad de desertar de la Cooperativa, aunque esto puede deberse a las transacciones que realizan cada vez que cancelan las cuotas del crédito y/o el PAP; siendo aquellas personas con entre 23 y 43 transacciones voluntarias en el 2014 las que menos probabilidad de permanecer como asociados muestran al finalizar el seguimiento con 30% de probabilidad de desertar de la Cooperativa.

**Figura 37.**

*Curva de supervivencia por cantidad de productos y transacciones*



**Saldo promedio de crédito, ahorros, PAP y CDAT durante 2014**

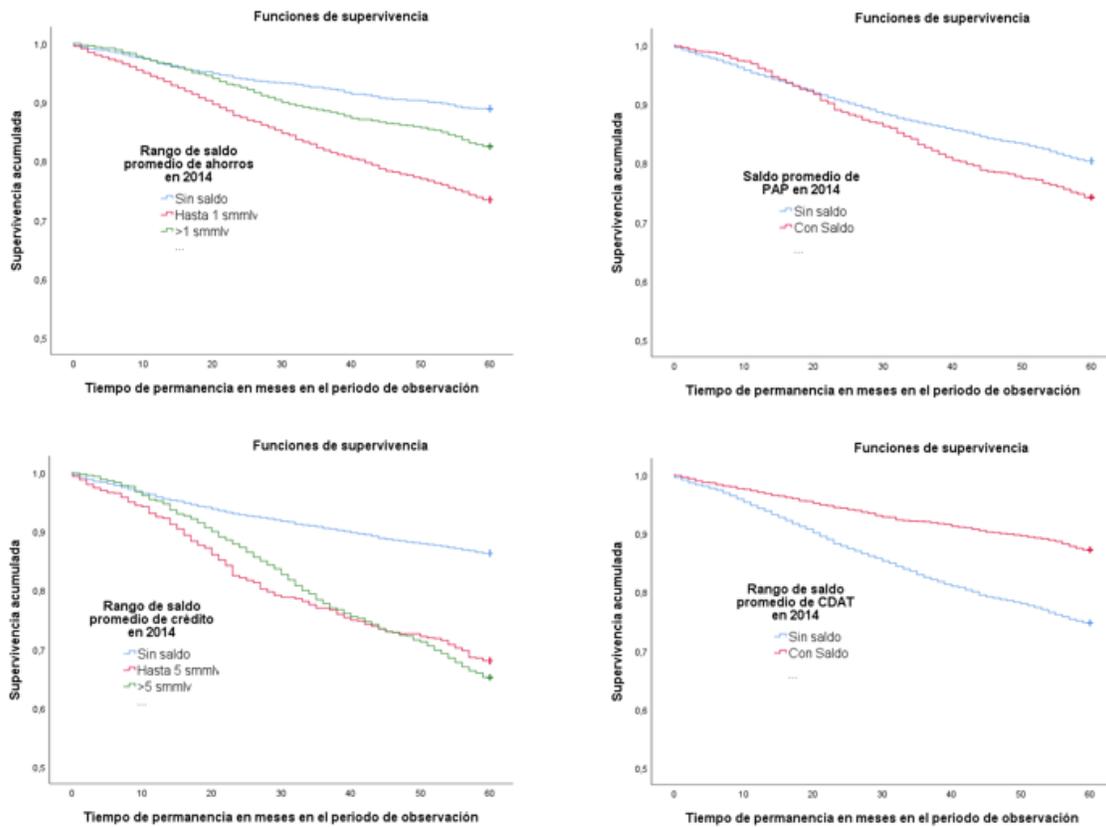
Finalmente, las variables que hacen alusión al saldo de los productos denotan que tanto en el caso de ahorros, como de crédito y PAP, el hecho de no haber tenido saldo a lo largo del 2014 aumenta la probabilidad de retirarse como asociado; todo lo contrario en cuanto a CDAT, donde las personas con saldo son aquellas con más opciones de permanecer en la entidad. En el caso de crédito las personas que mantuvieron saldo a lo largo de 2014 terminaron el estudio con alrededor del 30% de probabilidad de retirarse como asociado; para ahorros las personas con hasta 1 smmlv finalizaron con más del 25% de probabilidad de irse de la Cooperativa, y aquellos con más de 1 smmlv culminaron con un 80% de probabilidad de permanecer como asociados.

En PAP las personas con saldo promedio durante el 2014, finalizaron el estudio con casi un 75% de probabilidad de mantenerse en la entidad, 5 puntos porcentuales debajo de los asociados sin esta condición, quienes culminaron el periodo de observación con el 80% de probabilidad de permanecer como asociados de la Cooperativa.

Finalmente en CDAT las personas con saldo a lo largo del 2014 culminaron con casi un 20% menos de probabilidad de desertar que aquellos que no contaron con el producto en el año en mención. Lo anterior se aprecia en la Figura 38 y permite inferir que las personas con perfil inversor son menos propensas a desertar de la Cooperativa tal y como se había apreciado en los perfiles inversionistas.

**Figura 38.**

*Curva de supervivencia por nivel rango de promedio de saldos de productos en 2014*



Las demás variables categóricas que se tuvieron en cuenta en este proyecto como el estado civil, la inactividad de la cuenta de ahorros y la tenencia de personas a cargo no presentaron diferencias estadísticamente significativas respecto a las categorías en las que se clasificaron.

### **A manera de conclusión**

Según lo apreciado en las curvas de supervivencia de Kaplan-Meier el perfil de los asociados más propensos a permanecer en la Cooperativa después de 5 años de seguimiento es el siguiente: Mujeres, con edad superior a 35 años, habitantes en vivienda propia con una ocupación que se encuentre dentro de la categoría “OTRAS”, residentes en estrato 1 y 2, que su nivel educativo máximo sea la primaria. También que cuente con más de 10 años como asociada de la Cooperativa e ingresos inferiores a 1 smmlv y egresos entre 1 y 2 smmlv; sin ahorro, crédito y/o PAP vigentes al inicio del periodo de observación pero con CDAT vigente en el mismo periodo. Finalmente personas sin productos vigentes al momento de iniciar el seguimiento y que hayan hecho hasta 2 transacciones voluntarias durante el 2014, sin saldo promedio en los productos de ahorro, crédito y PAP pero con saldo en la variable de CDAT.

Por otra parte, los perfiles con mayor probabilidad de abandonar la Cooperativa, tomando como referencia el 25% al concluir el seguimiento son: Asociados entre los 18 y 35 años, con ocupación de “OPERARIOS”, con menos de 1 años de antigüedad como asociados de la entidad con crédito y/o PAP vigente al momento de iniciar el periodo de observación. Además con saldo promedio de ahorros durante el año anterior al comienzo del estudio de hasta 1 smmlv, y que hayan realizado entre 23 y 43 transacciones voluntarias en este mismo lapso.

**2.1.2.4.2 Modelo de regresión de Cox.** A continuación se presentan y analizan los resultados arrojados por la aplicación del modelo de regresión de Cox, por medio del cual se

buscaba identificar la influencia de las variables explicativas en la probabilidad de deserción de los asociados de la Cooperativa.

**Tabla 27.**

*Resultado general regresión de Cox*

Resumen de procesamiento de casos		N	Porcentaje
Casos disponibles en el análisis	Evento <sup>a</sup>	1269	21,0%
	Censurado	4760	79,0%
	Total	6029	100,0%
Casos eliminados	Casos con valores perdidos	0	0,0%
	Casos con tiempo negativo	0	0,0%
	Casos censurados antes del evento más cercano en un estrato	0	0,0%
	Total	0	0,0%
Total		6029	100,0%
a. Variable dependiente: Tiempo de permanencia en meses en el periodo de observación			

En la tabla 27 se exponen la totalidad de casos ingresados al modelo de regresión de Cox, aquí, se aprecia que de los 6.029 casos, el 21% de estos presentaron el evento de interés, es decir desertaron de la Cooperativa y el 79% restante continuaron siendo asociados al terminar el seguimiento.

**Tabla 28.**

*Prueba Ómnibus del modelo*

Paso	Logaritmo de la verosimilitud -2	Pruebas ómnibus de coeficientes de modelo <sup>P</sup>								
		Global (puntuación)			Cambiar respecto al paso anterior			Cambiar respecto al bloque anterior		
		Chi-cuadrado	df	Sig.	Chi-cuadrado	df	Sig.	Chi-cuadrado	df	Sig.
1 <sup>a</sup>	21494,291	344,267	1	,000	314,507	1	,000	314,507	1	,000
2 <sup>b</sup>	21319,699	519,845	3	,000	174,592	2	,000	489,099	3	,000
3 <sup>c</sup>	21261,708	566,905	4	,000	57,991	1	,000	547,090	4	,000
4 <sup>d</sup>	21222,701	597,180	6	,000	39,007	2	,000	586,097	6	,000
5 <sup>e</sup>	21200,500	631,421	9	,000	22,201	3	,000	608,298	9	,000
6 <sup>f</sup>	21178,622	660,909	13	,000	21,878	4	,000	630,175	13	,000
7 <sup>g</sup>	21161,496	681,756	16	,000	17,127	3	,001	647,302	16	,000
8 <sup>h</sup>	21155,559	690,235	17	,000	5,936	1	,015	653,238	17	,000
9 <sup>i</sup>	21147,219	693,800	19	,000	8,340	2	,015	661,578	19	,000
10 <sup>j</sup>	21141,240	702,309	20	,000	5,979	1	,014	667,557	20	,000
11 <sup>k</sup>	21131,480	705,117	23	,000	9,761	3	,021	677,318	23	,000
12 <sup>l</sup>	21120,988	712,181	24	,000	10,492	1	,001	687,810	24	,000
13 <sup>m</sup>	21121,811	711,656	23	,000	,823	1	,364	686,987	23	,000
14 <sup>n</sup>	21116,506	714,085	24	,000	5,306	1	,021	692,292	24	,000
15 <sup>o</sup>	21120,584	711,043	22	,000	4,078	2	,130	688,214	22	,000

La Tabla 28 muestra la prueba ómnibus de los coeficientes del modelo en el que se evidencia que después de 15 pasos se encontró el mejor de estos y apreciando que su bondad es estadísticamente significativa, aquí se recalca que cuanto más próximo a cero sea el valor del estadístico, más próxima a 1 será la verosimilitud y por ende mejor será el modelo.

Después de incluir cada una de las variables categóricas, tal y como se apreció en las curvas de supervivencia de Kaplan-Meier, se ajustó el modelo y se identificaron once (11) variables significativas con nivel de significancia del 5% de las veintidós (22) incluidas en total. En la tabla 28 se observa la codificación de cada una de ellas, existen tantas variables ficticias como categorías menos una tiene la variable en cuestión. Por lo que finalmente se cuenta con cincuenta y nueve

(59) categorías divididas en las catorce (11) variables ya mencionadas que el modelo de regresión de Cox expuso como significativas después de quince (15) pasos.

Es importante destacar que la primera categoría de cada una de las variables se tomó como la de referencia, y a partir de ella se interpretan los coeficientes de regresión. Por ejemplo en la variable que hace referencia al rango de edad, el correspondiente a los asociados de 18 a 35 años se asumió como el punto de referencia con respecto al cual se comparan las demás categorías, y para efectos de la interpretación de los coeficientes de las Tablas 29 y 30 entonces la categoría de 36 a 45 años pasar a ser la categoría número 1, la categoría de los 46 a 55 años se convierte en la categoría número 2 y la categoría de 56 a 65 años queda estipulada con el número 3, lo anterior se puede apreciar en los índices que se encuentran a la derecha de la frecuencia en la Tabla 29.

Tabla 29.

Codificación de variables categóricas

Codificaciones de variable categórica		Frecuencia	(1)	(2)	(3)	(4)
Rango de edad <sup>a</sup>	1=18 a 35 años	1648	0	0	0	0
	2=36 a 45 años	1207	1	0	0	0
	3=46 a 55 años	1696	0	1	0	0
	4=56 a 65 años	1398	0	0	1	0
Género <sup>b</sup>	1=Femenino	3540	0			
	2=Masculino	2489	1			
Estado civil <sup>b</sup>	1=Soltero	2369	0	0		
	2=Casado	2197	1	0		
	3=Otros	1463	0	1		
Tipo de vivienda <sup>b</sup>	1=Familiar	2634	0	0		
	2=Propia	2492	1	0		
	3=Alquiler	703	0	1		
Ocupación <sup>b</sup>	1=Independiente	2300	0	0	0	
	2=Operarios	1003	1	0	0	
	3=Oficinistas	569	0	1	0	
	4=Otras	2077	0	0	1	
Estrato socioeconómico <sup>b</sup>	1=1 y 2	2312	0	0		
	2=3	2231	1	0		
	3=4, 5 y 6	1486	0	1		
Nivel educativo <sup>b</sup>	1=Ninguno y Primaria	1401	0	0		
	2=Secundaria	2463	1	0		
	3=Técnica, Universitario y Posgrado	2165	0	1		
Rango de antigüedad <sup>b</sup>	1=Hasta 1 año	1679	0	0		
	2=>1 hasta 10 años	2316	1	0		
	3=>10 años	2034	0	1		
Rango de ingresos <sup>b</sup>	1=Hasta 1 smmlv	1145	0	0	0	
	2=>1 hasta 2 smmlv	1680	1	0	0	
	3=>2 hasta 5 smmlv	2005	0	1	0	
	4=>5 smmlv	1192	0	0	1	
Rango de egresos <sup>b</sup>	1=Hasta 1 smmlv	3518	0	0		
	2=>1 hasta 2 smmlv	1075	1	0		
	3=>2 smmlv	1436	0	1		
Personas a cargo <sup>b</sup>	0=NO	3529	0			
	1=SI	2505	1			
Crédito vigente a dic 2014 <sup>b</sup>	0=NO	4029	0			
	1=SI	2000	1			
Cuenta de ahorros vigente a dic 2014 <sup>b</sup>	0=NO	1064	0			
	1=SI	4965	1			
PAP vigente a dic 2014 <sup>b</sup>	0=NO	4897	0			
	1=SI	1032	1			
CDAT vigente a dic 2014 <sup>b</sup>	0=NO	4862	0			
	1=SI	1167	1			
Cuenta de ahorros inactiva en 2014 <sup>b</sup>	0=NO	4287	0			
	1=SI	1742	1			
Rango de número de productos <sup>b</sup>	1=Ninguno	638	0	0	0	
	2=1 producto	2306	1	0	0	
	3=2 productos	1946	0	1	0	
	4=> 3 productos	1139	0	0	1	
Rango de saldo promedio de ahorros en 2014 <sup>b</sup>	1=Sin saldo	1341	0	0		
	2=Hasta 1 smmlv	3305	1	0		
	3=>1 smmlv	1383	0	1		
Rango de saldo promedio de crédito en 2014 <sup>b</sup>	1=Sin saldo	3813	0	0		
	2=Hasta 5 smmlv	873	1	0		
	3=>5 smmlv	1343	0	1		
Saldo promedio de PAP en 2014 <sup>b</sup>	1=Sin saldo	4666	0			
	2=Con Saldo	1363	1			
Rango de saldo promedio de CDAT en 2014 <sup>b</sup>	1=Sin saldo	3896	0			
	2=Con Saldo	2033	1			
Rango de número de transacciones <sup>b</sup>	1=Hasta 2	1408	0	0	0	0
	2=3 hasta 9	1085	1	0	0	0
	3=10 hasta 22	1165	0	1	0	0
	4=23 hasta 43	1168	0	0	1	0
	5=> 44	1203	0	0	0	1

Una vez se tiene clara la codificación que se tiene en cuenta en el modelo podemos observar los resultados el modelo tal y como se expone en las tablas 30 y 31:

**Tabla 30.**

*Variables en la ecuación modelo de regresión de Cox (Parte 1)*

		Variables en la ecuación							
		B	SE	Wald	df	Sig.	Exp(B)	95,0% CI para Exp(B)	
								Inferior	Superior
Paso 15	Rango de edad			15,340	3	,002			
	Rango de edad(1)	-,185	,077	5,757	1	,016	,831	,715	,967
	Rango de edad(2)	-,271	,080	11,531	1	,001	,762	,652	,892
	Rango de edad(3)	-,292	,093	9,840	1	,002	,747	,623	,896
	Ocupación			16,632	3	,001			
	Ocupación(1)	,274	,076	13,171	1	,000	1,316	1,134	1,525
	Ocupación(2)	-,027	,105	,067	1	,796	,973	,792	1,196
	Ocupación(3)	,045	,074	,359	1	,549	1,046	,904	1,209
	Rango de antigüedad			64,049	2	,000			
	Rango de antigüedad(1)	-,339	,069	23,873	1	,000	,713	,622	,816
	Rango de antigüedad(2)	-,703	,090	61,461	1	,000	,495	,415	,590
	Crédito vigente a dic 2014	,803	,191	17,661	1	,000	2,232	1,535	3,245
	Cuenta de ahorros vigente a dic 2014	,714	,152	22,115	1	,000	2,042	1,517	2,750
	PAP vigente a dic 2014	,524	,111	22,116	1	,000	1,689	1,357	2,101
	Cuenta de ahorros inactiva en 2014	,170	,069	6,043	1	,014	1,185	1,035	1,357
	Rango de número de productos			32,015	3	,000			
	Rango de número de productos(1)	,018	,215	,007	1	,935	1,018	,668	1,551
	Rango de número de productos(2)	-,504	,275	3,373	1	,066	,604	,353	1,034
	Rango de número de productos(3)	-,978	,327	8,946	1	,003	,376	,198	,714

En el caso de la edad, para las personas entre 36 y 45 años se puede afirmar que tienen 1-0,831= 16,9% menor riesgo de desertar de la Cooperativa que los asociados entre 18 y 35 años, las personas entre 46 y 55 años tienen 1-0,762= 23,8% menor riesgo de desertar que aquellos que pertenecen a la categoría de referencia y las personas que tienen entre 56 y 65 años tienen menos

riesgo de la Cooperativa en un  $1-0,747= 25,3\%$ . Con respecto a los asociados que iniciaron el periodo de observación con edades entre los 18 y 35 años.

Al observar los niveles de significancia de las categorías de la ocupación comparados con la categoría de referencia, que en este caso es la categoría de “Independientes”, se aprecia que en el caso de las categorías “Oficinistas” y “Otras” son superiores a 0,10 lo cual implica que las diferencias observadas se deben al azar. Sin embargo, se puede afirmar que para los asociados con la ocupación de operarios el riesgo de desertar es 1,316 veces mayor comparado con la categoría de referencia.

En la antigüedad, la categoría de referencia es aquella que contiene a los asociados con menos de un año de antigüedad en la Cooperativa y se aprecia para los asociados con antigüedad  $>1$  hasta 10 años el riesgo de desertar es menor con respecto a los que llevan menos de 1 año en un  $1-0,713= 28,3\%$  y los asociados con más de 10 de años de antigüedad tienen menor riesgo de retirarse de la Cooperativa en un  $1-0,495= 50,5\%$ .

Con respecto a la tenencia de un crédito vigente al momento del inicio del periodo de observación, es decir diciembre de 2014, se aprecia que aquellas personas que tenían un préstamo a este corte tienen 2,232 veces el riesgo de desertar con respecto a quienes no contaban con esta condición. Similar a lo recientemente descrito pero esta vez con la tenencia de un producto de ahorro a diciembre de 2014, se puede afirmar que los que cuentan con producto de ahorro tienen 2,042 veces el riesgo de retirarse de la entidad que aquellos que no lo tienen. De igual forma las personas con por lo menos un PAP vigente al momento de iniciar el seguimiento son 68,9% más propensas a retirarse de la Cooperativa con respecto a quienes que no cumplen esta condición. Así mismo, aquellos asociados que tenían la cuenta de ahorro inactiva a diciembre de 2014, tienen

18,5% más probabilidad de retirarse de la entidad respecto a los que habían realizado movimiento en su cuenta de ahorros en los 6 meses previos y por tanto no tenían la cuenta inactiva.

En lo referente al número de productos, en donde la categoría de referencia es aquella que indica la no tenencia de productos al momento de iniciar el seguimiento, se aprecia que el nivel de significancia de la categoría de las personas que tenían un producto vigente en diciembre de 2014 es superior a 0,1 ( $p=0,935$ ), por lo que las diferencias entre esta y la categoría de referencia pueden deberse al azar. Por otra parte se aprecia que para los asociados con 2 productos vigentes al momento de iniciar el periodo de observación el riesgo de desertar es menor en un  $1-0,604= 39,6\%$  con respecto a los que no tenían productos vigentes y los asociados con 3 productos o más tienen menor riesgo de retirarse de la Cooperativa en un  $1-0,376= 62,4\%$ , en comparación con la no tenencia de productos.

**Tabla 31.**

*Variables en la ecuación modelo de regresión de Cox (Parte 2)*

		Variables en la ecuación							
		B	SE	Wald	df	Sig.	Exp(B)	95,0% CI para Exp(B)	
								Inferior	Superior
Paso 15	Rango de saldo promedio de crédito en 2014			10,719	2	,005			
	Rango de saldo promedio de crédito en 2014(1)	,427	,163	6,839	1	,009	1,533	1,113	2,111
	Rango de saldo promedio de crédito en 2014(2)	,252	,178	2,018	1	,155	1,287	,909	1,823
	Rango de saldo promedio de CDAT en 2014	-,232	,085	7,424	1	,006	,793	,671	,937
	Rango de número de transacciones			18,844	4	,001			
	Rango de número de transacciones(1)	,284	,121	5,493	1	,019	1,329	1,048	1,685
	Rango de número de transacciones(2)	,272	,126	4,640	1	,031	1,312	1,025	1,681
	Rango de número de transacciones(3)	,459	,126	13,371	1	,000	1,582	1,237	2,024
	Rango de número de transacciones(4)	,209	,134	2,449	1	,118	1,233	,949	1,602

En el caso del promedio del saldo de crédito durante el 2014, el rango de saldo > 5 smmlv muestra un valor de significancia superior a 0,10, por lo que cualquier diferencia observada entre estas categorías y la de referencia (Sin saldo) podría haberse presentado fortuitamente. Para el rango de hasta 5 smmlv se puede afirmar que tienen 1,533 veces el riesgo de desertar de la Cooperativa con respecto a la categoría de referencia.

En cuanto a los promedios del saldo de CDAT , se puede afirmar que el riesgo de los asociados que mantuvieron saldo en este producto en 2014 es  $1-0,793= 20,7\%$  menor con respecto a las personas que no cumplen esta condición; y para concluir en lo que se refiere al número de transacciones voluntarias a lo largo del 2014 que tiene como categoría de referencia hasta 2 transacciones, las personas que hicieron entre 3 y 9 transacciones a lo largo del 2014 tienen 1,329 veces el riesgo de desertar, para los asociados que hicieron entre 10 y 22 transacciones el riesgo es de 1,312 veces respecto a la categoría de referencia, aquellos que hicieron entre 23 y 43 transacciones en el 2014 tienen 1,582 veces el riesgo de abandonar la entidad y los que voluntariamente hicieron más de 43 transacciones tienen un nivel de significancia mayor a 0,10 por lo que las diferencias entre esta y la categoría de referencia pueden deberse a contingencias.

Por tanto, la función de riesgo, presentada en función de los exponenciales de las tablas 30 y 31, y usando la codificación de las categorías de la Tabla 29, queda de la siguiente manera:

$$\begin{aligned}
 h_t(X_{it}) = & h_0(t) * (0,831 \text{ rango\_edad}(2) + 0,762 \text{ rango\_edad}(3) + \\
 & 0,747 \text{ rango\_edad}(4) + 1,316 \text{ cargo}(2) + 0,713 \text{ rango\_antigüedad}(2) + \\
 & 0,495 \text{ rango\_antigüedad}(3) + 2,232 \text{ tenencia\_credito}(1) + 2,042 \text{ tenencia\_ahorro}(1) + \\
 & 1,689 \text{ tenencia\_pap}(1) + 1,185 \text{ inactividad\_ahorro}(1) + 0,604 \text{ rango\_producto}(3) + \\
 & 0,376 \text{ rango\_producto}(4) + 1,533 \text{ rango\_promedio\_credito}(2) +
 \end{aligned}$$

$$0,793 \text{ rango\_promedio\_cdat}(2) + 1,329 \text{ rango\_transacciones}(2) +$$

$$1,312 \text{ rango\_transacciones}(3) + 1,582 \text{ rango\_transacciones}(4)$$

Por lo tanto podemos observar que según el modelo de regresión de Cox, los perfiles con mayor probabilidad de desertar de la Cooperativa son: personas entre 18 y 35 años con ocupación de operarios que tengan hasta 1 año de antigüedad como asociados y que al momento de iniciar el periodo de observación no contaban con un CDAT vigente pero si con al menos 1 producto de crédito, ahorros y/o PAP. Son más propensas a retirarse aquellas sin productos vigentes al iniciar el seguimiento y con saldo promedio de crédito durante el 2014 hasta 5 smmlv, personas con saldo promedio en CDAT y que hayan realizado entre 3 y 43 transacciones voluntarias a lo largo del 2014.

Finalmente se valida la precisión del modelo comparando lo predicho por el modelo ajustado con los datos reales de la muestra. En la tabla 32 se aprecia que el modelo ajustado predice correctamente poco más del 80% de los casos en lo referente al retiro o no de la Cooperativa, por lo que se concluye que el modelo desarrollado satisface las expectativas planteadas.

**Tabla 32.**

*Predicción del modelo vs Datos reales en la muestra*

Resultados para el campo de resultado fugado		
Comparando \$C-fugado con fugado		
Correctos	4.920	81,61%
Erróneos	1.109	18,39%
Total	6.029	

### 3. Conclusiones

Este estudio propuso diseñar un modelo predictivo de deserción de asociados de una la Cooperativa de Ahorro y Crédito. Para ello se utilizó el análisis de supervivencia y así poder identificar las variables indicadoras de riesgo. En este capítulo se presentan las conclusiones del estudio realizado.

Por medio de las curvas de supervivencia de Kaplan-Meier y del modelo de regresión de Cox se pudieron identificar las características sociodemográficas estadísticamente significativas que ayudan a explicar la deserción de asociados que tenía la entidad a corte de diciembre de 2014 realizando un seguimiento de 5 años.

La muestra con la que se trabajó contaba con 6.029 personas naturales que pertenecían a la agencia principal de la Cooperativa, tenían entre 18 y 65 años, se retiraron voluntariamente, no se volvieron a asociar durante el seguimiento, actualizaron sus datos en el año previo al inicio del periodo de observación, pertenecientes a la mayoría en lo que respecta a comportamiento crediticio y que no se quejaron durante el 2014.

En cuanto a las curvas de supervivencia de Kaplan-Meier, sólo en las variables que indicaban el estado civil, la inactividad en las cuentas de ahorro y la tenencia de personas a cargo no se apreciaron diferencias estadísticamente significativas en sus curvas. Las variables que mostraron diferencias estadísticamente significativas son las siguientes: sexo, edad, tipo de vivienda, ocupación, estrato, nivel educativo, antigüedad como asociado, ingresos, egresos, tenencia de productos de crédito, ahorro, PAP y/o CDAT, cantidad de productos vigente en

diciembre de 2014, cantidad de transacciones voluntarias y promedio de saldo durante el año anterior a iniciar el seguimiento en los productos de crédito, ahorro, PAP y CDAT.

Por medio de las curvas de supervivencia de Kaplan-Meier se identificó que los asociados más propensos a desertar de la Cooperativa tenían entre 18 y 35 años y menos de 1 años de antigüedad, con ocupación en labores operativas y con crédito y/o PAP vigente al iniciar el seguimiento. Además con saldo promedio de ahorros durante el año anterior al comienzo del estudio de hasta 1 smmlv, y que hicieron entre 23 y 43 transacciones voluntarias a lo largo del 2014.

Al ejecutar el modelo de regresión de Cox se construyó la ecuación que identificó catorce (14) variables de diversa índole (sociodemográficas y transaccionales) influyentes en la probable deserción de los asociados de la Cooperativa

Las categorías de las variables que aumentan la probabilidad de desertar de la Cooperativa son:

- Ocupación como operarios
- Tener crédito vigente a diciembre de 2014
- Tener cuenta de ahorros vigente a diciembre de 2014
- Tener PAP vigente a diciembre de 2014
- Tener la cuenta de ahorros inactiva en diciembre de 2014
- Saldo promedio de crédito en 2014 hasta 5 smmlv
- Entre 3 y 43 transacciones voluntarias realizadas durante el 2014

Mientras que las categorías de las variables que disminuyen este riesgo son:

- Edad entre los 36 y los 65 años
- Más de 1 año de antigüedad como asociado

- 2 o más productos vigentes al iniciar el periodo de observación
- Contar con saldo promedio de CDAT a lo largo del 2014

Teniendo en cuenta que el estudio, al compararlo con la totalidad de asociados, se realizó con una muestra reducida de personas de la oficina principal de la Cooperativa, el trabajo a seguir es aplicar análisis de supervivencia a todos los asociados a nivel nacional de la entidad en busca de proveer un insumo vital para retener no sólo los asociados pertenecientes a la muestra sino a la población en general de la Cooperativa. Este debe ser un estudio localizado: se haría en forma independiente por ciudades y por oficinas dentro de las ciudades. Un resultado interesante de este estudio futuro sería identificar los riesgos comunes y los propios de cada ciudad y cada oficina, en caso de que los hubiera, claro.

### Referencias Bibliográficas

- Ayala, María & Borges, Rafael & Colmenares, Gerardo. (2007). Análisis de supervivencia aplicado a la banca comercial venezolana, 1996 - 2004. REVISTA COLOMBIANA DE ESTADÍSTICA. 30.
- Banco de Bogotá (Sin fecha). Lo que debes saber de las Transacciones. Recuperado de <https://www.bancodebogota.com/wps/portal/banco-de-bogota/bogota/atencion-al-cliente/canales-electronicos/transacciones>
- Carmona Suarez, Enrique J. (2014). Tutorial sobre Máquinas de Vectores Soporte (SVM). Recuperado de [http://www.ia.uned.es/~ejcarmona/publicaciones/\[2013-Carmona\]%20SVM.pdf](http://www.ia.uned.es/~ejcarmona/publicaciones/[2013-Carmona]%20SVM.pdf)
- Cobeña Teran, Brenda (2016). Modelo de predicción de deserción de clientes de tarjetas de crédito. Recuperado de <http://www.dspace.espol.edu.ec/xmlui/bitstream/handle/123456789/32341/D-CD102143.pdf?sequence=-1&isAllowed=y>
- D. Van Den Poel & B. Larivière, 2003. "Customer Attrition Analysis For Financial Services Using Proportional Hazard Models," Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium 03/164, Ghent University, Faculty of Economics and Business Administration.
- Herrera Rangel, Laura (2016). Diseño de un programa de retención de usuarios de tarjetas de crédito. Recuperado de <http://bdigital.unal.edu.co/56307/1/1128447884.2016.pdf>
- IBM (Sin fecha). Using Cox regression to model to customer time to churn. Recuperado de [https://www.ibm.com/support/knowledgecenter/es/SSLVMB\\_25.0.0/statistics\\_casestudies\\_project\\_ddita/common/examples/coxreg\\_telco\\_intro.html](https://www.ibm.com/support/knowledgecenter/es/SSLVMB_25.0.0/statistics_casestudies_project_ddita/common/examples/coxreg_telco_intro.html)
- Kleinbaum, David & Klein, Mitchel (2012). Survival Analysis A Self Learning Text. New York, USA: Springer

- Lara Rubio, Juan & Liébana Cabanillas, Francisco & Fiestas, Myriam. (2013). Lealtad bancaria y la medida del riesgo de abandono de los clientes de las entidades financieras. Recuperado de [https://www.researchgate.net/publication/270019816\\_Lealtad\\_bancaria\\_y\\_la\\_medida\\_del\\_riesgo\\_de\\_abandono\\_de\\_los\\_clientes\\_de\\_las\\_entidades\\_financieras](https://www.researchgate.net/publication/270019816_Lealtad_bancaria_y_la_medida_del_riesgo_de_abandono_de_los_clientes_de_las_entidades_financieras)
- Miranda, Jaime & Rey, Pablo & Weber, Richard (2005). Predicción de Fugas de Clientes para una Institución Financiera mediante Support Vector Machines. Recuperado de <http://www.dii.uchile.cl/ris/RISXIX/RISXIXpaper4.pdf>
- Perez Rodriguez, Marcela & Rivas Ruiz, Rodolfo & Palacios Cruz, Lino & Talavera, Juan. (2014). Del juicio clínico al modelo de riesgos proporcionales de Cox. Rev Med Inst Mex Seguro Soc Recuperado de [http://revistamedica.imss.gob.mx/es/system/files/recurso\\_diverso/rm-recop-caic-11-rm2014-4-16-xxii.pdf](http://revistamedica.imss.gob.mx/es/system/files/recurso_diverso/rm-recop-caic-11-rm2014-4-16-xxii.pdf)
- Ruiz Londoño, Daniel Felipe (2015). Aplicación de un modelo de predicción de fuga de clientes en la mesa de dinero de una entidad bancaria. Recuperado de <https://intellectum.unisabana.edu.co/handle/10818/21165>