

**CARACTERIZACIÓN Y PREDICCIÓN DE LA SINIESTRALIDAD EN EL  
SEGURO DE VIDA DEUDORES DE UNA COOPERATIVA DE AHORRO Y  
CRÉDITO**

**LEIDY MAGALLY RUÍZ MACHUCA**



**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE CIENCIAS  
ESCUELA DE MATEMÁTICAS  
ESPECIALIZACIÓN EN ESTADÍSTICA  
BUCARAMANGA**

**2016**

**CARACTERIZACIÓN Y PREDICCIÓN DE LA SINIESTRALIDAD EN EL  
SEGURO DE VIDA DEUDORES DE UNA COOPERATIVA DE AHORRO Y  
CRÉDITO**

**LEIDY MAGALLY RUÍZ MACHUCA**

**Trabajo de Tesis para optar el grado académico de Especialista en  
Estadística**

**Director**

**Dr. GABRIEL YÁÑEZ CANAL**

**Matemático**



**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE CIENCIAS  
ESCUELA DE MATEMÁTICAS  
ESPECIALIZACIÓN EN ESTADÍSTICA  
BUCARAMANGA**

**2016**

## **AGRADECIMIENTOS**

*En primera instancia quiero agradecer a Dios porque me dio la fuerza, la salud y la sabiduría para sacar adelante este proyecto.*

*Agradezco a mi familia y a mi novio por estar a mi lado incondicionalmente, por esa voz de aliento que me brindaron siempre.*

*A mis amigos y compañeros de trabajo, en especial a Luddy por su apoyo incondicional.*

*Un agradecimiento especial a mi Director de proyecto el profesor Gabriel Yañez Canal, por su dedicación y apoyo constante.*

*A Financiera Comultrasan, por esta oportunidad para demostrar mis capacidades y crecer intelectualmente.*

*En general a todos los que creyeron en mí y me dieron fuerzas siempre para seguir adelante.*

*Hoy les doy gracias y les presento mi trabajo final, fruto de la dedicación y el esfuerzo constante.*

## TABLA DE CONTENIDO

	Pág.
INTRODUCCIÓN .....	13
1. JUSTIFICACIÓN.....	15
<b>2.1. OBJETIVO GENERAL .....</b>	<b>16</b>
<b>2.2. OBJETIVOS ESPECÍFICOS .....</b>	<b>16</b>
3. ANTECEDENTES.....	17
4. MARCO TEÓRICO .....	21
<b>4.1. SERIES DE TIEMPO .....</b>	<b>21</b>
<b>4.2. ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLE (ACM).....</b>	<b>23</b>
5. ANÁLISIS DE DATOS E INTERPRETACIÓN DE RESULTADOS.....	25
<b>5.1. ANÁLISIS DESCRIPTIVOS.....</b>	<b>25</b>
5.1.1. Análisis descriptivos de las variables cualitativas .....	26
5.1.1.1 Variable: <i>ESTADO</i> .....	26
5.1.1.2 Variable: <i>CAUSA</i> .....	27
5.1.1.3 Variable: <i>OFICINA</i> .....	29
5.1.1.4 Variable: <i>OCUPACIÓN</i> .....	31
5.1.2. Análisis descriptivos de las variables cuantitativas .....	32
5.1.2.1 Variable: RECLAMACIONES.....	32
5.1.2.2 Variable: <i>CUANTÍA</i> .....	37
5.1.2.3 Variable: <i>EDAD</i> .....	40

<b>5.2. ANÁLISIS DE CORRESPONDENCIA MÚLTIPLE (ACM)</b> .....	<b>42</b>
<b>5.3. ANÁLISIS DE SERIES DE TIEMPO</b> .....	<b>47</b>
5.3.1 Análisis de la serie temporal .....	47
5.3.2. Comparación de modelos .....	49
5.3.3. Pronósticos del modelo.....	54
<b>6. CONCLUSIONES</b> .....	<b>57</b>
<b>7. RECOMENDACIONES</b> .....	<b>59</b>
<b>BIBLIOGRAFÍA</b> .....	<b>60</b>

## LISTA DE TABLAS

	<b>Pág.</b>
Tabla 1. Porcentaje de siniestros que superan un valor determinado. ....	19
Tabla 2. Frecuencias esperadas de siniestros por valores. ....	19
Tabla 3. Tabla de frecuencias y diagrama de barras para la variable ESTADO. ....	27
Tabla 4. Tabla de frecuencias y diagrama de barras para la variable CAUSA. ....	28
Tabla 5. Ubicación por departamentos de las oficinas de la cooperativa en Colombia.....	29
Tabla 6. Tabla de frecuencias y diagrama de barras para la variable OCUPACIÓN. ....	31
Tabla 7. Estadísticos descriptivos y diagramas de caja para la variable RECLAMACIONES.....	33
Tabla 8. Perfil de medias por año para la variable RECLAMACIONES. ....	35
Tabla 9. Perfil de medias por mes para la variable RECLAMACIONES. ....	36
Tabla 10. Estadísticos descriptivos y diagrama de caja para la variable CUANTÍA. ....	37
Tabla 11. Perfil de medias por mes para la variable CUANTÍA. ....	39
Tabla 12. Perfil de medias por año para la variable CUANTÍA. ....	40
Tabla 13. Estadísticos descriptivos, percentiles y diagrama de caja para la variable EDAD.....	41

Tabla 14. Categorización de las variables de estudio: zona, edad, cuantía, estado, causa y ocupación. ....	43
Tabla 15. Contribuciones y cosenos cuadrados de las variables activas. ....	44
Tabla 16. Estadísticos de error para el periodo de estimación y validación de los modelos de Tendencia Lineal y Arima. ....	51
Tabla 17. Resumen del modelo de Tendencia Lineal. ....	52
Tabla 18. Pronósticos del modelo de Tendencia Lineal. ....	54
Tabla 19. Pronostico para 12 periodos (ene 2016 - dic 2016). ....	56

## LISTA DE FIGURAS

	<b>Pág.</b>
Figura 1. Diagrama de barras para la variable RECLAMACIONES según oficina.	34
Figura 2. Primer plano factorial de las variables activas: estado, causa, cuantía, edad, oficina y ocupación (inercia 24,04%).	46
Figura 3. Gráfico temporal de la serie CUANTÍA.	47
Figura 4. Diagrama de cajas para CUANTÍA por año.	48
Figura 5 . Función de autocorrelación para los residuos del modelo de tendencia lineal (FAC).	52
Figura 6 . Función de autocorrelación parcial para los residuos del modelo de tendencia lineal (FACP).	53

## RESUMEN

### TÍTULO:

CARACTERIZACIÓN Y PREDICCIÓN DE LA SINIESTRALIDAD EN EL SEGURO DE VIDA DEUDORES DE UNA COOPERATIVA DE AHORRO Y CRÉDITO\*.

### AUTORES:

Leidy Magally Ruíz Machuca \*\*

### PALABRAS CLAVES:

Caracterización

Predicción

Siniestralidad

Serie de tiempo

Análisis de correspondencias

### DESCRIPCION:

En este estudio se realizaron análisis estadísticos alrededor de los datos de siniestralidad de la póliza vida deudores de una entidad financiera obtenidos en los últimos cinco años.

Esta póliza representa gran importancia para una entidad financiera, dado que es la que protege y resguarda los saldos de los créditos otorgados a los clientes, que en caso de no poder seguir respondiendo por su deuda por presentar una Incapacidad total y permanente o pérdida de la capacidad laboral, o por fallecer durante la vigencia del crédito, es esta póliza la que se encarga de pagar en dichos casos, siempre y cuando se cumplan con las condiciones de dicho contrato.

En el caso de que la aseguradora con la que se tiene contratada dicha póliza no dé una respuesta positiva a este pago, es la entidad financiera quien tiene que hacerse cargo de este valor. Es por esto que se requiere analizar el comportamiento que ha venido presentado y las características ya sean demográficas o laborales, que tienen los clientes y que influyen en el momento de presentar estas reclamaciones, con el ánimo de reducir este índice, y así mismo lograr estar dentro de los límites de control establecidos por las compañías aseguradoras.

En este trabajo se desea aplicar los métodos de series de tiempo y análisis de correspondencia, para lograr así una predicción de las futuras reclamaciones por este concepto y la caracterización de estos clientes para la implementación de acciones preventivas ante estos hechos.

---

\* Trabajo de grado.

\*\* Facultad de Ciencias. Especialización en Estadística. Dr. Gabriel Yáñez Canal

## SUMMARY

### TITLE:

\*CHARACTERIZATION AND PREDICTION OF ACCIDENT IN LIFE INSURANCE BEBTORS A COOPERATIVE SAVINGS AND CREDIT\*

### AUTHORS:

Leidy Magally Ruiz Machuca\*\*

### KEY WORDS:

Characterization  
Prediction  
Accident  
Time Series  
Correspondence analysis

### DESCRIPTION:

This work contain the statistical analyzes of data bases about the insurance claims of a financial company since five years ago, specifically from 2011 to 2015.

The policy of the credits represents the most important policy for a financial company, because it protects and preserves the balance of loans granted to customers, in case that they cannot continue to pay their debt because present a total and permanent disability or loss labor capacity, or death during the term of the loan, this policy is responsible for paying in such cases, as long as they meet the conditions of the contract.

In the event that the insurer who has contracted this policy does not give a positive response to this payment, the financial company has to take care of this value, because of it is required to analyze the behavior that has been presented and the demographics and occupation characteristics of customers that influencing these claims, with the aim of reducing this index, and likewise be achieved within the control limits established by the insurance companies.

In this work, was apply the methods of time series and correspondence analysis to achieve a prediction of future claims for this concept and characterization of these customers to implement preventive actions to these events.

---

\* Grade work.

\*\* Science Faculty. Specialization in Statistics. Dr. Gabriel Yáñez Canal.

## INTRODUCCIÓN

En los años recientes se ha observado un crecimiento importante en el ámbito de los seguros, debido a la necesidad de resguardarse ante riesgos naturales y económicos preservando así los bienes de las personas, empresas y sociedad en general.

A través de los años se han incrementado los distintos ramos de seguros; los patrimoniales, donde se encuentran los de vehículos, incendio y terremoto y responsabilidad civil, entre otros, y los seguros de personas que se subdividen en accidentes personales, vida individual, vida colectiva entre los que se encuentra el seguro de “Vida Grupo Deudores”.

Para las entidades financieras, este es uno de los más importantes, que consiste en una prima o valor incluido en la cuota de los créditos que se cobra al cliente para que en caso de incapacidad total y permanente o de fallecimiento, el saldo insoluto del crédito sea cubierto por la aseguradora.

Asociado con este seguro de deudores, es de vital importancia para la Cooperativa poder predecir el nivel de siniestralidad en un rango de tiempo relativamente amplio para poder ajustar las primas que se deben cobrar al cliente y no tenga la Cooperativa que cubrir el faltante que se pueda ocasionar por el cobro de la Compañía de Seguros. De otro lado, existe la inquietud de saber qué características posee un cliente que por incapacidad o fallecimiento no puede cumplir con el crédito realizado.

Es en este sentido que las preguntas que nos hemos planteado para resolver en este trabajo son las siguientes:

- ¿Cuál es el nivel de siniestralidad que se puede esperar en la Cooperativa en un periodo de seis meses en el seguro de Vida Deudores?

— ¿Qué características poseen las personas que por incapacidad o fallecimiento no terminan de pagar sus créditos?

Con base en los datos históricos que posee la Cooperativa y de la implementación de un modelo de series de tiempo se dará respuesta a la primera pregunta.

Para responder la segunda pregunta se realizará un análisis de correspondencia que permita estudiar las asociaciones que tienen un conjunto de variables demográficas y laborales con el monto del crédito y la existencia o no de incapacidad o fallecimiento.

## 1. JUSTIFICACIÓN

Uno de los problemas que se le presentan a una entidad financiera al momento de fijar los términos del contrato de renovación de la póliza Vida Grupo Deudores es el riesgo de que la aseguradora incremente la tasa anual cobrada por este concepto.

Cabe resaltar que en condiciones normales la aseguradora mantiene estable el pago de las primas cuando la siniestralidad es igual o inferior al 60%, cuando se sobrepasa este indicador las tarifas de la aseguradora empiezan a subir, poniendo en “aprietos” a la entidad financiera que tiene ya fijada la prima de seguro con los créditos vigentes y desembolsados en periodos anteriores.

El presente trabajo tiene como finalidad, realizar el análisis de la siniestralidad dado que permite obtener información sumamente valiosa no solo para estimar el riesgo al que se puede estar expuesto, sino para obtener el perfil de su siniestralidad en el que tienen en cuenta los parámetros cambiantes que influyen directamente en el incremento y gravedad de los siniestros. Lo anterior con el ánimo de reducir este índice, y así mismo lograr estar dentro de los límites de control establecidos por las compañías aseguradoras.

## **2. OBJETIVOS**

### **2.1. OBJETIVO GENERAL**

Caracterizar y Predecir la siniestralidad en el seguro de vida deudores de una Cooperativa de Ahorro y Crédito.

### **2.2. OBJETIVOS ESPECÍFICOS**

- Recopilar información histórica de forma mensual de los últimos 5 años sobre la presentación de reclamos por muerte o invalidez total y permanente de los asociados de una entidad de ahorro y crédito.
- Ajustar un modelo de series de tiempo a los datos históricos de la siniestralidad por invalidez o muerte de los seguros de vida grupo y realizar predicciones para los próximos 6 meses.
- Realizar un análisis de correspondencia de las variables demográficas y laborales de los clientes de la Cooperativa que permita caracterizar a los clientes que fallecen o se incapacitan antes de terminar de pagar sus créditos.

### 3. ANTECEDENTES

Se observa que el tema a tratar en el presente trabajo de grado ha sido poco aplicado y así mismo presenta carencia de recursos e investigaciones, lo que permite ver la importancia de trabajar este tema y desarrollar la metodología que se quiere implementar.

El análisis de la siniestralidad es de suma importancia para valorar el riesgo al que se está expuesto, así mismo para la toma de decisiones en cuanto tasas, coberturas y exclusiones que se puedan negociar dependiendo el comportamiento de este concepto en las distintos ramos de seguros.

Para apoyar este concepto se mencionan los siguientes análisis y metodologías similares aplicadas a la siniestralidad de la póliza de vehículos, realizada por Alejandro Román y a la póliza de Rotura de Maquinaria, realizada por Carlos Ríos Purón de la Gerencia de Riesgos del Grupo Endesa.

En la Tesis realizada por Alejandro Román Vásquez en Mayo de 2012, se analizan los datos del robo de vehículos asegurados en la República mexicana a través de los indicadores relativos regionalizados de prima de riesgo y proporción de robo.

Para el análisis de estos indicadores el autor Alejandro Román empleó series temporales mensuales de cada una de las regiones durante el periodo del 2008 al 2010 para generar así pronósticos a un año que le permitieran estimar y proyectar cuál sería el comportamiento a mediano plazo, para de esta forma poder cuantificar el riesgo.

Para realizar la clasificación usó herramientas estadísticas multivariadas de conglomerados, como el Agrupamiento Jerárquico de Aglomeración y la aplicación del modelo paramétrico ARIMA, dado que permite modelar comportamientos

estacionarios y no estacionarios. Así mismo emplearon la metodología propuesta por Box y Jenkins<sup>1</sup> para la aplicación de los modelos ARIMA.

Con la tesis anteriormente expuesta se observó que con su implementación y desarrollo se lograron conocer las necesidades y problemas del sector asegurador, logrando así proponer modelos de series de tiempo de acuerdo a esas necesidades. Un resultado importante logrado con esta metodología fue la generación de una regionalización de la república mexicana en zonas de comportamiento homogéneo de robo de los autos asegurados. Adicionalmente se logró la estimación de un modelo ARIMA ajustado a los indicadores de prima de riesgo y proporción de robo, lo que les permitirá seguir percibiendo su comportamiento a mediano plazo gracias a los pronósticos generados.

Adicionalmente se encuentra el método seguido por Carlos Ríos Purón de la Gerencia de Riesgos del Grupo Endesa (Empresa Nacional de electricidad de España) para estimar el número de siniestros esperados por averías que puedan presentar sus maquinarias, con el fin de poder valorar el riesgo al que la empresa pueda estar expuesta.

Para lo anterior implementan un modelo que les permita obtener el perfil de la siniestralidad y los parámetros que con el tiempo influyen directamente en el número y en la gravedad de los siniestros que se les presentan. Para este análisis tomaron la base de datos histórica de la siniestralidad que cuenta con 402 registros de siniestros ocurridos en un periodo de 14 años. Tomando como variables principales el valor de las pérdidas (importe de los daños a consecuencia de un siniestro) y la variable fecha de siniestro.

De esta manera obtienen el porcentaje de siniestros esperados por encima de un cierto valor o la probabilidad de que un siniestro supere una cifra determinada (Ver Tabla 1).

---

<sup>1</sup> Metodología destinada a identificar, estimar y diagnosticar modelos dinámicos de series temporales.

**Tabla 1. Porcentaje de siniestros que superan un valor determinado.**

<b>Valor de los daños ( en millones de pesetas)</b>	<b>Porcentaje de siniestros que superan el valor</b>
200	2,63
100	5,26
50	11,76
20	28,44
10	41,18
5	55,88

*Fuente: Tomado de artículo "Estimación de siniestralidad" Grupo Endesa, Pág. 12.*

Una vez registrados estadística y gráficamente los siniestros históricos, proceden a estimar el número de siniestros esperados para el año próximo, para esto tienen los siniestros ocurridos anualmente en los últimos 14 años.

De estos datos pueden concluir que existen dos conceptos que influyen en gran medida el número de siniestros: la capacidad de producción instalada y la producción anual. Así mismo obtienen la frecuencia esperada de siniestros en función de su magnitud:

**Tabla 2. Frecuencias esperadas de siniestros por valores.**

<b>Valor ( en millones de pesetas)</b>	<b>Frecuencia</b>
Mayor de 500	1 cada 11 años
Mayor de 100	1 al año
Entre 51 y 100	3 al año
Entre 20 y 49	5 al año
Entre 10 y 19	4 al año
Entre 5 y 9	5 al año
Menores de 5	14 al año

*Fuente: Tomado de artículo "Estimación de siniestralidad" Grupo Endesa, Pág. 16.*

Se observa que la siniestralidad se ubica en dos rangos, en la ocurrencia de siniestros con valores inferiores a los 50 millones de pesetas y los superiores a este.

Con relación a los siniestros de valores inferiores a 50 millones de pesetas presentan una probabilidad de ocurrencia de más de dos por año, lo que consideran estadísticamente suficiente, siendo una siniestralidad fija a lo largo de los años.

Para concluir se puede ver que los resultados obtenidos le permiten al Grupo Endesa, establecer los límites de retención propia en función de los diferentes factores y estimar la bondad de las primas determinando la parte destinada al siniestro<sup>2</sup>.

---

<sup>2</sup> Estimación de la siniestralidad esperada mediante el análisis de la siniestralidad histórica-Grupo Endesa-  
[https://www.fundacionmapfre.org/documentacion/.../i18n/catalogo\\_imagenes/grupo.cmd?path=1018790](https://www.fundacionmapfre.org/documentacion/.../i18n/catalogo_imagenes/grupo.cmd?path=1018790)

## **4. MARCO TEÓRICO**

En este capítulo se expone el marco teórico concerniente a la técnica estadística de Series de Tiempo y del Análisis de Correspondencia Múltiple (ACM) utilizadas en el desarrollo de este trabajo.

### **4.1. SERIES DE TIEMPO**

Una Serie de Tiempo son observaciones que se hacen de una variable específica de la que se recopilan valores a lo largo del tiempo, ya sean diarios, semanales, mensuales, semestrales, anuales, entre otros.

Por medio de este método se puede analizar el comportamiento de los datos históricos así como realizar predicciones futuras de la variable observada. Estas series pueden tener distintos comportamientos, una de ellas es cuando oscila alrededor de un nivel constante, la cual se puede llamar serie estacionaria o estable, cuando por lo contrario no se mantiene en un mismo nivel se dice que es una serie no estacionaria. Así mismo se tienen las series estacionales que se pueden diferenciar por presentar un nivel fijo o variable en el tiempo, o por tener un comportamiento que se repite a lo largo del mismo.

Al aplicar esta metodología es de gran ayuda y una herramienta fundamental la realización del gráfico de la serie trabajada, dado que permite identificar los distintos componentes que estas presentan. Uno de ellos es la Tendencia (cambios en la media en periodos largos), efectos estacionales (periodicidades), efectos cíclicos (cambios fácilmente predecibles), fluctuaciones irregulares (cambios en la varianza) y fluctuaciones aleatorias.

En este trabajo se va a construir un modelo que permita representar la evolución de un serie de tiempo y de igual forma generar predicciones de su comportamiento en el futuro. Será un modelo univariante, dado que solo se basará en los históricos de la propia serie. La predicción que se va a realizar con este modelo se centra en la hipótesis de que las predicciones serán similares a las históricas y útiles para la previsión a corto plazo.

En las series temporales se observan dos fenómenos. Los primeros son los fenómenos estacionarios, son los que toman valores estables en el tiempo alrededor de un nivel constante, sin observarse una tendencia creciente o decreciente a largo plazo. Los segundo son lo no estacionarios, que muestran estacionalidad, tendencia a crecer o decrecer, entre otros. Es importante recalcar que el clasificar una serie como estacionaria o no, puede depender del periodo de observación, dado que si es un periodo corto se puede mostrar estable, y no estacionaria si es un periodo mayor.

Es importante tener presente que al tener una serie no estacionaria se debe modelar para lograr así su comportamiento estacionario. Por lo tanto se requiere remover o quitar los efectos de tendencias, periodicidades y ciclos. Por lo que se deben calcular los logaritmos de esta serie, y aplicar las respectivas diferencias regulares tantas veces como sea necesario.

Una vez se tenga que la serie es estacionaria, se debe verificar con los gráficos de correlación *fas* (correlación simple) y *fap* (correlación parcial), si se cumple dicha condición. Si al obtener el correlograma se observa dependencia estacional en la serie, se deben aplicar diferencias estacionales tantas veces como se considere necesario.

Al lograr que la serie inicial cumpla con las condiciones anteriormente mencionadas, se procede a generar los diferentes modelos y lograr el que mejor se ajuste para la predicción deseada. Es importante tener presente los siguientes aspectos, al momento de realizar la validación de los resultados obtenidos en los

diferentes modelos: los coeficientes han de ser significativamente distintos de cero, las raíces han de ser mayores que 1 (estacionariedad), los residuos deben ser ruido blanco, ser estacionarios en media y varianza y proceder de una población normal.

Lo anteriormente expuesto encierra el procedimiento que se realiza al aplicar la metodología escogida en este trabajo, el método de Box-Jenkins<sup>3</sup>, el cual permite encontrar el mejor ajuste de la serie a fin de lograr que el pronóstico sea lo más acertado posible.

#### **4.2. ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLE (ACM)**

Para la caracterización de los clientes se implementará el método de análisis de correspondencia<sup>4</sup>. Esta técnica consiste en representar tablas de contingencia donde se agrupan las frecuencias de dos o más variables cualitativas, con el fin de resumir una cantidad de datos en un número reducido de dimensiones, con la menor pérdida de información posible, aplicado a variables categóricas u ordinales.

Cabe resaltar el concepto de tablas de contingencia, es un conjunto de números positivos dispuestos en una matriz, en el que el número de cada casilla representa la frecuencia absoluta observada para la combinación de las variables.

Existen distintos enfoques del análisis de correspondencia; el análisis de correspondencia simple y de correspondencia múltiple.

Para este caso en particular se aplicará el análisis de correspondencia múltiple o ACM, dado que se trabajará con más de dos variables. Para desarrollar esta técnica se deben crear diferentes categorías para cada una de las variables, lo

---

<sup>3</sup> Técnica propuesta por P. Box y M. Jenkins en 1976.

<sup>4</sup> Técnica desarrollada por Jean-Paul Benzécri.

que permitirá analizar a través de los mapas perceptuales las relaciones entre cada una de las categorías.

El Análisis de Correspondencia Múltiple presenta unos usos para las distintas áreas en las que se aplica. Uno de ellos es, profundizar el análisis de la asociación entre variables cualitativas (nominales u ordinales), analizar la estructura de la relación entre categorías de variables, identificar dimensiones que representen esquemas conceptuales de análisis y corroborar dependencia entre dos variables cualitativas.

Se debe cumplir con unas condiciones fundamentales para poder aplicar el método de Análisis de Correspondencia Múltiple; la condición principal es tener 2 o más variables para trabajar, estas variables deben ser nominales y ordinales. Si se tienen variables cuantitativas, estas se deben recodificar o categorizar para que se conviertan a ordinales. Y por último que se cumpla con la condición de que exista relación entre las variables, si no hay relación entre ellas no tiene sentido un aplicar un análisis de correspondencia.

Habiendo cumplido con estas condiciones se podrá realizar la representación gráfica de las variables relacionadas. Esta representación se realiza por medio de mapas (representaciones bidimensionales, simétricas o asimétricas), de las que se podrán sacar conclusiones con respecto a la cercanía de los puntos (variables) a cada eje y después con los puntos entre si dentro de la respectiva escala del eje.

Con la aplicación de este método se busca observar las relaciones entre cada variable trabajada, logrando así la interpretación de los datos de una forma más detallada, definiendo la importancia de cada variable en el estudio realizado.

## **5. ANÁLISIS DE DATOS E INTERPRETACIÓN DE RESULTADOS**

La exposición de los resultados de esta investigación se organiza en tres grandes secciones: en la primera parte se muestran los resultados obtenidos con los análisis descriptivos de las variables de estudio haciendo uso de tablas de frecuencias, diagramas de barras, diagramas de cajas, histogramas y los principales estadísticos descriptivos de tendencia central y dispersión. En la segunda parte se realiza el Análisis de Correspondencia Múltiple con el propósito de observar cómo correlacionan las distintas modalidades de las variables categóricas y en la tercera parte, se presentan los resultados de los análisis estadísticos de Series de Tiempo con el propósito de identificar un modelo que permita realizar predicciones para los próximos 6 meses.

### **5.1. ANÁLISIS DESCRIPTIVOS**

Para este estudio se han considerado las variables cualitativas y cuantitativas manejadas en la base de datos de los Siniestros de la Póliza Vida Deudores de una entidad cooperativa recopiladas desde enero de 2011 a diciembre de 2015.

Entre las variables cualitativas están:

- ESTADO: Se relaciona al estado en que se encuentra la reclamación del asociado: pagado, pendiente, objetado, prescrito.
- CAUSA: Hace referencia al motivo de la reclamación que realiza el asociado: Incapacidad Total y Permanente (I.T.P.) y fallecimiento (Muerte) del deudor.
- OFICINA: Corresponde a la ubicación geográfica de la sede de la cooperativa en donde es asociado el reclamante y en la cual se encuentra radicado el crédito.

- OCUPACIÓN: Se relaciona con la labor principal en la que se desempeña el asociado reclamante y que fue reportada al inicio de la relación contractual: trabajador independiente, pensionado, operario (transportadores, fuerza militar, etc.) y oficinista (empleados, docentes, etc.).

Las variables cuantitativas son:

- CUANTÍA: Corresponde al monto de dinero insoluto reportado a la aseguradora para reclamo (saldo de capital, más intereses corrientes, más intereses de mora, más honorarios jurídicos) de la deuda a la fecha del siniestro.
- RECLAMACIONES: Hace referencia al número de reclamos presentados por los asociados por periodo calendario de 30 días (mensual).
- EDAD: Se refiere a la edad en años cumplidos del asociado al momento de presentar la reclamación.

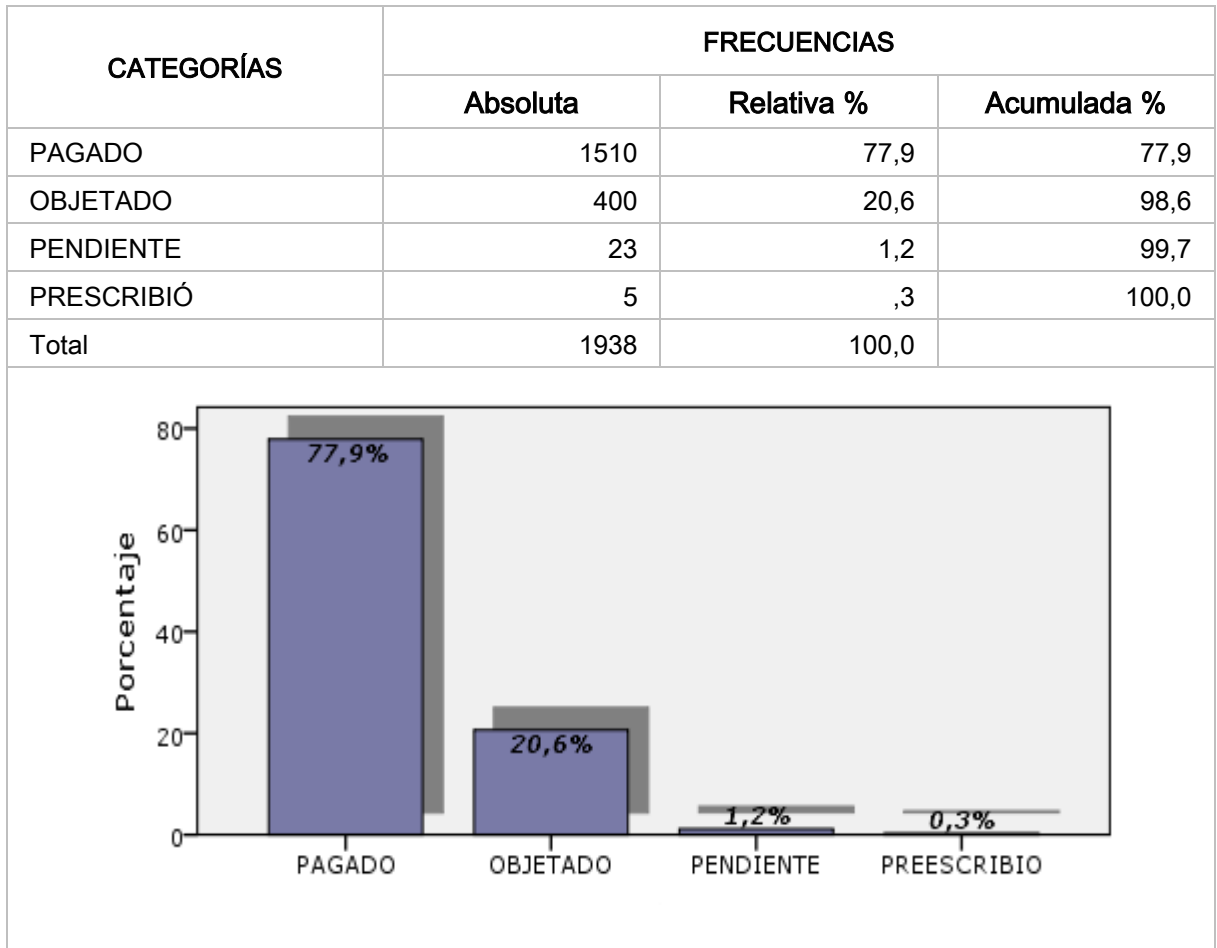
### **5.1.1. Análisis descriptivos de las variables cualitativas**

#### **5.1.1.1 Variable: *ESTADO***

En la Tabla 3 se presenta la tabla de frecuencias para la variable estado y en la parte inferior el correspondiente diagrama de barras.

Se evidencia en el diagrama que un poco más de las tres cuartas partes (77,9%; 1510) de los reclamos por siniestros en el quinquenio de estudio fueron pagados, una quinta parte (20,6%; 400) se objetaron, mientras que el restante 1,4% (28) o están pendientes (1,2%) o prescribieron.

**Tabla 3. Tabla de frecuencias y diagrama de barras para la variable ESTADO.**

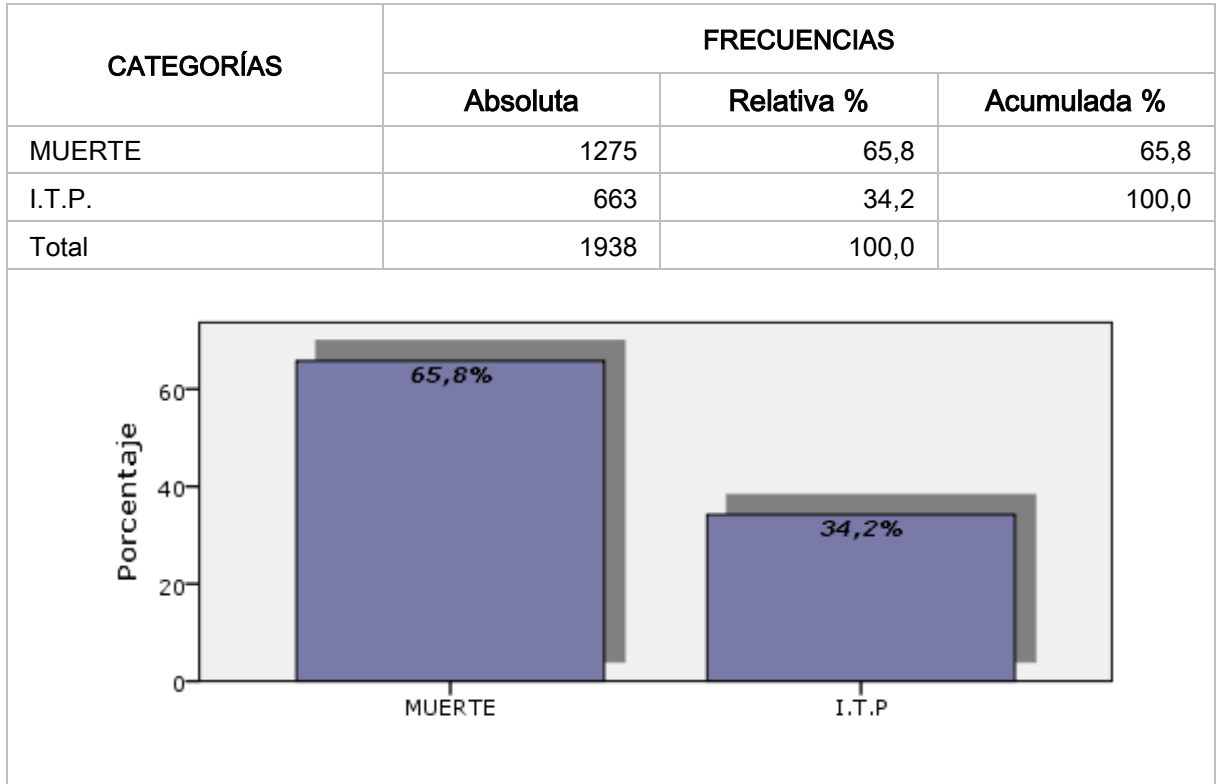


*Fuente: Autor*

### **5.1.1.2 Variable: CAUSA**

En la Tabla 4 se puede apreciar la distribución de las frecuencias en las categorías de la variable causa y en la parte inferior, el correspondiente diagrama de barras.

**Tabla 4. Tabla de frecuencias y diagrama de barras para la variable CAUSA.**



*Fuente: Autor*

La mayoría de las reclamaciones de los asociados (65,8%; 1275) tiene como causa la reclamación por el fallecimiento del deudor, el restante 34,2% (663) corresponde a las reclamaciones por Invalidez o Incapacidad Total y Permanente del deudor (I.T.P.), es decir, por pérdida de la capacidad laboral por enfermedad o accidente.

### 5.1.1.3 Variable: *OFICINA*

La Cooperativa cuenta con 51 oficinas ubicadas en seis departamentos de la zona centro oriental de Colombia.

En la Tabla 5 se aprecian los departamentos y los municipios donde se encuentran ubicadas las oficinas en el país.

**Tabla 5. Ubicación por departamentos de las oficinas de la cooperativa en Colombia.**

DEPARTAMENTO	OFICINA
CUNDINAMARCA	Bogotá (Kennedy)
	Bogotá (Avenida 68)
	Bogotá (calle 73)
BOYACÁ	Tunja
	Tunja Norte
	Duitama
	Sogamoso
ATLÁNTICO	Barranquilla (centro)
	Barranquilla (Prado)
CESAR	Bosconia
	Valledupar
	Codazzi
	Curumaní
	San Alberto
	Pelaya
	Aguachica
	San Martín
SANTANDER	El Playón
	Rionegro
	Calle 35
	Cabecera
	Cañaveral

	Carrera 11
	Centenario
	Kennedy
	La Cumbre
	Plaza Satélite
	San Francisco
	Girón (Poblado)
	Lebrija
	Barrancabermeja (comercio)
	Barrancabermeja (Nororient)
	Barrancabermeja (Torcoroma)
	Puerto Wilches
	Sabana de Torres
	San Rafael
	Floridablanca
	Piedecuesta
	San gil
	Socorro
	Barbosa
	Cimitarra
	Málaga
	Puente Nacional
	Zapatoca
	San Vicente de Chucuri
	Carmen de chucuri
<b>NORTE DE SANTANDER</b>	Cúcuta
	Pamplona
	Ocaña

*Fuente: Autor*

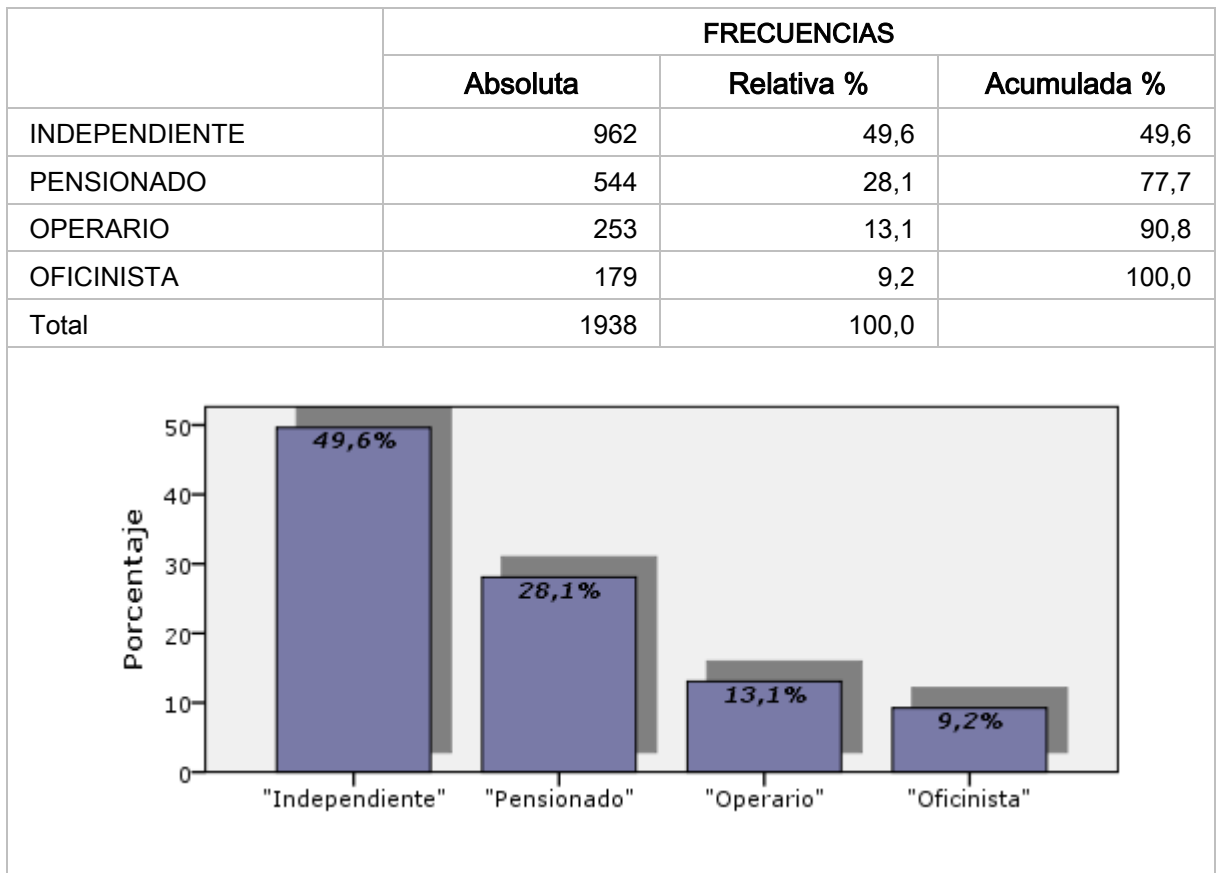
Cómo se puede ver, la mayor representatividad se encuentra en el departamento de Santander que cuenta con un poco más de la mitad de las oficinas que hay en todo el país (58,8%; 30); incluso en Bucaramanga y su área metropolitana se

localizan 12 oficinas que corresponden casi a una cuarta parte (23,52%; 12) del total de sedes con las que cuenta la cooperativa. En orden descendiente, el departamento del Cesar tiene 8 oficinas (15,7%); Boyacá 4 (7,8%); Cundinamarca y Norte de Santander, cada uno con 3 (5,8%) y finalmente Atlántico con 2(3,9%).

#### 5.1.1.4 Variable: **OCUPACIÓN**

La Tabla 6 da cuenta de la distribución de frecuencias de las categorías de la variable ocupación, en la parte inferior se presenta el correspondiente diagrama de barras.

**Tabla 6. Tabla de frecuencias y diagrama de barras para la variable OCUPACIÓN.**



*Fuente: Autor*

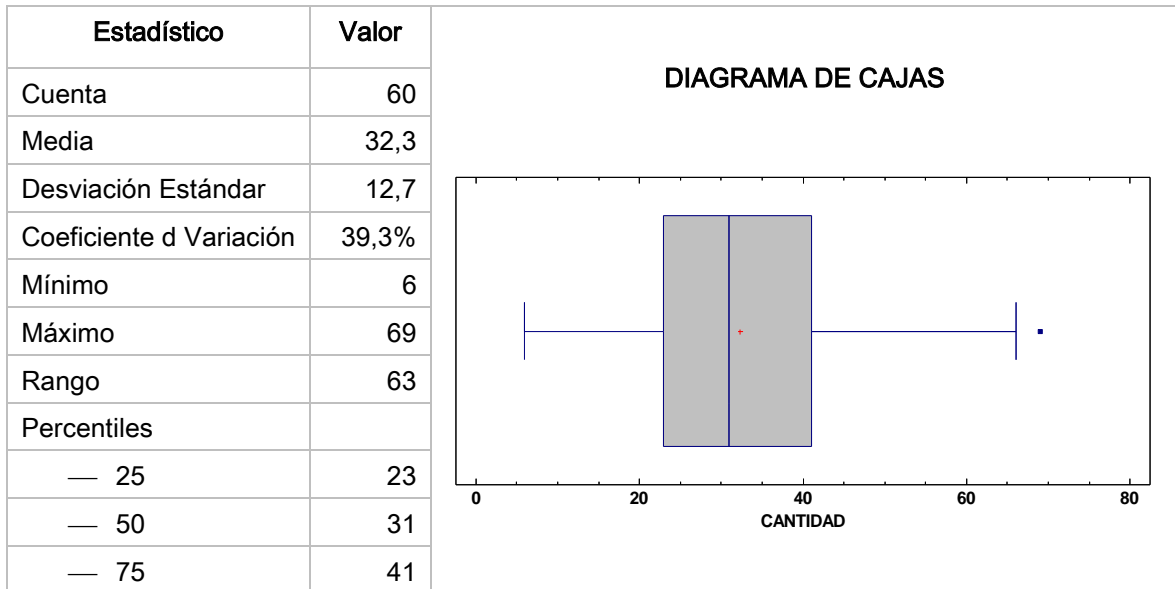
Casi la mitad (49,6%; 962) de los asociados que han hecho reclamaciones por siniestros a la cooperativa se han registrado como trabajadores independientes. En orden descendiente, el 28,1% de los reclamantes son pensionados y el restante se registran en la categoría “operarios” un 13,1% (253), donde se agrupa a las personas que realizan actividades de tipo manual que requieren esfuerzo físico, transportadores y empleados de la fuerza pública. Y en la categoría “oficinista” un 9,2% (179) donde se agrupan los empleados y los docentes.

### **5.1.2. Análisis descriptivos de las variables cuantitativas**

#### **5.1.2.1 Variable: *RECLAMACIONES***

Para el análisis descriptivo de la variable reclamaciones (cantidad de reclamos presentados) se aborda inicialmente desde el punto de vista univariado y posteriormente desde un enfoque bivariado en relación con las variables oficina, año y mes.

**Tabla 7. Estadísticos descriptivos y diagramas de caja para la variable RECLAMACIONES.**

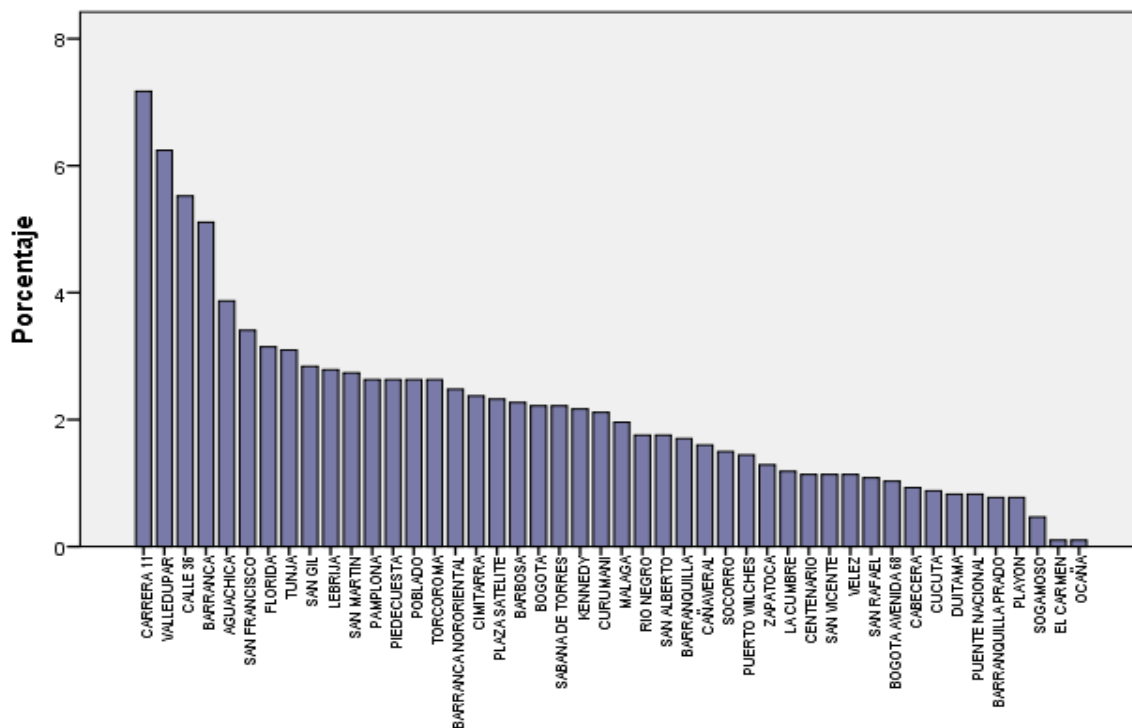


*Fuente: Autor*

En la Tabla 7, el diagrama de caja exhibe para la variable reclamaciones una distribución un poco asimétrica a la derecha con mediana de 31, una media de 32 reclamos con desviación de 12,7 reclamos; con un mínimo de 6 siniestros (diciembre de 2011) y un máximo muy alto de 69 siniestros (julio de 2015).

La Figura 1 da cuenta del diagrama de barras para la variable reclamaciones según la variable oficina.

Figura 1. Diagrama de barras para la variable RECLAMACIONES según oficina.



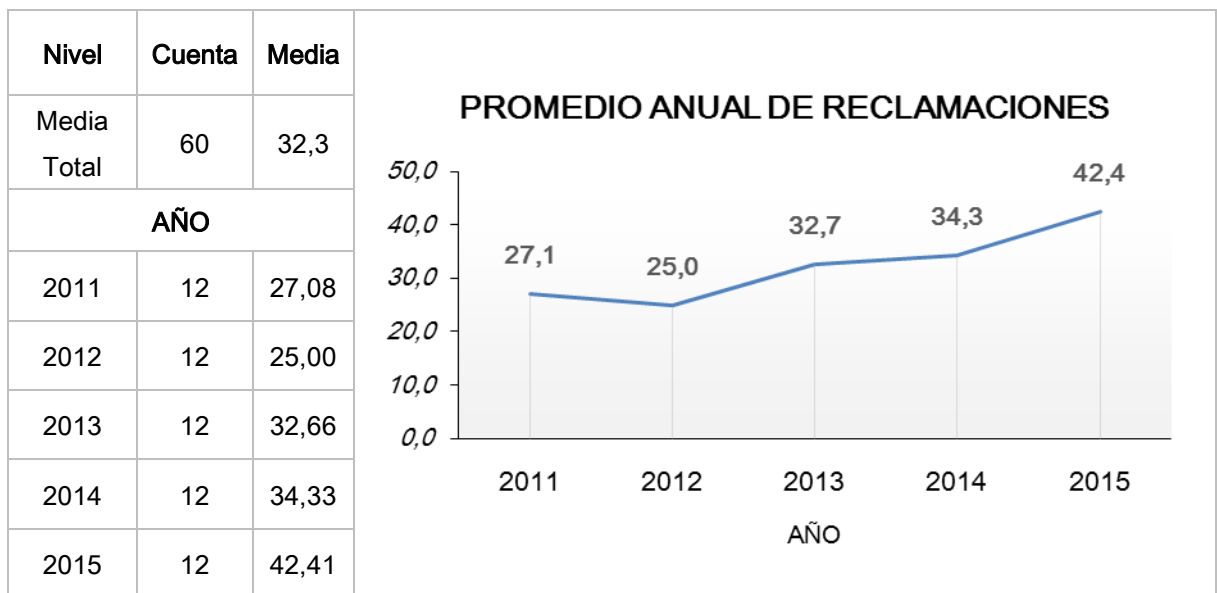
Fuente: Autor

Se puede ver que dos oficinas ubicadas en la ciudad de Bucaramanga están entre las cinco que muestran mayor número de reclamaciones. Estas corresponden, en primer lugar, a la oficina de la carrera 11 (139; 7,2%) y en tercer lugar a la oficina de la calle 35 (107; 5,5%). Acompañando a las anteriores están: la oficina de Valledupar (121; 6,2%), Barranca (99; 5,1%) y Aguachica (75; 3,9%) que unidas suman un poco más de la cuarta parte (27,9%) de las reclamaciones en el quinquenio. Por el contrario, tres oficinas se distinguen por tener el menor número de reclamaciones en el periodo de referencia: la oficina de Ocaña y El Carmen cada una con 2 (0,1%) y la oficina de Sogamoso con 9 reclamaciones que corresponde al 0,5%.

Las demás oficinas participan con porcentajes que van desde 3,4% (Oficina de san Francisco en Bucaramanga) hasta 0,8% (Oficinas de Duitama, Puente

Nacional, Barranquilla-Prado y el Playón). Para análisis posteriores, esta variable será categorizada en 3 niveles.

**Tabla 8. Perfil de medias por año para la variable RECLAMACIONES.**



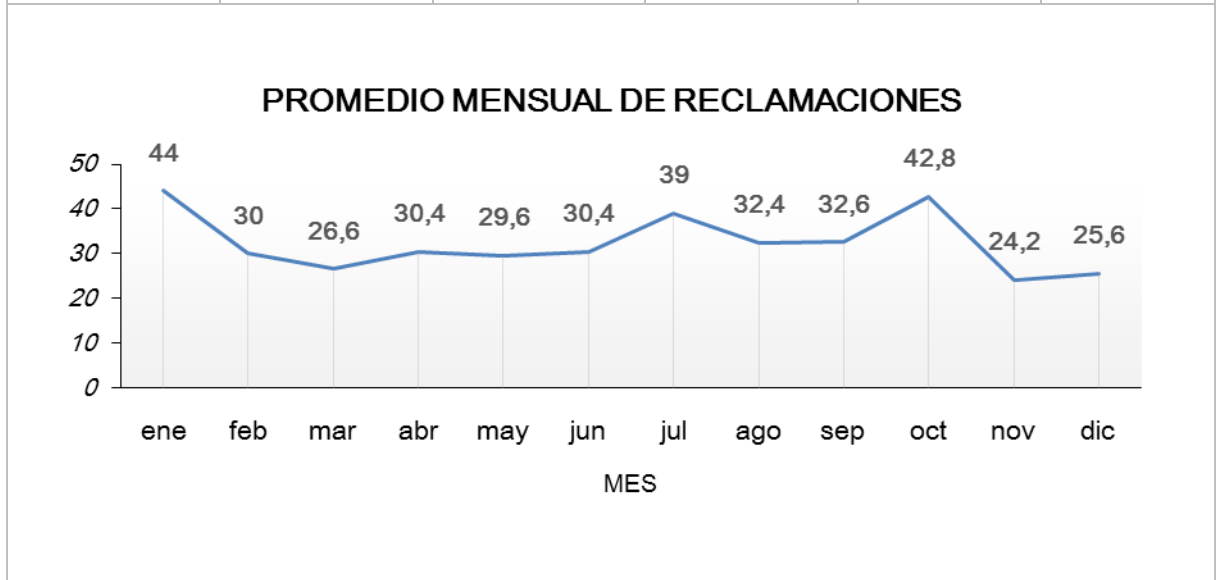
*Fuente: Autor*

En la Tabla 8 se cuantifica el promedio de la cantidad de reclamaciones en cada uno de los años considerados y a la derecha se presenta el diagrama de perfil de medias anuales correspondiente.

En el año 2012 se presenta en promedio el menor número de reclamos (25) y en el año 2015 el mayor promedio (42). A partir del año 2013 el incremento anual de la variable es con respecto al año inmediatamente anterior de 30,64%, 5,11% y 23,53% para los años 2013, 2014 y 2015 respectivamente. Se observa una tendencia creciente en el periodo 2012-2015 que obliga a buscar razones que la expliquen.

**Tabla 9. Perfil de medias por mes para la variable RECLAMACIONES.**

Media Total		60		32,3	
MES	CUENTA	MEDIA	MES	CUENTA	MEDIA
ene	5	44	jul	5	39
feb	5	30	ago	5	32,4
mar	5	26,6	sep	5	32,6
abr	5	30,4	oct	5	42,8
may	5	29,6	nov	5	24,2
jun	5	30,4	dic	5	25,6



*Fuente: Autor*

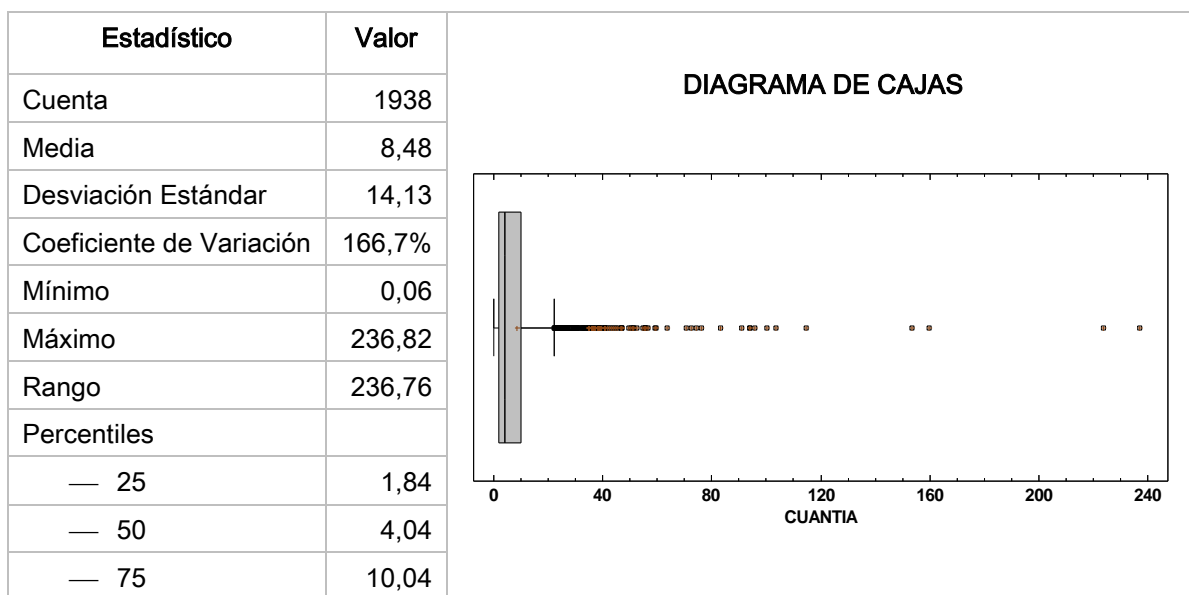
La Tabla 9 muestra los promedios mensuales del número de reclamaciones que realizan los asociados. Se advierte que el último bimestre del año presenta el menor promedio mensual con 24,2 (nov) y 25,6 (dic) reclamos. Siendo a su vez, los meses cercanos de oct (42,8) y ene (44) los de mayor número de reclamaciones. En general el promedio mensual de reclamaciones mensuales es de 32,3.

### 5.1.2.2 Variable: CUANTÍA

Para la descripción de la variable cuantía también se abordará inicialmente desde el punto de vista univariado y seguidamente desde su relación con las variables año y mes.

La Tabla 10 da cuenta de los principales estadísticos descriptivos, los percentiles 25, 50, 75 y en la parte derecha el diagrama de caja para la variable cuantía. Los valores de la cuantía reclamada por los asociados están expresados en millones de pesos.

**Tabla 10. Estadísticos descriptivos y diagrama de caja para la variable CUANTÍA.**



*Fuente: Autor*

El valor promedio de las reclamaciones que han hecho los asociados de la cooperativa en este quinquenio asciende a 8,48 millones con desviación estándar de 14,13 millones. El 50% de las reclamaciones han sido inferiores a 4,04

millones, siendo la mínima cuantía de 0,06 millones y la máxima de 236,82 millones de pesos. El valor del Coeficiente de Variación, CV=166,6%, indica una muy alta variabilidad de los datos con respecto a la media.

En el diagrama de caja se evidencian muchos puntos extremos que al identificar en la base de datos los 5 mayores y los 5 menores corresponden a:

#### VALORES MAYORES

FECHA	ESTADO	CAUSA	EDAD	ZONA-OFICINA	CUANTÍA
07-nov-14	OBJETADO	MUERTE	65	CALLE 35	236,82
04-ago-14	PAGADO	I.T.P	56	SABANA DE TORRES	223,59
15-abr-11	OBJETADO	I.T.P	52	CALLE 35	159,66
13-ago-11	OBJETADO	I.T.P	61	ZAPATOCA	153,45
28-oct-13	OBJETADO	MUERTE	42	SAN MARTIN	114,75

#### VALORES MENORES

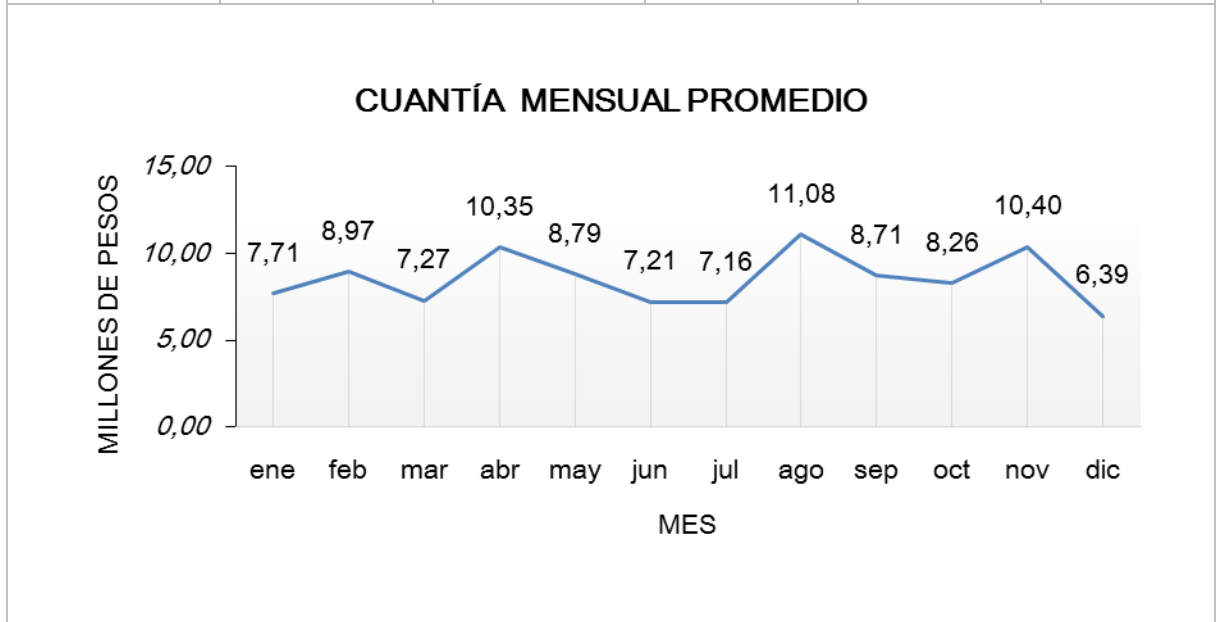
FECHA	ESTADO	CAUSA	EDAD	ZONA-OFICINA	CUANTÍA
18-ene-13	PAGADO	MUERTE	30	AGUACHICA	0,06
24-ene-11	PAGADO	MUERTE	48	TUNJA	0,06
30-oct-12	PAGADO	MUERTE	73	POBLADO	0,08
25-mar-15	PAGADO	MUERTE	22	PLAZA SATÉLITE	0,09
05-jun-13	PAGADO	MUERTE	62	SAN RAFAEL	0,1

Se advierte que de las cinco cuantías más altas reclamadas, cuatro se objetaron, dos fueron por muerte y las tres restantes por Incapacidad Total y Permanente. Y de las cinco cuantías más bajas, todas fueron reclamadas por muerte y todas fueron pagadas.

En la Tabla 11 se presenta el promedio mensual de los valores reclamados por los asociados en las distintas oficinas del país. Se puede ver que en promedio el mes de agosto es el que presenta mayor cuantía reclamada (11,08 millones), seguido por los meses de noviembre (10,40 millones) y abril (10,35 millones). Contrario a estos meses, diciembre presenta el menor valor promedio de cuantía reclamada.

**Tabla 11. Perfil de medias por mes para la variable CUANTÍA.**

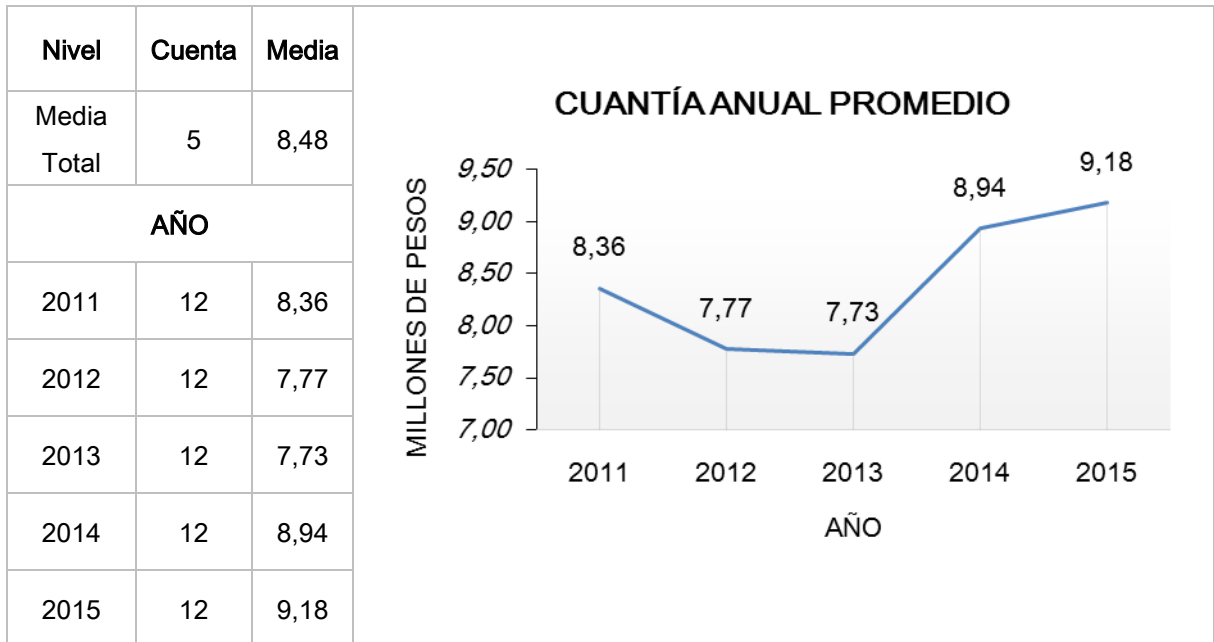
Media Total		60		8,48	
MES	CUENTA	MEDIA	MES	CUENTA	MEDIA
ene	5	7,71	jul	5	7,16
feb	5	8,97	ago	5	11,08
mar	5	7,27	sep	5	8,71
abr	5	10,35	oct	5	8,26
may	5	8,79	nov	5	10,40
jun	5	7,21	dic	5	6,39



*Fuente: Autor*

En la Tabla 12 que se presenta a continuación se muestra el promedio anual de las cuantías reclamadas por los asociados en las distintas oficinas y en la figura de la derecha el respectivo gráfico de perfil de medias por año.

**Tabla 12. Perfil de medias por año para la variable CUANTÍA.**



*Fuente: Autor*

Se ve claramente que en el periodo 2011 a 2013 las cuantías reclamadas presentan una trayectoria en descenso, pero a partir de este año, la tendencia cambia, y los siguientes periodos 2014 (8,94 millones) y 2015 (9,18 millones) exhiben una tendencia ascendente, siendo este último año el de mayor promedio anual en el valor de las cuantías reclamadas.

En general el promedio anual de las cuantías reclamadas es de 8,48 millones de pesos.

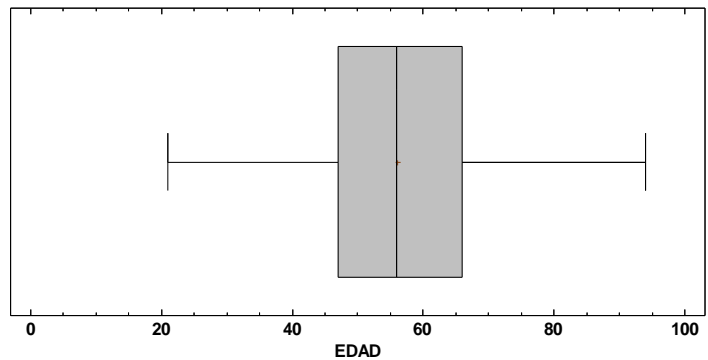
### 5.1.2.3 Variable: **EDAD**

En la Tabla 13 se presentan en la parte superior, los principales estadísticos de tendencia central y dispersión y los percentiles 25, 50, 75; en la parte inferior, el correspondiente diagrama de caja para la variable edad. Se evidencia que la variable edad del reclamante presenta una distribución bastante simétrica. El valor

del coeficiente de variación  $CV=24,3\%$  indica una moderada dispersión de los datos con relación a la media.

**Tabla 13. Estadísticos descriptivos, percentiles y diagrama de caja para la variable EDAD.**

Estadístico	Valor
Cuenta	1938
Media	56,16
Desviación Estándar	13,66
Coeficiente de Variación	24,3%
Mínimo	21
Máximo	94
Rango	73
Percentiles	
— 25	47
— 50	56
— 75	66



*Fuente: Autor*

El promedio de edad de los asociados que han presentado reclamaciones por siniestros a la cooperativa es de 56 años con desviación de 13,66 años. El asociado reclamante de menor edad en este quinquenio tenía 21 años y el de mayor edad 94 años.

El 25% de los asociados reclamantes más jóvenes tenían menos de 47 años y el 25% de los mayores, más de 66 años.

## 5.2. ANÁLISIS DE CORRESPONDENCIA MÚLTIPLE (ACM)

Realizada la exploración inicial de los datos, el interés en este segundo apartado se centra, en el análisis de correspondencia múltiple que permite conocer las asociaciones que existen entre las variables incluidas en este estudio. Como uno de los objetivos de este trabajo fue hallar la asociación existente entre las variables categóricas y las variables cuantitativas RECLAMACIONES (número de reclamos) y CUANTÍA (monto de los reclamos), se hizo necesario categorizar estas últimas tal como se describe más adelante.

De otro lado, al tener la variable oficina tan amplio número de modalidades, se cuenta con 51 sedes, se determinó de acuerdo a la ubicación y los intereses del estudio agruparlas en 3 grandes zonas a saber:

- ZONA A: Reúne las oficinas ubicadas en la ciudad de Bucaramanga y el área metropolitana.
- ZONA B: Aglomera las oficinas ubicadas en el departamento de Santander con excepción de las contempladas en la zona A, es decir, las oficinas del norte, centro, sur y oriente del departamento, así como el área del Magdalena medio.
- ZONA C: Agrupa el resto de oficinas que se encuentran en cinco departamentos del país.

Las variables continuas edad y cuantía se categorizan de acuerdo a las medidas descriptivas, los percentiles y la interpretación práctica de las modalidades. Según estos criterios las variables categorizadas para este análisis se presentan a continuación en la Tabla 14 donde se condensan todas las variables y sus respectivos niveles para cada una de ellas.

**Tabla 14. Categorización de las variables de estudio: zona, edad, cuantía, estado, causa y ocupación.**

VARIABLE	CATEGORÍAS DE LA VARIABLE		
	NIVEL	ETIQUETA	VALOR
OFICINA	1	zona A	Oficinas de Bucaramanga y Área Metropolitana
	2	zona B	Oficinas del Departamento de Santander
	3	zona C	Oficinas en el resto del país
EDAD	1	Menores de 47	[21 – 47) años de edad
	2	Entre 47 -56	[47 – 56) años de edad
	3	Entre 56 - 66	(56 – 66] años de edad
	4	Mayores de 66	(66 – 94] años de edad
CUANTÍA	1	Cuantía baja	[0,06 – 1,84] millones
	2	Cuantía media	(1,84 – 4,04] millones
	3	Cuantía mayor	(4,04 – 10,04] millones
	4	Cuantía alta	(10,04 – 236,82] millones
ESTADO	1	Pagado	
	2	Objetado	
	3	Pendiente	
	4	Prescribió	
CAUSA	1	Muerte	
	2	I.T.P.	
OCUPACIÓN	1	Independiente	
	2	Pensionado	
	3	Operario	
	4	Oficinista	

*Fuente: Autor*

La nueva base de datos obtenida con todas las variables categorizadas fue registrada con el software Microsoft® Excel 2010. El procedimiento estadístico con el que se pretende identificar las proximidades entre las modalidades de las

variables e identificar cómo está estructurada esta asociación se realizó con el software *Système Portable pour l'Analyse de Données*® (SPAD v56).

En la Tabla 15 se presentan las contribuciones y cosenos cuadrados de las variables activas en el análisis: ESTADO, CAUSA, CUANTÍA, EDAD, OFICINA Y OCUPACIÓN.

Al examinar la contribución de las variables a la inercia de los factores se observa que en el primer factor las variables ESTADO (34,7%), y CAUSA (39,4%) son las que más contribuyen a ese factor; mientras que las variables

**Tabla 15. Contribuciones y cosenos cuadrados de las variables activas.**

CATEGORIES			LOADINGS					CONTRIBUTIONS					SQUARED COSINES				
IDEN - LABEL	REL. WT.	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1 . ESTADO																	
1 - pagado	13.11	0.27	0.43	0.03	-0.01	-0.08	0.03	7.4	0.1	0.0	0.5	0.1	0.68	0.00	0.00	0.02	0.00
2 - objetado	3.56	3.68	-1.58	-0.10	0.03	0.29	-0.10	27.3	0.2	0.0	1.7	0.2	0.68	0.00	0.00	0.02	0.00
			CUMULATED CONTRIBUTION = 34.7					0.2 0.0 2.2 0.2									
2 . CAUSA																	
1 - muerte	10.96	0.52	0.63	-0.01	0.00	-0.08	0.01	13.5	0.0	0.0	0.4	0.0	0.77	0.00	0.00	0.01	0.00
2 - I.T.P	5.70	1.92	-1.22	0.01	0.00	0.15	-0.02	25.9	0.0	0.0	0.7	0.0	0.77	0.00	0.00	0.01	0.00
			CUMULATED CONTRIBUTION = 39.4					0.0 0.0 1.1 0.0									
3 . CUANT=A																	
1 - menos 1,84	4.19	2.98	0.44	0.04	0.65	-0.89	0.47	2.4	0.0	10.0	18.8	5.4	0.06	0.00	0.14	0.26	0.07
2 - [1,84-4,04]	4.14	3.03	0.19	-0.08	-0.07	0.60	-1.12	0.5	0.1	0.1	8.5	30.1	0.01	0.00	0.00	0.12	0.41
3 - (4,04-10,04]	4.18	2.99	0.05	-0.01	-0.38	0.72	0.44	0.0	0.0	3.4	12.3	4.7	0.00	0.00	0.05	0.17	0.07
4 - mBs 10,04	4.16	3.00	-0.68	0.05	-0.20	-0.43	0.20	5.8	0.0	1.0	4.3	0.9	0.15	0.00	0.01	0.06	0.01
			CUMULATED CONTRIBUTION = 8.8					0.2 14.5 43.9 41.1									
4 . EDAD																	
1 - menos 47 años	4.38	2.81	-0.45	0.16	0.88	0.48	0.04	2.7	0.6	19.3	5.7	0.0	0.07	0.01	0.28	0.08	0.00
2 - [47-56] años	4.06	3.11	-0.51	0.07	-0.42	-0.95	-0.01	3.2	0.1	4.0	21.1	0.0	0.08	0.00	0.06	0.29	0.00
3 - [56-66] años	4.14	3.03	0.18	-0.25	-0.92	0.20	0.42	0.4	1.3	20.1	1.0	4.3	0.01	0.02	0.28	0.01	0.06
4 - mBs 66 años	4.09	3.07	0.80	0.01	0.41	0.23	-0.46	8.1	0.0	3.8	1.3	4.9	0.21	0.00	0.05	0.02	0.07
			CUMULATED CONTRIBUTION = 14.5					2.0 47.3 29.0 9.2									
5 . ZONA																	
1 - zona A	6.53	1.55	0.03	0.54	-0.17	0.18	0.60	0.0	9.8	1.1	1.3	13.8	0.00	0.19	0.02	0.02	0.23
2 - zona B	5.67	1.94	-0.05	0.34	0.01	-0.17	-0.86	0.0	3.3	0.0	0.9	24.5	0.00	0.06	0.00	0.01	0.38
3 - zona C	4.47	2.73	0.01	-1.21	0.24	-0.05	0.21	0.0	33.8	1.4	0.1	1.2	0.00	0.54	0.02	0.00	0.02
			CUMULATED CONTRIBUTION = 0.1					46.9 2.5 2.3 39.5									
6 . OCUPACION																	
1 - independiente	8.27	1.01	-0.19	0.35	0.35	-0.31	-0.06	1.0	5.1	5.6	4.5	0.1	0.04	0.12	0.12	0.09	0.00
2 - pensionado	4.68	2.56	0.33	0.52	-0.60	0.45	0.11	1.5	6.4	9.6	5.5	0.3	0.04	0.10	0.14	0.08	0.00
3 - operario	2.18	6.66	-0.05	-1.57	-0.85	-0.50	-0.57	0.0	27.6	8.9	3.1	4.0	0.00	0.37	0.11	0.04	0.05
4 - oficinista	1.54	9.83	0.13	-1.21	1.16	0.98	0.77	0.1	11.6	11.7	8.5	5.3	0.00	0.15	0.14	0.10	0.06
			CUMULATED CONTRIBUTION = 2.6					50.7 35.8 21.5 9.8									

Fuente: salida del programa *Système Portable pour l'Analyse de Données*® (SPAD v56).

Ahora, al inspeccionar las contribuciones en el primer factor, las modalidades que lo definen son objetado (27,3%) en ESTADO y I.T.P (25,9%) en CAUSA lo cual indica una alta asociación entre estas modalidades y el primer eje factorial, así como una proximidad entre ellas con valores cercanos en sus contribuciones.

Opuesto a éstas, las contribuciones de las modalidades de OFICINA, OCUPACIÓN y CUANTÍA son mínimas en la conformación de este factor.

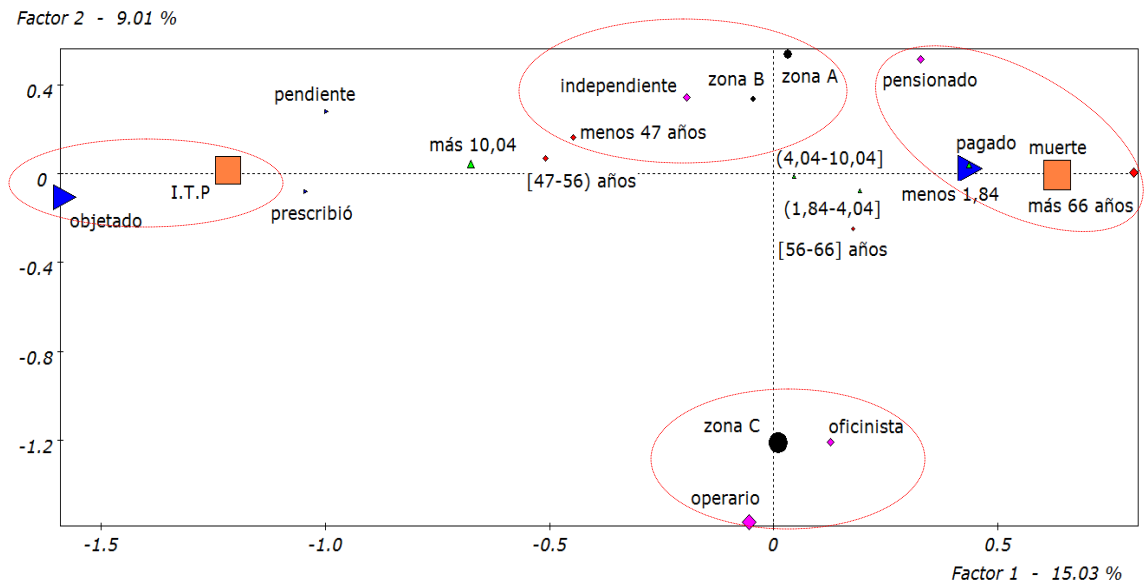
Al mirar en la Tabla 15 el último bloque (correspondiente a los cosenos cuadrados) las modalidades de ESTADO (0,68) y CAUSA (0,77) están muy bien representadas.

El segundo factor, según sus contribuciones, está determinado por las modalidades zona C (33,8%) de OFICINA y operario (27,6%) de OCUPACIÓN, lo cual indica una alta asociación entre estas modalidades y el segundo eje factorial, además una buena representación con valores de los cosenos cuadrados de 0,54 y 0,37 respectivamente.

A continuación se presenta en la Figura 2 el primer plano factorial de las variables activas. Se ha situado el primer eje principal horizontalmente y el segundo eje principal verticalmente. Junto a los ejes se indican los correspondientes porcentajes de las inercias (15,03% y 9,01% respectivamente). Al sumar estos valores el porcentaje de la inercia explicada por el plano es de 24,04%.

En el primer eje factorial se oponen las modalidades, objetado y pagado de la variable ESTADO, y las modalidades I.T.P. y muerte de la variable CAUSA. En el segundo eje factorial se oponen las modalidades zona A y zona B de la modalidad zona C en la variable OFICINA y las modalidades, independiente y pensionado de la variable OCUPACIÓN.

**Figura 2. Primer plano factorial de las variables activas: estado, causa, cuantía, edad, oficina y ocupación (inercia 24,04%).**



Fuente: Autor

Por consiguiente, del ACM se puede concluir que en principio se determinan los siguientes grupos:

- Un grupo formado por socios cuyos reclamos por Incapacidad Total y Permanente (I.T.P.) fueron objetados.
- Un grupo que reúne a socios mayores de 66 años, pensionados, que reclamaron por muerte del deudor en cuantía menor a 1,84 millones y les fueron pagados.
- Un grupo que aglomera a socios en oficinas ubicadas en el resto del país (con excepción de las oficinas en el departamento de Santander) cuya ocupación registrada es trabajador independiente u oficinista.
- Un grupo formado por socios en las oficinas del departamento de Santander, menores de 47 años y trabajadores independientes.

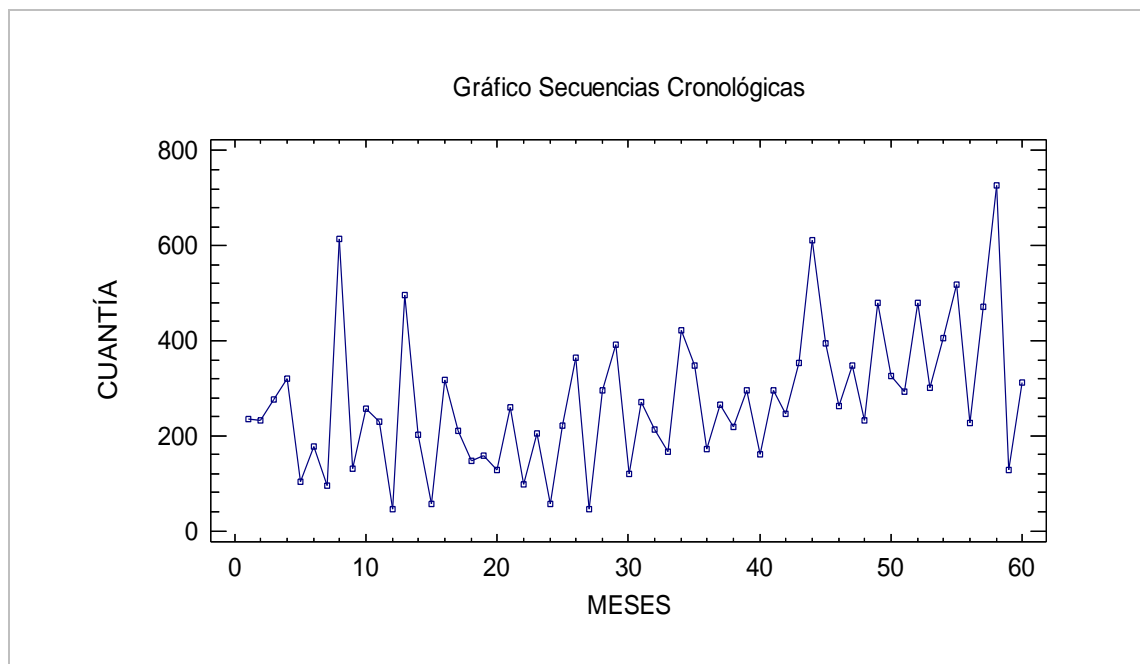
### 5.3. ANÁLISIS DE SERIES DE TIEMPO

Con el propósito de poder cumplir con los objetivos planteados se realiza en este apartado el análisis de la serie cuantía en el quinquenio 2011 – 2015. Los valores de la serie hacen referencia a la suma total de las cuantías reclamadas en cada uno de los meses del periodo considerado.

#### 5.3.1 Análisis de la serie temporal

El análisis de esta serie temporal univariada comienza con la representación gráfica de la variable a lo largo del tiempo.

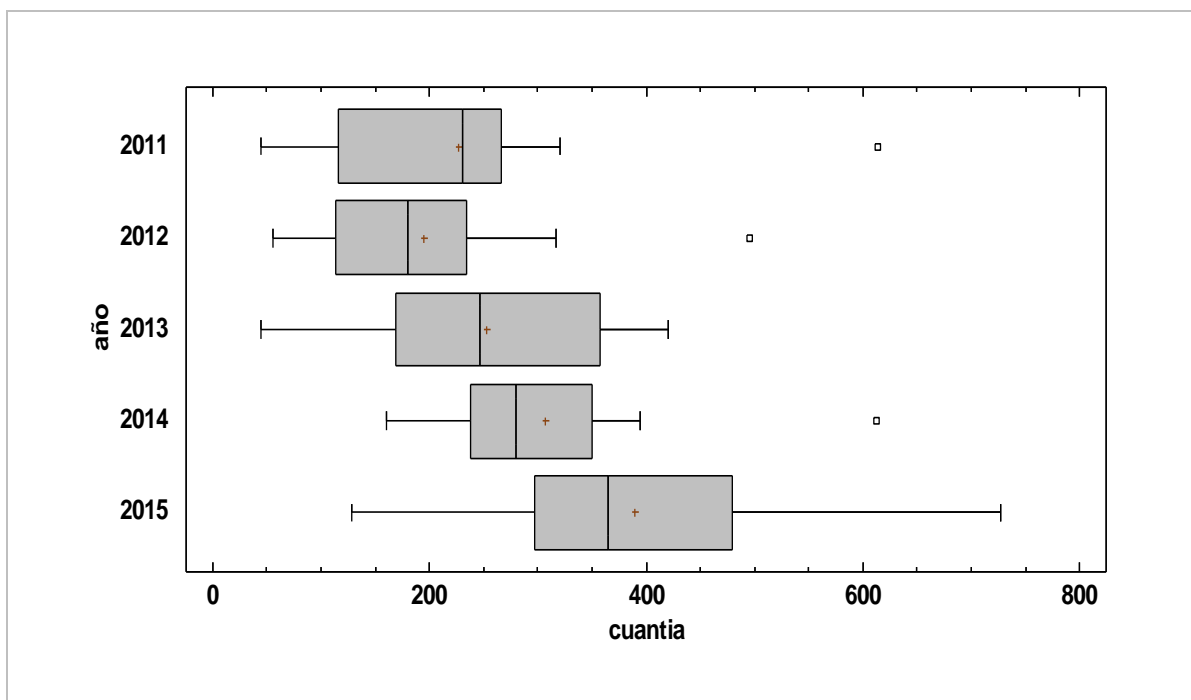
Figura 3. Gráfico temporal de la serie CUANTÍA.



Fuente: Autor.

En el gráfico temporal de la Figura 3 se evidencia una serie con tendencia creciente en el periodo 2012 - 2016. La varianza de la serie se muestra estable salvo periodos donde los valores fueron muy grandes.

**Figura 4. Diagrama de cajas para CUANTÍA por año.**



*Fuente: Autor.*

La Figura 4 confirma la tendencia creciente que existe en la serie cuantía donde se advierte que desde el año 2012 la mediana en la caja correspondiente a cada año se va incrementando.

A continuación se muestra la salida del programa con los resultados comparativos de varios modelos considerados para ajustar la serie CUANTÍAS. De acuerdo a las diferentes estadísticas que miden la magnitud del error, existen dos modelos con los errores menores, tanto en el periodo de estimación como en el de

validación, y además, muy semejantes entre ellos: el modelo de tendencia lineal (modelo C) y el modelo ARIMA (0,1,2) (modelo M).

### 5.3.2. Comparación de modelos

#### A. Comparación de Modelos

Variable de datos: cuantía

Número de observaciones = 60

Índice Inicial = 1/11

Intervalo de Muestra = 1,0 mes(es)

Número de periodos retenidos para validación: 6

#### B. Modelos

(A) Caminata aleatoria

(B) Media constante = 260,131

**(C) Tendencia lineal = -2023,67 + 3,00698 t**

(H) Suavización exponencial simple con alfa = 0,1085

(I) Suavización exp. De Brown con alfa = 0,0634

(J) Suavización exp. De Holt con alfa = 0,1116 y beta = 0,1099

**(M) ARIMA (0,1,2)**

(N) ARIMA (0,1,1) con constante

(O) ARIMA (1,1,1)

(P) ARIMA (0,1,1)

(Q) ARIMA (0,1,2) con constante

#### C. Periodo de Estimación

<b>Modelo</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>	<b>ME</b>	<b>MPE</b>	<b>AIC</b>
(A)	189,227	152,613	89,5492	6,97136E-15	-47,2319	10,4305
(B)	130,427	100,869	67,9648	1,89478E-14	-44,5679	9,74163
<b>(C)</b>	<b>122,859</b>	<b>92,5678</b>	<b>60,4197</b>	<b>1,56846E-13</b>	<b>-38,1272</b>	<b>9,65911</b>
(H)	127,104	95,3556	58,3348	17,9818	-30,5868	9,69001
(I)	127,239	95,9857	56,9233	20,8109	-27,6375	9,69214
(J)	127,783	96,8013	57,9386	12,0332	-30,6534	9,7377
<b>(M)</b>	<b>124,913</b>	<b>86,9437</b>	<b>51,7547</b>	<b>16,347</b>	<b>-26,8614</b>	<b>9,72932</b>
(N)	126,375	94,3006	60,9911	5,43139	-37,2151	9,75258
(O)	126,801	90,4361	54,5677	17,4431	-28,5636	9,75931
(P)	129,531	97,204	59,4876	17,7461	-31,344	9,76488
(Q)	125,216	92,2065	60,0827	2,55237	-37,544	9,77119

<i>Modelo</i>	<i>RMSE</i>	<i>RUNS</i>	<i>RUNM</i>	<i>AUTO</i>	<i>MEDIA</i>	<i>VAR</i>
(A)	189,227	OK	***	***	OK	OK
(B)	130,427	OK	OK	OK	**	OK
<b>(C)</b>	<b>122,859</b>	<b>OK</b>	<b>OK</b>	<b>OK</b>	<b>OK</b>	<b>OK</b>
(H)	127,104	OK	*	*	OK	OK
(I)	127,239	OK	*	*	OK	OK
(J)	127,783	OK	*	*	OK	OK
<b>(M)</b>	<b>124,913</b>	<b>OK</b>	<b>OK</b>	<b>OK</b>	<b>OK</b>	<b>OK</b>
(N)	126,375	OK	*	*	OK	OK
(O)	126,801	OK	OK	OK	OK	OK
(P)	129,531	OK	*	*	OK	OK
(Q)	125,216	OK	OK	OK	OK	OK

#### D. Periodo de Validación

<i>Modelo</i>	<i>RMSE</i>	<i>MAE</i>	<i>MAPE</i>	<i>ME</i>	<i>MPE</i>
(A)	320,575	280,261	127,356	-18,8476	-72,4097
(B)	242,084	192,415	48,7868	137,718	9,78827
<b>(C)</b>	<b>206,063</b>	<b>177,497</b>	<b>59,1116</b>	<b>47,5082</b>	<b>-21,8935</b>
(H)	215,598	189,109	66,3015	37,4748	-28,6429
(I)	214,929	191,692	72,2374	8,6328	-39,2826
(J)	216,255	191,579	78,1101	-27,6559	-51,7428
<b>(M)</b>	<b>198,382</b>	<b>173,633</b>	<b>55,3681</b>	<b>27,0357</b>	<b>-23,0629</b>
(N)	209,693	180,258	59,557	52,2838	-20,938
(O)	209,763	184,746	60,4397	28,1227	-26,0181
(P)	215,665	189,585	67,0298	34,5891	-29,8009
(Q)	200,668	170,718	51,3893	56,4383	-14,1493

Clave:

RMSE = Root Mean Squared Error (Raíz del Cuadrado Medio del Error)

RUNS = Prueba corridas excesivas arriba y abajo

RUNM = Prueba corridas excesivas arriba y abajo de la mediana

AUTO = Prueba de Box-Pierce para autocorrelación excesiva

MEDIA = Prueba para diferencia en medias entre la 1ª mitad y la 2ª mitad

VAR = Prueba para diferencia en varianza entre la 1ª mitad y la 2ª mitad

OK = no significativo ( $p \geq 0,05$ )

\* = marginalmente significativo ( $0,01 < p \leq 0,05$ )

\*\* = significativo ( $0,001 < p \leq 0,01$ )

\*\*\* = altamente significativo ( $p \leq 0,001$ )

Los valores de los estadísticos de los dos modelos considerados tanto para el ajuste del modelo como para la predicción se presentan a continuación en la Tabla 16.

**Tabla 16. Estadísticos de error para el periodo de estimación y validación de los modelos de Tendencia Lineal y Arima.**

Estadísticos de error	PERIODO			
	ESTIMACIÓN		VALIDACIÓN	
	Modelo Tendencia Lineal (C)	Modelo Arima (0,1,2) (M)	Modelo Tendencia Lineal (C)	Modelo Arima (0,1,2) (M)
<b>RMSE</b>	122,859	124,913	206,063	198,382
<b>MAE</b>	92,567	86,943	177,497	173,633
<b>MAPE</b>	60,419	51,754	59,111	55,368
<b>ME</b>	1,57E-13	16,347	47,508	27,035
<b>MPE</b>	-38,127	-26,861	-21,893	-23,062

*Fuente: Autor.*

Se observa que los estadísticos del error son ligeramente menores en el modelo Arima para el periodo de validación, además como se vio anteriormente estos dos modelos (C) y (M) satisfacen todas las pruebas que realiza el algoritmo. El AIC (Criterio de Información de Akaike) es ligeramente inferior en el modelo lineal respecto al ARIMA (0,1,2), argumento que permite seleccionar al modelo lineal como el mejor modelo.

En la Tabla 17 se resume la significancia estadística de los términos en el modelo de pronósticos. En este caso el valor\_p para los parámetros del modelos es menor que 0,05, de modo que son estadísticamente diferentes de cero.

**Tabla 17. Resumen del modelo de Tendencia Lineal.**

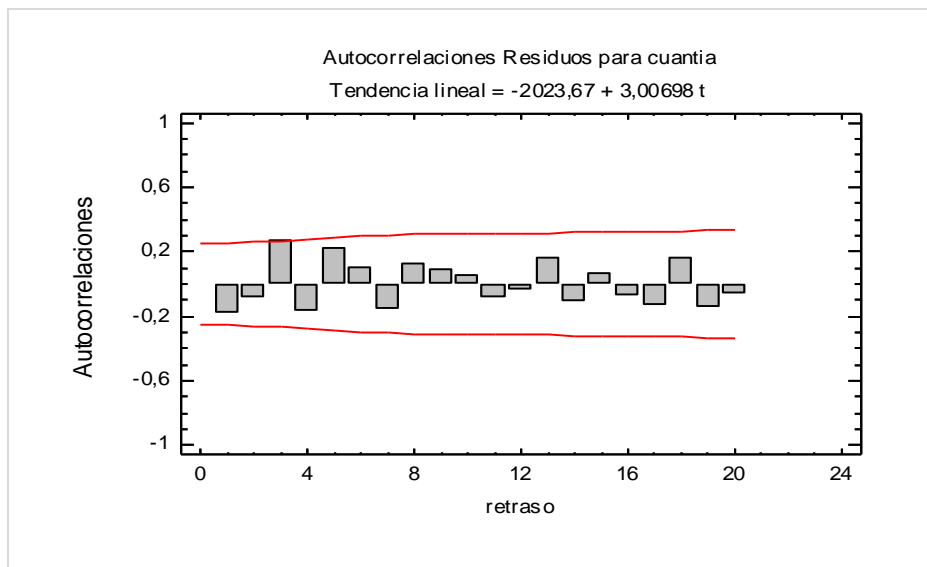
RESUMEN DEL MODELO	Parámetro	Estimado	Error Estd.	t	Valor-P
Modelo de Tendencia Lineal $-2023,67 + 3,00698 t$	Constante	-2023,67	822,691	-2,45982	0,017262
	Pendiente	3,00698	1,08297	2,7766	0,007619

*Fuente: Autor.*

Finalmente, antes de realizar las predicciones y aunque el modelo de regresión para la serie no incluye un elemento aleatorio, realizamos un análisis de los residuos con el ánimo de saber si el modelo podría mejorarse añadiéndole una componente que explicara la autocorrelación entre los residuos.

La Figura 5 da cuenta del gráfico de las autocorrelaciones estimadas entre los residuos a diferentes retrasos. El coeficiente de autocorrelación con retraso k mide la correlación entre los residuos al tiempo t y al tiempo t-k.

**Figura 5 . Función de autocorrelación para los residuos del modelo de tendencia lineal (FAC).**

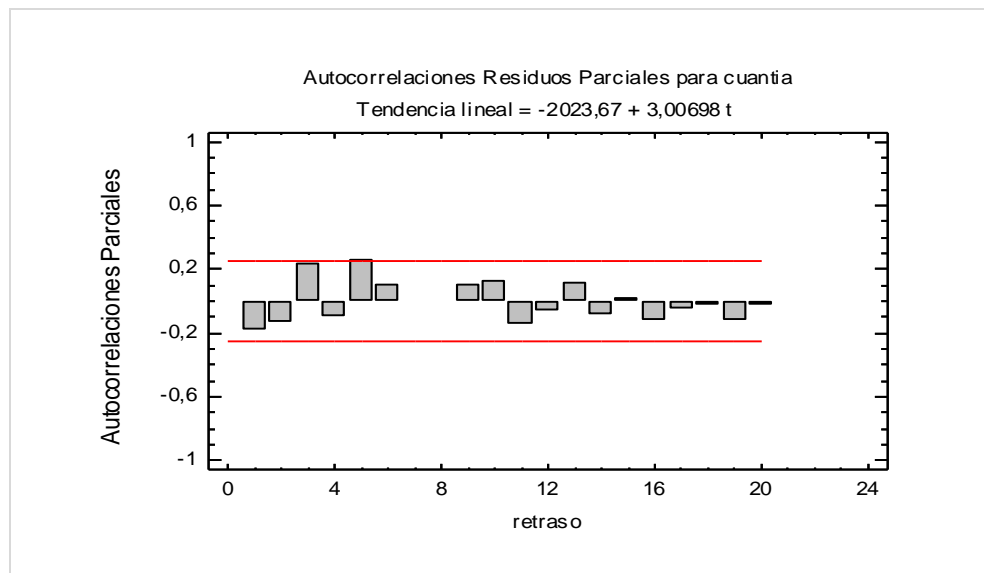


También se muestran límites de probabilidad del 95,0% alrededor de 0. En el gráfico se muestran unas franjas rojas que indican los límites de probabilidad del 95,0% alrededor de 0. Como se puede observar ninguno de los 24 coeficientes de autocorrelación parcial es estadísticamente significativo con un 95,0% de nivel de confianza implicando que la serie de tiempo bien puede ser un ruido blanco (completamente aleatoria).

Si los límites de probabilidad a un retraso particular no contienen el coeficiente estimado, hay una correlación estadísticamente significativa a ese retraso al nivel de confianza del 95,0%. En este caso, ninguno de los 24 coeficientes de autocorrelación son estadísticamente significativos.

La Figura 6 muestra el gráfico de las autocorrelaciones parciales estimadas entre los residuos a diferentes retrasos. El coeficiente de autocorrelación parcial al retraso  $k$  mide la correlación entre los residuos al tiempo  $t$  y al tiempo  $t+k$ , habiendo descontado por la correlación a todos los retrasos menores.

**Figura 6 . Función de autocorrelación parcial para los residuos del modelo de tendencia lineal (FACP).**



En el gráfico se muestran unas franjas rojas que indican los límites de probabilidad del 95,0% alrededor de 0. Como se puede observar ninguno de los 24 coeficientes de autocorrelación parcial es estadísticamente significativo con un 95,0% de nivel de confianza.

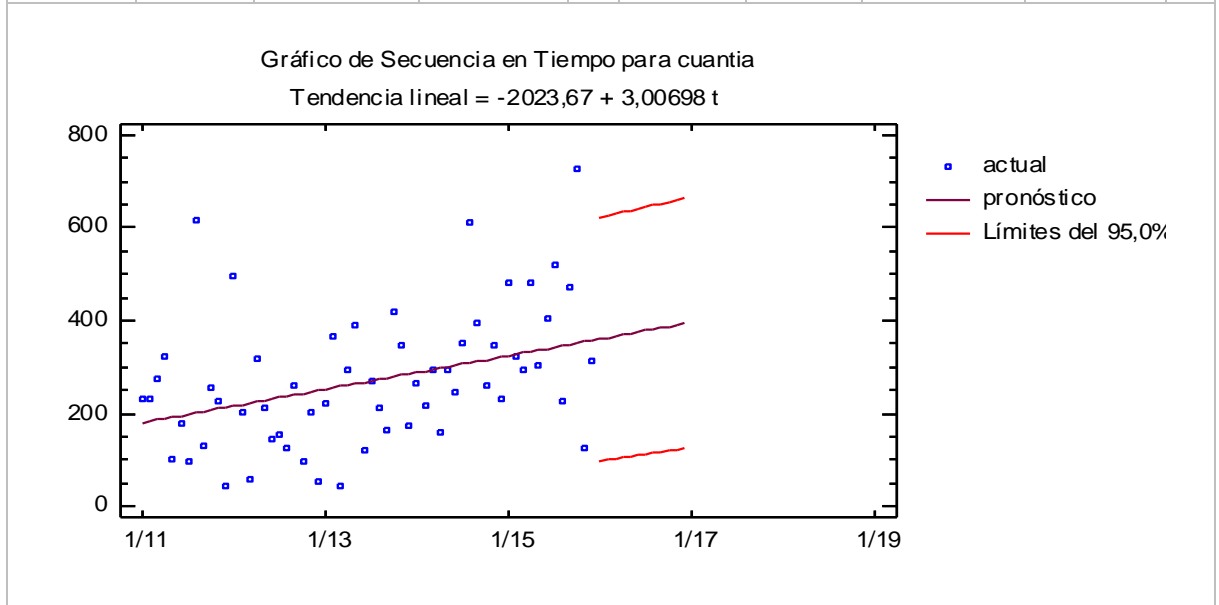
### 5.3.3. Pronósticos del modelo

La Tabla 18 da cuenta de los valores pronosticados para CUANTÍA por el modelo de tendencia lineal el periodo en donde hay disponibles datos, se muestran también los valores predichos del modelo ajustado y los residuos (datos – pronóstico). Los últimos 6 datos de la serie fueron retenidos para validar el modelo.

**Tabla 18. Pronósticos del modelo de Tendencia Lineal.**

Modelo de Tendencia Lineal									
-2023,67 + 3,00698 t									
Periodo	Datos	Pronóstico	Residuo		Periodo	Datos	Pronóstico	Residuo	
1/11	234,17	180,445	53,7245		7/13	271,55	270,655	0,89499	
2/11	232,93	183,452	49,4776		8/13	212,24	273,662	-61,422	
3/11	275,68	186,459	89,2206		9/13	165,73	276,669	-110,939	
4/11	321,01	189,466	131,544		10/13	420,88	279,676	141,204	
5/11	103,0	192,473	-89,4734		11/13	349,14	282,683	66,4571	
6/11	178,51	195,48	-16,9704		12/13	172,49	285,69	-113,2	
7/11	96,06	198,487	-102,427		1/14	265,9	288,697	-22,7969	
8/11	614,27	201,494	412,776		2/14	218,89	291,704	-72,8139	
9/11	129,87	204,501	-74,6313		3/14	296,2	294,711	1,48912	
10/11	258,07	207,508	50,5617		4/14	160,89	297,718	-136,828	
11/11	229,06	210,515	18,5447		5/14	294,45	300,725	-6,27485	
12/11	44,78	213,522	-168,742		6/14	245,08	303,732	-58,6518	

1/12	495,08	216,529	278,551		7/14	352,04	306,739	45,3012	
2/12	202,97	219,536	-16,5663		8/14	612,83	309,746	303,084	
3/12	57,25	222,543	-165,293		9/14	394,23	312,753	81,4772	
4/12	316,35	225,55	90,7998		10/14	262,26	315,76	-53,4998	
5/12	210,47	228,557	-18,0872		11/14	347,07	318,767	28,3032	
6/12	146,73	231,564	-84,8342		12/14	231,87	321,774	-89,9037	
7/12	157,13	234,571	-77,4412		1/15	480,05	324,781	155,269	
8/12	127,47	237,578	-110,108		2/15	324,58	327,788	-3,20772	
9/12	258,63	240,585	18,0448		3/15	293,22	330,795	-37,5747	
10/12	98,54	243,592	-145,052		4/15	480,05	333,802	146,248	
11/12	204,96	246,599	-41,6391		5/15	301,82	336,809	-34,9887	
12/12	56,03	249,606	-193,576		6/15	406,03	339,816	66,2143	
1/13	221,27	252,613	-31,3431		7/15	519,35	342,823	176,527	V
2/13	365,4	255,62	109,78		8/15	227,95	345,83	-117,88	V
3/13	44,91	258,627	-213,717		9/15	471,51	348,837	122,673	V
4/13	295,62	261,634	33,986		10/15	727,66	351,844	375,816	V
5/13	391,02	264,641	126,379		11/15	128,22	354,851	-226,631	V
6/13	120,32	267,648	-147,328		12/15	312,4	357,858	-45,4576	V



Fuente: Autor.

El gráfico de secuencia en el tiempo al final de la Tabla 18, muestra con la línea en color café, el pronóstico del modelo, y las dos líneas en rojo al final, señalan los límites del intervalo al 95% de confianza para los 12 periodos pronosticados de la serie y que se presentan a continuación numéricamente en la Tabla 19.

Para los 12 periodos de tiempo, más allá de la serie, se muestran en la Tabla 19 los límites del 95% de predicción para los pronósticos. Estos límites muestran en donde podría estar el valor verdadero del dato, al tiempo futuro seleccionado, con 95% de confianza, asumiendo que el modelo ajustado es apropiado para los datos.

**Tabla 19. Pronostico para 12 periodos (ene 2016 - dic 2016).**

Periodo	Pronóstico	Límite en 95,0%	
		Inferior	Superior
1/16	360,865	99,3392	622,39
2/16	363,872	101,733	626,01
3/16	366,879	104,11	629,647
4/16	369,886	106,471	633,30
5/16	372,892	108,815	636,97
6/16	375,899	111,144	640,655
7/16	378,906	113,456	644,357
8/16	381,913	115,752	648,075
9/16	384,92	118,033	651,808
10/16	387,927	120,298	655,557
11/16	390,934	122,547	659,322
12/16	393,941	124,781	663,102

*Fuente: Autor.*

## 6. CONCLUSIONES

Como resultado de los análisis descriptivos efectuados a las variables de estudio se tiene las siguientes características:

La mayoría de las reclamaciones (cerca del 80%) que hacen los asociados por siniestros son pagadas por la aseguradora.

En mayor porcentaje (66%) la causa de las reclamaciones de los asociados es por muerte.

La mayor representatividad (59%) se encuentra en el departamento de Santander que cuenta con un poco más de la mitad de las oficinas que hay en todo el país.

Casi la mitad de los asociados que han hecho reclamaciones por siniestros a la cooperativa se han registrado como trabajadores independientes y un poco menos de la tercera parte (28%) como pensionados.

Dos oficinas ubicadas en la ciudad de Bucaramanga están entre las cinco que muestran mayor número de reclamaciones. Estas corresponden a la oficina de la carrera 11 y a la oficina de la calle 35.

El 25% de los asociados reclamantes más jóvenes tenían menos de 47 años y el 25% de los mayores, más de 66 años. El promedio de edad de los asociados que han presentado reclamaciones por siniestros a la cooperativa es de 56 años con desviación de 13,66 años.

El análisis de correspondencias múltiple permitió evidenciar las siguientes asociaciones:

- Un grupo formado por asociados cuyos reclamos por Incapacidad Total y Permanente (I.T.P.) fueron objetados.

- Un grupo que reúne a socios mayores de 66 años, pensionados, que reclamaron por muerte del deudor en cuantía menor a 1,84 millones y les fueron pagados.
- Un grupo que aglomera a socios en oficinas ubicadas en el resto del país (con excepción de las oficinas en el departamento de Santander) cuya ocupación registrada es trabajador independiente u oficinista.
- Un grupo formado por socios en las oficinas del departamento de Santander, menores de 47 años y trabajadores independientes.

El análisis de series de tiempo permitió escoger el Modelo de Tendencia Lineal ( $-2023,67 + 3,00698 t$ ) con el menor AIC y estadísticos del error no muy diferentes de los demás modelos con el que se comparó.

## **7. RECOMENDACIONES**

Se recomienda en la fijación de las tasas del seguro del crédito tener en cuenta la ocupación que registra el asociado al momento de su vinculación, así como la oficina o la zona en la que se encuentra ubicado.

Es recomendable en la aplicación de la serie de tiempo tener en cuenta la presentación de siniestros sobre la colocación de créditos, para poder identificar si presentan un comportamiento independiente o dependiente entre ellas.

## BIBLIOGRAFÍA

De la Fuente F. Santiago, Análisis de correspondencia simples y múltiples -Fac. Ciencias Económicas y Empresariales UAM- Universidad Autónoma de Madrid,, (2011), disponible en:  
<http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/REDUCIR-DIMENSION/CORRESPONDENCIAS/correspondencias.pdf>

Murcia Diana, & Rodríguez Giovanna, (2002), Seguro Vida Deudores comentarios a la sentencia T-1165 de 2001 Corte Constitucional, disponible en:  
<http://www.javeriana.edu.co/biblos/tesis/derecho/dere4/Tesis-60.pdf>

Peña D., Análisis de series temporales (2005), Alianza Editorial (2005). 592 paginas. ISBN: 9788420691282.

Ríos Carlos, P., (n.d.), Estimación de la siniestralidad esperada mediante el análisis de la siniestralidad histórica-Grupo Endesa-, disponible en:  
[https://www.fundacionmapfre.org/documentacion/.../i18n/catalogo\\_imagenes/grupo.cmd?path=1018790](https://www.fundacionmapfre.org/documentacion/.../i18n/catalogo_imagenes/grupo.cmd?path=1018790)

Román Alejandro, V., (2012), Tesis, Análisis de los datos de robo de vehículos asegurados: una aplicación de las series de tiempo, disponible en: <http://mat.izt.uam.mx/mcmai/documentos/tesis/Gen.09-O/Vazquez-AL-Tesis.pdf>

Román Salmerón, (n.d.), Guía breve de análisis de procesos ARIMA, disponible en: <http://www.ugr.es/~romansg/material/WebEco/Eco3-Gretl1.pdf>

Súper Financiera de Col., Boletín Jurídico N°14 (2008) Seguro vida Grupo Deudores, disponible en: <https://www.superfinanciera.gov.co/jsp/loader.jsf?!Servicio=Publicaciones&ITipo=publicaciones&IFuncion=loadContenidoPublicacion&id=15943>.