



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

FACULTAD DE CIENCIAS

CARRERA ESTADÍSTICA

**“CLUSTERS ESPACIALES DE LA ECONOMÍA DE LOS SOCIOS
DE LA COOP. “EDUCADORES DE CHIBORAZO” LTDA (2016-
2020)”**

Trabajo de Titulación

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

INGENIERA EN ESTADÍSTICA INFORMÁTICA

AUTORAS: CRISTINA ELIZABETH LÓPEZ GUILCAPI

MAGDALENA ESTEFANÍA MONTERO ARIAS

DIRECTORA: Ing. AMALIA ISABEL ESCUDERO VILLA MSc.

Riobamba – Ecuador

2022

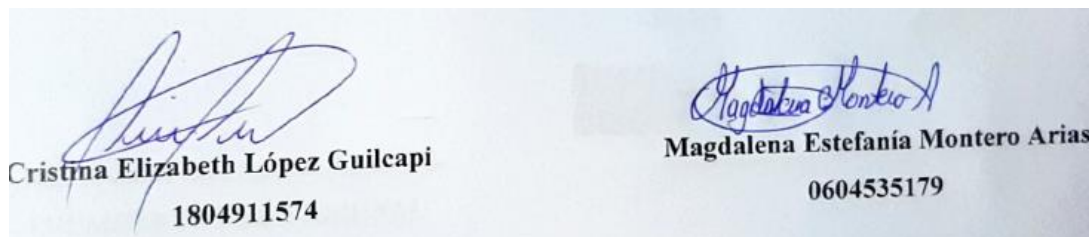
© 2022, Cristina Elizabeth López Guilcapi & Magdalena Estefanía Montero Arias

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo cita bibliográfica del documento, siempre y cuando se reconozca el Derecho de Autor.

Nosotras, CRISTINA ELIZABETH LÓPEZ GUILCAPI y MAGDALENA ESTEFANÍA MONTERO ARIAS, declaramos que el presente Trabajo de Titulación es de nuestra autoría y los resultados del mismo son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autoras asumimos la responsabilidad legal y académica de los contenidos de este Trabajo de Titulación; el patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo.

Riobamba, 18 de Julio de 2022






The image shows two handwritten signatures in blue ink. The signature on the left is for Cristina Elizabeth López Guilcapi, and the signature on the right is for Magdalena Estefanía Montero Arias. Below each signature is the corresponding printed name and a numerical identifier.

Cristina Elizabeth López Guilcapi
1804911574

Magdalena Estefanía Montero Arias
0604535179

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

El Tribunal del Trabajo de Titulación certifica que: El trabajo de titulación; Tipo: Proyecto de Investigación, **CLUSTERS ESPACIALES DE LA ECONOMÍA DE LOS SOCIOS DE LA COOP. “EDUCADORES DE CHIMBORAZO” LTDA (2016 -2020)**, realizado por las señoritas: **CRISTINA ELIZABETH LÓPEZ GUILCAPI Y MAGDALENA ESTEFANÍA MONTERO ARIAS**, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Titulación, el mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal Autoriza su presentación.

	FIRMA	FECHA
Ing. Johanna Enit Aguilar Reyes MSc. PRESIDENTE DEL TRIBUNAL	 Firmado digitalmente por: JOHANNA ENITH AGUILAR REYES	2022-07-18
Ing. Amalia Isabel Escudero Villa MSc. DIRECTORA DE TRABAJO DE TITULACIÓN	 Firmado digitalmente por AMALIA ISABEL ESCUDERO VILLA - NIF: Y8507318T Escudero Villa, Amalia Isabel VILLA - NIF: Y8507318T	2022-07-18
Dra. Jaqueline Elizabeth Balseca Castro MSc. MIEMBRO DEL TRIBUNAL	 Firmado digitalmente por JAQUELINE ELIZABETH BALSECA CASTRO Fecha: 2022.10.24 22:35:25 -05'00'	2022-07-18

DEDICATORIA

Esta tesis está dedicada a:

A mis padres Luis y Mayra quienes con su amor, paciencia y esfuerzo me han permitido llegar a cumplir hoy un sueño más, gracias por inculcar en mí el ejemplo de esfuerzo y valentía, de no temer las adversidades porque Dios está conmigo siempre. A mis hermanos Alexander y María Eliza por su cariño y apoyo incondicional, durante todo este proceso, por estar conmigo en todo momento gracias. A toda mi familia porque con sus oraciones, consejos y palabras de aliento hicieron de mí una mejor persona y de una u otra forma me acompañan en todos mis sueños y metas.

Finalmente quiero dedicar esta tesis a mis mejores amigas Roxana y Magdalena, por apoyarme cuando más las necesito, por extender su mano en momentos difíciles y por el amor brindado cada día, de verdad mil gracias hermanitas, siempre las llevo en mi corazón.

Cristina

El presente trabajo investigativo lo dedico principalmente a Dios, por ser el inspirador y darme fuerza para continuar en este proceso de obtener uno de los anhelos más deseados. A mis padres Juan y Sonia, por su amor, trabajo y sacrificio en todos estos años, gracias a ustedes he logrado llegar hasta aquí y convertirme en lo que soy. Ha sido un orgullo y privilegio de ser su hija, son los mejores padres. A mis hermanos por siempre estar presentes acompañandome y por el apoyo moral, que me brindaron a lo largo de esta etapa de mi vida. Finalmente quiero agradecer a mi mejor amiga Elizabeth por haberme acompañado a recorrer este camino apoyando mis decisiones siendo mi apoyo y fortaleza.

Magdalena

AGRADECIMIENTO

A nuestros padres, por ser los principales promotores de nuestros sueños y por la confianza depositada en nuestras expectativas profesionales.

A las autoridades y catedráticos que hacen el alma mater de la Escuela Superior Politécnica de Chimborazo, por la confianza y disposición para realizar todo el proceso educativo en la Facultad de Ciencias, en especial a la Ing. Isabel Escudero, Dra. Jaqueline Balseca, Ing. Alexandra Marcatoma, Ing. Patricia Paredes, Ing. Alexandra Viñan, por la paciencia, empatía, amistad y sobre todo por ser ejemplo a seguir por su profesionalismo y pasión al instruir a los estudiantes. Finalmente expresar un sincero agradecimiento al Ing. Fabián Tobar por el apoyo, colaboración brindada y disponibilidad de todos los miembros que conforman la Cooperativa de Ahorro y Crédito “Educadores de Chimborazo” Ltda.

Cristina & Magdalena

TABLA DE CONTENIDO

ÍNDICE DE TABLAS.....	viii
ÍNDICE DE FIGURAS.....	ix
ÍNDICE DE GRÁFICOS.....	x
ÍNDICE DE ECUACIONES	xii
ÍNDICE DE ANEXOS	xiii
ÍNDICE DE ABREVIATURAS.....	xiv
RESUMEN.....	xv
SUMMARY	xvi
INTRODUCCIÓN	1

CAPÍTULO I

1. MARCO TEÓRICO	7
1.1. Economía y conceptualizaciones fundamentales.....	7
1.1.1. Microeconomía	7
1.1.2. Macroeconomía	8
1.1.3. Ingresos	8
1.1.4. Egresos	8
1.1.5. Tasa de crecimiento anual.....	9
1.2. Demografía	9
1.2.1. Fuentes de datos demográficos	9
1.2.2. Variables Demográficas	10
1.3. Datos espaciales.....	10
1.3.1. Variables Espaciales	11
1.3.2. Geoestadística.....	12
1.3.3. Principios básicos en el tratamiento de datos espaciales	12
1.3.4. Heterogeneidad espacial.....	13
1.3.5. Autocorrelación espacial	14
1.3.6. Covarianza espacial	15
1.3.7. Tipos de análisis de datos espaciales.....	15
1.3.8. Estadísticos globales y locales de autocorrelación espacial	16
1.3.9. Técnicas de visualización de la presencia de efectos espaciales	19
1.3.10. Autocorrelación espacial en el modelo de regresión	20
1.4. Cartografía	23

1.4.1.	<i>Elementos esenciales en un mapa</i>	24
1.4.2.	<i>Los mapas y su clasificación</i>	24
1.5.	Técnicas Multivariantes	25
1.5.1.	<i>Análisis de componentes principales (PCA)</i>	26
1.5.2.	<i>Análisis de Clúster (Conglomerados)</i>	26
1.5.3.	<i>Medidas de similitud</i>	34
1.5.4.	<i>Espacios métricos</i>	34
1.5.5.	<i>Métricas</i>	35
1.5.6.	<i>Distancias</i>	36
1.5.7.	<i>Validación de resultados</i>	36
1.6.	Etapas del análisis de clúster o conglomerado	37
1.7.	Software R	38
1.7.1.	<i>Librerías para clúster</i>	38
1.8.	Requisitos para pertenecer a la CACECH	40

CAPÍTULO II

2.	MARCO METODOLÓGICO	41
2.1.	Tipo y diseño de la investigación	41
2.2.	Descripción del Área de estudio	41
2.3.	Población de estudio	42
2.4.	Operacionalización de las variables	43

CAPÍTULO III

3.	MARCO DE RESULTADOS Y DISCUSIÓN DE RESULTADOS	44
3.1.	Análisis Exploratorio de datos	44
3.2.	Análisis Exploratorio Espacial	59
3.3.	Análisis de Clustering Jerárquico	60

CONCLUSIONES	85
--------------	-------	----

RECOMENDACIONES	86
-----------------	-------	----

GLOSARIO

BIBLOGRAFÍA

ANEXOS

ÍNDICE DE TABLAS

Tabla 1-1:	Índice de autocorrelación espacial.....	17
Tabla 2-1:	Test de autocorrelación espacial.....	22
Tabla 3-1:	Funciones del paquete factextra.....	39
Tabla 1-2:	Operacionalización de las variables de los socios de la CACECH.....	43
Tabla 1-3:	Frecuencia de los socios de la CACEH en los cantones de Chimborazo.....	44
Tabla 2-3:	Frecuencia de los socios de la CACECH según el sexo.....	45
Tabla 3-3:	Frecuencia de los socios de la CACECH según la zona residencial.....	46
Tabla 4-3:	Frecuencia de los socios de la CACECH según su actividad económica.....	47
Tabla 5-3:	Frecuencia de los socios de la CACECH según la parroquia.....	47
Tabla 6-3:	Descriptivo anual de los ingresos de los socios de la CACECH (2016-2020)....	49
Tabla 7-3:	Descriptivo anual de los egresos de los socios de la CACECH (2016-2020).....	50
Tabla 8-3:	Ingresos promedios de los socios de CACECH (2016-2020).....	51
Tabla 9-3:	Egresos por actividad económica de los socios de CACECH (2016-2020).....	51
Tabla 10-3:	Egresos por zona de los socios de la CACECH (2016-2020).....	54
Tabla 11-3:	Egresos por sexo de los socios de la CACECH (2016-2020).....	55
Tabla 12-3:	Tasa de movimiento anual de los ingresos de los socios de la CACECH.....	58
Tabla 13-3:	Tasa de movimiento anual de los egresos de los socios de la CACECH.....	58
Tabla 14-3:	Índice de autocorrelación espacial del periodo 2016-2020.....	59
Tabla 15-3:	Presencia de ruido en los egresos de los socios de la CACECH (2016-2020)....	60
Tabla 16-3:	Clúster de los ingresos de los socios de la CACECH año 2020.....	63
Tabla 17-3:	Clúster de los egresos de los socios de la CACECH del año 2016.....	67
Tabla 18-3:	Clúster de los egresos de los socios de la CACECH del año 2017.....	70
Tabla 19-3:	Clúster de los egresos de los socios de la CACECH del año 2018.....	74
Tabla 20-3:	Clúster de los egresos de los socios de la CACECH del año 2019.....	77
Tabla 21-3:	Clúster de los egresos de los socios de la CACECH del año 2020.....	81

ÍNDICE DE FIGURAS

Figura 1-1: Parejas de clusters	30
Figura 2-1: Dendrograma.....	31

ÍNDICE DE GRÁFICOS

Gráfico 1-2:	Ubicación geográfica de la CACECH.....	42
Gráfico 2-2:	Mapa geográfico de la República del Ecuador.....	42
Gráfico 1-3:	Densidad cantonal de socios de la CACECH en Chimborazo	45
Gráfico 2-3:	Sexo de los socios de la CACECH.....	46
Gráfico 3-3:	Zona domiciliaria de los socios de la CACECH	46
Gráfico 4-3:	Frecuencia de socios de la CACECH según la actividad económica.....	47
Gráfico 5-3:	Frecuencia de socios de la CACECH por parroquias.....	49
Gráfico 6-3:	Ingresos por actividad económica de los socios de la CACECH (2016-2020)	51
Gráfico 7-3:	Egresos de docentes universitarios de la CACECH (2016-2020)	52
Gráfico 8-3:	Egresos de empleados privados de la CACECH (2016-2020).....	52
Gráfico 9-3:	Egresos de empleados públicos de la CACECH (2016-2020)	53
Gráfico 10-3:	Egresos de jubilados de la CACECH (2016-2020)	53
Gráfico 11-3:	Egresos de pensiones alimenticias de la CACECH (2016-2020).....	54
Gráfico 12-3:	Egresos por zona de los socios de la CACECH (2016-2020)	54
Gráfico 13-3:	Egresos por sexo de los socios de la CACECH (2016-2020).....	55
Gráfico 14-3:	Serie temporal de los ingresos de los socios de la CACECH (2016-2020).....	56
Gráfico 15-3:	Ingresos de los socios de la CACECH (2016-2020)	56
Gráfico 16-3:	Serie temporal de los egresos de los socios de la CACECH (2016-2020)	57
Gráfico 17-3:	Egresos de los socios de la CACECH (2016-2020)	57
Gráfico 18-3:	Ubicación geográfica de los socios en el cantón Riobamba.....	59
Gráfico 19-3:	PCA de los ingresos de los socios de la CACECH del año 2020.....	61
Gráfico 20-3:	Homogeneidad de los ingresos de los socios de la CACECH del año 2020	61
Gráfico 21-3:	Biplot de los ingresos de los socios de la CACECH del año 2020	62
Gráfico 22-3:	Dendrograma de los ingresos de los socios de la CACECH del año 2020.....	62
Gráfico 23-3:	Clúster de los ingresos de los socios de la CACECH del año 2020.....	63
Gráfico 24-3:	Lugar geográfico de los ingresos de los socios de la CACECH del año 2020.	64
Gráfico 25-3:	PCA de los egresos de los socios de la CACECH del año 2016.....	65
Gráfico 26-3:	Homogeneidad de los egresos de los socios de la CACECH del año 2016	65
Gráfico 27-3:	Biplot de los egresos de los socios de la CACECH del año 2016.....	66
Gráfico 28-3:	Dendrograma de los egresos de los socios de la CACECH del año 2016.....	66
Gráfico 29-3:	Clúster de los egresos de los socios de la CACECH del año 2016	67
Gráfico 30-3:	Lugar geográfico de los egresos de los socios de la CACECH del año 2016 ..	68
Gráfico 31-3:	PCA de los egresos de los socios de la CACECH del año 2017	68

Gráfico 32-3:	Homogeneidad de los egresos de los socios de la CACECH del año 2017	69
Gráfico 33-3:	Biplot de los egresos de los socios de la CACECH del año 2017	69
Gráfico 34-3:	Dendrograma de los egresos de los socios de la CACECH del año 2017	70
Gráfico 35-3:	Clúster de los egresos de los socios de la CACECH del año 2017	71
Gráfico 36-3:	Lugar geográfico de los egresos de los socios de la CACECH del año 2017 ..	71
Gráfico 37-3:	PCA de los egresos de los socios de la CACECH del año 2018	72
Gráfico 38-3:	Homogeneidad de los egresos de los socios de la CACECH del año 2018	72
Gráfico 39-3:	Biplot de los egresos de los socios de la CACECH del año 2018	73
Gráfico 40-3:	Dendrograma de los egresos de los socios de la CACECH del año 2018	73
Gráfico 41-3:	Clúster de los egresos de los socios de la CACECH del año 2018	74
Gráfico 42-3:	Lugar geográfico de los egresos de los socios de la CACECH del año 2018 ..	75
Gráfico 43-3:	PCA de los egresos de los socios de la CACECH del año 2019	75
Gráfico 44-3:	Homogeneidad de los egresos de los socios de la CACECH del año 2019	76
Gráfico 45-3:	Biplot de los egresos de los socios de la CACECH del año 2019	76
Gráfico 46-3:	Dendrograma de los egresos de los socios de la CACECH del año 2019	77
Gráfico 47-3:	Clúster de los egresos de los socios de la CACECH del año 2019	78
Gráfico 48-3:	Lugar geográfico de los egresos de los socios de la CACECH del año 2019 ..	78
Gráfico 49-3:	PCA de los egresos de los socios de la CACECH del año 2020	79
Gráfico 50-3:	Homogeneidad de los egresos de los socios de la CACECH del año 2020	79
Gráfico 51-3:	Biplot de los egresos de los socios de la CACECH del año 2020	80
Gráfico 52-3:	Dendrograma de los egresos de los socios de la CACECH del año 2020	80
Gráfico 53-3:	Clúster de los egresos de los socios de la CACECH del año 2020	81
Gráfico 54-3:	Lugar geográfico de los egresos de los socios de la CACECH del año 2020 ..	82

ÍNDICE DE ECUACIONES

Ecuación 1-1:	Tasa de crecimiento anual.....	9
Ecuación 2-1:	Covarianza espacial	15
Ecuación 3-1:	Índice Global de Moran	18
Ecuación 4-1:	Variable endógena	20
Ecuación 5-1:	Ruido blanco	20
Ecuación 6-1:	Proceso de medias móviles espacial de orden 1	21
Ecuación 7-1:	Función de verosimilitud logarítmica	22
Ecuación 8-1:	Método de Ward	29
Ecuación 9-1:	Distancia Euclídea	36
Ecuación 10-1:	Distancia Manhattan	36
Ecuación 11-1:	Test estadístico de Hopkins	37

ÍNDICE DE ANEXOS

ANEXO A: CÓDIGO EN R DE CLÚSTER.

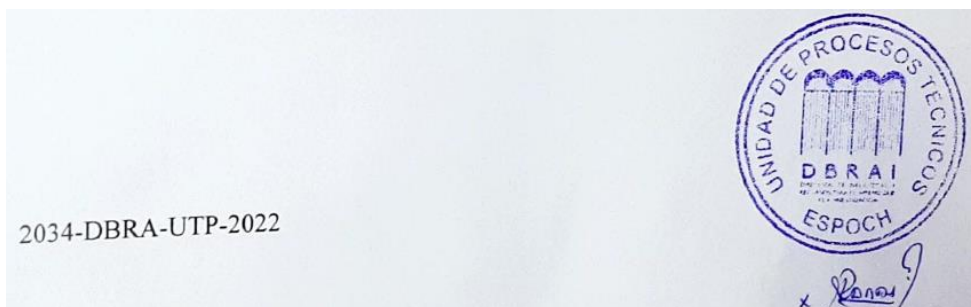
ÍNDICE DE ABREVIATURAS

AEDE	Análisis Exploratorio de Datos Espaciales
CA	Análisis de Correspondencias
CACECH	Cooperativa de Ahorro y Crédito “Educadores de Chimborazo”
FAMD	Análisis Factorial de Datos Mixtos
HMFA	Análisis Jerárquico de Factores Múltiple
MCA	Análisis de Correspondencia Múltiple
MFA	Análisis de Factores Múltiples
PAM	Particiones Alrededor Medoids
PCA	Análisis de Componentes Principales

RESUMEN

El presente trabajo tuvo como objetivo encontrar patrones de comportamiento de la economía de los socios de la Cooperativa de Ahorro y Crédito “Educadores de Chimborazo” durante el periodo 2016 - 2020, a través de la aplicación de técnicas estadísticas. La matriz de información fue proporcionada por el analista informático del departamento de sistemas de la entidad financiera, donde se almacena el historial económico, demográfico y geográfico de los asociados. Para el estudio se consideraron las variables: ingresos, egresos, coordenadas de longitud y latitud, sexo, parroquia, zona y actividad económica. Se realizó un análisis descriptivo de las variables en estudio para observar el comportamiento de los datos, seguidamente se aplicó un análisis temporal para verificar el procedimiento histórico de las cifras económicas a través del tiempo, luego se realizó un análisis espacial para observar si existe relación o no entre las variables económicas y geográficas objeto de estudio, por último se aplicó un análisis de clúster con técnicas multivariantes con el objetivo de representar la clasificación de conglomerados según las características del comportamiento existente entre las variables económicas y demográficas propuestas en el presente estudio. Se obtuvo como resultados que las variables ingresos y egresos presentaron cambios inesperados en el año 2020 debido al confinamiento y la crisis sanitaria por la pandemia Covid-19; el índice de Moran identificó que la economía de los accionistas de la CACECH presentó una autocorrelación negativa perfecta; la aplicación del análisis de clúster a través del mapeo de clasificación generó resultados óptimos para el estudio ya que, se logró identificar mediante conglomerados que en las parroquias Lizarzaburu y Velasco existe una mayor concentración de egresos. Se concluye que la clasificación de conglomerados fue mediante la ubicación geográfica y la actividad económica. Se recomienda aplicar técnicas multivariantes en investigaciones futuras para contribuir con un enfoque espacial en la toma de decisiones.


Palabras clave: <ANÁLISIS ESPACIAL>, <CLUSTERS ESPACIALES>, <CONGLOMERADO>, <ECONOMÍA>, <ESTADÍSTICA ESPACIAL>, <PATRONES DE COMPORTAMIENTO>, <TÉCNICAS MULTIVARIANTES>.



SUMMARY

The objective of this work was to find patterns of behavior of the economy of members of the Savings and Credit Cooperative "Educadores de Chimborazo" during the period 2016 - 2020, through the application of statistical techniques. The information matrix was provided by the computer analyst of the systems department of the financial institution, where the economic, demographic and geographical history of the associates is stored. For the study, the variables: income, expenses, coordinates of longitude and latitude, sex, parish, area and economic activity were considered. A descriptive analysis of the variables in study to observe the behavior of the data, then a temporal analysis was applied to verify the historical procedure of the economic figures through time, then performed a spatial analysis to observe whether or not there is a relationship between the economic variables and geographical areas under study, finally a cluster analysis was applied through multivariate techniques with the aim of representing the classification of conglomerates according to the characteristics of the existing behavior between the economic and demographic variables proposed in this study. It was obtained as results that the variables income and expenses presented unexpected changes in 2020 due to confinement and the health crisis due to the Covid-19 pandemic; the Moran's I identified that the economy of the shareholders of the CACECH presented a perfect negative autocorrelation; the application of cluster analysis through classification mapping generated optimal results for the study since it was identified through conglomerates that in the Lizarzaburu and Velasco parishes there is a higher concentration of expenses. It is concluded that the classification of conglomerates was by geographic location and economic activity. It is recommended to apply multivariate techniques in future research to contribute to a spatial approach to decision making.

Keywords: <SPATIAL ANALYSIS>, <SPATIAL CLUSTERS>, <CLUSTER>, <ECONOMICS>, <SPATIAL STATISTICS>, <PATTERNS OF BEHAVIOR>, <MULTIVARIATE TECHNIQUES>.



Edgar Mesias Jaramillo Moyano
0603497397

INTRODUCCIÓN

En los últimos años, el Ecuador ha sufrido varias transformaciones económicas, sociales, políticas y de salud. La población se vio aún más afectada de manera directa ante la presencia del COVID-19. Este problema ha generado incertidumbre económica y de salud, debido a su vulnerabilidad ante la pérdida de seres queridos que tienen como consecuencias el empeoramiento de las condiciones sociales de vida, mismo que produce: inseguridad económica, falta de recursos y desempleo (Tulcanazo, 2022, p.1).

La motivación principal de este trabajo de investigación es conocer el comportamiento de la economía de los socios de la Cooperativa de Ahorro y Crédito “Educadores de Chimborazo”, y como se ha regido antes y durante el primer periodo de pandemia, dando paso a un análisis espacial enfocado en las correlaciones de las ubicaciones geográficas y los movimientos financieros de los accionistas, considerando principios fundamentales de análisis multivariado con datos espaciales, para el estudio de clúster que permite identificar en que sector de la ciudad se encuentra la mayor concentración de los asociados de la institución financiera y sus características.

Las técnicas para el análisis estadístico espacial son herramientas que utilizan la ubicación geográfica de los datos, que describen los métodos desde el abordaje descriptivo y analítico, haciendo énfasis en la observación, correlación espacial, identificación de clusters sociodemográficos y uso de mapas (Valbuena y Rodríguez, 2018: p.359).

Para la identificación de conglomerados espaciales (clusters) del estado económico en la Cooperativa de Ahorro y Crédito “Educadores de Chimborazo” se integra en tres secciones: en la primera se argumentan la distribución y naturaleza de los procesos abordados en esta investigación desde la perspectiva de la ciencia de los planteamientos de la geoestadística sobre la económica. El segundo apartado introduce al tema de estudio a través de un diagnóstico de la dependencia espacial. En la tercera se analiza la heterogeneidad espacial y una exploración geográfica para identificar los clusters espaciales jerárquicos aglomerativos. Finalmente, se resumen los hallazgos de los conglomerados económicos de los socios de la cooperativa en el periodo 2016-2020.

ANTECEDENTES

La estadística espacial en sus inicios apareció como la aplicación de métodos detallados clásicos a la geografía. Uno de sus pioneros fue David en 1977, quien más tarde desarrolló un trabajo basado en sus propios métodos estadísticos, y los aplicó a datos geográficos con referencia a Brian Ripley, en 1981 con Furingham y Robert Hoening. Se aplican métodos espaciales, clasificación de tres áreas principales (Cressie, 1991), datos geoestadísticos, correlación espacial (Índice de Moran) y clusters espaciales (Bohórquez y Ceballos, 2008: p.12).

El análisis de clusters es el estudio de algoritmos y métodos cuyo objetivo es encontrar una forma conveniente y válida de organizar un conjunto de datos en grupos para una mejor comprensión de la distribución según características similares partiendo de esta idea múltiples aplicaciones, se encuentran la segmentación de imágenes, la clasificación automática de documentos o archivos multimedia, la detección de comunidades en redes sociales, la identificación de genes con funciones similares, entre otros. El desarrollo de la metodología de clustering ha sido interdisciplinario, investigadores como taxonomistas, psicólogos, biólogos, estadísticos, cuentistas sociales, ingenieros e informáticos, etc., de muchas áreas han contribuido en la propagación de investigación de clasificación de datos. Por ello es por lo que el clustering se puede encontrar bajo diferentes nombres en diferentes contextos, como aprendizaje no supervisado (*en Machine Learning y Pattern Recognition*), taxonomía numérica (*en Biología, Ecología*), tipología (*en Ciencias Sociales*), partición (*en Teoría de Grafos*), mecanismos de segmentación (*Marketing*) y algoritmos de clasificación (*en Estadística*). Además de diferentes terminologías, estas comunidades tienen otros supuestos acerca de los componentes del proceso y el contexto en el que se usa (Vidal y Favaro, 2014: p.49).

En el siglo XIX emergió la economía espacial como la distribución geográfica en espacio y tiempo desde un enfoque de similitud económica, en un principio, el análisis del vínculo entre geografía y economía se estudió haciendo alusión a elementos geográficos como el clima, la vegetación y una serie de condicionantes geográficas que tienen influencia en el crecimiento; sin embargo, en años recientes el estudio de esta relación ha adquirido una connotación metodológica diferente. Es decir, la reconsideración y revalorización del espacio en el análisis económico ha traído la aparición de un nuevo campo teórico que retoma los estudios pioneros de la teoría de la localización, la geografía económica y la ciencia regional. La síntesis de estos enfoques ha madurado en una nueva formulación metodológica que considera un conjunto de técnicas basadas en la geoestadística, la econometría

espacial, la simulación matemática y los Sistemas de Información Geográfica (SIG). Debido a lo anterior, el crecimiento económico regional es reconsiderado como un proceso espacial que se esparce por el territorio y está influenciado por el desempeño económico de las áreas vecinas, es decir, existe correlación espacial. Por ello, para conocer un poco más sobre cómo este proceso se contagia a regiones vecinas se realiza un Análisis Exploratorio de Datos Espaciales (ESDA, por sus siglas en inglés) para investigar la relación existente entre geografía y crecimiento económico (Rodríguez y Cabrera, 2017: p.709).

Alex Edward Tristán Gómez indica en la tesis *“Identificación de conglomerados espaciales de acuerdo con niveles de morosidad de empresas en el Perú”* la importancia de detectar los niveles de riesgo de morosidad en las empresas, enfocando la identificación de conglomerados de provincias del Perú, en función de la tasa de incumplimiento de pagos. Para ello propone un modelamiento en dos niveles. En el primer nivel se usan modelos aglomerativo jerárquicos para seleccionar n conglomerados candidatos a priori, donde el número final de aglutinados se escoge mediante criterios de selección de modelos y en un segundo nivel, modela el nivel de riesgo haciendo uso del modelo de Poisson y prioris condicionales autoregresivas en base a los conglomerados definidos en el primer nivel. Finalmente, como resultado de la aproximación se obtienen conglomerados de provincias de acuerdo con sus niveles de morosidad, permitiendo clasificar las provincias en conglomerado de alto, medio y bajo nivel de riesgo de morosidad (Gómez, 2021, p.4).

Gilberto Martínez Sidón y Salvador Corrales comentan que los principales encadenamientos y clusters a través de un análisis del tejido sectorial en la economía, en donde asumió la hipótesis que en Nuevo León se han creado importantes cadenas productivas y clusters entre sus principales subsectores económicos, estrategias de organización y de fomento económico fundamentales para generar mayor valor agregado, resultado final de una economía regional exitosa (Martínez y Corrales, 2017: pp.41-42).

Carlos Vilalta comenta en el artículo científico *“¿Se pueden predecir geográficamente los resultados electorales? Una aplicación del análisis de clusters y outliers espaciales”* que el estudio demuestra que al aplicar la estadística espacial en la geografía electoral es posible predecir los resultados electorales. Se utilizan los conceptos geográficos de clúster y outlier espaciales, y como variable predictiva la segregación espacial socioeconómica. Las técnicas estadísticas que se emplean son los índices globales y locales de autocorrelación espacial de Moran y el análisis de regresión lineal (Vilalta 2008, p.572).

Juan Barón Rivera identifica las regiones económicas de Colombia a través del análisis de clusters. Aplicando esta metodología sobre una variable que refleja el desempeño económico de los departamentos a través del tiempo enfatizando que una característica fundamental de esta clasificación es que los departamentos que integran una región económica no necesariamente son vecinos geográficos (Barón y Rivera, 2002: p.3).

La literatura en Ecuador es limitada así lo definía Cisneros et al. en su trabajo:

En donde desarrollan clúster de tipo jerárquico en base a valores procesados que permite detectar y agrupar variables heterogéneas, logrando homogenizarlas en grupos con características similares. Para analizar la expansión competitiva de calzado en el cantón Cevallos y así, especializar su producción, mejorar sus procesos e innovar a un nivel industrial con competidores nacionales e internacionales (Cisneros et al., 2018: pp.300-310).

Capas y otros autores explican en el artículo científico “*El Clúster: una alternativa para la competitividad de las PYMES de banano orgánico en Ecuador*” que el clúster es una forma de organización múltiple que puede resistir la competencia del mercado y ocupar un lugar importante en el desarrollo regional donde el objetivo es analizar sistemáticamente cómo las estrategias de clúster pueden ser una alternativa para la competitividad de las PYMES de banano orgánico. Los métodos que emplean son la historia lógica, el análisis y la síntesis, así como el análisis documental. Como resultado, el clúster es una alternativa de desarrollo y competitividad para las PyMEs del Distrito 7 que presenta una perspectiva de análisis de conglomerados para promover la competitividad de las PYMES bananeras orgánicas ecuatorianas (Benítez et al., 2017: p.682).

PLANTEAMIENTO DEL PROBLEMA

Enunciado del problema

En la actualidad la economía de todo el mundo, como la del Ecuador se encuentra afectada por la presencia de la pandemia provocada por el SARS-CoV-2 (Síndrome respiratorio agudo severo coronavirus 2), con el aislamiento y el confinamiento inició un declive económico en la sociedad, provocando el desempleo y falta de recursos, presentando un fenómeno que motiva investigar y caracterizar el efecto de la pandemia sobre la economía de los socios de la Cooperativa de Ahorro y Crédito “Educadores de Chimborazo”.

Formulación (Incógnita)

¿Los clusters espaciales permitirán caracterizar el efecto económico en los socios de la Cooperativa de Ahorro y Crédito “Educadores de Chimborazo” debido al SARS-CoV-2 (Síndrome respiratorio agudo severo coronavirus 2)?

JUSTIFICACIÓN

La crisis financiera generada y avivada por la presencia del SARS-CoV-2 (Síndrome respiratorio agudo severo coronavirus 2) ha generado el incremento de desempleo, pérdidas monetarias obligando a empresarios y a microempresarios a cerrar sus empresas paralizando la economía local y nacional.

La Cooperativa de Ahorro y Crédito “Educadores de Chimborazo” en la actualidad no cuenta con estudios específicos sobre el efecto que ha producido el SARS-CoV-2 (Síndrome respiratorio agudo severo coronavirus 2) en su primera etapa (2020), es relevante un análisis estadístico basado en correlaciones espaciales que permita identificar si existe relación entre la ubicación geográfica, la actividad económica del cliente y la influencia del covid-19.

El análisis mediante clúster es fundamental para identificar similitudes en la economía de los socios de la Cooperativa de Ahorro y Crédito “Educadores de Chimborazo” de acuerdo a la ubicación geográfica y características demográficas, además de la segmentación por conglomerados de ingresos y egresos del periodo 2016-2020 generando una clasificación de acuerdo a las características de los asociados de la institución financiera con el uso de técnicas estadísticas enfocadas en un análisis espacial.

OBJETIVOS

General

Identificar patrones de comportamiento en la economía de los socios de la Cooperativa “Educadores de Chimborazo” mediante un análisis de clúster espaciales en función de las variables demográficas y geográficas.

Específicos

Identificar la posición geográfica de los socios de la Cooperativa “Educadores de Chimborazo”.

Realizar un análisis estadístico descriptivo de las variables económicas, demográficas y espaciales de los socios de la cooperativa.

Identificar la relación de las variables económicas y demográficas en función de la ubicación demográfica de los socios.

Identificar patrones de comportamiento mediante clusters espaciales.

CAPÍTULO I

1. MARCO TEÓRICO

1.1. Economía y conceptualizaciones fundamentales

La economía nació en 1776, con la publicación del libro “La riqueza de las naciones” de Adam Smith, desde entonces se han desarrollado múltiples teorías útiles; pero sigue la búsqueda de respuestas para muchos problemas financieros. La economía es el estudio de la manera en que los agentes deciden asignar recursos limitados y cómo esas decisiones afectan a la sociedad (Acemoglu et al., 2017: p.41).

Como era de esperar, esta definición pone el énfasis en las decisiones. La definición también tiene en cuenta que esas decisiones afectan a la sociedad. Por ejemplo, la venta de un coche deportivo no sólo tiene repercusiones en la persona que sale montado en él del concesionario. La venta genera un impuesto sobre las compras, que recauda el Estado, el cual a su vez financia proyectos como autopistas y hospitales. La adquisición del coche también produce algo de congestión del tráfico, pues se incorporará a los atascos de las horas punta. Y es otro vehículo que tal vez ocupe el último aparcamiento que quedaba en tu calle. Si se conduce de forma temeraria, el coche puede implicar riesgos para otros conductores. Además, será fuente de contaminación. En economía se estudia la decisión original y sus múltiples consecuencias para otras personas. Para comprender el alcance de la economía hay que hacer otra distinción. La economía puede dividirse en dos grandes disciplinas de estudio, aunque muchos economistas trabajan en ambas (Acemoglu et al., 2017: pp.41-42).

1.1.1. *Microeconomía*

Es el estudio de cómo toman decisiones los individuos, las familias, las empresas y los gobiernos y de cómo esas decisiones afectan a los precios, la asignación de recursos y el bienestar de otros agentes. Por ejemplo, en microeconomía se diseñan políticas para reducir la contaminación. Como el calentamiento global está causado en parte por las emisiones de carbono producidas por el carbón, el petróleo y otros combustibles fósiles, los microeconomistas diseñan políticas para reducir su uso. Los microeconomistas tienen el encargo de diseñar impuestos al carbono y de determinar cómo afectarían al uso de la energía por parte de familias y empresas. En general, se les consulta siempre que se desea comprender una parte pequeña de la economía general (Acemoglu et al., 2017: p.43).

1.1.2. Macroeconomía

Es el estudio de la economía en su conjunto, es decir, de los fenómenos que afectan a toda una economía, como la tasa de crecimiento de la producción económica total de un país, o el aumento porcentual del conjunto de los precios (la tasa de inflación), o la proporción de población activa que busca trabajo pero no lo encuentra (el nivel de desempleo). Los macroeconomistas diseñan políticas públicas que mejoren el comportamiento general, o «agregado», de una economía. Por ejemplo, la macroeconomía intenta identificar las mejores políticas para estimular una economía que viene sufriendo un periodo prolongado de crecimiento negativo, es decir, una recesión. Así, durante la crisis financiera de 2007–2009, cuando los precios de la vivienda caían en picado y los bancos quebraban o eran intervenidos, los macroeconomistas estaban muy ocupados tratando de explicar por qué la economía se contraía y de recomendar políticas que la reactivaran (Acemoglu et al., 2017: p.43).

1.1.3. Ingresos

Se asigna a las percepciones que se obtienen por la explotación del negocio, podemos decir que es la recaudación de dinero por la ejecución de las operaciones de venta o servicios que realiza el negocio, es decir, se percibe en forma regular y permanente, y se representa en la cuenta llamada “Ventas”. Las otras partidas, que también constituyen una entrada de dinero, pero que se perciben como resultado de operaciones aisladas y ajenas a la naturaleza de este, se califican como otros ingresos (Ramírez, 2018, p.36).

1.1.4. Egresos

Se entienden como los gastos que el negocio realiza para la ejecución de su actividad económica. Existen clasificaciones específicas de gastos propios del negocio, se conocen como gastos de operación que, a su vez, se dividen en gastos de venta, pago de comisionistas y gastos por el mantenimiento de las unidades de transporte. También existen los gastos que no están directamente relacionados con la actividad de venta, se les conoce como gastos de administración (Ramírez, 2018, p.37).

1.1.5. Tasa de crecimiento anual

La tasa de crecimiento, también conocida como tasa de variación (positiva), es el cambio positivo en porcentaje de una variable entre dos momentos distintos del tiempo (López, 2019, p.1).

Tasa de variación o crecimiento anual: Expresa la variación media acumulada en porcentaje por cada subperiodo entre dos fechas. Sin embargo, a diferencia de la tasa de variación del periodo que expresa la variación total, la tasa de variación anual expresa cuánto ha variado por subperiodo durante dos fechas.

Ecuación 1-1: Tasa de crecimiento anual

$$\text{Tasa de crecimiento anual} = \left(\left(\frac{f}{s} \right)^{\frac{1}{y}} - 1 \right) * 100$$

Fuente: Lewis, 2018, p.4-6

donde f es el valor final, s es el valor inicial y y es la cantidad de años (Lewis, 2018, pp.4-6).

1.2. Demografía

Es la ciencia que se encarga de la cuantificación de aspectos poblacionales, tales como: volumen, estructura y distribución geográfica; así como los cambios que estos aspectos tienen en el tiempo. Estos cambios son debidos a tres fenómenos básicos: fecundidad, mortalidad y migraciones (Palladino, 2010, pp.2).

1.2.1. Fuentes de datos demográficos

- Fuentes Primarias: Es cuando los datos han sido obtenidos y elaborados por el propio investigador o usuario de la información. En salud esta situación se da cuando se conocen los “límites” de la población de referencia; es decir, se puede determinar esta población por delimitación geográfica o por conocerse a los usuarios de un servicio (Arretx et al., 1983: pp.59-65).
- Fuentes Secundarias: Son a las que con mayor frecuencia se recurren. Básicamente, los censos de población y los registros. A los primeros, en nuestros países (América Latina) se los denominan Censo de Población y Vivienda, por recoger datos acerca de ambas características poblacionales. En ellos se recogen información “individualizada” de todas las personas que han pernoctado la noche anterior en el domicilio censado (esto se llama “de juri”). Además de las características

básicas de la ciudad y de vivienda, se recoge datos respecto a ocupación, mortalidad, fecundidad, etc. Estas últimas sirven para completar las que proveen los registros y, eventualmente, corregirlas. El censo es una fuente de datos poblacionales muy completa; aunque su limitante es la periodicidad de su ejecución. Los registros, en tanto, recogen información de manera permanente (Arretx et al., 1983: pp.59-65).

1.2.2. Variables Demográficas

La segmentación demográfica consiste en dividir el mercado en grupos más pequeños, este tipo de desmembración toma en cuenta variables como la edad, ingresos, grado de estudios, nacionalidad, raza, religión, ocupación, etc. Esto ayuda a las organizaciones a comprender con exactitud el comportamiento del consumidor, lo que a su vez les ayuda a desempeñarse mejor. Uno de los tipos de segmentación que está estrechamente relacionado con la desmembración demográfica es la segmentación geográfica. Dentro de esta, existen varias variables. Estas variables son las siguientes (Zavala, 2019, p.21):

- Segmentación demográfica basada en la edad: Una de las variables más importantes para la segmentación demográfica es la edad. Una generación es un conjunto de personas que nacieron durante la misma época, crecieron con el mismo tipo de experiencias con cierta segmentación geográfica (Zavala, 2019, p.22).
- Segmentación demográfica basada en el género: Los hombres y las mujeres tienen diferencias en cómo perciben las formas en que función del factor económico basado en ingresos y egresos mensuales. Incluso los investigadores son muy claros sobre la diferencia en el razonamiento de ambos sexos y, por lo tanto, fabrican productos específicos para cada género (Zavala, 2019, p.22).
- Segmentación demográfica basada en los ingresos: Los ingresos también son una de las variables importantes, ya que esta ayuda a determinar la economía del individuo (socios).

1.3. Datos espaciales

Un dato espacial es aquel dato que tiene asociada una referencia geográfica directa a través de unas coordenadas o una referencia geográfica indirecta como un código postal, de tal modo que se puede localizar exactamente dónde sucede dentro de un mapa. Los datos espaciales aportan información sobre hechos vinculados con objetos que tienen una extensión en el espacio, entendiendo por

extensión el área que cubre un objeto y que se caracteriza por un tamaño, una posición y una forma (Valbuena y Rodríguez, 2018: p.359).

Los datos espaciales deben ser creados de manera que se facilite su disponibilidad, acceso, interoperabilidad y aplicación para infinidad de propósitos y que, para ello, constituye en el proceso productivo y en el uso de los datos, un prerequisite, el estricto cumplimiento de la aplicación de normas, como las normas ISO/TC 211, sin las cuales los datos espaciales terminarían siendo subutilizados y las bases de datos de las que formen parte, elementos aislados y de utilidad restringida (Valbuena y Rodríguez, 2018: p.360).

- Latitud (ϕ o \emptyset): Es la distancia angular que existe entre el paralelo del Ecuador y un punto cualquiera de la superficie terrestre medida sobre el arco de meridiano que pasa por dicho punto. Dependiendo del hemisferio en el que se encuentre el punto considerado, la latitud puede ser norte o sur. Máxima latitud: coincide con el polo (90° norte o sur). Mínima latitud: coincide con el Ecuador (0°) (Auletta et al., 2020: p.19).
- Longitud (λ): Es la distancia angular que existe ente el meridiano de Greenwich y un punto cualquiera de la superficie terrestre medida sobre el arco del paralelo que pasa por dicho punto. Dependiendo del hemisferio en el que se encuentre el punto considerado, la longitud puede ser oeste o este. Máxima longitud: coincide con el antimeridiano de Greenwich (180°).Mínima longitud: coincide con el meridiano de Greenwich (0°) (Auletta et al., 2020: p.19).

1.3.1. Variables Espaciales

Dentro del conjunto de variables que inciden en la delimitación del espacio personal, el entorno físico y sus características pueden jugar un papel determinante. En efecto, ciertas disposiciones ambientales pueden “fijar” distancias interpersonales y, por tanto, definir las interacciones sociales (Pérez, 2006, p.134).

Se debe considerar que el dato espacial es un dato cualquiera sin ninguna peculiaridad supone no realizar sobre él un análisis óptimo. Las características propias de los datos espaciales dotan a estos de una gran potencialidad de estudio, al tiempo que condicionan o limitan otras operaciones. Asimismo, estas particularidades son el origen de una gran parte de los retos aún existentes dentro de la exposición geográfica, y por sus implicaciones directas no pueden desestimarse sin más. Su conocimiento es, por tanto, imprescindible para todo tipo de investigación espacial. El carácter

especial del dato espacial deriva de la existencia de posición. Esta se ha de entender tanto en términos absolutos (posición de una entidad en el espacio expresada por sus coordenadas) como relativos (relación con otras entidades también en dicho espacio). Las consecuencias de que todo dato espacial se halle por definición localizado a través de coordenadas son diversas, y deben enfocarse desde los distintos puntos de vista del análisis espacial (Pérez, 2006, p.134).

1.3.2. Geoestadística

La geoestadística se refiere al conjunto de técnicas estadísticas utilizadas para el estudio de variables numéricas distribuidas en el espacio. Para Journel y Huijbregts (1978), es simplemente la aplicación de la teoría de funciones aleatorias al reconocimiento y estimación de fenómenos naturales. La geoestadística se ha convertido en una herramienta útil para explicar la variación de propiedades en el espacio y su uso se ha generalizado recientemente donde se analizan hechos geográficos, debido a la posibilidad de conocer, entre otros, el rango de influencia de la propiedad, estimar su valor en sitios donde no existe información, por último, calcular el error de la estimación efectuada (González et al., 2007: p.82).

1.3.3. Principios básicos en el tratamiento de datos espaciales

Paelinck y Klaassen (1979) destacan cuatro principios básicos en el campo de la econometría espacial y el tratamiento de datos de corte transversal en general:

- **Interdependencia:** todo modelo espacial ha de caracterizarse por su interdependencia, es decir, deben incorporarse relaciones mutuas entre las observaciones de las variables monetarias, sociales, demográficas, etc. Por ejemplo, se puede pensar en la interdependencia espaciales de gastos y de ingresos por la que las maximizaciones de ingresos en un área pueden conllevar mayores gastos no sólo en esa área sino también en áreas vecinas, favoreciendo el crecimiento financiero de éstas últimas.
- **Asimetría:** las relaciones espaciales son, en principio, asimétricas. Basta con pensar en los comportamientos de compra allá donde exista una jerarquía de centros, lo que implica que la probabilidad de que un residente de las afueras de la urbe vaya a comprar al centro es superior a la posibilidad de que un residente en la metrópolis vaya a comprar a las afueras.
- **Alotopía:** se ha de buscar a priori “la causa” de un fenómeno espacial en otro lugar. Así, por ejemplo, los fenómenos migratorios se explican por la comparación de ventajas e inconvenientes no

sólo en los espacios de origen y de destino sino en causas que se encuentran en otros lugares del espacio.

- No linealidad: de soluciones espaciales óptimas ex-ante conduce a modelos econométricos ex-post que requieren una atención particular en lo que respecta a su especificación, la cual generalmente será no lineal.

Según Paelinck y Klaassen (1979), no siempre es posible observar estos cuatro principios de construcción de modelos espaciales y probablemente puede haber otros además de los aquí especificados (Serrano y Valcarce, 2000: pp.19-20).

1.3.4. Heterogeneidad espacial

La heterogeneidad espacial consiste en la variación de las relaciones en el espacio. De forma general, se puede decir que existen dos aspectos distintos de diversidad espacial como la inestabilidad estructural y la heteroscedasticidad. En el primer caso, la complejidad espacial se refiere a la falta de estabilidad en el espacio del comportamiento de la variable bajo estudio que lleva a que la forma funcional y los parámetros de una regresión puedan variar según la localización siendo, por tanto, no homogéneos en toda la muestra. El segundo aspecto, la heteroscedasticidad, proviene de la omisión de variables u otras formas de errores de especificación que llevan a la aparición de errores de medida.

El tratamiento de la heterogeneidad espacial puede realizarse por medio de las técnicas econométricas estándar, entre otras, la consideración explícita de parámetros variantes, coeficientes aleatorios (Hildreth, y otros, 1968) y cambiar regresiones o las técnicas de filtraje adaptativo cósmica. Sin embargo, en el caso específico de la heterogeneidad espacial, en donde ésta viene causada por cuestiones inherentes a la localización de las observaciones, el conocimiento teórico de la estructura espacial de los datos puede conducirnos a procedimientos más complejos, pero más eficientes como la expansión cósmica de parámetros (Casetti, 1972) o las regresiones ponderadas geográficamente (Fotheringham, y otros, 1998). Sin embargo, dado que este efecto puede tratarse en la mayoría de los casos mediante las técnicas econométricas tradicionales, el presente trabajo se centra en la problemática en torno a la autocorrelación espacial (Serrano y Valcarce, 2000: pp.20-21).

1.3.5. Autocorrelación espacial

La dependencia o autocorrelación espacial aparece como consecuencia de la existencia de una relación funcional entre lo que ocurre en un punto determinado del espacio y lo que ocurre en otro lugar (Serrano y Valcarce, 2000: p.21).

La autocorrelación espacial puede ser positiva o negativa. Si la presencia de un fenómeno determinado en una región lleva a que se extienda ese mismo anómalo hacia el resto de las regiones que la rodean, favoreciendo así la concentración de este, se halla ante un caso de auto reciprocidad positiva. Por el contrario, existirá auto reciprocidad negativa cuando la presencia de un fenómeno en una región impida o dificulte su aparición en las regiones vecinas a ella, es decir, cuando unidades geográficas cercanas sean netamente más disímiles entre ellas que entre regiones alejadas en el espacio. Por último, cuando la variable analizada se distribuya de forma aleatoria, no existirá autocorrelación espacial. Dos son las principales causas que pueden inducir a la aparición de dependencia espacial: la existencia de errores de medida y de anomalías de interacción espacial, de efectos desbordamiento y de jerarquías espaciales (Serrano y Valcarce, 2000: p.22).

Tras la definición anterior, es posible detectar una cierta similitud entre los conceptos de autocorrelación espacial y temporal en la medida en que, en ambos casos, se produce un incumplimiento de la hipótesis de independencia entre las observaciones muestrales, se hallen éstas referidas a unidades de corte transversal o a series temporales. Sin embargo, una importante diferencia aparece entre ellas: la dependencia temporal es únicamente unidireccional (el pasado explica el presente), mientras que la dependencia espacial es multidireccional (una región puede no sólo estar afectada por otra región contigua a ella sino por otras muchas que la rodean, al igual que ella puede influir sobre aquéllas) (Serrano y Valcarce, 2000: p.22).

La solución al problema de la multidireccionalidad en el contexto espacial pasa por la definición de la denominada matriz de pesos espaciales, de retardos o de contactos, W , una matriz cuadrada no estocástica cuyos elementos w_{ij} reflejan la intensidad de la interdependencia existente entre cada par de regiones i y j . Con respecto a cómo definir los mencionados pesos, cabe destacar que no existe una definición unánimemente aceptada, si bien se ha de cumplir que dichos pesos sean no negativos y finitos (Anselin, 1980). A pesar de ello, de forma habitual se recurre al concepto de contigüidad física de primer orden, utilizado inicialmente por Moran (1948) y Geary (1954), donde w_{ij} es igual a 1 si las regiones i y j son físicamente adyacentes o a 0 en caso contrario (se asume por definición que

$w_{ii} = 0$). Siguiendo el criterio de la proximidad física, diversos autores han propuesto definiciones de W basadas en la utilización de la distancia entre regiones (Cliff, y otros, 1968, 1973, 1981). En este sentido, Anselin (1980) propone la utilización de una matriz inversa de distancias al cuadrado, de manera que claramente la intensidad de la interdependencia entre dos regiones disminuye con la distancia que separa sus respectivos centros.

Por último, cabe destacar que, de forma habitual, se recurre a la estandarización de la matriz W , dividiendo cada elemento w_{ij} por la suma total de la fila a la que pertenece, de forma que la suma de cada fila de la matriz estandarizada sea igual a la unidad. Finalmente, cabe comentar que, a partir de la matriz de pesos espaciales, es posible obtener el denominado *retardo espacial* de una variable x sin más que premultiplicar la matriz W por la citada variable (Wx). Así, cada elemento de dicha variable retardada espacialmente se muestra como un promedio ponderado de los valores de la variable en el subgrupo de observaciones vecinas S_i , dado que $w_{ij} = 0$ para $j \notin S_i$ (Serrano y Valcarce, 2000: p.23).

1.3.6. Covarianza espacial

En estadística, la covarianza espacial o temporal describe un proceso o campo de variable aleatoria. Para un campo aleatorio o proceso estocástico $Z(x)$ en un dominio D , una función de covarianza $C(x, y)$ da la covarianza de los valores del campo de azar en los dos lugares x y y :

Ecuación 2-1: Covarianza espacial

$$C(x, y) = cov(Z(x), Z(y)) = E[\{Z(x) - E[Z(x)]\} * \{Z(y) - E[Z(y)]\}].$$

Fuente: Hans, 2019, p.1

1.3.7. Tipos de análisis de datos espaciales

1.3.7.1. Análisis exploratorio de datos espaciales (AEDE)

En las investigaciones aplicadas de economía regional y urbana y, muy especialmente en aquellos centrados en temas relativos a la localización de la actividad y la desigualdad regional, es habitual proceder inicialmente a un análisis descriptivo de la distribución en el espacio de la variable (o variables) objeto de estudio a través de la representación en un mapa de la citada variable, complementándose dicha etapa descriptiva con el cómputo de diversos índices de

concentración/ desigualdad como el índice de Gini, de Theil, de Atkinson o la σ -convergencia. Sin embargo, dichas técnicas presentan importantes limitaciones. Así, si bien la observación de un mapa podría ofrecernos una idea intuitiva acerca de, por ejemplo, la posible correlación entre regiones vecinas o la presencia de algunas regiones próximas con valores elevados o bajos de la citada variable, la información así obtenida, aunque útil, es subjetiva y altamente dependiente, entre otras cosas, del número de intervalos seleccionados para representar a la variable.

Asimismo, muy diferentes distribuciones espaciales, con distintos patrones de dependencia espacial, podrían suministrar valores similares de los citados índices de desigualdad (Vayá, 1998) dado que no consideran de forma explícita el espacio en el cual están localizadas las regiones. Por ello queda justificada la necesidad de un estudio complementario, como el análisis exploratorio de datos espaciales (AEDE), centrado de forma explícita en los efectos espaciales y consistente en el conjunto de técnicas que permiten describir distribuciones espaciales, identificar localizaciones atípicas (outliers espaciales), descubrir esquemas de asociación espacial (cluster espacial) y sugerir diferentes regímenes espaciales u otras formas de inestabilidad espacial. El centro de estos conceptos lo ocupa la noción de autocorrelación espacial (Serrano y Valcarce, 2000: p.25).

1.3.8. Estadísticos globales y locales de autocorrelación espacial

Los estadísticos globales de autocorrelación constituyen la aproximación más tradicional al efecto de dependencia espacial, permitiendo contrastar la presencia o ausencia de un esquema de dependencia espacial a nivel univariante, es decir, contrastar si se cumple la hipótesis de que una variable se encuentra distribuida de forma totalmente aleatoria en el espacio o si, por el contrario, existe una asociación significativa de valores similares o disímiles entre regiones vecinas. Entre dichos estadísticos globales, los cuales resumen en un indicador único el esquema general de dependencia, se encuentran la I de Moran (Moran, 1948) y la G(d) (Getis, y otros, 1992). De forma general, todos ellos permiten contrastar la suposición nula de no autocorrelación espacial (Serrano y Valcarce, 2000: p.27).

Dado que los tests de autocorrelación espacial global analizan todas las observaciones de la muestra de forma conjunta, no resultan sensibles a situaciones donde predomine una importante inestabilidad en la distribución espacial de la variable objeto de estudio (procesos no estacionarios espacialmente), es decir, no contemplan la posibilidad de que el esquema de dependencia detectado a nivel global (por ejemplo, ausencia de autocorrelación espacial) pueda no mantenerse en todas las unidades del espacio analizado. Dicha limitación es fácilmente superable por medio del cálculo de los

denominados contrastes locales de asociación espacial entre los que se encuentran el estadístico local de Moran, I_i , (Anselin, 1980). En este caso, se obtendrá un valor de dichos estadísticos para cada observación de la muestra, pudiendo así analizar la situación de cada unidad espacial por separado. Finalmente, es necesario resaltar la complementariedad de la información suministrada por estos índices, tanto globales como locales (Serrano y Valcarce, 2000: p.27).

Tabla 1-1: Índice de autocorrelación espacial

I de Moran
$I = \frac{N}{S_0} \cdot \frac{\sum_{i \neq j}^N W_{ij} (X_i - \bar{X})(X_j - \bar{X})}{S_{i=1}^N (X_i - \bar{X})^2}$
<p>X_i: valor de la variable x en la región i; \bar{x}: media muestral de la variable x; w_{ij}: pesos de la matriz W; N: tamaño muestral; $S_0 = \sum_i \sum_j w_{ij}$</p>
<p><i>Interpretación tras su estandarización:</i> $Z(I) > 0$ y signif.: autocorrelación positiva $Z(I) < 0$ y signif.: autocorrelación negativa</p>

Fuente: Serrano y Valcarce, 2000, p.27

1.3.8.1. Índice Global de Moran

El Índice Global de Moran es una medida estadística desarrollada por Alfred Pierce Moran en 1950 que analiza de forma integral las variaciones de autocorrelación espacial entre valores vecinos más cercanos, los mismos que pueden clasificarse como positivo, negativo y sin autocorrelación espacial. Cuando los valores tienden a agruparse, se habla de una autocorrelación espacial positiva, pero si estos valores se dispersan, entonces se convierte en una autocorrelación negativa, y si los valores se encuentran dispersos o distribuidos de forma aleatoria, entonces no hay autocorrelación espacial entre los valores analizados. El Índice Global de Moran consiste en la medición de la presencia o ausencia de autocorrelación espacial de una variable. La autocorrelación espacial está indicada por medio de valores que oscilan entre +1 y -1, en donde +1 indica autocorrelación positiva perfecta, -1 expresa autocorrelación negativa perfecta y un valor 0 muestra la presencia de patrones completamente aleatorios en su distribución espacial. El Índice Global de Moran se obtiene de una fórmula generada por un algoritmo complejo para lo cual es necesario usar la estructura de un SIG con capacidad geoestadística (Bucheli, 2019, pp.97-98).

Ecuación 3-1: Índice Global de Moran

$$I = \frac{N}{S_0} \cdot \frac{\sum_{ij}^N W_{ij} (X_i - \bar{X})(X_j - \bar{X})}{S_{i=1}^N (X_i - \bar{X})^2} \quad i \neq j$$

Fuente: Bucheli, 2019, p. 97-98

En dónde;

n; número de unidades geográficas en el mapa.

W_{ij}: matriz de distancia que define si las áreas geográficas *i* y *j* son contiguas o no.

El coeficiente de Índice de Moran, se ajusta a la prueba de significancia estadística de valores Z, suponiendo una distribución normal. El valor z es una desviación estándar, medida entre la diferencia de un valor de la variable y el promedio. El resultado de esta fórmula arroja una regla de decisión que valida la hipótesis de investigación de la siguiente forma:

H₀: Índice Moran = 0 → *H₀*: Índice Global de Moran es igual a 0,

entonces no existe autocorrelación espacial. Es así como para verificar el nivel de significancia se suele utilizar como referencia el valor de 0.05, ya que si el valor de la probabilidad p es menor que o igual a 0.05, se rechaza la hipótesis. Adicionalmente el valor P es una probabilidad y se refiere a aproximaciones numéricas del área debajo de la curva de una distribución conocida. Dentro de las herramientas de análisis existe la probabilidad de que el patrón espacial observado se haya creado mediante algún proceso aleatorio.

H₁: Índice Moran ≠ 0 → *H₁*: Índice Global de Moran no es igual a 0,

por tanto, si existe autocorrelación espacial, sino pueden encontrarse dispersos o agrupados en el espacio geográfico. Es así como mientras más pequeño sea el valor de p, mayor probabilidad hay que exista autocorrelación espacial (Bucheli, 2019, pp.97-98).

1.3.9. Técnicas de visualización de la presencia de efectos espaciales

Los estadísticos de dependencia espacial analizados en el apartado anterior se muestran como instrumentos útiles para detectar la existencia de patrones de asociación espacial específicos en la distribución de una variable en el territorio. No obstante, el AEDE incluye otras herramientas complementarias que permiten, por medio de un entorno gráfico, enriquecer los resultados obtenidos tras el cómputo de los citados contrastes. Entre ellas, hay que destacar el denominado box map, una extensión del mapa cuartil junto con los outliers superiores e inferiores (definidos como las observaciones fuera de los límites en un box plot), el cual se muestra como una forma sencilla de identificar outliers espaciales (Serrano y Valcarce, 2000: p.20).

Una aproximación más formal a la visualización de la dependencia espacial se basa en el concepto del scatterplot de Moran y su asociado scatter map. El scatterplot de Moran es un tipo de gráfico que muestra en el eje de abscisas las observaciones de la variable objeto de estudio normalizada y en el de ordenadas el retardo espacial de la misma también normalizado. De este modo, los cuatro cuadrantes reproducen diferentes tipos de dependencia espacial. Si la nube de puntos está dispersa en los cuatro cuadrantes es indicio de ausencia de correlación espacial. Si por el contrario los valores se encuentran concentrados sobre la diagonal que cruza los cuadrantes I (derecha superior) y III (izquierda inferior), existe una elevada correlación espacial positiva de la variable, coincidiendo su pendiente con el valor de la I de Moran. La dependencia será negativa si los valores se concentran en los dos cuadrantes restantes. Finalmente, el scattermap es la traslación de la información del scatterplot a un mapa del territorio, donde cada uno de los cuatro cuadrantes es resaltado con un color diferente (Serrano y Valcarce, 2000: p.20).

Como se puede derivar de las herramientas comentadas, la introducción del mapa como una posible vista adicional de la información es básica. De esta forma, muchas de las técnicas de análisis exploratorio de datos espaciales aprovechan la capacidad de visualización y manipulación de estos que proporcionan los Sistemas de Información Geográfica (SIG). Por ejemplo, el SIG ArcView (ESRI, 1995), se organiza en torno a varias vistas vinculadas de las fichas (un mapa, una tabla y varios tipos de gráficos), lo cual permite un cierto grado de interacción dinámica en el sentido de que una selección determinada en una de las vistas se refleja de forma inmediata en el resto (Serrano y Valcarce, 2000: p.30).

1.3.10. Autocorrelación espacial en el modelo de regresión

En el apartado anterior se han analizado los estadísticos espaciales habitualmente utilizados para contrastar la presencia de dependencia espacial a nivel univariante. No obstante, de igual modo, es posible que el citado efecto espacial esté presente en el contexto de un modelo de regresión, ya sea como consecuencia de la existencia de variables sistemáticas (endógena y/o exógenas) correlacionadas espacialmente o como consecuencia de la existencia de un esquema de dependencia espacial en el término de perturbación (Serrano y Valcarce, 2000: p.31).

En caso de que la variable endógena de un modelo de regresión lineal esté correlacionada espacialmente, la solución pasa por especificar el siguiente modelo:

Ecuación 4-1: Variable endógena

$$y = \rho Wy + X\beta + \mu$$

Fuente: Serrano y Valcarce, 2000: p. 31

$$\mu \sim N(0, \sigma^2 I)$$

donde y es un vector ($N \times 1$), Wy el retardo espacial de la variable y , X una matriz de K variables exógenas, u un término de perturbación ruido blanco, N el número de observaciones y , por último, ρ el parámetro autorregresivo que recoge la intensidad de las interdependencias entre las observaciones muestrales.

De igual forma, la autocorrelación espacial podría estar únicamente presente en el término de perturbación:

$$y = X\beta + \varepsilon$$

Ecuación 5-1: Ruido blanco

$$\varepsilon = \lambda W\varepsilon + \mu$$

Fuente: Serrano y Valcarce, 2000: p.31

$$\mu \sim N(0, \sigma^2 I)$$

donde u es un término ruido blanco y λ el parámetro autorregresivo que refleja la intensidad de las interdependencias (Serrano y Valcarce, 2000: p.31).

Si bien la estructura autorregresiva es la más utilizada a nivel tanto teórico como empírico, es posible modelizar el término de perturbación de forma diferente. Así, (Cliff , y otros) sugieren utilizar un proceso de medias móviles espacial de orden 1:

Ecuación 6-1: Proceso de medias móviles espacial de orden 1

$$y = X\beta + \varepsilon$$

Fuente: Serrano y Valcarce, 2000: p.31

$$\varepsilon = \theta W_1 \mu + \mu$$

$$u \sim N(0, \sigma^2 I)$$

donde θ es el coeficiente de medias móviles espacial y u es el término de perturbación incorrelacionado.

A partir de la expresión es fácil observar cómo, en caso de omitir de forma errónea un retardo espacial de la variable endógena y/o exógenas, la dependencia espacial se trasladaría directamente al término de perturbación, el cual pasaría a estar correlacionado espacialmente. Este tipo de auto reciprocidad espacial es conocido con el nombre de autocorrelación espacial sustantiva y su solución pasa por la inclusión en el modelo de un retardo espacial de la variable correlacionada espacialmente. Por el contrario, cuando la dependencia espacial residual no está causada por la omisión errónea de un retardo de alguna de las variables sistemáticas, se presenta un caso de auto correspondencia espacial residual como nuisance o perturbación. Dicha situación podría verse explicada por la omisión de variables no cruciales que se hallen correlacionadas espacialmente o por la existencia de errores de medida. En este caso, la solución implica la consideración explícita de un esquema de dependencia espacial en el término de error (Serrano y Valcarce, 2000: p.32).

Existe una amplia batería de estadísticos espaciales para contrastar las anteriores estructuras: los tests ad-hoc, los cuales son válidos para contrastar la hipótesis nula de no autocorrelación espacial residual (si bien no presentan una hipótesis alternativa definida que explícitamente describa el proceso generador del término de perturbación); y aquéllos basados en el principio de máxima verosimilitud (los cuales están rigurosamente estructurados en términos de una suposición nula y alternativa específicas). Entre los primeros destacan los tests de la I de Moran y el test K-R, mientras que en el segundo grupo se incluyen el contraste de Wald, el de razón de verosimilitud y el de los multiplicadores de Lagrange, si bien los más utilizados son estos últimos (Serrano y Valcarce, 2000: p.32).

Tabla 2-1: Test de autocorrelación espacial

I de Moran	$I = \frac{N}{S} \frac{e'We}{e'e}$	e: residuos MCO, N: tamaño muestral, S: suma de todos los elementos w_{ij} de la matriz W. Debe estandarizarse.
------------	------------------------------------	--

Fuente: Serrano y Valcarce, 2000, p.32

En relación con las consecuencias derivadas de la autocorrelación espacial en la estimación habitual por MCO, es preciso distinguir entre los dos tipos de autocorrelación comentados: la dependencia espacial residual y sustantiva. Por lo que hace referencia a las consecuencias de la presencia de residuos correlacionados espacialmente en la estimación MCO, éstas son similares a las conocidas en el contexto temporal. Todo ello lleva a sesgar la inferencia basada en los tests de significación de la t-student y en el coeficiente de determinación R^2 (Serrano y Valcarce, 2000: p.32).

De todo ello se desprende que, con independencia del esquema de dependencia espacial detectado, la estimación MCO no es la adecuada. En este sentido, la aproximación máximo verosímil (MV) es una de las alternativas más ampliamente utilizadas. Con respecto a dicho método de estimación, idéntica la forma que, en el contexto temporal, los estimadores MV deberán ser obtenidos a partir de la maximización del logaritmo de la función de probabilidad asociada al modelo espacial especificado. En este sentido, partiendo de una función de densidad normal conjunta para los errores del modelo más general con autocorrelación tanto residual como sustantiva (retardo espacial de la endógena), se obtiene una función de verosimilitud cuya expresión es (en logaritmos):

Ecuación 7-1: Función de verosimilitud logarítmica

$$\ln L(\theta) = -\frac{N}{2} \ln 2\pi - 1n|\Omega| + 1n|B| + 1n|A| - \frac{1}{2} v'v$$

Fuente: Serrano y Valcarce, 2000: p.32

donde $\theta = [\rho, \beta', \lambda, \sigma^2]$, $A = (I - \rho W)$, $B = (I - \lambda W)$, Ω es la matriz de varianzas y covarianzas del término de perturbación y $v'v = (Ay - X\beta)'B'\Omega^{-1}B(Ay - X\beta)$. A partir de la expresión, los estimadores MV podrán ser obtenidos igualando a 0 el vector de derivadas parciales de la ecuación respecto a θ (Serrano y Valcarce, 2000: p.32).

1.4. Cartografía

La cartografía ha desempeñado un papel fundamental a lo largo de la historia de la Humanidad. En el Siglo XX los mapas aunados a la tecnología de los sistemas de información geográfica son excelentes herramientas que permiten comparar, escoger y tomar decisiones basados en información actualizada e integral.

La palabra cartografía tiene su origen en los vocablos *charta* del latín que significa papel que sirve para comunicarse o carta y *grapho* del griego que significa descripción, estudio o tratado. La cartografía es la rama del grafismo que se ocupa de los métodos e instrumentos utilizados para exponer y expresar ideas, formas y relaciones en un espacio bi o tridimensional. La cartografía parte del principio de que los seres vivos, los fenómenos físicos y sus interrelaciones ocurren en un contexto temporal y espacial y que por lo tanto es posible mapearlos (González et al., 2007: p.2).

Un mapa es la representación gráfica a una escala reducida de una porción de la superficie terrestre que muestra sólo algunos rasgos o atributos de la realidad. En este sentido el plano es un sustituto de la porción de la superficie terrestre que se va a estudiar. La proyección también puede definirse como un instrumento analógico diseñado para el registro, cálculo, exposición, análisis y, en general, la comprensión de los hechos geográficos y de sus relaciones espaciales. Su función es representar visualmente una imagen. Tres de las características más importantes de los mapas son su control geodésico y su precisión horizontal y vertical, los cuales responden a los estándares utilizados en cada país (González et al., 2007: p.2).

Desde un punto de vista geométrico los mapas pueden concebirse como una representación bidimensional de la superficie terrestre que muestra atributos tales como distancias, direcciones, tamaños y formas. Los mapas son elaborados normalmente para mostrar la distribución espacial de uno o más fenómenos geográficos.

En el mapa se usa signos convencionales para representar detalles de la superficie terrestre que dada la escala del plano no es posible dibujar utilizando sus formas y proporciones reales. Las proyecciones son elaboradas en muy diferentes estilos y escalas y cada uno de ellos cumple una función específica. Sin embargo, para facilitar su uso todos deben poseer ciertos elementos comunes. La omisión de cualquiera de estos elementos reduce su utilidad (González et al., 2007: p.2).

1.4.1. Elementos esenciales en un mapa

- **Título:** El título expresa la esencia del mapa o sea su tema principal. Debe incluir el área o zona geográfica que representa y el objeto de estudio.
- **Fecha de los datos:** Los mapas son representaciones estáticas de un fenómeno temporal y por lo tanto debe indicarse claramente la fecha en que fueron recolectados los datos.
- **Leyenda:** En los mapas se utilizan símbolos, tramados, colores o tonos de gris para expresar cantidades, gradientes o aun cuando algunos símbolos se explican por sí mismos es necesario incluir una leyenda explicativa en una esquina del mapa.
- **Proyección y datum:** La proyección y el datum son dos atributos del mapa que definen sus características y propiedades geométricas. Esta información es esencial para referenciar y posteriormente manipular un mapa utilizando un sistema de información geográfica.
- **Escala:** Indique la escala gráfica y/o numérica de su mapa.
- **Autor(a)/Fuente:** Indique si usted es el autor o si la información utilizada para elaborar el mapa proviene de otra fuente. Recuerde que usted debe dar crédito al autor(es) de la información original (González et al., 2007: p.3).

1.4.2. Los mapas y su clasificación

La información contenida en los mapas es muy diversa, sin embargo, puede clasificarse de la siguiente manera (González et al., 2007: p.4):

1.4.2.1. Mapas generales, base o topográficos

Los mapas generales muestran diversos atributos de un área geográfica y su función es ubicar al lector en su área de trabajo. Los planos topográficos son un ejemplo de proyecciones de uso general ya que muestran tanto detalles planimétricos como altimétricos de una determina zona. Elementos típicos de estos son: carreteras, elevaciones, ríos, lagos y asentamientos humanos. Su elaboración es mediante métodos y técnicas fotogramétricas de alta precisión (González et al., 2007: p.4).

1.4.2.2. Mapas cualitativos

Estos mapas expresan variables de carácter nominal u ordinal y normalmente se utilizan para representar características del paisaje tales como uso-cobertura del suelo, geología, geomorfología o suelos (González et al., 2007: p.4).

1.4.2.3. Mapas cuantitativos de superficie

Los mapas cuantitativos de superficie proporcionan tanto información cuantitativa del fenómeno es estudio, como sobre su distribución espacial. La información puede mapearse utilizando líneas de igual valor denominadas isopletas, isoarritmas o isolíneas o valores medios por unidad de área (coropletas). (González et al., 2007: p.4)

Los mapas coropléticos muestran valores por unidad de área y se utilizan frecuentemente con unidades administrativas tales como fincas, distritos, cantones, provincias o países (unidades estadísticas). Los planos coropléticos exhiben las características del área en forma simple y concisa y tienen como objetivo transmitir una impresión concreta de la realidad a partir del mapa. (González et al., 2007: p.4)

En resumen, los mapas generales o temáticos deben utilizarse en forma conjunta y seleccionarse en función de su uso; ya que es prácticamente imposible lograr una comunicación eficiente manipulando en forma aislada los elementos de cada uno de ellos. Los planos son considerados por especialistas de diversas áreas del saber cómo un excelente medio para organizar, analizar y expresar datos y conceptos. (González et al., 2007: p.4)

1.5. Técnicas Multivariantes

Las técnicas multivariantes hacen referencia a cualquier método estadístico que analice simultáneamente múltiples características en cada uno de los individuos o muestras objeto de la investigación. Una de las dificultades en definir qué es el análisis multivariante reside en el hecho de que el término multivariante (o multivariado) no ha sido usado de manera consistente en la literatura. Algunos investigadores usan el término multivariado simplemente para referirse a las relaciones existentes entre más de dos variables. Sin embargo, para que un análisis sea considerado verdaderamente multivariante, todas las variables deben de ser aleatorias y deben de estar

interrelacionadas de tal manera que los diferentes efectos no puedan ser interpretados significativamente de manera independiente (Cayuela, 2011, p.4).

Podemos considerar como técnicas multivariantes, entre otras:

- Análisis de componentes principales
- Análisis discriminante
- Análisis cluster (técnica de agrupación)
- Análisis de correspondencias
- Escalamiento multidimensional
- Modelo de ecuaciones estructurales (análisis causal)
- Análisis de la varianza multivariado (incluyendo la regresión multivariada)

1.5.1. Análisis de componentes principales (PCA)

El análisis de componentes principales (PCA) es una técnica estadística de síntesis de la información, o reducción de la dimensión (número de variables). Es decir, ante un banco de datos con muchas variables, el objetivo será reducirlas a un menor número perdiendo la menor cantidad de información posible. Los nuevos componentes principales o factores serán una combinación lineal de las variables originales, y además serían independientes entre sí (Cayuela, 2011, pp.4-5).

Un aspecto clave en PCA es la interpretación de los factores, ya que ésta no viene dada a priori, sino que será deducida tras observar la relación de los factores con las variables iniciales (habrá, pues, que estudiar tanto el signo como la magnitud de las correlaciones). Esto no siempre es fácil, y será de vital importancia el conocimiento que el experto tenga sobre la materia de investigación (Cayuela 2011, pp.4-5).

1.5.2. Análisis de Clúster (Conglomerados)

El análisis de cluster es una técnica cuya idea básica es agrupar un conjunto de observaciones en un número dado de clusters o grupos. Este agrupamiento se basa en la idea de distancia o similitud entre las observaciones. La obtención de dichos clusters depende del criterio o distancia considerados. Por ejemplo, una baraja de cartas españolas se podría dividir de distintos modos: en cuatro clusters (los

cuatro palos), en ocho clusters (los cuatro palos y según sean figuras o números), en dos clusters (figuras y números). Es decir, todo depende de lo que consideremos como similar (Marin, 2018, p.1).

El número posible de combinaciones de grupos y de elementos que integran los posibles grupos se hace intratable desde el punto de vista computacional, aún con un número escaso de observaciones. Se hace necesario, pues, encontrar métodos o algoritmos que infieran el número y componentes de los clusters más aceptable, aunque no sea el óptimo absoluto. Previamente es necesario considerar el concepto de medida de similitud (Marin, 2018, p.1).

Finalmente, el análisis cluster depende en su totalidad, de las variables que se utilizan, de la adición o destrucción de variables relevantes que pueden tener un impacto significativo sobre la solución resultante. Existen dos grandes tipos de análisis de clusters: Jerárquicos y No Jerárquicos (Arita et al., 2021, p.1).

1.5.2.1. Métodos de clúster jerárquicos

En la práctica, no se pueden examinar todas las posibilidades de agrupar los elementos, incluso con los ordenadores más rápidos. Una solución se encuentra en los llamados métodos jerárquicos que se basan en el cálculo de una matriz de distancias y se aplican con $n < 200$, ya que los cálculos y resultados se complican al aumentar el tamaño de la muestra. Estos métodos se caracterizan por no permitir una reasignación en los grupos, es decir, que dos clusters o individuos unidos durante el proceso, no pueden separarse en etapas decisivas (Arita et al., 2021: p.1).

Se pueden aplicar a variables o a observaciones:

- ✓ Si se agrupan variables se necesitan de tres o más variables numéricas.
- ✓ Si se agrupan observaciones se necesita al menos una variable numérica.

Dada la matriz, se tienen dos posibles formas de actuar:

Métodos jerárquicos aglomerativos

También se denominan ascendentes. Parten de objetos singulares (cada objeto es un conglomerado) de este modo, se tienen tantos clústeres iniciales como objetos. Luego se van agrupando de modo que los primeros en hacerlo son los más similares y al final, todos los subgrupos se unen en un único clúster. Requieren menor tiempo y son los más utilizados (Arita et al., 2021: p.1).

Se considera aquí los métodos aglomerativos con diferentes métodos de unión (linkage methods). Los más importantes son:

- Simple Linkage (vecino más próximo)

Considera que la distancia entre dos clúster es la distancia más corta desde un miembro de un clúster a otro miembro de otro clúster. Si los datos se basan en similitudes entre dos clúster, se considera la mayor similitud desde cualquier miembro de un clúster a otro miembro de otro clúster. Por tanto, el proceso de unión de estos dos clusters sería para minimizar las distancias o maximizar las similitudes (Arita et al., 2021: p.2).

- Complete Linkage (Vecino más lejano)

Considera la distancia más grande desde cualquier miembro de un clúster a otro miembro de otro clúster. Si la medida es la distancia, se recoge la distancia máxima de los individuos del grupo al nuevo individuo. Si la medida es la similitud entre el grupo formado y el nuevo individuo, se toma la mínima de los individuos del grupo al nuevo individuo. Por tanto, el proceso de unión de los dos clusters sería para minimizar las distancias o para maximizar las similitudes (Arita et al., 2021: p.2).

- Average Linkage (Promedio entre Grupos)

Mide la proximidad de dos grupos, calculando la media de las distancias entre objetos de ambos grupos o las medias de las similitudes entre objetos de ambos grupos. Según algunos autores, afirman que este método está sesgado a formar conglomerados con aproximadamente la misma varianza (Arita et al., 2021: p.2).

- Método de Ward

Aquí se unen clusters si la variación dentro de ellos no ha aumentado de manera significativa, haciendo más homogéneos los nuevos cluster formados. Este método se basa en la suma de cuadrados y a crear grupos de tamaño similar, dando paso a realizar buenos análisis de varianza por la producción de clusters definidos (Arita et al., 2021: p.2). Notemos por :

- ✓ x_{ij}^k al valor de la j -ésima variable sobre el i -ésimo individuo del k -ésimo cluster, suponiendo que dicho cluster posee n_k individuos.
- ✓ m_j^k al centroide del cluster k , con componentes m_j^k .
- ✓ E_k a la suma de cuadrados de los errores del cluster k , o sea, la distancia euclídea al cuadrado entre cada individuo del cluster k a su centroide.

Ecuación 8-1: Método de Ward

$$E_k = \sum_{i=1}^{n_k} \sum_{j=1}^n (x_{ij}^k - m_j^k)^2 = \sum_{i=1}^{n_k} \sum_{j=1}^n (x_{ij}^k)^2 - n_k \sum_{j=1}^n (m_j^k)^2$$

Fuente: Arita et al., 2021: p.2

- ✓ E a la suma de cuadrados de los errores para todos los clusters, o sea, si suponemos que hay h clusters

$$E = \sum_{k=1}^h E_k$$

Se puede observar que, de este modo, se define una posible distancia entre dos clústeres: la correspondiente a la pareja de elementos más cercana, la más lejana o la media de todas las posibles parejas de elementos de ambos clústeres (Marin, 2018, p.8).

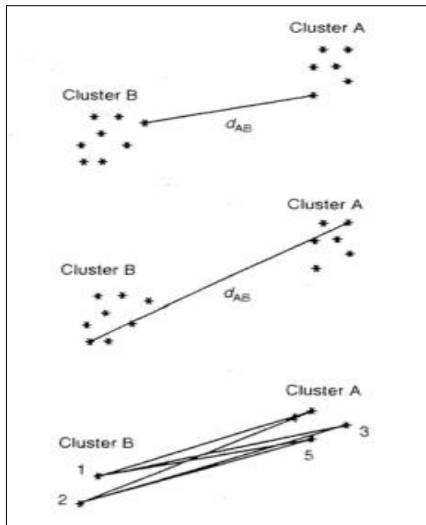


Figura 1-1: Parejas de clusters

Fuente: Marin, 2018

Definidas las distancias anteriores, se puede considerar el algoritmo básico, dados N objetos o individuos:

- ✓ Empezar con N clústeres (el número inicial de elementos) y una matriz $N \times N$ simétrica de distancias o similitudes. $D = [d_{ik}]_{ik}$.
- ✓ Dentro de la matriz de distancias, buscar aquella entre los clústeres U y V (más próximos, más distantes o en media más cercanos) que sea la menor entre todas, d_{uv} .
- ✓ Juntar los clústeres U y V en uno solo. Actualizar la matriz de distancias:
 - i) Borrando las filas y columnas de los clústeres U y V .
 - ii) Formando la fila y columna de las distancias del nuevo clúster (UV) al resto de clústeres.
- Repetir los pasos (2) y (3) un total de $(N - 1)$ veces.

Al final, todos los objetos están en un único clúster cuando termina el algoritmo. Además, se guarda la identificación de los clústeres que se van uniendo en cada etapa, así como las distancias a las que se unen. Finalmente se construye un dendrograma (Marin, 2018, pp.9-10).

Métodos jerárquicos divididos

Se actúa, al contrario. Parten del conjunto único de todas las observaciones y se van dividiendo según lo lejanos que estén, los elementos que se incluyen en un grupo no se pueden reasignar (Marin, 2018, p.7).

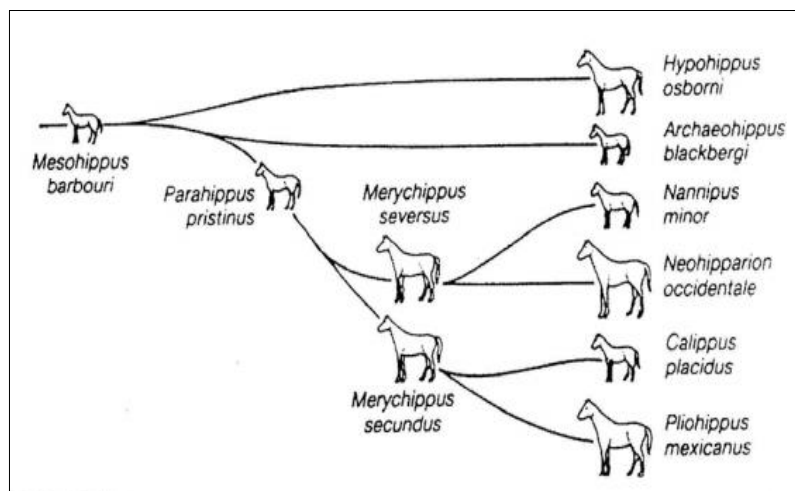


Figura 2-1: Dendrograma

Fuente: Marin, 2018

En cualquier caso, de ambos métodos se deriva un dendrograma, que es un gráfico que ilustra cómo se van haciendo las subdivisiones o los agrupamientos, etapa a etapa.

1.5.2.2. Métodos de clúster no jerárquicos

También conocidos como métodos de “optimización.” Estos métodos si permiten una reasignación en los grupos, aunque se vuelve necesario fijar de antemano el número de clusters deseado. Se usan para agrupar objetos, pero no variables, en un conjunto de k clústeres ya predeterminado. No se tiene que especificar una matriz de distancias ni se tienen que almacenar las iteraciones. Todo esto permite trabajar con un número de datos mayor que en el caso de los métodos jerárquicos (Arita et al., 2021: p.2).

Se utilizan cuando:

- ✓ Los métodos de reasignación permiten reasignar objetos a distintos conglomerados en cada fase.
- ✓ Los métodos de búsqueda de densidad se agrupan mediante la búsqueda de altas densidades (modas).

- ✓ Los métodos directos permiten clasificar de forma simultánea individuos y variables.

Se parte de un conjunto inicial de clústeres elegidos al azar, que son los representantes de todos ellos; luego se van cambiando de modo iterativo. Se usa habitualmente el método de las k-medias (Marin, 2018, p.12).

- Método K-medias

Es un método que permite asignar a cada observación el clúster que se encuentra más próximo en términos del centroide (media). En general, la distancia empleada es la euclídea. Pasos:

- ✓ Se toman al azar k clústeres iniciales.
- ✓ Para el conjunto de observaciones, se vuelve a calcular las distancias a los centroides de los clústeres y se re-asignan a los que estén más próximos. Se vuelven a recalcular los centroides de los k clústeres después de las reasignaciones de los elementos.
- ✓ Se repiten los dos pasos anteriores hasta que no se produzca ninguna reasignación, es decir, hasta que los elementos se estabilicen en algún grupo.

Usualmente, se especifican k centroides iniciales y se procede al paso (2) y, en la práctica, se observan la mayor parte de reasignaciones en las primeras iteraciones (Marin, 2018, p.13).

- Método K-medoids clustering (PAM)

Una definición más exacta del término medoid es: elemento dentro de un cluster cuya distancia (diferencia) promedio entre él y todos los demás elementos de este cluster es lo mejor posible. El hecho de utilizar medoids en lugar de centroides hace de K-medoids un método más robusto que K-means, viéndose menos afectado por outliers o ruido (Yaselga, 2020, p. 2).

El algoritmo más empleado para aplicar K-medoids se conoce como PAM (Partitioning Around Medoids) y sigue los siguientes pasos:

- ✓ Seleccionar K observaciones aleatorias como medoids iniciales. También es posible identificarlas de forma específica.
- ✓ Calcular la matriz de distancia entre todas las observaciones si esta no se ha calculado anteriormente.

- ✓ Asignar cada observación a su medoid más cercano.
- ✓ Para cada uno de los clusters creados, comprobar si seleccionando otra observación como medoides consigue reducir la distancia promedio del cluster, si esto ocurre, seleccionar la observación que consigue una mayor reducción como nuevo medoid.
- ✓ Si al menos un medoid ha cambiado en el paso 4, volver al paso 3, de lo contrario, se termina el proceso (Yaselga, 2020, p.2).

A diferencia del algoritmo K-means, en el que se minimiza la suma total de cuadrados intra-cluster (suma de las distancias al cuadrado de cada observación respecto a su centroide), el algoritmo PAM minimiza la suma de las diferencias de cada observación respecto a su medoid. Por lo general, el método de K-medoids se utiliza cuando se conoce o se sospecha de la presencia de outliers. Si esto ocurre, es recomendable utilizar como medida de similitud la distancia de Manhattan, ya que es menos sensible a outliers que la euclídea (Yaselga, 2020, p.2).

- Método Clara

Una de las limitaciones del método K-medoids-clustering es que su algoritmo requiere mucha memoria RAM, lo que impide que se pueda aplicar cuando el set de datos contiene varios miles de observaciones. CLARA (Clustering Large Applications) es un método que combina la idea de K-medoids con el resampling para que pueda aplicarse a grandes volúmenes de datos (Yaselga, 2020, p.3).

A continuación, se describen los pasos del algoritmo CLARA:

- ✓ Se divide aleatoriamente el set de datos en n partes de igual tamaño, donde n es un valor que determina el analista.
- ✓ Para cada una de las n partes:
 - Aplicar el algoritmo PAM e identificar cuáles son los k medoids.
 - Utilizando los medoids del paso anterior agrupar todas las observaciones del set de datos.
 - Calcular la suma total de las distancias entre cada observación del set de datos y su correspondiente medoid (suma total de distancias intra-clusters).
- ✓ Seleccionar como clustering final aquel que ha conseguido menor suma total de distancias intra-clusters en el paso (Yaselga, 2020, p.3).

1.5.3. Medidas de similitud

En realidad, es bastante subjetivo el hecho de elegir una medida de similitud ya que depende de las escalas de medida. Se pueden agrupar observaciones según la similitud expresada en términos de una distancia. Si se agrupan variables, es habitual utilizar como medida de similitud los coeficientes de correlación en valor absoluto. Para variables categóricas existen también criterios basados en la posesión o no de los atributos (tablas de presencia-absencia) (Marin, 2018, pp.1-2).

Dados dos vectores x_i, x_j pertenecientes a \mathbb{R}^k , se dice que se ha establecido una distancia entre ellos si se define una función d con las propiedades siguientes:

- $d : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^+$, es decir $d(x_i, x_j) \geq 0$;
- $d(x_i, x_i) = 0 \forall i$, la distancia entre un elemento y sí mismo es cero.
- $d(x_i, x_j) = d(x_j, x_i)$, la distancia es simétrica
- $d(x_i, x_j) \leq d(x_i, x_p) + d(x_p, x_j)$, la distancia verifica la propiedad triangular.

Estas propiedades generalizan la noción intuitiva de distancia euclídea entre dos puntos (Marin, 2018, pp.1-2).

1.5.4. Espacios métricos

La geometría del espacio tridimensional en el que se ha sumergido resulta muy natural. Conceptos tales como distancia, longitud, ángulo, perpendicularidad son de uso cotidiano. En matemáticas frecuentemente se agrupa ciertos objetos en espacios abstractos y define entre ellos relaciones semejantes a las existentes entre los puntos del espacio ordinario. El paralelismo que se establece así entre los espacios abstractos y el espacio euclideo permite visualizar y lograr un entendimiento más profundo de estos objetos. En algunas aplicaciones el planteo más simple que puede usarse es el de espacio métrico. Un espacio métrico es un conjunto de puntos en los que está definido la noción de distancia entre puntos. Se usa la función métrica para definir los conceptos fundamentales del análisis, tales como convergencia, continuidad y compacidad (Hilbert, 2019, p.1).

Un espacio métrico no necesita tener ninguna clase de estructura algebraica definida en él, es decir, puede no tener sentido la suma de elementos del espacio o la multiplicación de un elemento por un

número real o complejo. Sin embargo, es muy frecuente el uso de espacios métricos que son a su vez áreas vectoriales, con una métrica derivada de una norma que mide la longitud de un vector. Tales espacios serán llamados zonas normados. La intuición geométrica derivada de los espacios euclidianos de dimensiones finitas es esencial, aunque características completamente nuevas surgen en los espacios de infinitas dimensiones (Hilbert, 2019, p.1).

1.5.5. Métricas

Si a y b son dos números reales, puede pensarse al número real no negativo $|a - b|$ como la distancia que separa a de b . Esta operación de asignar distancias a pares de puntos es precisamente lo que da origen a los espacios métricos. La teoría básica que emana del concepto de distancia tiene que ver con las propiedades de subconjuntos (abiertos, cerrados, compactos, conexos), sucesiones (convergentes, Cauchy) y funciones (continuas), y la relación entre estas nociones (Hilbert, 2019, p.1).

Un espacio métrico es un par (X, d) donde X es un conjunto no vacío y d es una función real definida en $X \times X$, llamada distancia o métrica, y que satisface los siguientes axiomas:

- $d(x, y) \geq 0 \forall x, y \in X$, y $d(x, y) = 0 \Leftrightarrow x = y$,
- $d(x, y) = d(y, x) \forall x, y \in X$, y
- $d(x, z) \leq d(x, y) + d(y, z) \forall x, y, z \in X$.

- ✓ Las distancias son no negativas y el único punto a distancia cero de x es el mismo x ;
- ✓ la distancia es una función simétrica;
- ✓ la distancia satisface la desigualdad triangular: la longitud de un lado de un triángulo es menor que la suma de las longitudes de los otros dos lados.

Para un conjunto dado X es posible definir más de una métrica. Cuando la métrica del espacio se da por sobreentendida, habla simplemente del sitio métrico X aunque se sabe que realmente este es un par (X, d) . A los elementos de X los llaman puntos del espacio métrico (Hilbert, 2019, pp.1-2).

1.5.6. Distancias

1.5.6.1. Distancia Euclidiana

Es la distancia en línea recta o la trayectoria más corta posible entre dos puntos:

Ecuación 9-1: Distancia Euclídea

$$D_{Euc}(x_i, x_j) = \sqrt{\sum_{r=1}^p (x_{ri} - x_{rj})^2}$$

Fuente: Hilbert, 2019, p.3

1.5.6.2. Distancia Manhattan

La función de la distancia Manhattan calcula la distancia que se puede recorrer para llegar de un punto de datos a la otra si un camino en forma de rejilla es seguido. El trayecto de Manhattan entre dos elementos es la suma de las diferencias de sus correspondientes componentes:

Ecuación 10-1: Distancia Manhattan

$$D_{Man}(x, y) = \sum_{i=1}^k |X_i - Y_i|$$

Fuente: Hilbert, 2019, p.3

1.5.7. Validación de resultados

1.5.7.1. Dendrograma

Un dendrograma es una representación gráfica con forma de árbol que sirve para resumir el proceso de agrupación del análisis de clusters. Los objetos similares están conectados por enlaces y su posición, en el diagrama, está determinada por el nivel de similitud o disimilitud de los objetos (Arita et al., 2021: p.1).

1.5.7.2. Estadístico de Hopkins -Tendencia de agrupamiento

El estadístico *Hopkins* permite evaluar la tendencia de clustering de un conjunto de datos mediante el cálculo de la probabilidad de que dichos datos procedan de una distribución uniforme, es decir, estudia la distribución espacial aleatoria de las observaciones. La forma de calcular este estadístico es la siguiente:

- Extraer una muestra uniforme de n observaciones (p_1, \dots, p_n) del set de datos estudiado.
- Para cada observación p_i seleccionada, encontrar la observación vecina más cercana p_j y calcular la distancia entre ambas, $x_i = \text{dist}(p_i, p_j)$ (Amat, 2017, p. 12).
- Simular un conjunto de datos de tamaño n (q_1, \dots, q_n) extraídos de una distribución uniforme con la misma variación de los datos originales.
- Para cada observación simulada q_i , encontrar la observación vecina más cercana q_j y calcular la distancia entre ambas $y_i = \text{dist}(q_i, q_j)$.
- Calcular el estadístico Hopkins (H) como la medida de las distancias de vecinas más cercanos en el set de datos simulados, dividida por la suma de las medias de las distancias vecinas más cercanas del set de datos original y el simulado (Amat, 2017, p.12).

Ecuación 11-1: Test estadístico de Hopkins

$$H = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i + \sum_{i=1}^n y_i}$$

Fuente: Amat, 2017, p.12

Los valores de H en torno a 0.5 indican que $\sum_{i=1}^n x_i$ y $\sum_{i=1}^n y_i$ son muy cercanas el uno al otro, es decir, que los datos estudiados se distribuyen uniformemente y que por lo tanto no tiene sentido aplicar clustering. Cuando más se aproxime a 0 el estadístico H, más evidencias se tienen a favor de que existen agrupaciones en los datos y de que, si se aplica clustering correctamente, los grupos resultantes serán reales (Amat, 2017, p.12).

1.6. Etapas del análisis de clúster o conglomerado

- Elección de las variables.
- Elección de la técnica de clúster.
- Elección de la medida de asociación.
- Validación de resultados.

1.7. Software R

R es un software libre que permite realizar análisis estadísticos y el más usado en la comunidad científica. Este programa está disponible en la página web: <http://www.r-project.org> y consta de una aplicación central y de librerías de multitud de temas que se pueden instalar según necesidad. R es un programa de instrucciones, y por tanto, no resulta del todo “amigable” para los usuarios que no están acostumbrados a este tipo de manejo (Botella et al., 2018: p.3).

1.7.1. Librerías para clúster

1.7.1.1. Library(factoextra)

Es un paquete de R que facilita la extracción y visualización de los resultados de los análisis exploratorios de datos multivariados, que incluyen:

- Análisis de componentes principales (PCA), que se utiliza para resumir la información contenida en datos multivariados continuos (es decir, cuantitativos) al reducir la dimensionalidad de los datos sin perder información importante (Kassambara, 2020, p.1).
- Análisis de correspondencias (CA), que es una extensión del análisis de componentes principales adecuado para analizar una gran tabla de contingencia formada por dos variables cualitativas (o datos categóricos).
- Análisis de Correspondencia Múltiple (MCA), que es una adaptación de CA a una tabla de datos que contiene más de dos variables categóricas.
- Análisis de factores múltiples (MFA), dedicado a conjuntos de datos donde las variables se organizan en grupos (variables cualitativas y/o cuantitativas) (Kassambara, 2020, p.1).
- Análisis jerárquico de factores múltiples (HMFA), es una extensión de MFA en una situación en la que los datos se organizan en una estructura jerárquica.
- Análisis factorial de datos mixtos (FAMD), un caso particular del MFA, dedicado a analizar un conjunto de datos que contiene variables tanto cuantitativas como cualitativas.

Hay varios paquetes de R que implementan métodos de componentes principales. Estos paquetes incluyen: FactoMineR, ade4, stats, ca, MASS y ExPosition. Sin embargo, el resultado se presenta de forma diferente según los paquetes utilizados. FactoMineR es un paquete elegante creado por Sebastián Le en el 2008 que permite calcular PCA, MCA, FAMD y MFA (Kassambara, 2020, p.1).

Ventajas

- El paquete R factoextra tiene métodos flexibles y fáciles de usar para extraer rápidamente, en un formato de datos estándar legible por humanos, los resultados del análisis de los diferentes paquetes mencionados anteriormente.
- Produce una elegante visualización de datos basada en ggplot2 con menos tipeo.
- También contiene muchas funciones que facilitan el análisis y la visualización de agrupamiento (Kassambara, 2020, p.1).

Principales funciones del paquete factoextra

- Visualización de resultados de análisis de reducción de dimensiones
- Extracción de datos de resultados de análisis de reducción de dimensiones
- Análisis y visualización de agrupamiento

Tabla 3-1: Funciones del paquete factoextra.

Función	Descripción
fviz_screepplot	Grafique los valores propios/varianzas contra el número de dimensiones.
fviz_pca_var	Gráfico de variables.
fviz_pca	Gráfica de individuos/variables de la salida del Análisis de Componentes Principales (PCA).
fviz_pca_biplot	Biplot de individuos y variables.
fviz_dend	Visualización mejorada del dendrograma.
fviz_cluster	Visualizar resultados de agrupamiento.

Fuente: Kassambara, 2020, p.1

Realizado por: López, Cristina; Montero, Magdalena. 2022

1.8. Requisitos para pertenecer a la CACECH

Para ser pertenecer a la Cooperativa de Ahorro y Crédito “Educadores de Chimborazo” las características que deben cumplir o reunir una persona natural es:

- Copia de cédula y papeleta de votación a color.
- Foto tamaño carnet.
- Contrato o nombramiento del magisterio.
- Apertura de la cuenta con \$45 dólares.

CAPÍTULO II

2. MARCO METODOLÓGICO

2.1. Tipo y diseño de la investigación

Según el método se considera una investigación mixta, ya que la información se centra en variables estadísticas numéricas y categóricas; como objetivo: teórico, se enfoca en el análisis de clusters espaciales, pues se identifica el clúster de jerarquía más adecuado y óptimo entre la relación económica y geográfica de los accionistas; las mudables utilizadas no fueron manipuladas, por lo que se estima como una investigación no experimental; según el tipo de inferencia es deductiva partiendo desde las bases generales hasta las específicas, en donde se utilizó clúster espacial jerárquico aglomerativo, calificada como transversal según el periodo temporal 2016-2020 porque la información adquirida se encuentran en una región específica del tiempo sobre la provincia de Chimborazo que contiene a los socios de la cooperativa de Ahorro y Crédito “Educadores de Chimborazo” Ltda.

2.2. Descripción del Área de estudio

La presente investigación se ubica en Ecuador al centro de la región interandina, en la provincia de Chimborazo con una población total 458581 habitantes según el Instituto Ecuatoriano de Estadísticas y Censos (INEC) (Censo, 2010), siendo la novena provincia más poblada del Ecuador, dividida en 10 cantones que son: Alausí, Chambo, Chunchi, Colta, Cumanda, Guamote, Guano, Pallatanga, Penipe y Riobamba; 59 parroquias de las cuales 42 son rurales y 17 urbanas. El cantón Riobamba es considerado como la urbe más grande y poblada de la provincia. Por la importante presencia de muestras arquitectónicas coloniales y en especial republicanas.

La ciudad es estimada como Patrimonio Cultural de la nación, la Cooperativa de Ahorro y Crédito “Educadores de Chimborazo” Ltda. se encuentra ubicada dentro de una infraestructura colonial con las siguientes coordenadas geográficas 78° 38'41” longitud; 1°40'32” latitud en la provincia de Chimborazo, cantón Riobamba en las calles Eugenio Espejo y Veloz.



Gráfico 1-2: Ubicación geográfica de la CACECH

Fuente: Googlemaps



Gráfico 2-2: Mapa geográfico de la República del Ecuador

Realizado por: López, Cristina; Montero, Magdalena. 2022

2.3. Población de estudio

El estudio se realizó con los datos económicos, demográficos y geográficos de los accionistas proporcionados por el analista informático de la Cooperativa de Ahorro y Crédito “Educadores de Chimborazo” de los registros durante el periodo 2016-2020.

La población de estudio son los socios de la Cooperativa de Ahorro y Crédito “Educadores de Chimborazo”, tomando en cuenta a 3216 individuos que están distribuidos en los diferentes cantones de la provincia de Chimborazo: Colta con 600, Guano con 75, Cumanda con 16, Guamote con 14,

Alausí y Chambo con 8 respectivamente cada uno, Penipe con 2, Chunchi con 1 accionista, finalmente Pallatanga no cuenta con beneficiarios de la institución; existiendo una mayor concentración de estos en el cantón Riobamba con 2492 asociados.

Las variables de interés para el respectivo análisis son: económicas (ingresos y egresos), demográficas (residencia, sexo y profesión) y geográficas (longitud y latitud) de los socios, debido a que son los factores que más influyen al desarrollo económico de la provincia, además por la necesidad de proporcionar información confiable para la sociedad. Se considera las direcciones domiciliarias como variables geográficas para el estudio a través de la localización geográfica en coordenadas GPS y UTM.

2.4. Operacionalización de las variables

Tabla 4-2: Operacionalización de las variables de los socios de la CACECH

Tipo de variable	Variable	Unidad de medida	Tipo	Escala de medición	Descripción
Económicas	Ingresos	Dólares	Cuantitativas Continuas	Intervalo	Se refleja los movimientos económicos de los socios.
	Egresos				
Demográficas	Residencia	—	Cualitativa Dicotómica	Nominal	Característica que identifica la zona residencial de los socios.
	Sexo	—	Cualitativa Dicotómica	Nominal	Es aquella característica que identifica en dos grupos determinados a los socios.
	Profesión	—	Cualitativa Categórica	Ordinal	Es la característica que determina el nivel de educación de cada socio.
Geográficas	Longitud	Grados	Cuantitativas Discretas	Intervalo	Ángulo formado desde el centro de la tierra sobre el Ecuador y el meridiano que contiene en punto y origen.
	Latitud	Grados	Cuantitativas Discretas	Intervalo	Valor medido sobre la superficie terrestre que sirve para determinar la posición específica de cualquier punto.

Realizado por: López, Cristina; Montero, Magdalena. 2022

CAPÍTULO III

3. MARCO DE RESULTADOS Y DISCUSIÓN DE RESULTADOS

3.1. Análisis Exploratorio de datos

De acuerdo con el último Censo de población del Instituto Nacional de Estadísticas y Censos (INEC) en el año 2010, el Ecuador tiene 17.37 millones de habitantes los cuales están dispersos alrededor de todo el país distribuidos en las 24 provincias, siendo Chimborazo el noveno distrito con un 2,69% de la población total del país.

Con la información proporcionada por la cooperativa de ahorro y crédito “Educadores de Chimborazo” Ltda., para la presente investigación se consideró un análisis de clúster jerárquico aglomerativo. La matriz de información original reflejaba 3592 socios reduciendo a 3216 accionistas a través de una limpieza de datos tomando en consideración a los beneficiarios que residen dentro de la provincia, con 4 tipos de variables cuantitativas: ingresos, egresos, coordenadas de latitud y longitud, y con 4 variables cualitativas: sexo, parroquia, zona y actividad económica, datos tomados por la entidad financiera.

Tabla 5-3: Frecuencia de los socios de la CACEH en los cantones de Chimborazo

Cantón	Frecuencia	Porcentaje
Alausí	8	0.25%
Chambo	8	0.25%
Chunchi	1	0.03%
Colta	600	18.65%
Cumanda	16	0.50%
Guamote	14	0.44%
Guano	75	2.33%
Pallatanga	0	0.00%
Penipe	2	0.06%
Riobamba	2492	77.00%
TOTAL	3216	100%

Fuente: CACECH, 2016-2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

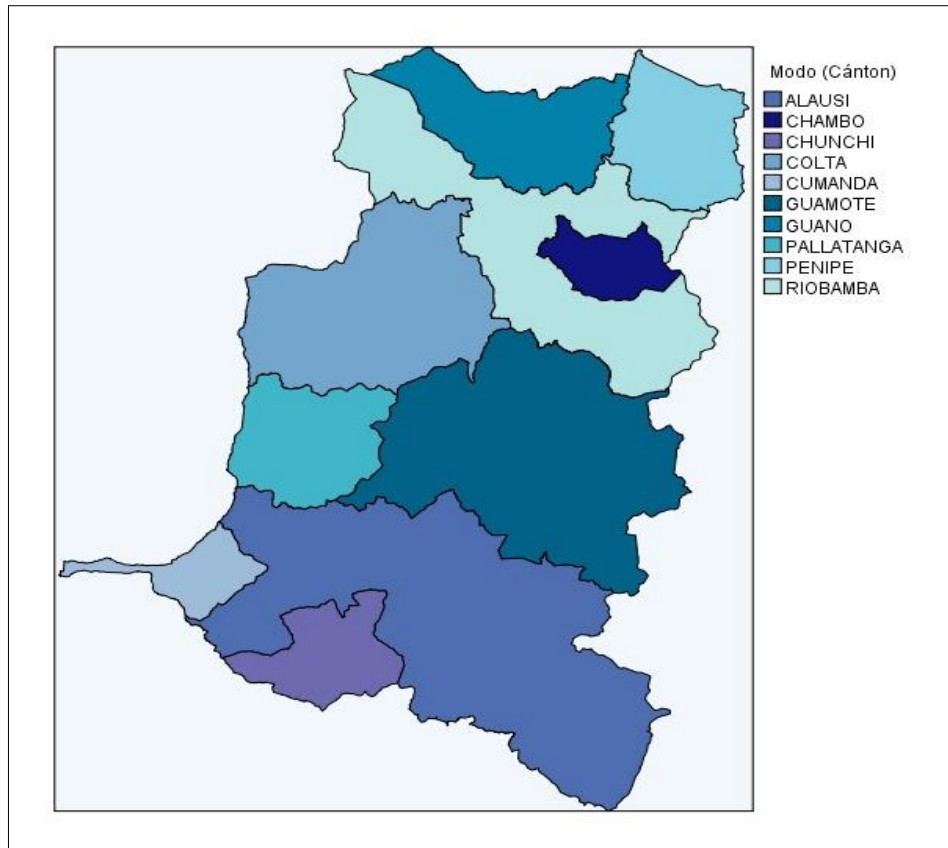


Gráfico 3-3: Densidad cantonal de socios de la CACECH en Chimborazo

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 3-3 presenta la densidad de socios de la CACECH distribuidos en 10 cantones de la provincia, no se considera Pallatanga al no presentar asociados, siendo el cantón Riobamba la localidad más grande que contiene al 77% de la población a comparación del cantón Chunchi donde se alberga el 0.03% de su totalidad.

Análisis exploratorio de las variables cualitativas

Tabla 6-3: Frecuencia de los socios de la CACECH según el sexo

Sexo	Frecuencia	Porcentaje
Femenino	2134	66.40%
Masculino	1082	33.60%

Fuente: CACECH, 2016-2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

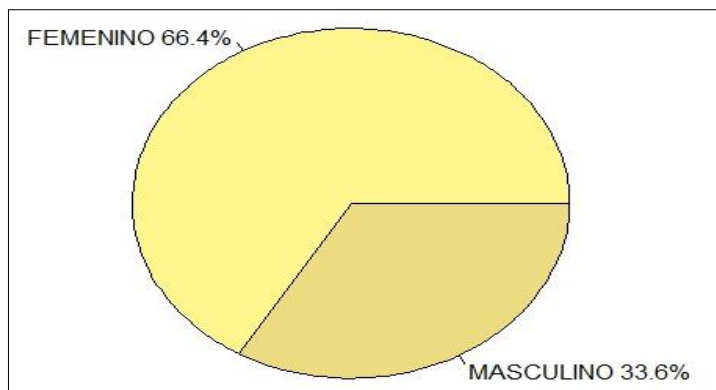


Gráfico 4-3: Sexo de los socios de la CACECH

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 4-3 muestra que el 66.40% de socios de la CACECH son accionistas de sexo femenino y el 33.60% pertenece al sexo masculino.

Tabla 7-3: Frecuencia de los socios de la CACECH según la zona residencial

Zona	Frecuencia	Porcentaje
Rural	123	4%
Urbana	3093	96%

Fuente: CACECH, 2016-2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

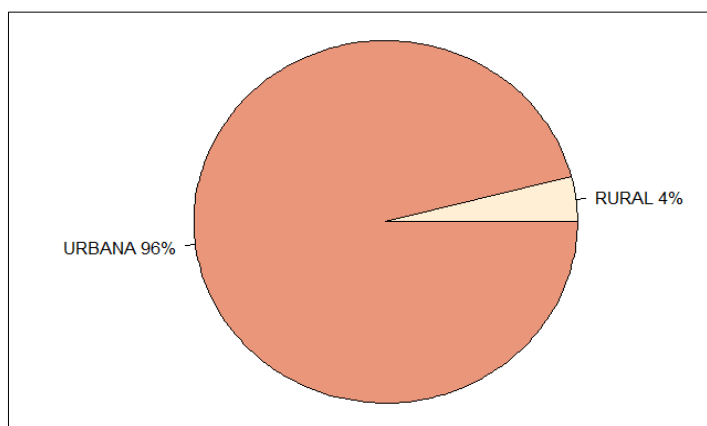


Gráfico 5-3: Zona domiciliaria de los socios de la CACECH

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 5-3 muestra que el 96% de socios de la CACECH residen en la zona urbana y el 4% de los accionistas en la zona rural de la provincia.

Tabla 8-3: Frecuencia de los socios de la CACECH según su actividad económica

Actividad Económica	Frecuencia	Porcentaje
Docente Universitario	62	1,94%
Empleado Privado	2146	67,29%
Empleado Público	758	23,77%
Jubilado	217	6,80%
Pensión Alimenticia	6	0,19%

Fuente: CACECH, 2016-2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

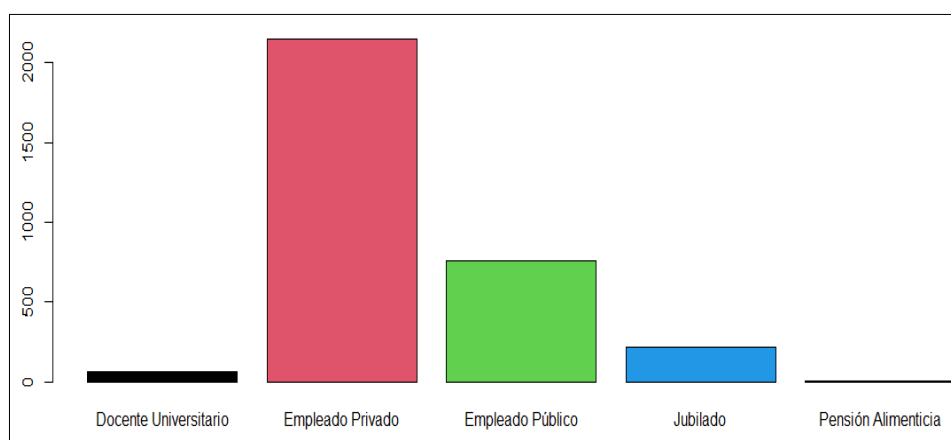


Gráfico 6-3: Frecuencia de socios de la CACECH según la actividad económica

Realizado por: López, Cristina; Montero, Magdalena. 2022

Las actividades económicas de los socios que se presentan en el gráfico 6-3 se dividen en 5 diligencias laborales alrededor de la provincia teniendo una mayor agrupación la contratación de empleados privados con 67,29%, seguida por 23,77 % que representa la inserción profesional en trabajos en sector público del país y una concentración mínima enfocada en el servicio de pensión alimenticia.

Tabla 9-3: Frecuencia de los socios de la CACECH según la parroquia

Parroquias	Frecuencia	Porcentaje
Alausí	2	0.06%
Cajabamba	641	19.93%
Calpi	22	0.68%
Chambo	7	0.22%
Chunchi	1	0.03%

Columbe	24	0.75%
Cubijies	3	0.09%
Cumanda	17	0.53%
El Rosario	1	0.03%
Flores	1	0.03%
Guamote	12	0.37%
Guano	48	1.49%
Huigra	4	0.12%
La Matriz	10	0.31%
La Providencia	1	0.03%
Lican	22	0.68%
Licto	1	0.03%
Lizarzaburu	610	18.97%
Maldonado	146	4.54%
Multitud	1	0.03%
Palmira	2	0.06%
Penipe	1	0.03%
Pungala	1	0.03%
Quimiag	2	0.06%
San Andrés	11	0.34%
San Gerardo de Pacaicagan	1	0.03%
San Isidro De Patulu	4	0.12%
San Juan	8	0.25%
San Luis	14	0.44%
Sicalpa	13	0.40%
Tixan	1	0.03%
Valparaiso	1	0.03%
Velasco	1424	44.28%
Veloz	126	3.92%
Yaruquies	33	1.03%

Fuente: INEC, 2010

Realizado por: López, Cristina; Montero, Magdalena. 2022

En la tabla 7-3 se muestra 35 de las 59 parroquias de la provincia de Chimborazo debido a que en estas localidades se encuentran ubicadas las residencias de los socios de la Cooperativa de Ahorro y Crédito “Educadores de Chimborazo”.

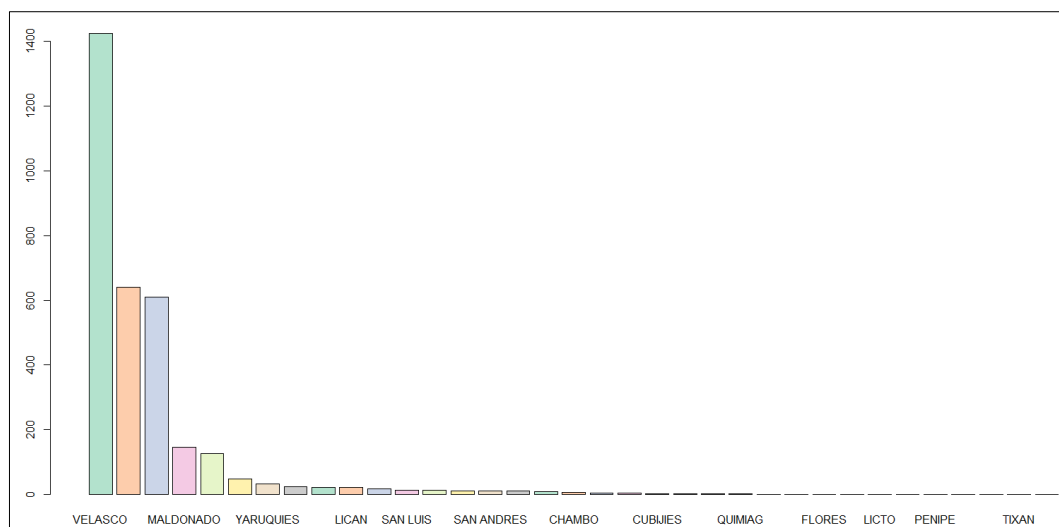


Gráfico 7-3: Frecuencia de socios de la CACECH por parroquias

Realizado por: López, Cristina; Montero, Magdalena. 2022

La distribución residencial de los socios que se muestran en el gráfico 7-3 está dividida en 35 parroquias en los diferentes cantones de Chimborazo, observando una concentración de los asociados en la congregación Velasco con un 44.28%, seguida por Cajabamba con 19,93% y Lizarzaburu con el 18,97% siendo estas tres localidades las más pobladas en la provincia; mientras que 9 de estas parroquias están representadas con el 0.03% de la población. Se observó que los cantones Chambo, Cumanda y Pallatanga no tienen parroquias.

Análisis exploratorio de las variables cuantitativas

Tabla 10-3: Descriptivo anual de los ingresos de los socios de la CACECH (2016-2020)

Descriptivos	Año 2016	Año 2017	Año 2018	Año 2019	Año 2020
Media	11503.04	11503.04	11447.70	9295.36	9295.36
Mediana	11021.00	11021.00	10824.92	8872.62	8872.62
Moda	11021.00	11021.00	10824.92	8872.62	8872.62
Desv.Estándar	2074.74	2074.74	2445.97	1739.23	1739.23
Varianza	4304549.87	4304545.71	5982784.27	3024936.45	3024936.45
Curtosis	53.92	53.92	41.66	44.10	44.10
Cof.Asimetría	5.81	5.81	5.07	5.20	5.20
Rango	29899.00	29899.00	40082.00	26457.20	26457.20
Mínimo	6250.00	6250.00	6142.00	4887.00	4887.00
Máximo	36149.00	36149.00	46224.00	31344.20	31344.20

Realizado por: López, Cristina; Montero, Magdalena. 2022

La tabla 8-3 muestra los ingresos promedio anuales de los socios de la CACECH de 2016-2020 los cuales oscilan entre \$9295.36 y \$ 11503.04 con una variabilidad de \$1739.23 y máxima de \$2074.74 con respecto a la media, con ingresos mínimos de \$4887.00 a \$6250.00 y máximos de \$31344.20 y \$46224.00. Los ingresos se distribuyen de forma leptocúrtica y con coeficientes de asimetría positivos para los 5 años.

Tabla 11-3: Descriptivo anual de los egresos de los socios de la CACECH (2016-2020)

Descriptivos	Año 2016	Año 2017	Año 2018	Año 2019	Año 2020
Media	5753.05	5983.53	7358.55	7830.71	10917.95
Mediana	5697.00	5982.00	7302.36	7831.81	12320.95
Moda	5804.00	6050.00	7974.60	7547.28	12498.20
Desv.Estándar	716.94	290.31	949.65	520.38	2171.20
Varianza	514003.62	84280.76	901835.29	270790.29	4714094.82
Curtosis	18.84	-0.07	13.00	3.20	-1.64
Cof.Asimetría	2.71	0.03	2.04	-0.60	-0.43
Rango	9095.00	2076.00	12054.12	4986.56	8045.72
Mínimo	3853.00	4921.00	4518.30	4903.03	5880.08
Máximo	12948.00	6997.00	16572.42	9889.59	13925.80

Realizado por: López, Cristina; Montero, Magdalena. 2022

La tabla 9-3 muestra los egresos promedio anuales de los socios de la CACECH de 2016-2020 los cuales oscilan entre \$5753.05y \$ 10917.95 con una variabilidad de \$290.31 y máxima de \$2171.20 con respecto a la media, con egresos mínimos de \$3853.00 a \$5880.08 y máximos de \$6997.00 y \$13925.80. Los egresos presentan asimetría negativa en el último año por la presencia de la pandemia covid-19 ya que genero mayores gastos en los hogares de los asociados.

Los movimientos bancarios de los socios de la CACECH en los años (2016-2018) reflejan remuneraciones de 11 mil con circulaciones de 5 a 7 mil dólares, es decir, obtienen un ahorro en sus cuentas, razón por la que en el año 2019 a través de la devaluación del peso Colombia y el libre comercio en la frontera Ecuador-Colombia los ecuatorianos viajaron a adquirir insumos a menor precio y en el año 2020 se produjo la pandemia covid-19 con ingresos que permanecen en 9 mil dólares e incrementan los egresos debido a salidas financieras no previstos en el presupuesto de los accionistas provocando un daño colateral a su economía (Primicias, 2018).

Análisis Exploratorio de los ingresos de los socios de la CACECH (2016 -2020)

Tabla 12-3: Ingresos promedios de los socios de CACECH (2016-2020)

Año 2016	Año 2017	Año 2018	Año 2019	Año 2020
11503,04	11503,04	11447,70	9295,36	9295,36

Realizado por: López, Cristina; Montero, Magdalena. 2022

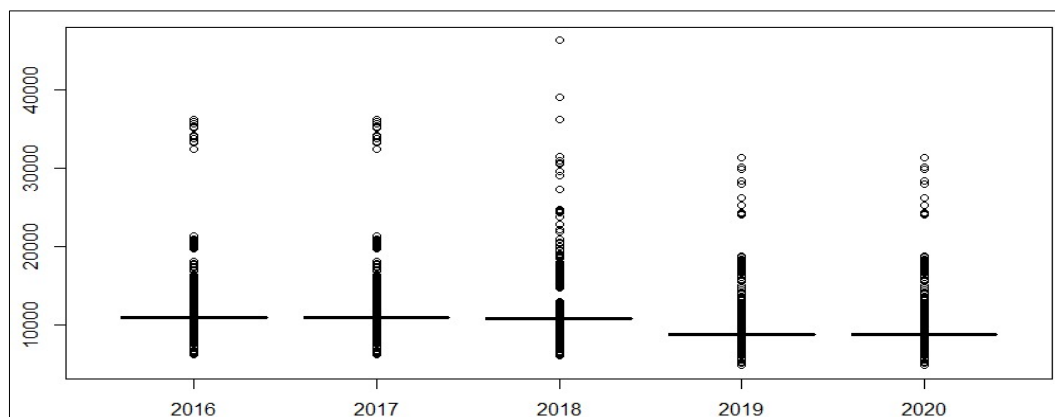


Gráfico 8-3: Ingresos por actividad económica de los socios de la CACECH (2016-2020)

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 8-3 muestra la media anual de los ingresos de los socios de cooperativa de Ahorro y Crédito “Educadores de Chimborazo” en el periodo 2016-2020, donde los 3 primeros años muestra valores similares a diferencia de los 2 últimos años con un descenso. La disminución de los ingresos es el reflejo del efecto colateral producido por la pandemia, ya que para salvaguardar la economía del país redujo un 20% los salarios de los empleados públicos y privados.

Análisis exploratorio de los egresos de los socios de la CACECH (2016 – 2020)

Tabla 13-3: Egresos por actividad económica de los socios de CACECH (2016-2020)

Actividad Económica	Egresos 2016	Egresos 2017	Egresos 2018	Egresos 2019	Egresos 2020
Docente Universitario	\$ 5655.32	\$ 5927.85	\$ 7224.37	\$ 7783.77	\$ 10420.67
Empleado Privado	\$ 5697.94	\$ 5983.65	\$ 7226.75	\$ 7840.54	\$ 11466.99
Empleado Público	\$ 5688.18	\$ 5973.19	\$ 7253.73	\$ 7841.10	\$ 11319.64
Jubilado	\$ 5701.69	\$ 5999.52	\$ 7179.81	\$ 7870.15	\$ 12118.94
Pensión Alimenticia	\$ 5734.60	\$ 6060.67	\$ 7561.62	\$ 7847.16	\$ 10718.05

Realizado por: López, Cristina; Montero, Magdalena. 2022

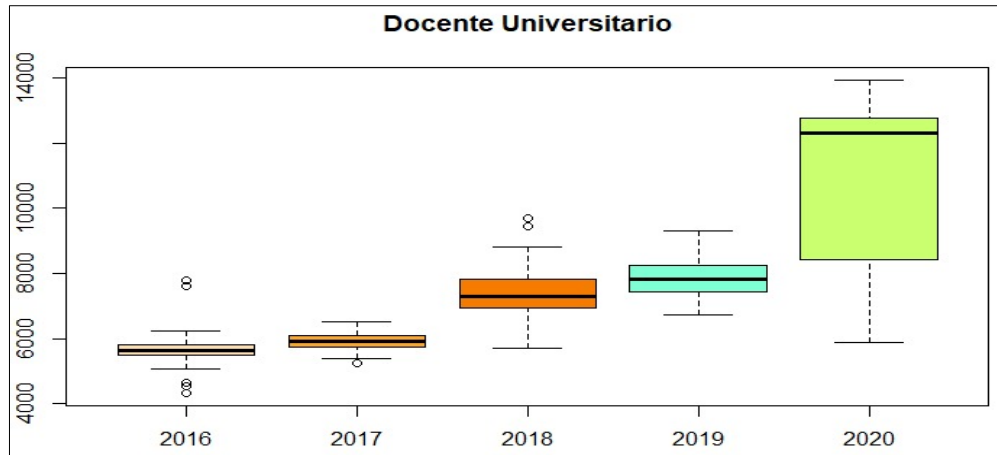


Gráfico 9-3: Egresos de docentes universitarios de la CACECH (2016-2020)

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 9-3 muestra la media anual de los egresos de los docentes universitarios socios de la CACECH en el periodo 2016-2020 donde los egresos oscilan entre \$5655.32 y \$10420.67 dólares, reflejando un aumento en el año 2020 ocasionado por la presencia de la pandemia ya que por el confinamiento y factores externos los accionistas generaron consumos mayores a los años anteriores.

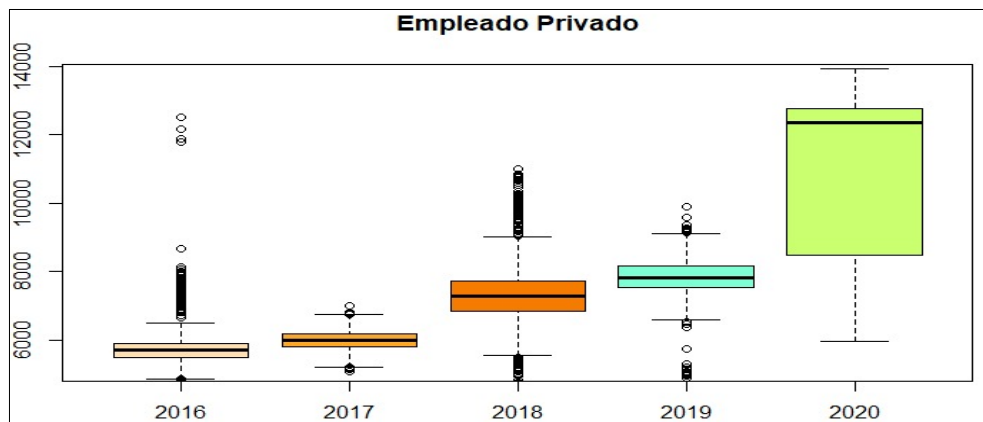


Gráfico 10-3: Egresos de empleados privados de la CACECH (2016-2020)

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 10-3 muestra la media anual de los egresos de los empleados privados socios de la CACECH en el periodo 2016-2020 donde los egresos oscilan entre \$5697.94 y \$11466.99 dólares, reflejando un aumento en el año 2020 ocasionado por la presencia de la pandemia ya que por el confinamiento y factores externos los accionistas generaron consumos mayores a los años anteriores.

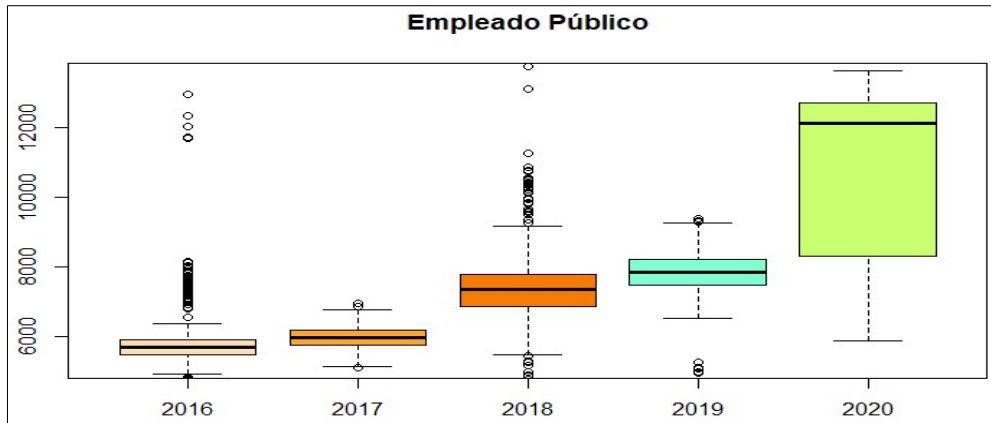


Gráfico 11-3: Egresos de empleados públicos de la CACECH (2016-2020)

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 11-3 muestra la media anual de los egresos de los empleados públicos socios de la CACECH en el periodo 2016-2020 donde los egresos oscilan entre \$5688.18 y \$11319.64 dólares, reflejando un aumento en el año 2020 ocasionado por la presencia de la pandemia ya que por el confinamiento y factores externos los accionistas generaron consumos mayores a los años anteriores.

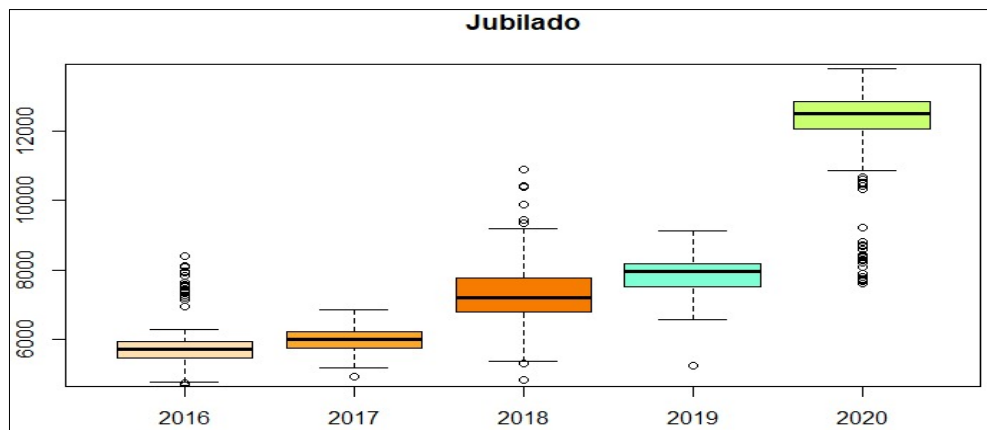


Gráfico 12-3: Egresos de jubilados de la CACECH (2016-2020)

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 12-3 muestra la media anual de los egresos de los pensionistas jubilados socios de la CACECH en el periodo 2016-2020 donde los egresos oscilan entre \$5701.69 y \$12118.94 dólares, reflejando un aumento en el año 2020 ocasionado por la presencia de la pandemia ya que por el confinamiento y factores externos los accionistas generaron consumos mayores a los años anteriores.

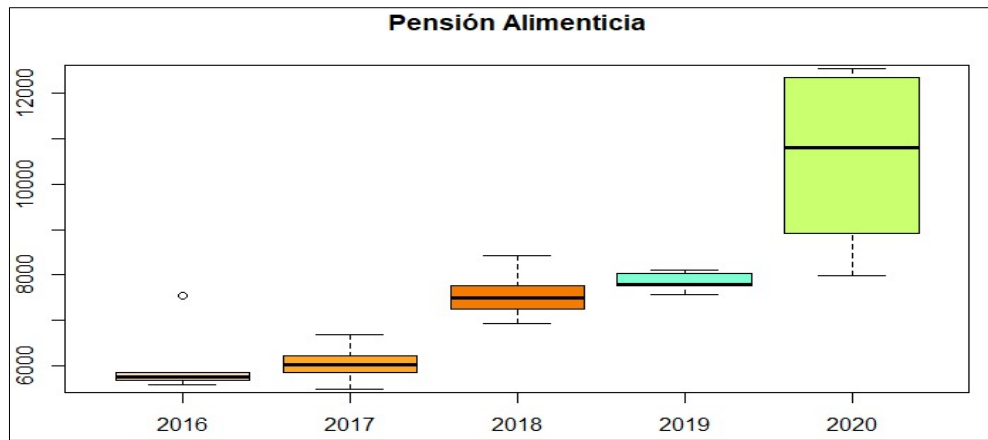


Gráfico 13-3: Egresos de pensiones alimenticias de la CACECH (2016-2020)

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 13-3 muestra la media anual de los egresos de los socios beneficiarios de pensiones de alimentos de la CACECH en el periodo 2016-2020 donde los egresos oscilan entre \$5734.60 y \$10718.05 dólares, reflejando un aumento en el año 2020 ocasionado por la presencia de la pandemia ya que por el confinamiento y factores externos los accionistas generaron consumos mayores a los años anteriores.

Tabla 14-3: Egresos por zona de los socios de la CACECH (2016-2020)

Zona	Egresos 2016	Egresos 2017	Egresos 2018	Egresos 2019	Egresos 2020
Rural	\$ 5682.00	\$ 6042.06	\$ 6042.06	\$ 7914.64	\$ 11514.42
Urbana	\$ 5695.35	\$ 5979.83	\$ 5979.83	\$ 7893.88	\$ 11457.63

Realizado por: López, Cristina; Montero, Magdalena. 2022

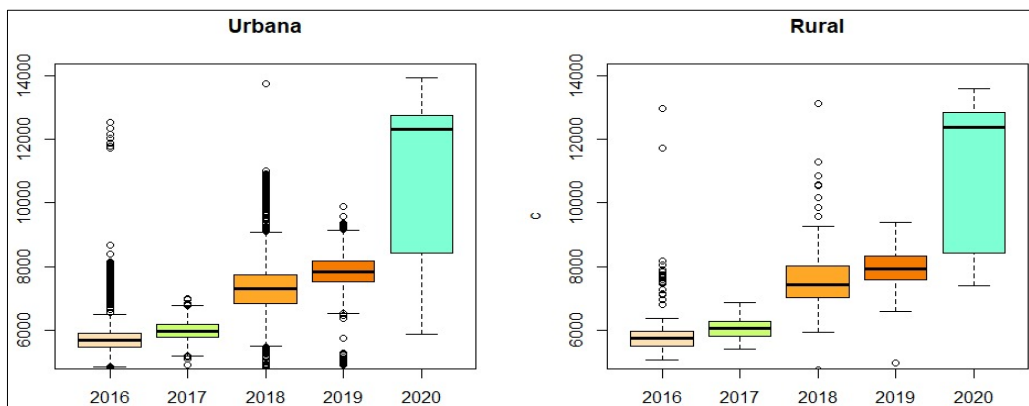


Gráfico 14-3: Egresos por zona de los socios de la CACECH (2016-2020)

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 14-3 muestra la media anual de los egresos de los socios de la CACECH en el periodo 2016-2020 según su zona residencial, donde los egresos en la zona rural oscilan entre \$5682.00 y \$11514.42 dólares y en la urbana de \$ 5695.35 a \$ 11457.63, reflejando un aumento en el año 2020 ocasionado por la presencia de la pandemia ya que por el confinamiento y factores externos los accionistas generaron consumos mayores a los años anteriores.

Tabla 15-3: Egresos por sexo de los socios de la CACECH (2016-2020)

Género	Egresos 2016	Egresos 2017	Egresos 2018	Egresos 2019	Egresos 2020
Femenino	\$ 5692.85	\$ 5980.83	\$ 7237.03	\$ 7832.27	\$ 11564.69
Masculino	\$ 5699.65	\$ 5982.75	\$ 7214.87	\$ 7862.51	\$ 11229.47

Realizado por: López, Cristina; Montero, Magdalena. 2022

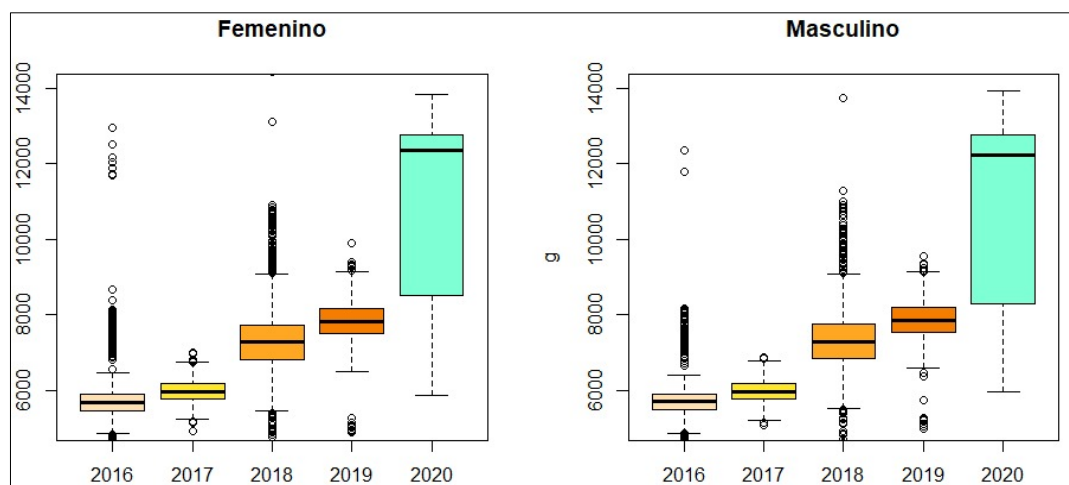


Gráfico 15-3: Egresos por sexo de los socios de la CACECH (2016-2020)

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 15-3 muestra la media anual de los egresos de los socios de la CACECH en el periodo 2016-2020 según el sexo, donde los egresos de las mujeres oscilan entre \$5692.85 y \$11564.69 dólares y de los hombres de \$ 5699.65 a \$ 11229.47, reflejando un aumento en el año 2020 ocasionado por la presencia de la pandemia ya que por el confinamiento y factores externos los accionistas generaron consumos mayores a los años anteriores.

Análisis exploratorio temporal 2016-2020

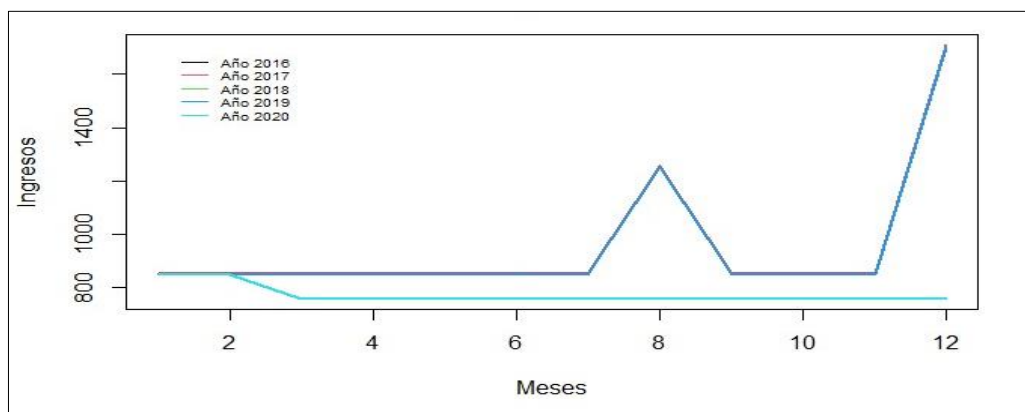


Gráfico 16-3: Serie temporal de los ingresos de los socios de la CACECH (2016-2020)

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 16-3 de los ingresos de los socios de la CACECH del periodo 2016-2020 presentan movimientos constantes en los 4 primeros años, se evidencia un descenso en el año 2020 a causa de la pandemia generada por el covid-19 donde existió una reducción del 20% de los sueldos en varios sectores laborales.

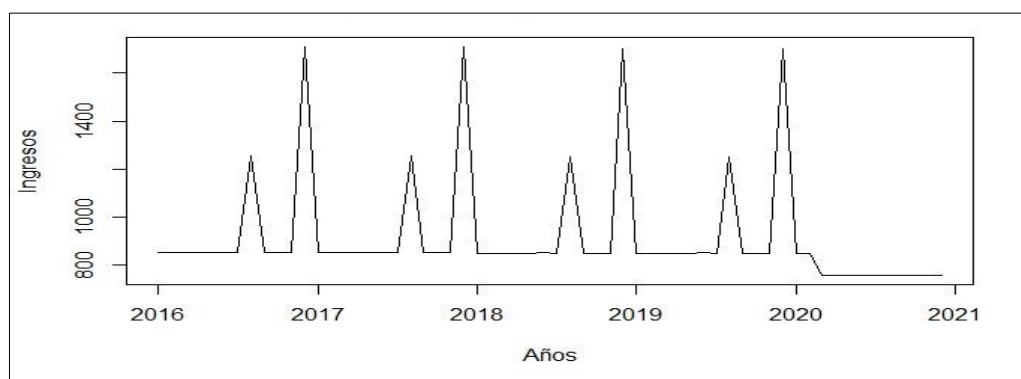


Gráfico 17-3: Ingresos de los socios de la CACECH (2016-2020)

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 17-3 de serie temporal de los ingresos de los socios de la CACECH del periodo 2016-2019 presenta los ingresos constantes año tras año en la institución con alteraciones en el mes de agosto y diciembre por el décimo tercero y cuarto sueldo, mientras que para el año 2020 con la reducción de sueldos del 20% se refleja un decrecimiento considerable en la economía de los asociados.

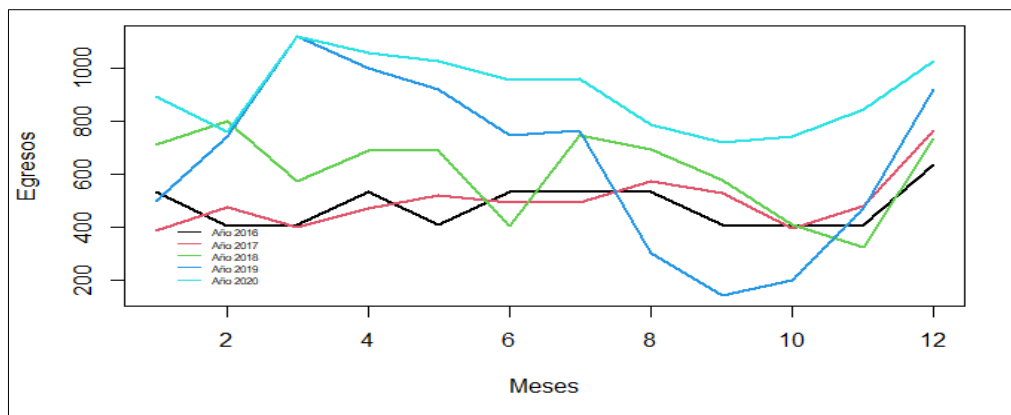


Gráfico 18-3: Serie temporal de los egresos de los socios de la CACECH (2016-2020)

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 18-3 de los egresos de los socios de la CACECH del periodo 2016-2020 durante los 3 primeros años presentan movimientos que oscilan entre los \$400 y \$800 dólares mensuales, con un cambio en el primer cuatrimestre del año 2019 con retiros de \$500 a \$1100 dólares por el libre comercio que se dio en la frontera Ecuador – Colombia (Primicias, 2018) y el último cuatrimestre con transacciones bancarias limitadas por el paro de octubre 2019 (González, 2019) y el conocimiento de la pandemia covid-19. En el año 2020 a partir del mes de marzo las salidas de dinero de los socios son más altos por el confinamiento y gastos externos generados por la crisis sanitaria a nivel nacional.

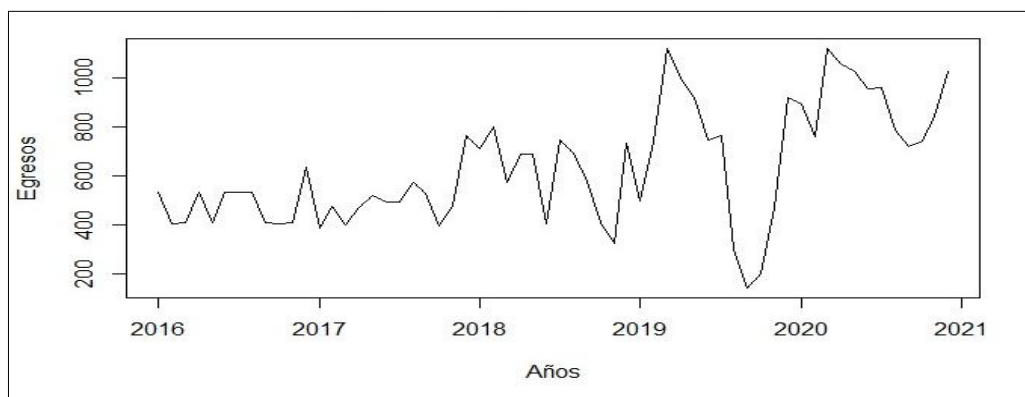


Gráfico 19-3: Egresos de los socios de la CACECH (2016-2020)

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 19-3 de serie temporal de los egresos de los socios de la CACECH del periodo 2016-2020 presenta tendencia creciente debido a la pandemia, el confinamiento y medidas de prevención generaron gastos no estipulados en los presupuestos de los socios.

Tabla 16-3: Tasa de movimiento anual de los ingresos de los socios de la CACECH

2016	-----
2017	0.00%
2018	-0.49%
2019	-18.80%
2020	0.00%

Realizado por: López, Cristina; Montero, Magdalena. 2022

Entre el año 2016 al 2017 no existió un movimiento de los ingresos de los socios por lo que se mantuvieron constantes con ingresos estables, entre 2017 al 2018 disminuyeron en 0.49%, en el año 2019 a comparación del 2018 los ingresos decrecieron en un 18.80% fenómeno provocado por el declive en el precio del barril del petróleo que afecto la economía del país (Reuters, 2017); a pesar de estos acontecimientos se manteniendo este nivel bajo de ingresos en el 2020 ya que no se observa una variación de los ingresos de los socios de la CACECH ya que por decisiones del gobierno los sueldos fueron estandarizados para evitar una crisis financiera generada por la pandemia covid-19.

Tabla 17-3: Tasa de movimiento anual de los egresos de los socios de la CACECH

2016	-----
2017	4.00%
2018	22.96%
2019	6.42%
2020	39.15%

Realizado por: López, Cristina; Montero, Magdalena. 2022

Los egresos decrecen la economía de los socios de la CACECH, en el año 2017 el 4% de crecimiento no representa salidas de efectivo de forma masiva manteniendo una estabilidad en la finanzas de la institución, para el 2018 se muestra un porcentaje considerable del 22.46% por gastos fuertes debido a la devaluación del peso colombiano y el libre comercio entre Colombia y Ecuador (Primicias, 2018); y con un cambio drástico en el año 2020 del 39.15% provocado por la crisis sanitaria y el confinamiento ocasionado por la pandemia covid-19 ya que los accionistas se encontraron en la necesidad de adquirir insumos alimenticios, de higiene, médicos y funerarios de manera apresurada.

3.2. Análisis Exploratorio Espacial

La información proporcionada por la cooperativa de ahorro y crédito “Educadores de Chimborazo” Ltda., presenta la mayor cantidad de socios en el cantón Riobamba con un 77% de la población total, por lo que, se realizó el análisis espacial y de conglomerados exclusivamente sobre la zona urbana de la ciudad.

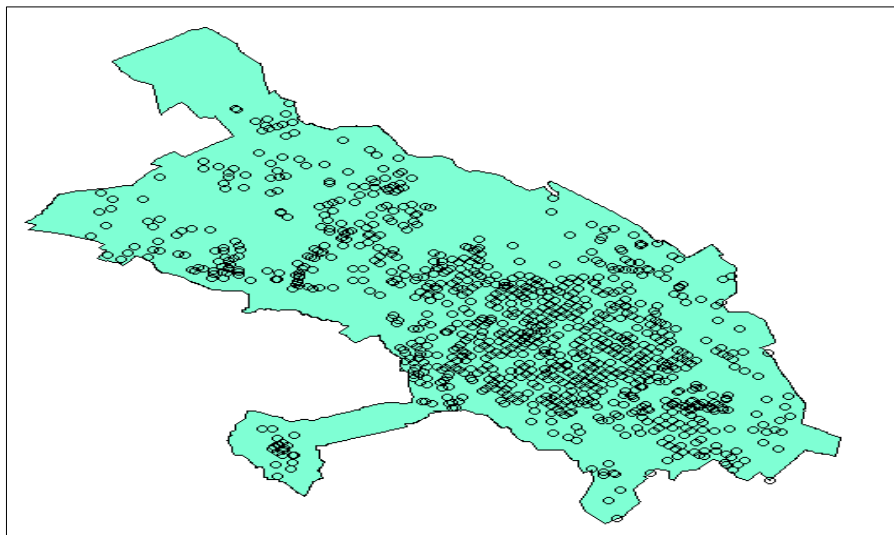


Gráfico 20-3: Ubicación geográfica de los socios en el cantón Riobamba

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 20-3 muestra la ubicación geográfica de los socios en el cantón Riobamba mediante su localización geográfica en coordenadas UTM.

Tabla 18-3: Índice de autocorrelación espacial del periodo 2016-2020

Variable	Índice de Moran	p- value
Ingresos	-0.01	0.47
Egreso 2016	-0.03	0.07
Egreso 2017	0.001	0.92
Egreso 2018	-0.02	0.07
Egreso 2019	-0.02	0.10
Egreso 2020	-0.003	0.87

Realizado por: López, Cristina; Montero, Magdalena. 2022

A través del análisis exploratorio se identificó la presencia de datos anómalos por lo cual fue necesario realizar una limpieza. Se halló el índice de Moran, que determinó que las economías de los socios de

la CACECH tienen una autocorrelación negativa perfecta con una confiabilidad superior al nivel de significancia, que presenta un equilibrio entre los valores positivos y negativos del estudio espacial.

3.3. Análisis de Clustering Jerárquico

Al realizar el análisis de correlación espacial, se consideró que la economía presenta autocorrelación espacial negativa, eliminando la presencia de ruido generado por agrupaciones no específicas que distorsiona los conglomerados. Para aplicar la metodología de clúster jerárquico aglomerativo es fundamental estandarizar la matriz de información, para ello se usó como referencia la ubicación geográfica, las actividades económicas y los movimientos financieros (ingresos – egresos) de los accionistas de la CACECH durante los 5 años de estudio.

Tabla 19-3: Presencia de ruido en los egresos de los socios de la CACECH (2016-2020)

Año	Porcentaje
2016	15,13%
2017	0,08%
2018	4,37%
2019	0,68%
2020	0,04%

Realizado por: López, Cristina; Montero, Magdalena. 2022

La tabla 17-3 muestra los porcentajes de ruido debido a la no existencia de relación estadística en 2 tiempos diferentes de los movimientos financieros (egresos) de los socios, que fueron separados por cada año para evitar el sesgo en el análisis de los conglomerados. El año 2016 presenta el porcentaje más alto de ruido que se produjo por la caída del precio del petróleo afectando la estabilidad económica del país y además del terremoto suscitado en costas ecuatorianas en abril del mismo año (Reuters, 2017).

Análisis de Conglomerados de los ingresos de los socios de la CACECH del periodo 2020

Se realiza el análisis de clúster de ingresos solo del año 2020, debido a que los sueldos de los asociados se presentan en su mayoría de manera constante, generando valores similares año tras año. Para este estudio se aplica el proceso de componentes principales que muestra la variabilidad de las mudables que influyen en el estudio.

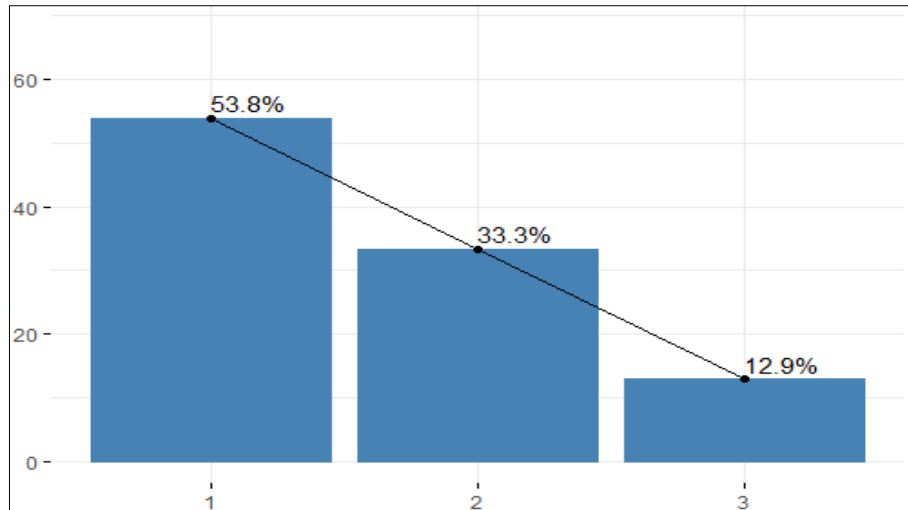


Gráfico 21-3: PCA de los ingresos de los socios de la CACECH del año 2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

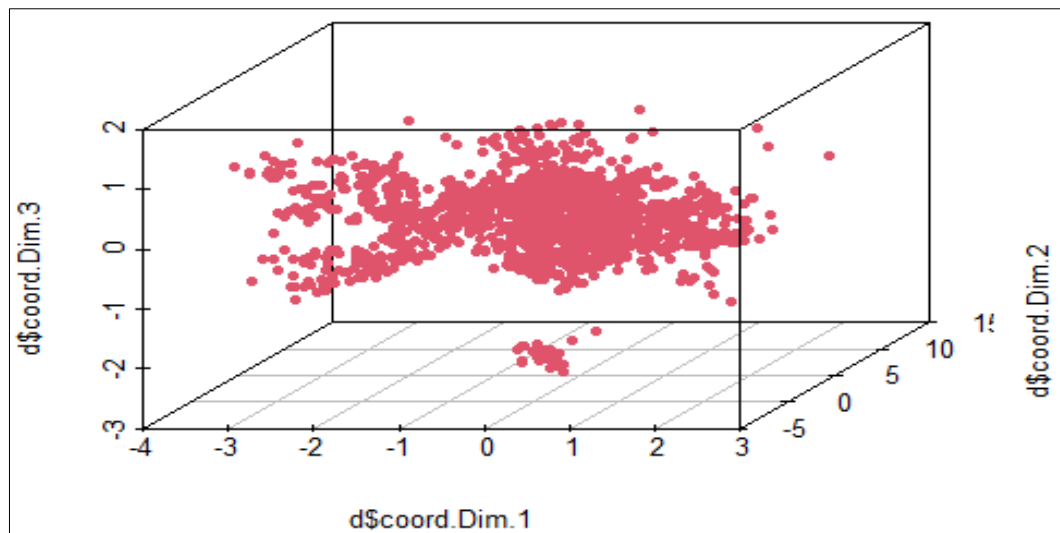


Gráfico 22-3: Homogeneidad de los ingresos de los socios de la CACECH del año 2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el método PCA presentado en el gráfico 21-3 se determina que las 2 primeras componentes explican un 87.14% de variabilidad total de los ingresos de los socios de la CACECH en el periodo 2020, juntamente con un test de homogeneidad Chi-cuadrado de 1.49 proporcionando una distribución homogénea explicada en el gráfico 22-2.

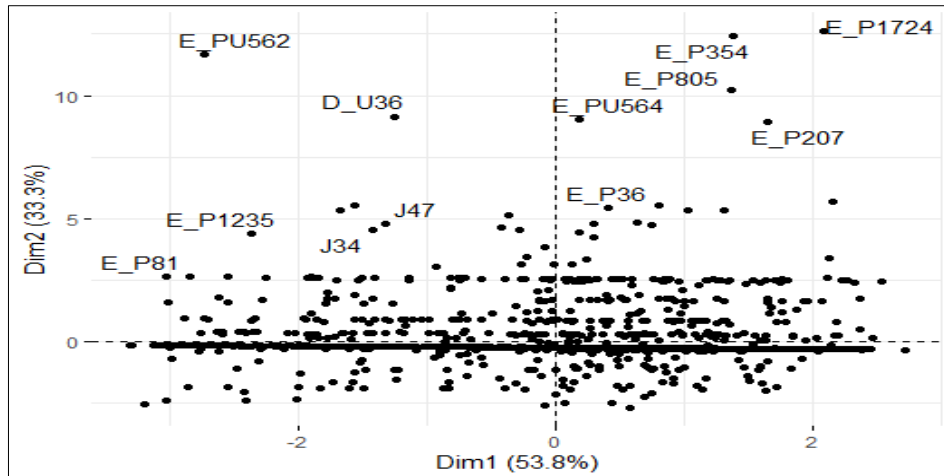


Gráfico 23-3: Biplot de los ingresos de los socios de la CACECH del año 2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el gráfico 23-3 se observa el aglutinamiento de la información en los cuatro cuadrantes, que indica la existencia de asociación mediante clúster. Además, el coeficiente estadístico de Hopkins con un índice de 0.08 que comprueba que la base de datos es clusterizable, con una concentración invariable en el tercer y cuarto cuadrante provocada por ingresos constantes.

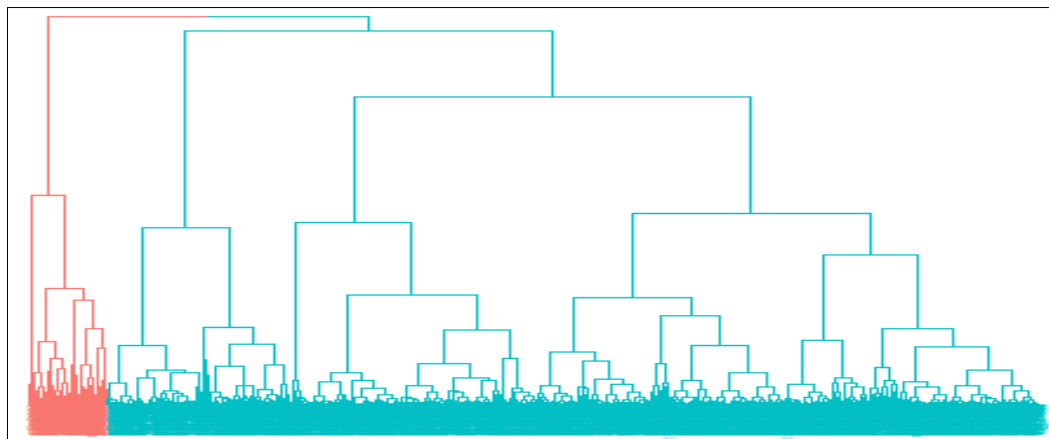


Gráfico 24-3: Dendrograma de los ingresos de los socios de la CACECH del año 2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 24-3 explica la distribución de 2 clústeres con una mayor concentración de los ingresos de los socios de la CACECH en el primer conglomerado según la actividad económica, mostrando similitudes entre la actividad laboral y la ubicación geográfica con cortes simétricos entre las dos agrupaciones siendo la razón por la que se intersecan varios puntos.

Tabla 20-3: Clúster de los ingresos de los socios de la CACECH año 2020

Actividad Económica	CLÚSTER	
	1	2
Docente Universitario	52	2
Empleado Privado	1543	117
Empleado Público	529	49
Jubilado	171	23
Pensión Alimenticia	5	1

Realizado por: López, Cristina; Montero, Magdalena. 2022

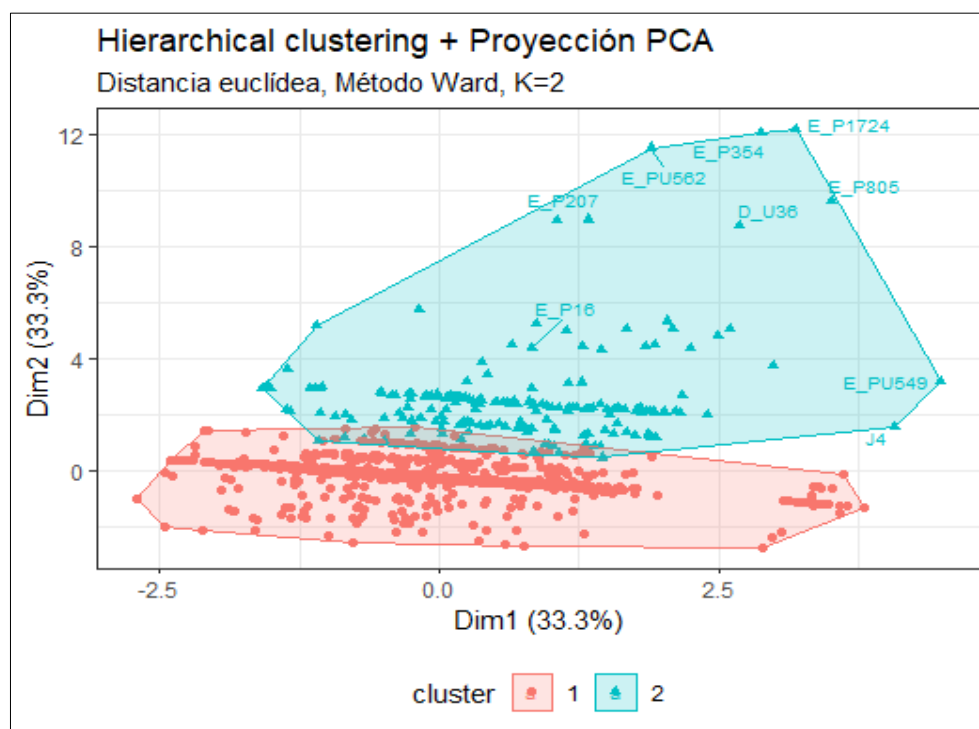


Gráfico 25-3: Clúster de los ingresos de los socios de la CACECH del año 2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el gráfico 25-3 que muestran los clústers de ingresos de los socios de la CACECH para año 2020 con un coeficiente de aglomeración del 0.72 se identifica 2 clústers explicados en la tabla 18-3 distribuidos según la actividad económica, con una mayor concentración en el primer clúster con 2300 asociados, estas agrupaciones son generadas por las ubicaciones geográficas y transacciones bancarias de los accionistas que en su mayoría se ubican en la parte izquierda inferior debido a la presencia de ingresos similares y por medidas financieras de emergencia como la modificación de sueldos y reducción de dividendos mensuales en algunos casos del sector privado.

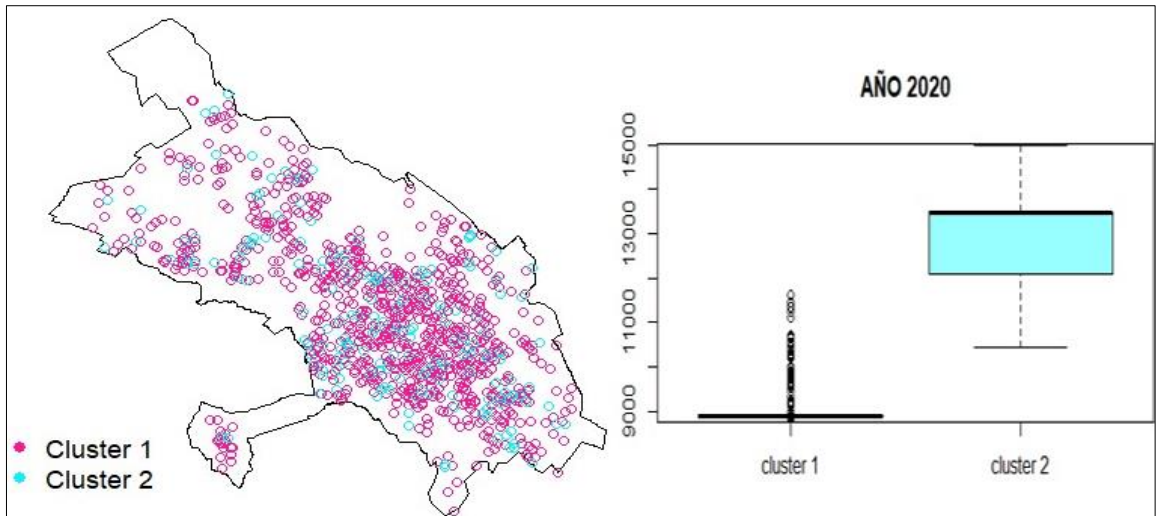


Gráfico 26-3: Lugar geográfico de los ingresos de los socios de la CACECH del año 2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 26-3 muestra la ubicación geográfica de los socios mediante la clasificación de clúster enfocada en los ingresos del año 2020, se detecta 2 grupos que se forman por similitudes tanto de los ingresos como ubicaciones geográficas de los socios, a partir de la distancia (euclídea). Y de algún modo se ratifica la agrupación expuesta en el gráfico 25-3 y la independencia espacial. Además, los diagramas de caja indican que los accionistas que integran el clúster 1 tienen ingresos promedios anuales de \$8884.95, mientras que el clúster 2 de \$13992.65 dólares por año, definiendo que el grupo 2 de los asociados presenta mayores ingresos económicos, que está relacionada al tipo de actividad laboral que realizan.

Análisis de Conglomerados de los egresos de los socios de la CACECH del periodo 2016

A continuación, se muestra el análisis clúster de los egresos por año debido a que en cada año los movimientos financieros de los asociados son diferentes generando la necesidad de realizar un estudio por separado, para este estudio se aplica el proceso de componentes principales que muestra la variabilidad de las mudables que influyen en el estudio.

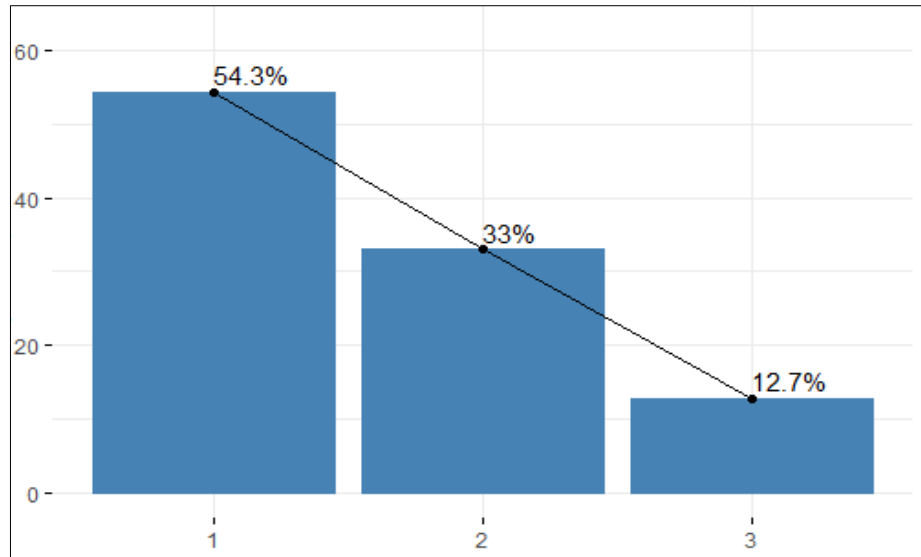


Gráfico 27-3: PCA de los egresos de los socios de la CACECH del año 2016

Realizado por: López, Cristina; Montero, Magdalena. 2022

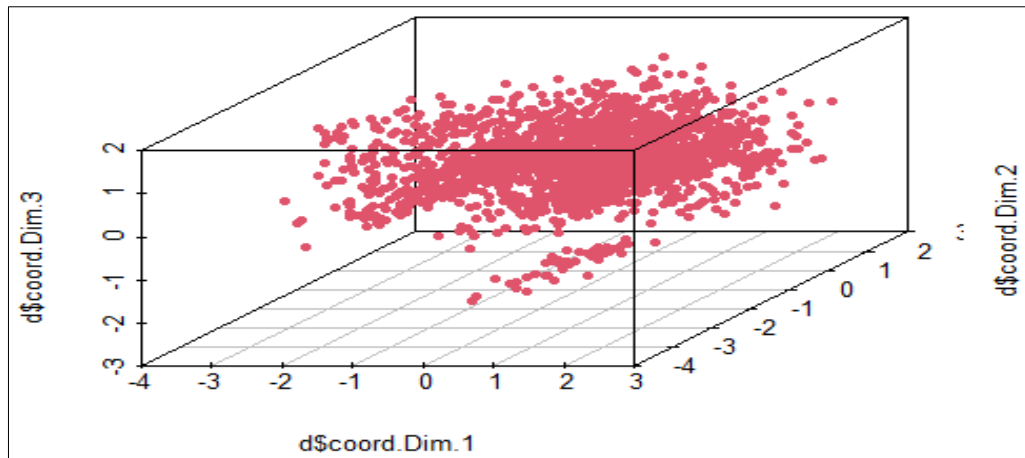


Gráfico 28-3: Homogeneidad de los egresos de los socios de la CACECH del año 2016

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el método PCA presentado en el gráfico 27-3 se determina que las 2 primeras componentes explican un 87.28% de variabilidad total de los egresos de los socios de la CACECH en el año 2016, a la vez con un test de homogeneidad Chi-cuadrado de 3.39 proporcionando una distribución homogénea explicada en el gráfico 28-3.

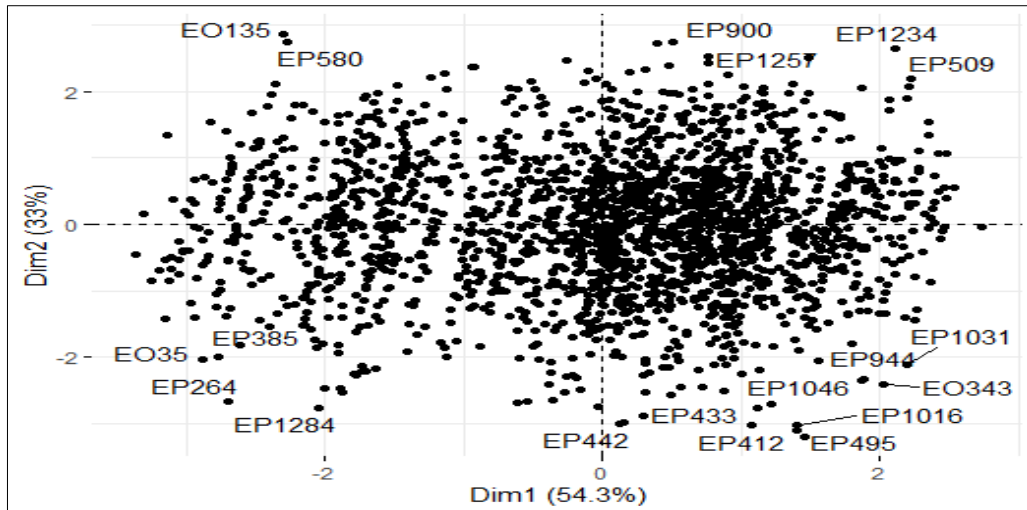


Gráfico 29-3: Biplot de los egresos de los socios de la CACECH del año 2016

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el gráfico 29-3 se observa el aglutinamiento de la información en los cuatro cuadrantes, que indica la existencia de asociación mediante clúster. Además, el coeficiente estadístico de Hopkins con un índice de 0.14 que evidencia que la base de datos es clusterizable.

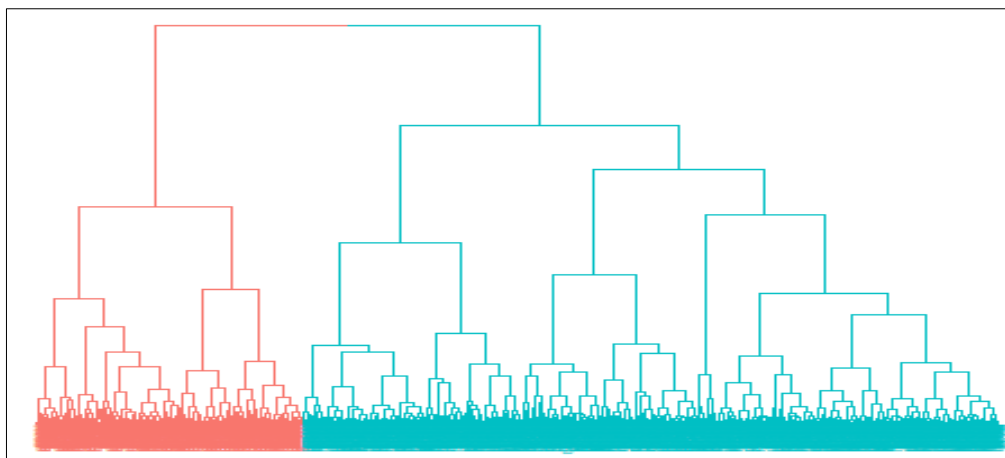


Gráfico 30-3: Dendrograma de los egresos de los socios de la CACECH del año 2016

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 30-3 de los egresos de los socios de la CACECH del año 2016 explica la distribución en 2 clústers con una mayor concentración en el primer conglomerado según la actividad económica, mostrando similitudes entre la actividad laboral y la ubicación geográfica con cortes simétricos entre las dos agrupaciones siendo la razón por la que se intersecan varios puntos.

Tabla 21-3: Clúster de los egresos de los socios de la CACECH del año 2016

Actividad económica	Clúster	
	1	2
Docente Universitario	37	13
Empleado Privado	1037	378
Empleado Público	343	149
Jubilado	107	45
Pensión Alimenticia	5	0

Realizado por: López, Cristina; Montero, Magdalena. 2022

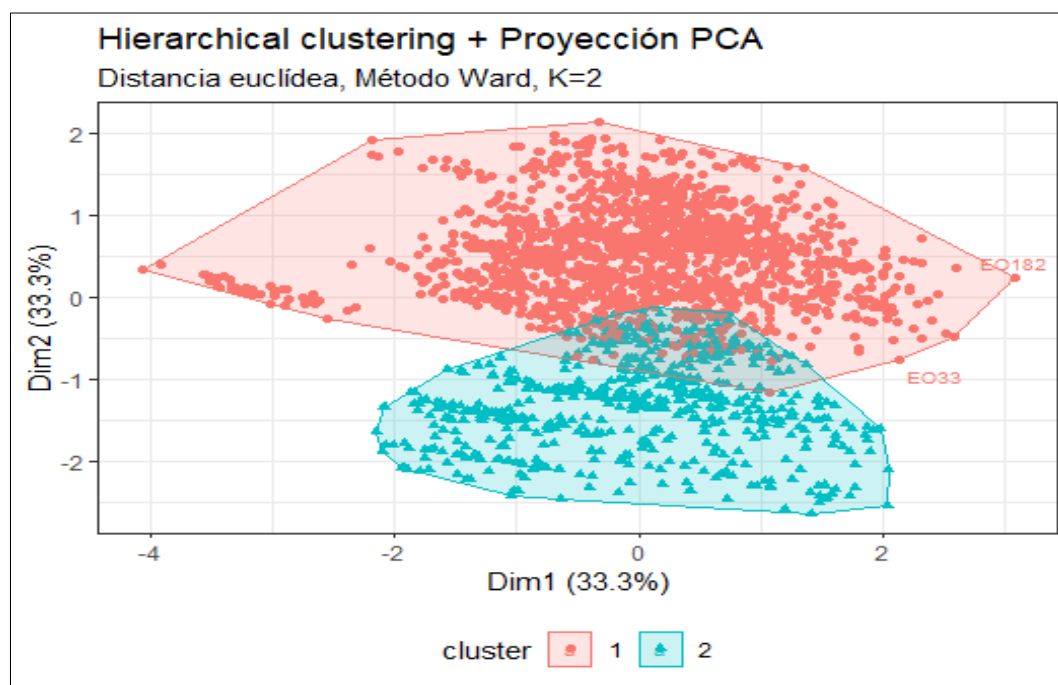


Gráfico 31-3: Clúster de los egresos de los socios de la CACECH del año 2016

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el gráfico 31-3 que muestran los clústeres de egresos de los socios de la CACECH para año 2016 con un coeficiente de aglomeración del 0.66 que identifica 2 clústeres explicados en la tabla 19-3 distribuidos según la actividad económica, con una mayor concentración en el primer clúster con 1529 asociados, estas agrupaciones son generadas por las ubicaciones geográficas y transacciones bancarias de los accionistas ubicando a los que presentan mayores retiros en la parte derecha, los de menor en la izquierda y en el centro los que muestran egresos similares y que se encuentran dentro del mismo perímetro zonal.

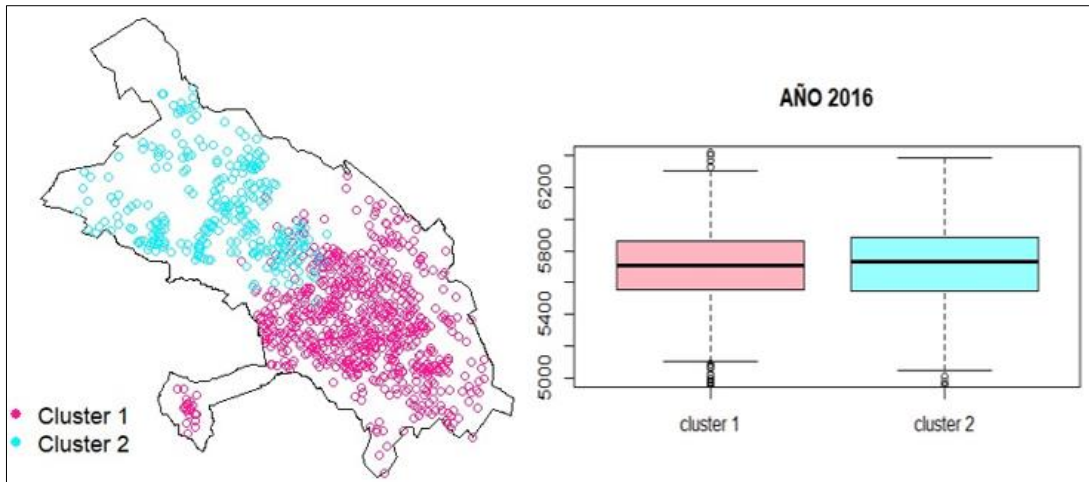


Gráfico 32-3: Lugar geográfico de los egresos de los socios de la CACECH del año 2016

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 32-3 muestra la ubicación geográfica de los socios mediante el mapa de clasificación de clúster enfocado en los egresos del año 2016, se detecta 2 grupos que se forman por similitudes territoriales y económicas a partir de la distancia (euclídea), mostrando mayor concentración en las parroquias Lizarzaburu y Velasco con un egreso promedio anual de \$5646.25 a \$5679.67 dólares debido a la existencia de centros de entretenimiento y establecimientos turísticos de la ciudad. Además, los diagramas de caja indican que el clúster 1 tiene un egreso promedio anual de \$5710.32, mientras que el clúster 2 con \$5655.02 dólares, especificando que el grupo 1 de los asociados presenta mayores egresos según su ubicación geográfica.

Análisis de Conglomerados de los egresos de los socios de la CACECH del periodo 2017

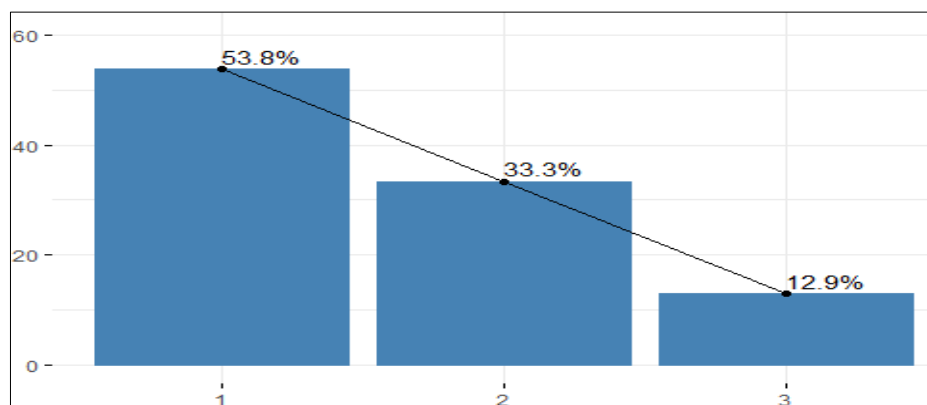


Gráfico 33-3: PCA de los egresos de los socios de la CACECH del año 2017

Realizado por: López, Cristina; Montero, Magdalena. 2022

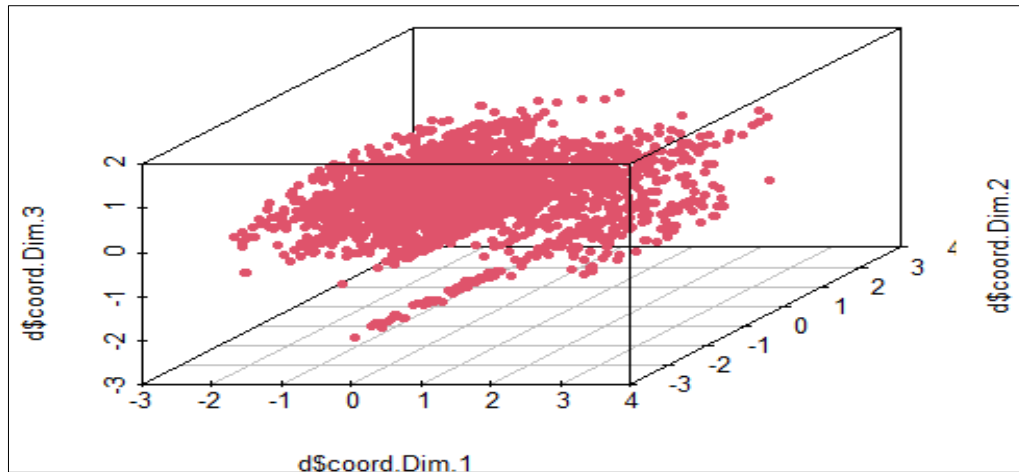


Gráfico 34-3: Homogeneidad de los egresos de los socios de la CACECH del año 2017

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el método PCA presentado en el gráfico 33-3 se determina que las 2 primeras componentes explican un 87.13% de variabilidad total de los egresos de los socios de la CACECH en el periodo 2017, a la vez con un test de homogeneidad Chi-cuadrado de 9.96 proporcionando una distribución homogénea explicada en el gráfico 34-3.

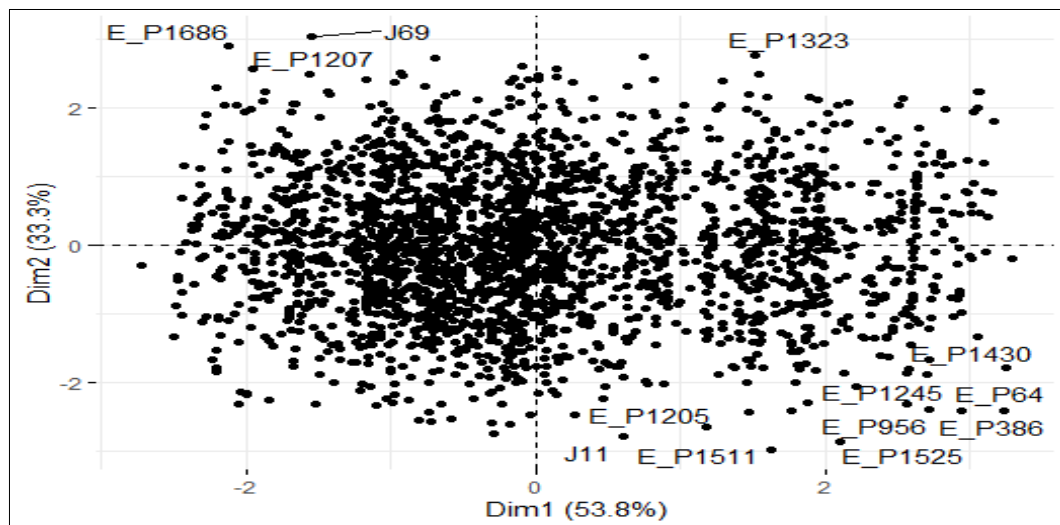


Gráfico 35-3: Biplot de los egresos de los socios de la CACECH del año 2017

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el gráfico 35-3 se observa el aglutinamiento de la información en los cuatro cuadrantes, que indica la existencia de asociación mediante clúster. Además, el coeficiente estadístico de Hopkins con un índice de 0.18 comprobando que la base de datos es clusterizable.

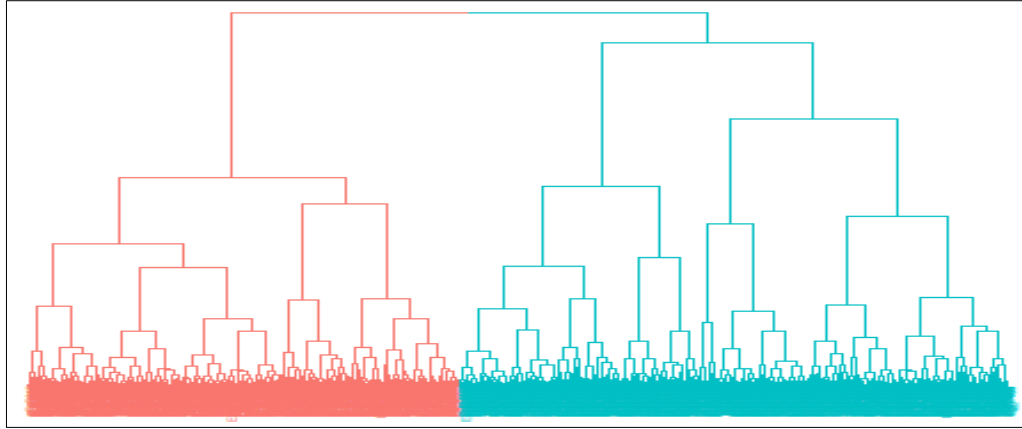


Gráfico 36-3: Dendrograma de los egresos de los socios de la CACECH del año 2017

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 36-3 de los egresos de los socios de la CACECH del año 2017 explica la distribución en 2 clústers con una mayor concentración en el segundo conglomerado según la actividad económica, mostrando similitudes entre la actividad laboral y la ubicación geográfica con cortes simétricos entre las dos agrupaciones siendo la razón por la que se intersecan varios puntos.

Tabla 22-3: Clúster de los egresos de los socios de la CACECH del año 2017

Actividad Económica	CLÚSTER	
	1	2
Docente Universitario	26	28
Empleado Privado	728	931
Empleado Público	263	314
Jubilado	77	116
Pensión Alimenticia	3	3

Realizado por: López, Cristina; Montero, Magdalena. 2022

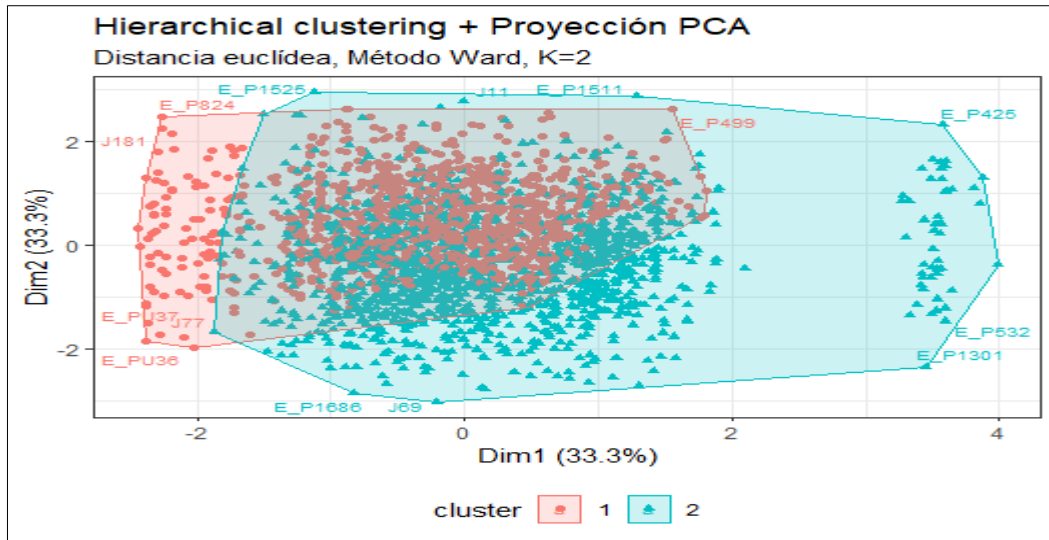


Gráfico 37-3: Clúster de los egresos de los socios de la CACECH del año 2017

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el gráfico 37-3 que muestra los clústers de los egresos de los socios de la CACECH para el año 2017 con un coeficiente de aglomeración del 0.60 donde se identifica 2 clústers explicados en la tabla 20-3 distribuidos según la actividad económica, con una mayor concentración en el segundo cluster con 1387 asociados, estas agrupaciones son generadas por la ubicaciones geográficas y transacciones bancarias de los accionistas ubicando a los que presentan mayores retiros en la parte derecha, los de menor en la izquierda y en el centro los que muestran egresos similares y que se encuentran dentro del mismo perímetro zonal.

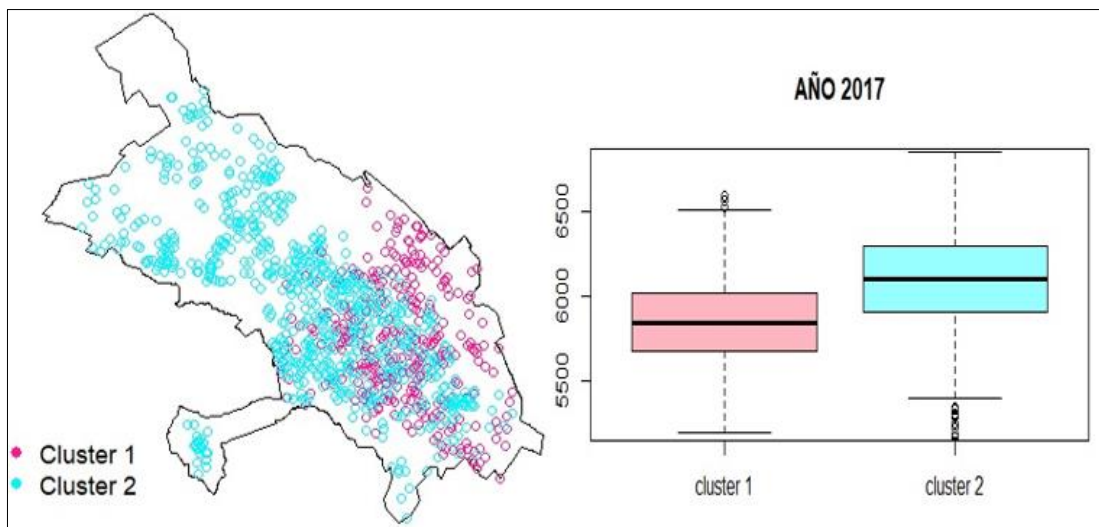


Gráfico 38-3: Lugar geográfico de los egresos de los socios de la CACECH del año 2017

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 38-3 muestra la ubicación geográfica de los socios mediante el mapa de clasificación de clúster enfocado en los egresos del año 2017, se detecta 2 grupos que se forman por similitudes territoriales y económicas a partir de la distancia (euclídea), mostrando mayor concentración en las parroquias Lizarzaburu y Velasco con un egreso promedio anual de \$5830.83 a \$6116.45 dólares debido a la existencia de centros de entretenimiento y establecimientos turísticos de la ciudad. Además, los boxplots indica que el clúster 1 tiene un egreso promedio anual de \$5840.56, mientras que el clúster 2 de \$6105.52 dólares definiendo que el grupo 2 de los asociados presenta mayores egresos según su ubicación geográfica.

Análisis de Conglomerados de los egresos de los socios de la CACECH del periodo 2018

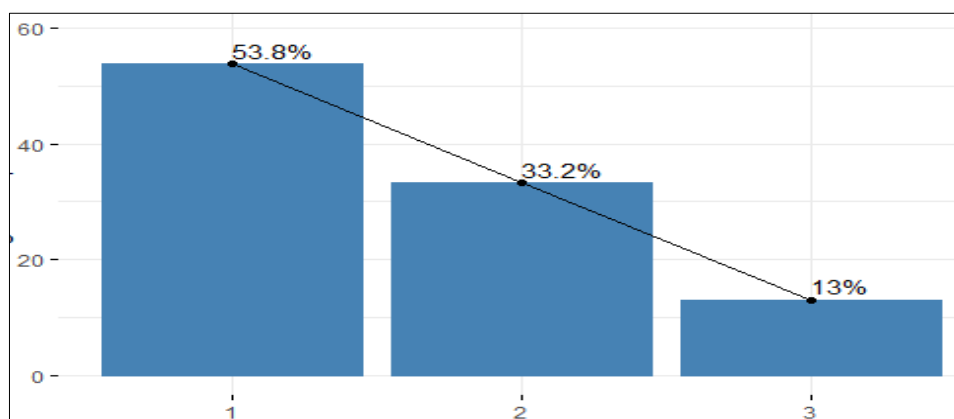


Gráfico 39-3: PCA de los egresos de los socios de la CACECH del año 2018

Realizado por: López, Cristina; Montero, Magdalena. 2022

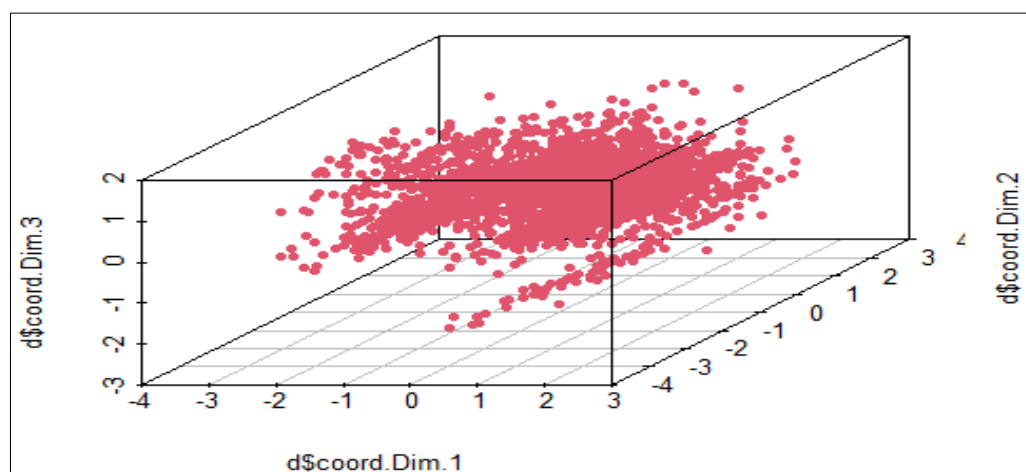


Gráfico 40-3: Homogeneidad de los egresos de los socios de la CACECH del año 2018

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el método PCA presentado en el gráfico 39-3 se determina que las 2 primeras componentes explican un 86.97% de variabilidad total de los egresos de los socios de la CACECH en el periodo 2018, a la vez con un test de homogeneidad Chi-cuadrado de 1.16 proporcionando una distribución homogénea explicada en el gráfico 40-3.

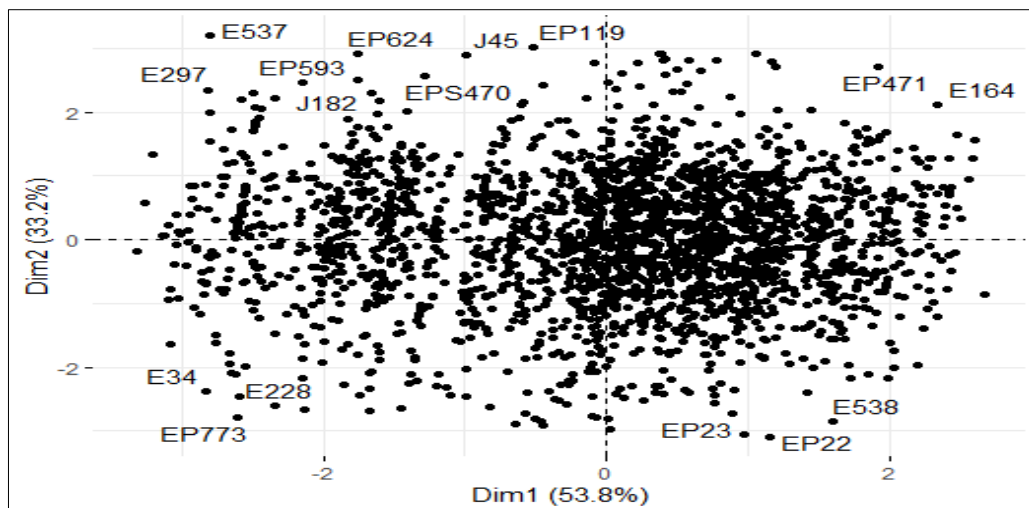


Gráfico 41-3: Biplot de los egresos de los socios de la CACECH del año 2018

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el gráfico 41-3 se observa el aglutinamiento de la información en los cuatro cuadrantes, que indica la existencia de asociación mediante clúster. Además, el coeficiente estadístico de Hopkins con un índice de 0.12 que evidencia que la base de datos es clusterizable.

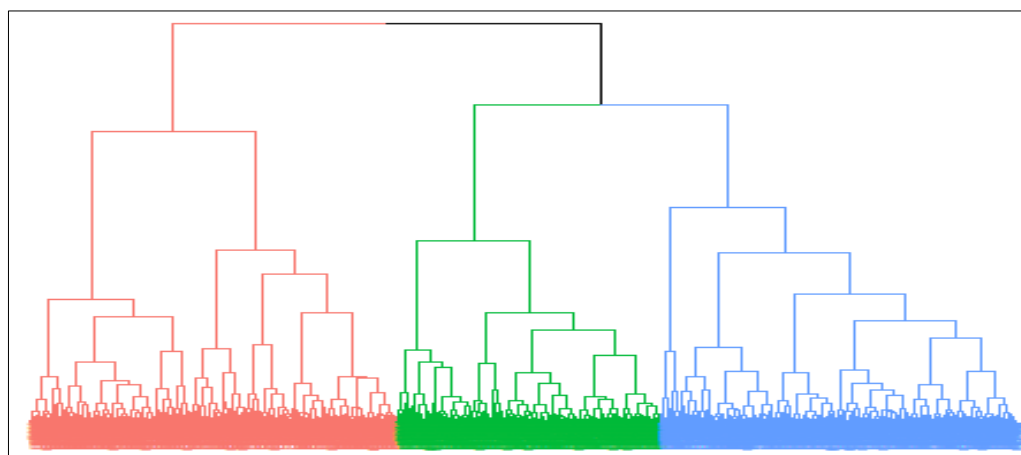


Gráfico 42-3: Dendrograma de los egresos de los socios de la CACECH del año 2018

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 42-3 de los egresos de los socios de la CACECH del año 2018 explica la distribución en 3 clústers con una mayor concentración en el tercer conglomerado según la actividad económica, mostrando similitudes entre la actividad laboral y la ubicación geográfica con cortes simétricos entre las tres agrupaciones siendo la razón por la que se intersecan varios puntos.

Tabla 23-3: Clúster de los egresos de los socios de la CACECH del año 2018

Actividad Económica	CLÚSTER		
	1	2	3
Docente Universitario	19	14	20
Empleado Privado	591	428	580
Empleado Público	186	146	209
Jubilado	59	49	76
Pensión Alimenticia	4	1	1

Realizado por: López, Cristina; Montero, Magdalena. 2022

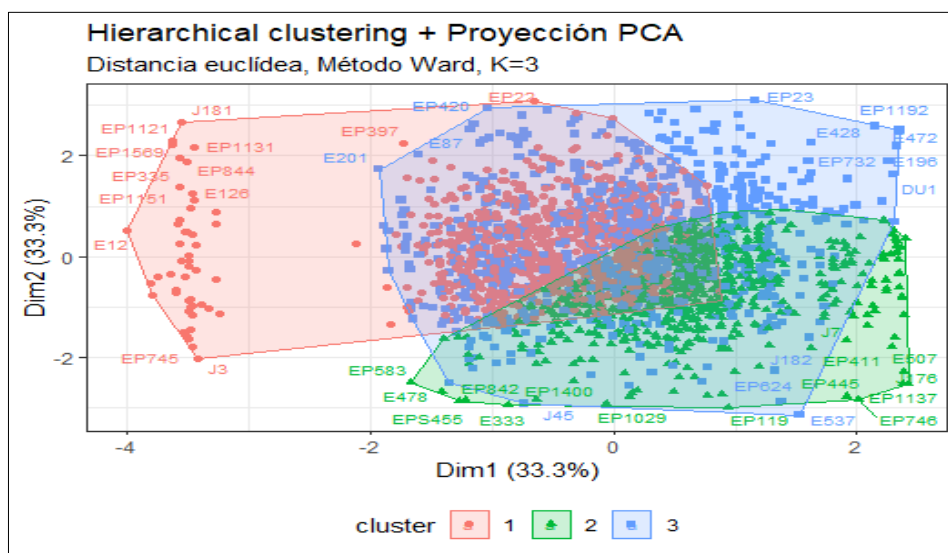


Gráfico 43-3: Clúster de los egresos de los socios de la CACECH del año 2018

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el gráfico 43-3 que muestran los clústers de los egresos de los socios de la CACECH para el año 2018 con un coeficiente de aglomeración del 0.57 donde se identifica 3 clústers explicados en la tabla 21-3 distribuidos según la actividad económica, con una mayor concentración en el tercer cluster con 886 asociados, estas agrupaciones son generadas por la ubicaciones geográficas y transacciones bancarias de los accionistas ubicando a los que presentan mayores retiros en la parte derecha, los de menor en la izquierda y en el centro los que muestran egresos similares y que se encuentran dentro del mismo perímetro zonal.

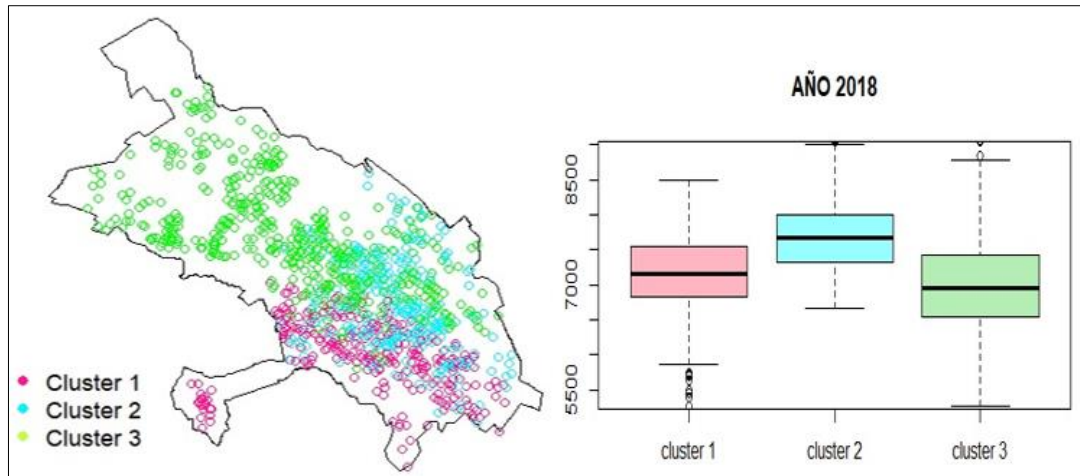


Gráfico 44-3: Lugar geográfico de los egresos de los socios de la CACECH del año 2018

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 44-3 muestra la ubicación geográfica de los socios mediante el mapa de clasificación de clúster enfocado en los egresos del año 2018, se detecta 3 grupos que se forman por similitudes territoriales y económicas a partir de la distancia (euclídea), mostrando mayor concentración en las parroquias Lizarzaburu y Velasco con un egreso promedio anual de \$6831.78 a \$7698.58 dólares debido a la existencia de centros de entretenimiento y establecimientos turísticos de la ciudad. Además, los boxplots indica que el clúster 1 tiene un egreso promedio anual de \$7142.63, el clúster 2 con \$7685.03 y el clúster 3 de \$6916.76 dólares, definiendo que el grupo 2 de los asociados presenta mayores egresos según su ubicación geográfica.

Análisis de Conglomerados de los egresos de los socios de la CACECH del periodo 2019

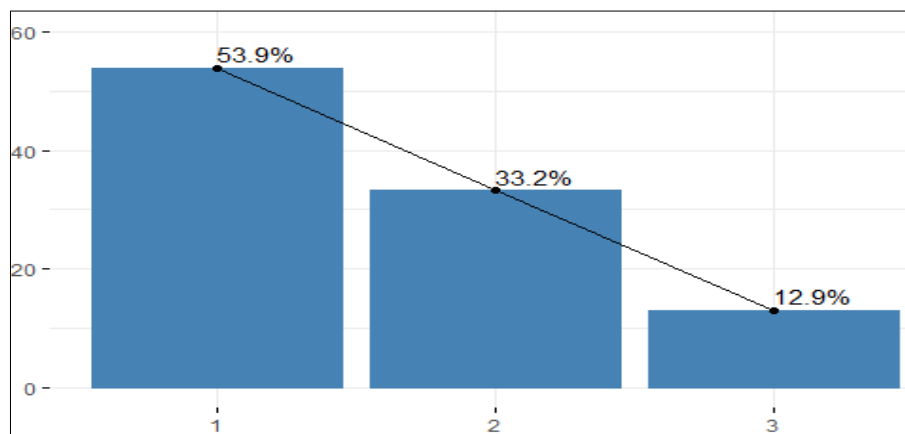


Gráfico 45-3: PCA de los egresos de los socios de la CACECH del año 2019

Realizado por: López, Cristina; Montero, Magdalena. 2022

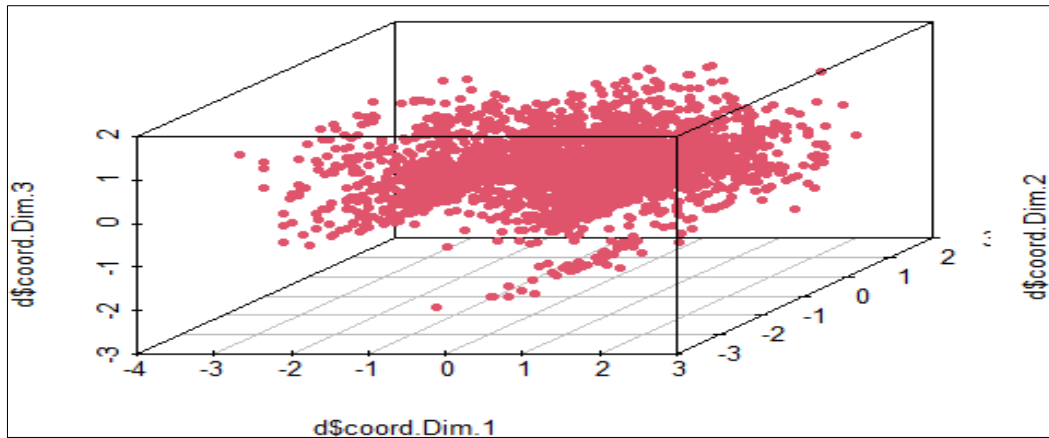


Gráfico 46-3: Homogeneidad de los egresos de los socios de la CACECH del año 2019

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el método PCA presentado en el gráfico 45-3 se determina que las 2 primeras componentes explican un 87.10% de variabilidad total de los egresos de los socios de la CACECH en el año 2019, a la vez con un test de homogeneidad Chi-cuadrado de 2.83 proporcionando una distribución homogénea explicada en el gráfico 46-3.

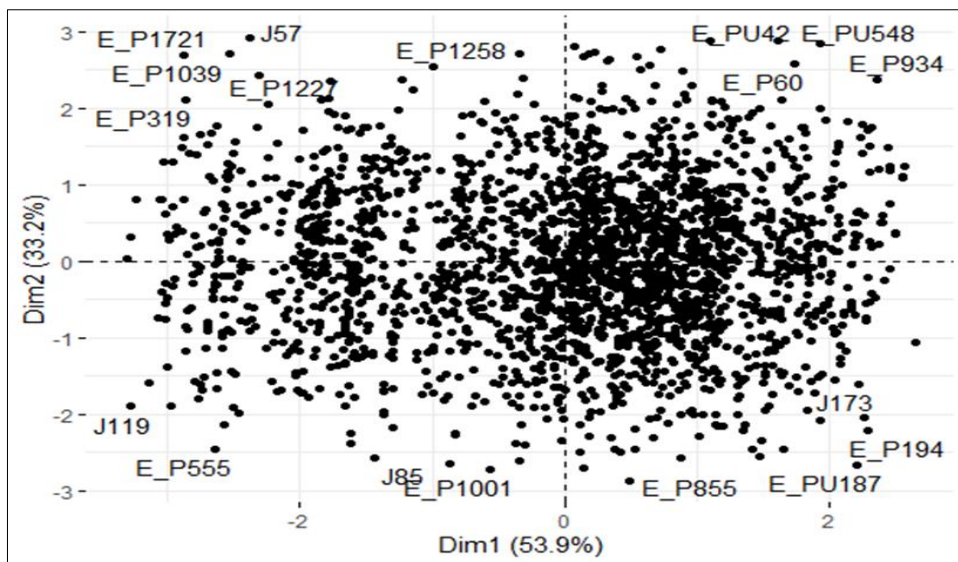


Gráfico 47-3: Biplot de los egresos de los socios de la CACECH del año 2019

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el gráfico 47-3 se observa el aglutinamiento de la información en los cuatro cuadrantes, que indica la existencia de asociación mediante clúster. Además, el coeficiente estadístico de Hopkins con un índice de 0.14 comprueba que la base de datos es clusterizable.

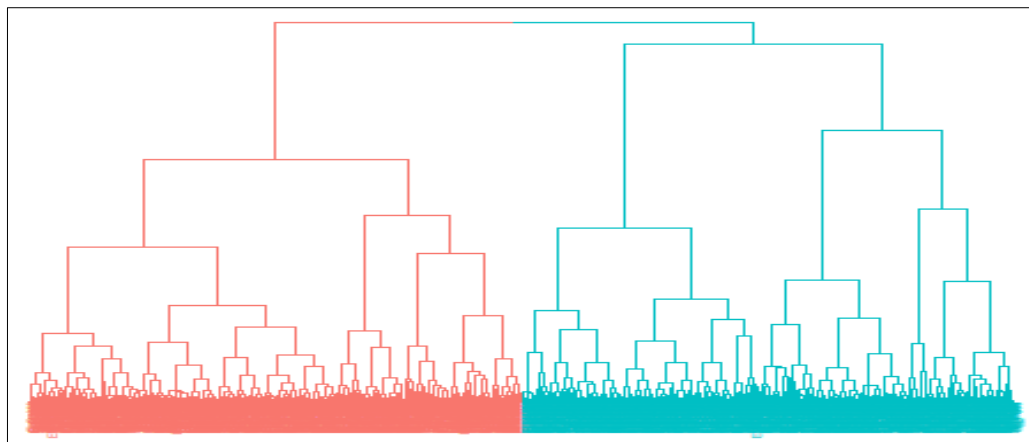


Gráfico 48-3: Dendrograma de los egresos de los socios de la CACECH del año 2019

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 48-3 de los egresos de los socios de la CACECH del año 2019 explica la distribución en 2 clústers con una mayor concentración en el primer conglomerado según la actividad económica, mostrando similitudes entre la actividad laboral y la ubicación geográfica con cortes simétricos entre las tres agrupaciones siendo la razón por la que se intersecan varios puntos.

Tabla 24-3: Clúster de los egresos de los socios de la CACECH del año 2019

Actividad Económica	CLÚSTER	
	1	2
Docente Universitario	30	23
Empleado Privado	829	820
Empleado Público	288	285
Jubilado	87	106
Pensión Alimenticia	5	1

Realizado por: López, Cristina; Montero, Magdalena. 2022

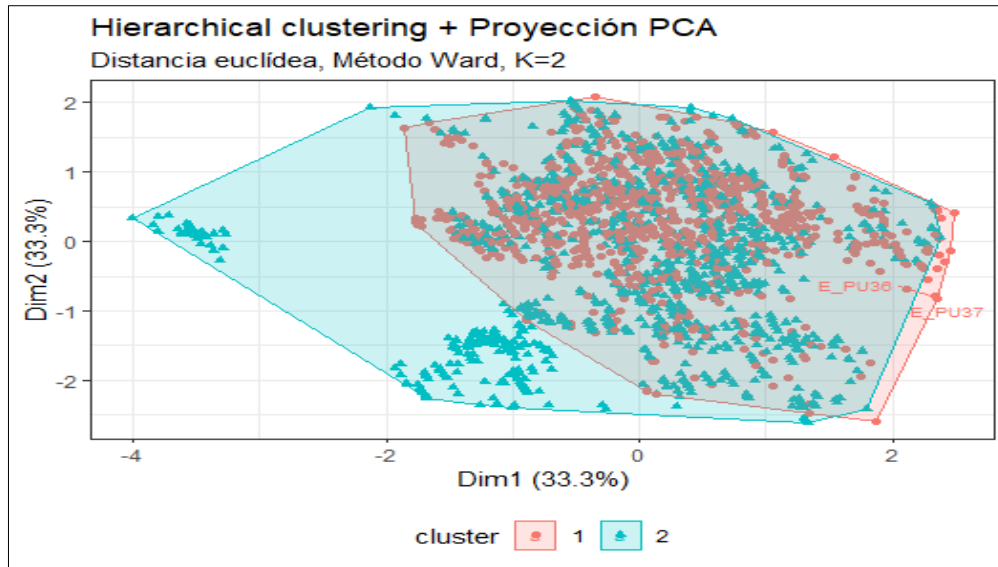


Gráfico 49-3: Clúster de los egresos de los socios de la CACECH del año 2019

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el gráfico 49-3 que muestran los clústers de los egresos de los socios de la CACECH para el año 2019 con un coeficiente de aglomeración del 0.61 donde se identifica 2 clústers explicados en la tabla 22-3 distribuidos según la actividad económica, con una mayor concentración en el primer cluster con 1239 asociados, estas agrupaciones son generadas por la ubicaciones geográficas y transacciones bancarias de los accionistas ubicando a los que presentan mayores retiros en la parte derecha, los de menor en la izquierda y en el centro los que muestran egresos similares y que se encuentran dentro del mismo perímetro zonal.

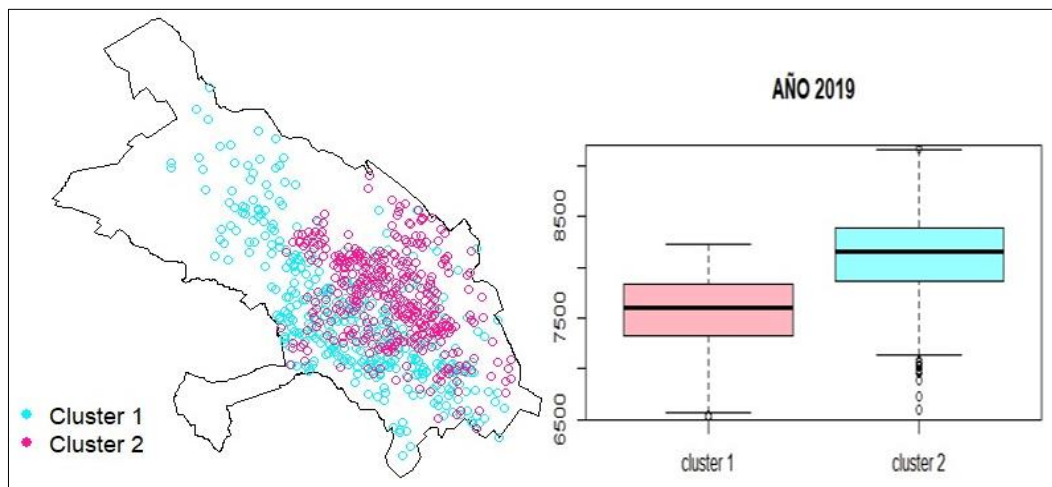


Gráfico 50-3: Lugar geográfico de los egresos de los socios de la CACECH del año 2019

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 50-3 muestra la ubicación geográfica de los socios mediante el mapa de clasificación de clúster enfocado en los egresos del año 2019, se detecta 2 grupos que se forman a partir de la distancia (euclídea), mostrando mayor concentración en el centro histórico de la ciudad, con egresos promedios de \$7571.02 a \$8153.36 dólares debido a la existencia de centros de entretenimiento y establecimientos turísticos. Además, los diagramas de caja indican que el clúster 1 tiene un egreso promedio anual de \$7576.90, mientras que el clúster 2 de \$8139.59 dólares, definiendo que el grupo 2 de los asociados presenta mayores egresos.

Análisis de Conglomerados de los egresos de los socios de la CACECH del periodo 2020

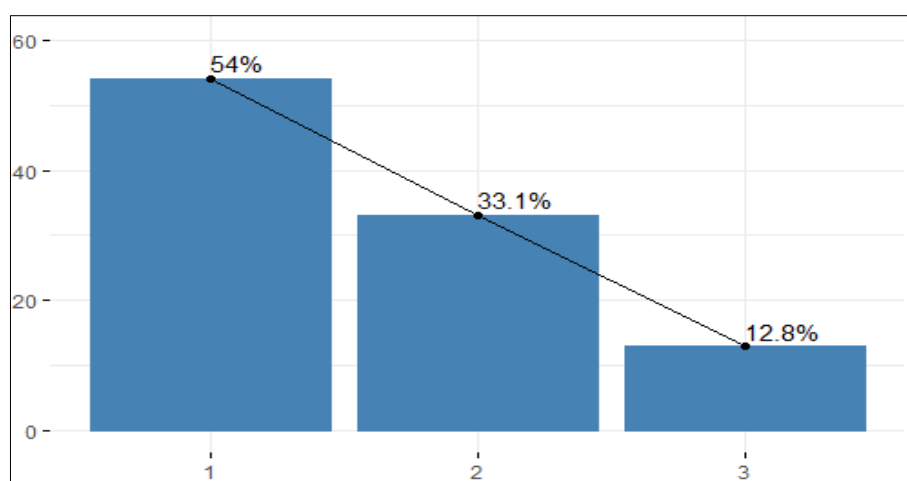


Gráfico 51-3: PCA de los egresos de los socios de la CACECH del año 2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

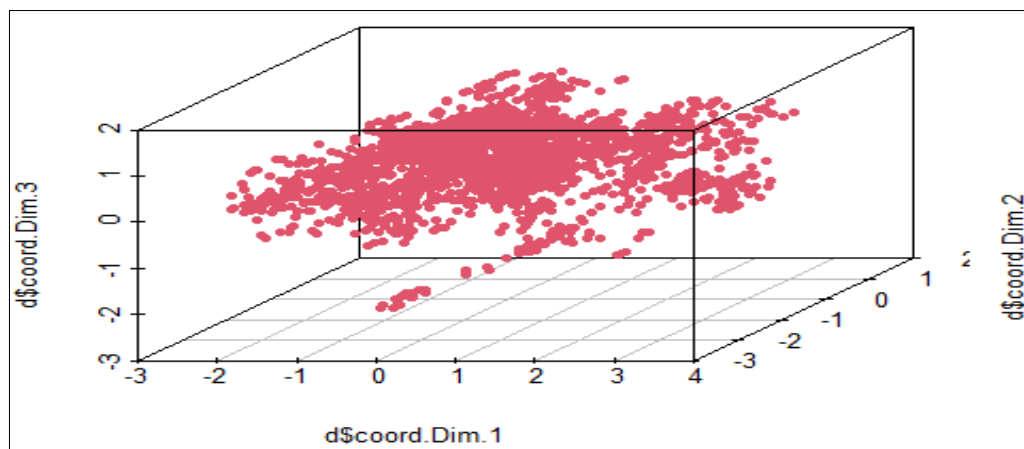


Gráfico 52-3: Homogeneidad de los egresos de los socios de la CACECH del año 2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el método PCA presentado en el gráfico 51-3 se determina que las 2 primeras componentes explican un 87.15% de variabilidad total de los egresos de los socios de la CACECH en el periodo 2020, a la vez con un test de homogeneidad Chi-cuadrado de 5.15 proporcionando una distribución homogénea explicada en el gráfico 52-3.

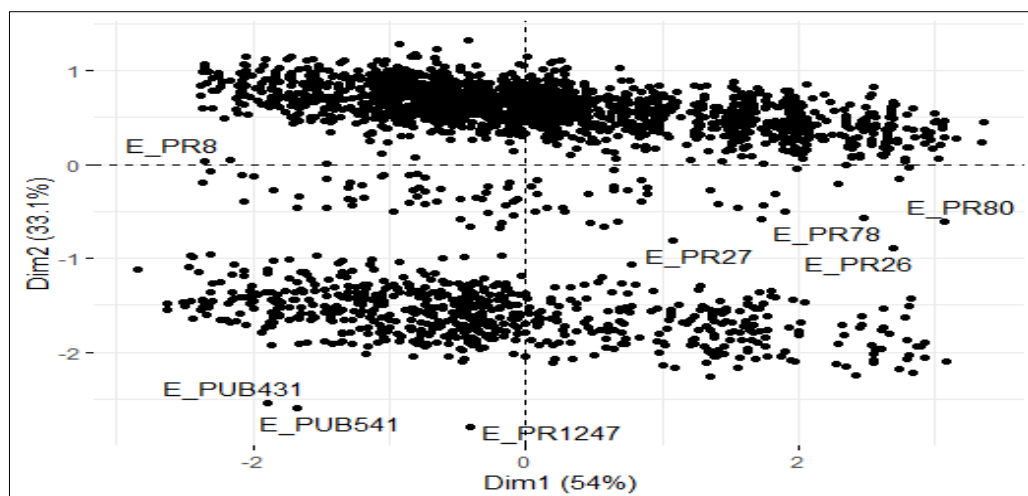


Gráfico 53-3: Biplot de los egresos de los socios de la CACECH del año 2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el gráfico 53-3 se observa el aglutinamiento de la información en los cuatro cuadrantes, que indica la existencia de asociación mediante clúster. Además, el coeficiente estadístico de Hopkins con un índice de 0.15 comprueba que la base de datos es clusterizable.

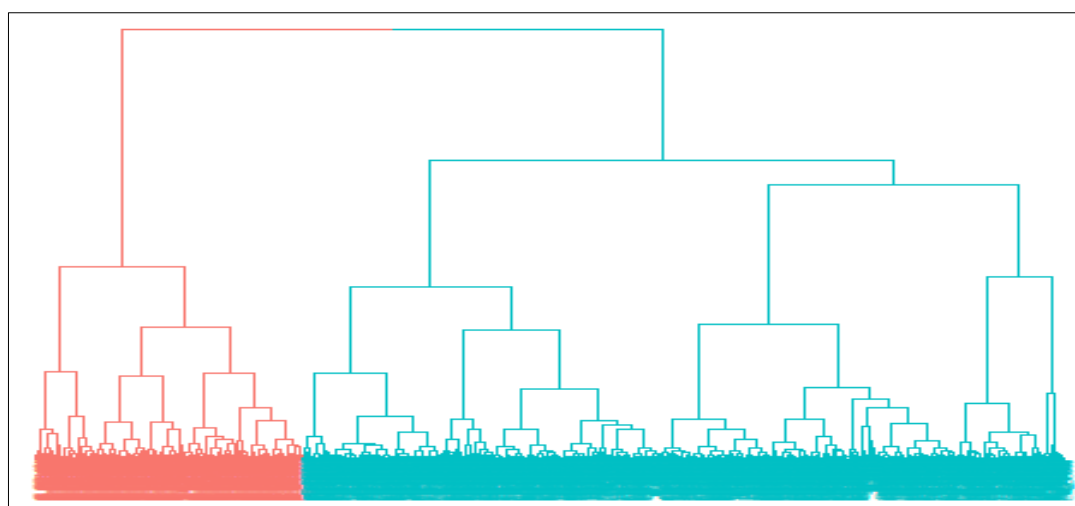


Gráfico 54-3: Dendrograma de los egresos de los socios de la CACECH del año 2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 54-3 de los egresos de los socios de la CACECH del año 2020 explica la distribución en 2 clústers con una mayor concentración en el primer conglomerado según la actividad económica, mostrando similitudes entre la actividad laboral y la ubicación geográfica con cortes simétricos entre las tres agrupaciones siendo la razón por la que se intersecan varios puntos.

Tabla 25-3: Clúster de los egresos de los socios de la CACECH del año 2020

Actividad Económica	CLÚSTER	
	1	2
Docente Universitario	26	28
Empleado Privado	1234	426
Empleado Público	408	170
Jubilado	175	19
Pensión Alimenticia	3	3

Realizado por: López, Cristina; Montero, Magdalena. 2022

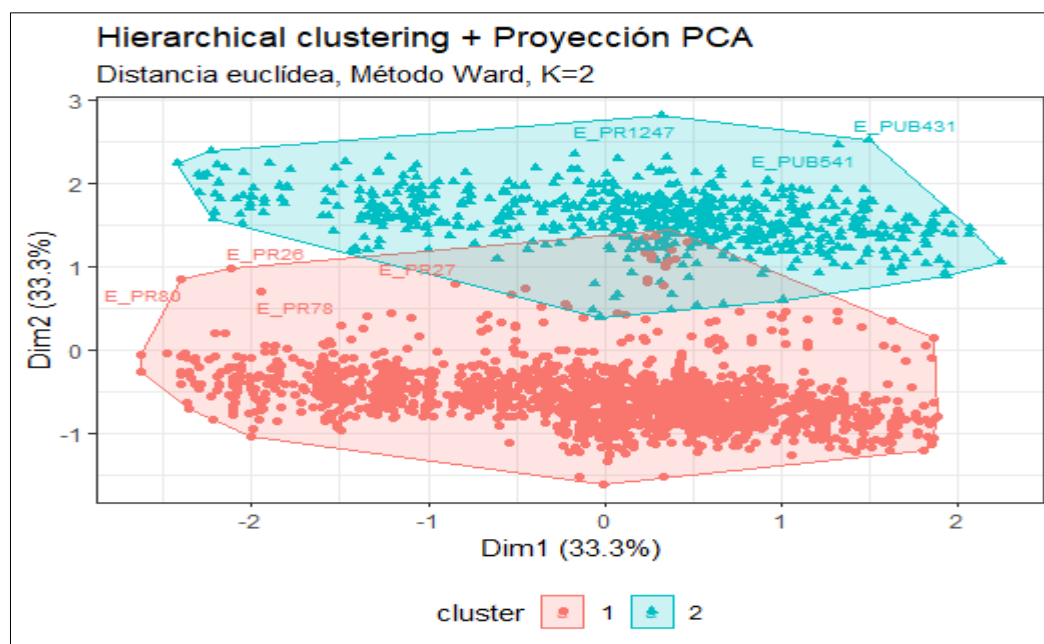


Gráfico 55-3: Clúster de los egresos de los socios de la CACECH del año 2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

Mediante el gráfico 55-3 que muestran los clústers de los egresos de los socios de la CACECH para el año 2020 con un coeficiente de aglomeración del 0.66 donde se identifica 2 clústers explicados en

la tabla 23-3 distribuidos según la actividad económica, con una mayor concentración en el primer cluster con 1846 asociados, estas agrupaciones son generadas por la ubicaciones geográficas y transacciones bancarias de los accionistas ubicando a los que presentan mayores retiros en la parte derecha, los de menor en la izquierda y en el centro los que muestran egresos similares y que se encuentran dentro del mismo perímetro zonal.

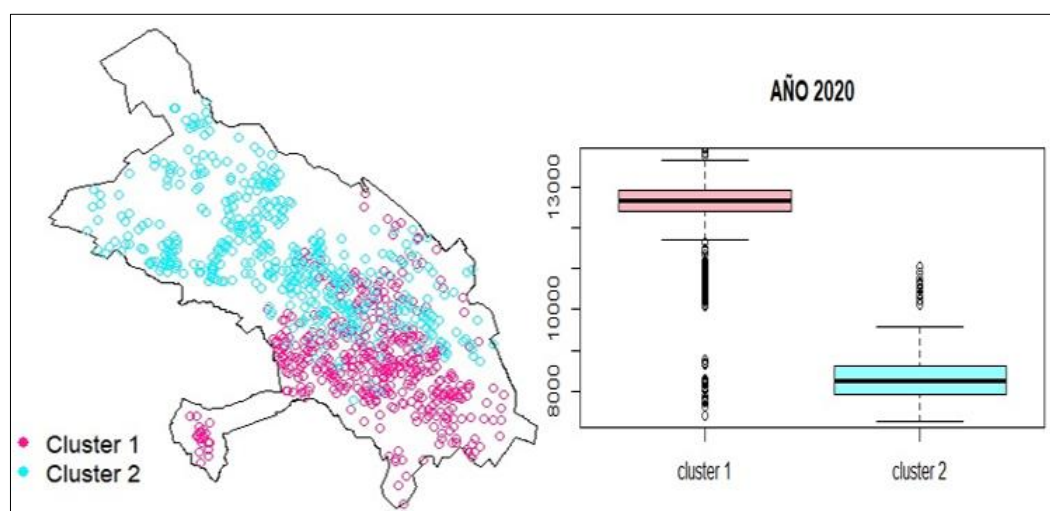


Gráfico 56-3: Lugar geográfico de los egresos de los socios de la CACECH del año 2020

Realizado por: López, Cristina; Montero, Magdalena. 2022

El gráfico 56-3 muestra la ubicación geográfica de los socios mediante el mapa de clasificación de clúster enfocado en los egresos del año 2020, se detecta 2 grupos, que se forman por similitudes territoriales y económicas a partir de la distancia (euclídea), mostrando mayor concentración en las parroquias Lizarzaburu y Velasco con un egreso promedio anual de \$8227.36 a \$12595.52 dólares, se cree que es debido a la existencia de centros de entretenimiento y establecimientos turísticos de la ciudad. Además, los boxplots indica que el clúster 1 tiene un egreso promedio anual de \$12559.83, mientras que el clúster 2 de \$8291.60 dólares, mostrando que el grupo 1 realiza mayores egresos.

Discusión de Resultados

La presente investigación tuvo como meta identificar los patrones de comportamiento de la economía de los socios de la Cooperativa de Ahorro y Crédito “Educadores de Chimborazo” mediante el uso y aplicación de metodologías espaciales enfocadas en la clasificación de conglomerados (clúster), abordando desde un enfoque descriptivo de las variables cualitativas se encontró que el cantón Riobamba contiene al 77% de los asociados de la institución financiera distribuidas en las diferentes parroquias con la mayor localización en la congregación Velasco con un 44.28%, con una concentración del 96 % en la zona urbana y el 4 % en la rural, con un 66.4% de mujeres a diferencia del 33.6% de hombres; la economía de los accionistas se ve reflejada a través de instituciones privadas teniendo una presencia 62.84% como la actividad laboral más frecuente entre socios.

El enfoque exploratorio de las variables cuantitativas presento un ingreso constante de los socios durante los 5 años, caso contrario ocurrido en los egresos ya que mostró un cambio abrupto en el último año porque genero gastos en insumos alimenticios, médicos y de vivienda. El comportamiento de los ingresos a través de los 12 meses del año presenta un valor constante durante el periodo (2016-2019) mostrando un ligero cambio en el año 2020 por la crisis sanitaria provocada por el covid-19, mientras que los egresos muestran alteraciones en el primer cuatrimestre del año 2019 por el libre comercio en la frontera Ecuador-Colombia y en el año 2020 por el confinamiento generando mayores gastos en los hogares de los accionistas.

De acuerdo a las direcciones domiciliarias facilitadas por miembros de la institución financiera se obtuvo las coordenadas geográficas de los socios en escalas GPS y UTM, abordando desde un punto inferencial espacial se halló el índice de Morán con el que se indica que el ingreso del último año con respecto a la localidad de los socios tiene una correlación de -0.73%, determinando una autocorrelación negativa perfecta de -1.83 hasta una positiva perfecta de 1.10 demostrando dependencia espacial en los datos permitiendo un análisis de conglomerados desde un enfoque espacial (cluster).

A través del análisis de clúster de los ingresos del último año se clasifican en dos grupos según la actividad económica donde la mayor agrupación se encuentra en el cluster 1 con un total de 2300 asociados con similitudes en varios puntos por la afinidad geográfica y económica, mientras que los egresos para los periodos (2016,2017,2019,2020) presenta 2 clústers obteniendo una mayor concentración el primer, tercero y el último año en el conglomerado 1 con 1529 ,1239 y 1846 respectivamente, en el segundo año el agrupamiento con mayor presencia es el conglomerado 2 con 1384 accionistas , esto se comprobó mediante la validación de clústers con los métodos gráficos de dendrograma, cubo de homogeneidad y diagrama de dispersión los cuales indican la asignación

óptima de clúster; además se aplicó el estadístico Hopkins el cual determinó que la información es clusterizable corroborando que los análisis expuestos se realizaron de manera correcta. De la misma manera para el año 2018 se generó un proceso similar obteniendo que el número óptimo de clúster son 3 el cual tiene la mayor concentración en la tercera agrupación con 886 asociados; de la misma se comprobó mediante los métodos gráficos y el test de Hopkins que la información es clusterizable. La mayoría de agrupaciones en los clústers generados de los egresos en el periodo 2016- 2020 presentan similitudes por semejanzas geográficas y económicas presentándose como puntos de intersección en las gráficas. Obteniendo de esta manera que la economía de la institución financiera se ha mantenido en equilibrio a pesar de los cambios económicos y sociales del país, observando el mayor movimiento de egresos promedios anuales en las parroquias Lizarzaburu y Velasco en todo el periodo de estudio.

CONCLUSIONES

- Se identificó la ubicación geográfica de los socios mediante Google maps, y se obtuvo coordenadas geográficas (GPS), que fueron transformadas a coordenadas métricas (UTM-Z17-SUR) para procesos de análisis en la investigación, el mismo se llevó a cabo con las direcciones de cada socio de la Cooperativa de Ahorro y Crédito “Educadores de Chimborazo”. Ltda., cabe mencionar que esta variable fue proporcionada bajo juramento de confidencialidad.
- El análisis exploratorio demográfico mostró que el 77% de los socios de la cooperativa se encuentran ubicados en el cantón Riobamba, en las diferentes parroquias con el 96% en la zona urbana y el 4% en la rural; predominando el sexo femenino con un 66.40% a diferencia del masculino con un 33.60% del total de asociados, además con actividades económicas variadas siendo la predominante el servidor privado con un 62.84% que representa a docentes con un ingreso mensual fijo manteniendo una economía estable en la entidad.
- A través del análisis exploratorio temporal de los egresos de los socios, se identificó un aumento abrupto en los egresos de los accionistas a comparación de los 4 años anteriores, con una tasa de crecimiento de 22.46% del 2018 al 2019 y del 39.15% del 2019 al 2020. Este fenómeno fue provocado por el confinamiento declarado por el gobierno ante la presencia de COVID-19 en el país.
- Al determinar que el 77% de la población se encuentra reunida en el cantón Riobamba se toma únicamente a esta localidad para el análisis espacial, el cual arroja un índice de Moran de los ingresos y egresos de los socios de la CACECH con autocorrelación negativa perfecta; dando paso a realizar un estudio de clúster.
- El análisis de clúster jerárquico aglomerativo permitió apreciar que los ingresos de los socios de la Cooperativa de Ahorro y Crédito “Educadores de Chimborazo durante los años 2016-2020 no muestran dependencia espacial, aun así, permitió formar grupos de acuerdo con las similitudes de ingresos promedios anuales con una clara diferencia entre conglomerados con ingresos más bajos en el grupo 1 a contraste con el grupo 2 con mayores entradas. Por otro lado los egresos presentan dependencia espacial, que otorgó establecer grupos de acuerdo con sus similitudes en función de la ubicación geográfica (distancias euclídeas) y movimientos financieros; comprobando mediante mapas geográficos que las parroquias con mayores egresos promedios anuales son Lizarzaburu y Velasco debido a los centros de entretenimiento y establecimientos turísticos de la ciudad.

RECOMENDACIONES

- La información obtenida por las entidades financieras debe ser manejada con cautela, para evitar cualquier tipo de error humano, tecnológico y técnico, reduciendo de esta manera tiempo y recursos de la cooperativa, además de proporcionar información de calidad para los futuros estudios.
- Para la precisión de resultados es recomendable una depuración de la información y un análisis exploratorio previo determinando de esta manera el comportamiento de los datos, y así seleccionar las técnicas estadísticas adecuadas para cumplir los objetivos del estudio.
- Es de vital importancia para la creación de mapas conocer el tipo de información geométrica (shapefile) que se debe usar según las necesidades del investigador, facilitando de esta manera el manejo de polígonos, líneas y puntos espaciales en el software R que permite optimizar el análisis espacial.
- Es primordial para los futuros profesionales de la carrera de Estadística, así como para sus catedráticos, investigar sobre técnicas espaciales, sobre todo, el manejo de conglomerados (clústers), fundamentando el análisis univariado y bivariado para una mejor comprensión e interpretación de la investigación.
- Se requiere conocer los tipos de metodologías existentes para el análisis de clústeres espaciales y los requerimientos necesarios para su aplicación, para evitar la presencia de ruido y outliers que pueden distorsionar los resultados del estudio, además de realizar una comprobación a través de mapas que queda a consideración del investigador.

GLOSARIO

Aleatorio: Pertenciente o relativo al juego de azar. (ASALE & RAE, 2022a: p.1)

Análisis de clúster: Es una técnica cuya idea básica es agrupar un conjunto de observaciones en un número dado de clústeres o grupos. Este agrupamiento se basa en la idea de distancia o similitud entre las observaciones. (Marin, 2018, p.1)

Análisis de componentes principales: Es un método estadístico que permite simplificar la complejidad de espacios muestrales con muchas dimensiones a la vez que conserva su información. Supóngase que existe una muestra con n individuos cada uno con p variables (X_1, X_2, \dots, X_p) , es decir, el espacio muestral tiene p dimensiones. (Amat, 2017, p.1)

Autocorrelación espacial: Mide el grado de asociación que una variable desarrolla a través de un espacio definido como marco geográfico. (Siabato et al., 2019: p.2)

Economía: Ciencia que estudia los métodos más eficaces para satisfacer las necesidades humanas materiales, mediante el empleo de bienes escasos. (ASALE & RAE, 2022b: p.5)

Estadística espacial: Se refiere a un conjunto de técnicas estadísticas que cuantifican aspectos relacionados con la estructura de las distribuciones espaciales. (Rodríguez et al., 2008: p.3)

Estadística multivariante: Se refiere a diferentes métodos que estudian y examinan el efecto simultáneo de múltiples variables. Los métodos estadísticos multivariados se utilizan para analizar el comportamiento conjunto de más de una variable aleatoria. (García, 2021, p.1)

Geoestadística: es la rama de la estadística especializada en el análisis y la modelización de la variabilidad espacial de fenómenos que tienen una componente espacial, es decir, de los cuales pueden obtenerse datos referidos a localizaciones específicas mediante coordenadas. (Geoinnova, 2019, p.1)

Medoid: Es el objeto de un grupo cuya disimilaridad media a todos los objetos en el grupo es mínima. Es el punto ubicado más hacia el centro en todo el grupo. (Rousseeuw, 2019, p.1)

Serie de tiempo: Es una secuencia de datos u observaciones, medidos en determinados momentos y ordenados cronológicamente. Visualmente, es una curva que evoluciona en el tiempo. (Montenegro, 2018, p.1)

BIBLOGRAFÍA

ACEMOGLU, Daron; et al. *Economía* [en línea]. España: Pearson Education, 2017. [Consulta: 25 de Julio 2022]. Disponible en: <https://elibro.net/en/ereader/epoch/129484>

AMAT, J. *RPubs—Clustering y heatmaps: Aprendizaje no supervisado con R* [blog]. [Consulta: 27 de Julio 2022]. Disponible en: https://rpubs.com/Joaquin_AR/310338

ARITA, Brian; et al. *RPubs—Análisis de Clúster* [blog]. [Consulta: 27 de Julio 2022]. Disponible en: https://rpubs.com/Eunice_Ramirez/analisis_cluster

ARRETX, C; et al. *Demografía histórica en América Latina: Fuentes y métodos*. 4ª ed. San José, Costa Rica: CELADE, 1983, pp. 59-65.

ASALE, R. *Aleatorio, aleatoria | Diccionario de la lengua española*. Edición del Tricentenario. [Consulta: 27 de Julio 2022]. Disponible en: <https://dle.rae.es/aleatorio>

ASALE, R. *Economía | Diccionario de la lengua española*. Edición del Tricentenario. [Consulta: 27 de Julio 2022]. Disponible en: <https://dle.rae.es/economía>

AULETTA, Cintia; et al. *Geografía* [en línea]. Ciudad Autónoma de Buenos Aires. Eudeba, 2020. [Consulta: 27 de Julio 2022]. Disponible en: <https://elibro.net/en/ereader/epoch/153598>

BARÓN-RIVERA, Juan. "Las regiones económicas de Colombia: Un análisis de clusters". *Documentos de Trabajo Sobre Economía Regional y Urbana*, n° 23 (2002), (Colombia) pp. 83-118.

BENITEZ, Lenny; et al. "El clúster: Una alternativa para la competitividad de las pymes de banano orgánico en Ecuador". *Conference Proceedings* [en línea], 2017, (Ecuador) 1(1), pp. 682-694. [Consulta: 20 julio 2022]. ISSN 2588-056X. Disponible en: <https://investigacion.utmachala.edu.ec/proceedings/index.php/utmach/article/view/214/185>

BOHÓRQUEZ, Ingrid. A., & CEBALLOS, Ermilson. "Algunos conceptos de la econometría espacial y el análisis exploratorio de datos espaciales". *Ecos de Economía* [en línea], 2008, (Colombia) 12(27), pp. 9-34. [Consulta: 20 julio 2022]. ISSN 2462-8107. Disponible en: <https://publicaciones.eafit.edu.co/index.php/ecos-economia/article/view/705/627>

BOTELLA, P; et al. *Instalación e introducción al software estadístico R y la librería R-Commander. Estadística descriptiva* [en línea]. Madrid-España: Universidad Cardenal Herrera, 2018. [Consulta: 20 de Julio 2022]. Disponible en: <https://www.uv.es/~mamtnz/IRCED.pdf>

BUHELLI, G. "Uso del Índice de Moran y LISA para explicar el ausentismo electoral rural en Ecuador". *Revista Geográfica* [en línea], 2019, (Ecuador), pp. 91-108. [Consulta: 27 de Julio 2022]. ISSN 2663-399X. Disponible en: <https://doi.org/10.35424/regeo.160.2019.746>

CAYUELA, L. *Análisis multivariante* [en línea]. Madrid-España: 2011. [Consulta: 27 de Julio 2022]. Disponible en: <https://botanicaargentina.org.ar/wp-content/uploads/2017/10/6-Analisis-multivariante.pdf>

CISNEROS MARTÍNEZ, M; et al. " Eficiencia colectiva de mipymes productoras de calzado ecuatorianas mediante el desarrollo del cluster". *Polo del Conocimiento* [en línea], 2018, (Ecuador) 3(1), pp. 300-313. [Consulta: 25 de Julio 2022]. ISSN 2550-682X. Disponible en: <https://polodelconocimiento.com/ojs/index.php/es/article/view/673>

GARCÍA, M. *Análisis multivariado: ¿Qué es y para qué sirve?* [blog]. México: 2 de Septiembre, 2021. [Consulta: 27 de Julio 2022]. Disponible en: <https://www.crehana.com/blog/desarrollo-web/analisis-multivariado/>

GEOINNOVA, A. *¿Qué es la geoestadística y cuáles son los principales análisis geoestadísticos?* [blog]. España, 11 de Julio, 2019. [Consulta: 27 de Julio 2022]. Disponible en: <https://geoinnova.org/blog-territorio/que-es-la-geoestadistica-analisisgeoestadisticos/>

GÓMEZ, Tristán. Identificación de Conglomerados Espaciales de Acuerdo a Niveles de Morosidad de Empresas en el Perú [en línea]. (Trabajo de Titulación) (Maestría). Pontificia Universidad Católica del Perú, Ciencias básicas y aplicadas, Escuela de Posgrado. (Lima-Perú). 2021. pp. 1-75. [Consulta: 2022-07-20]. Disponible en: https://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/20819/TRISTAN_GOMEZ_ALEX_IDENTIFICACION_CONGLOMERADOS_ESPACIALES.pdf?sequence=1&isAllowed=y

GONZÁLEZ, Julio; et al. "Conceptos básicos de geoestadística en geografía y ciencias de la tierra: Manejo y aplicación". *Geoenseñanza* [en línea], 2007, (Venezuela) 12(1), pp. 81-90. [Consulta: 20 julio 2022]. ISSN 1316-6077. Disponible en: <https://www.redalyc.org/pdf/360/36014577008.pdf>

HANS, W. *Función de covarianza Admisibilidad y Simplificaciones con estacionariedad* [wiki]. [Consulta: 27 de Julio 2022]. Disponible en: https://hmong.es/wiki/Covariance_function

HILBERT, D. *Matemática Aplicada 2* [en línea]. Argentina: 2019. [Consulta: 27 de Julio 2022]. Disponible en: <https://www.fceia.unr.edu.ar/~fismat2/apuntes/apun3-fismat2.pdf>

KASSAMBARA, A. *Extract and Visualize the Results of Multivariate Data Analyses* [blog]. [Consulta: 27 de Julio 2022]. Disponible en: <https://rpkg.s.datanovia.com/factoextra/index.html>

LEWIS, M. *Cómo calcular la tasa de crecimiento anual: 7 Pasos* [wiki]. [Consulta: 25 de Julio 2022]. Disponible en: <https://es.wikihow.com/calcular-la-tasa-de-crecimiento-anual>

LÓPEZ, J. *Tasa de crecimiento* [blog]. [Consulta: 25 de Julio 2022]. Disponible en: <https://economipedia.com/definiciones/tasa-de-crecimiento.html>

MARIN, J. *Estadística* [blog]. [Consulta: 27 de Julio 2022]. Disponible en: <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/AMult/AMult.html>

MARTÍNEZ, Gilberto., & CORRALES, Salvador. "Cadenas productivas y clusters en la economía regional de Nuevo León. Un análisis con matrices de insumo-producto". *Economía: teoría y práctica* [en línea], 2017, (México) 46, pp. 41-69. [Consulta: 20 julio 2022]. ISSN 2448-7481. Disponible en: https://www.scielo.org.mx/scielo.php?pid=S0188-33802017000100041&script=sci_arttext

MONTENEGRO, I. *Pricing: Series de Tiempo* [blog]. [Consulta: 27 de Julio 2022]. Disponible en: <https://www.pricing.cl/conocimiento/series-de-tiempo/>

PALLADINO, A. *Introducción a la demografía* [en línea]. Corrientes-Argentina: Universidad Nacional del Nordeste, 2010. [Consulta: 20 de Julio 2022]. Disponible en: <https://med.unne.edu.ar/sitio/multimedia/imagenes/ckfinder/files/files/aps/INTRODUCCION%20A%20LA%20DEMOGRAFIA%20APS.pdf>

PÉREZ, Jorge. "Econometría espacial y ciencia regional". *Investigación económica* [en línea], 2006, (México) 65(258), pp. 129-160. [Consulta: 20 de Julio 2022]. ISSN 0185-1667. Disponible en: https://www.scielo.org.mx/scielo.php?pid=S0185-16672006000400129&script=sci_arttext

RAMÍREZ, M. *Cómo entender contabilidad sin ser contador* [en línea]. México: Instituto Mexicano de Contadores Públicos, 2018. [Consulta: 25 de Julio 2022]. Disponible en: <https://elibro.net/en/ereader/epoch/116943>

RODRÍGUEZ, Milagros; et al. "Las técnicas de estadística espacial en la investigación salubrista. Caso síndrome de Down". *Revista Cubana de Salud Pública* [en línea], 2008, (Cuba), 34(4), pp. 1-11. [Consulta: 27 de Julio 2022]. ISSN 0864-3466. Disponible en: <https://www.redalyc.org/articulo.oa?id=21419854003>

RODRÍGUEZ, Liz., & CABRERA, José. "Análisis espacial de las dinámicas de crecimiento económico en México (1999-2009)". *Economía, Sociedad y Territorio* [en línea], 2017, (México) 17(55), pp. 709-741. [Consulta: 20 julio 2022]. ISSN 2448-6183. Disponible en: <https://doi.org/10.22136/est2017913>

ROUSSEUW, P. *K-medoides Algoritmos y Software* [blog]. [Consulta: 27 de Julio 2022]. Disponible en: <https://hmong.es/wiki/K-medoids>

SERRANO, Rosina., & VALCARCE, Esther. "Técnicas econométricas para el tratamiento de datos espaciales: La econometría espacial". *Investigaciones Regionales - Journal of Regional Research* [En línea], 2000, (España) (1), pp. 211-228 . [Consulta: 20 julio 2022]. ISSN 2340-2717. Disponible en: <https://dialnet.unirioja.es/servlet/libro?codigo=21033>

SIABATO, Willington; et al. "La autocorrelación espacial y el desarrollo de la geografía cuantitativa". *Cuadernos de Geografía: Revista Colombiana de Geografía* [en línea], 2019, (Colombia) 28(1), pp. 1-22. [Consulta: 27 de Julio 2022]. ISSN 0121-215X. Disponible en: <https://doi.org/10.15446/rcdg.v28n1.76919>

TULCANAZO, Tamia. La resiliencia económica en Ecuador en el contexto de la crisis sanitaria Covid-19 [en línea]. (Trabajo de Titulación) (Pregrado). Universidad Técnica del Norte, Facultad de Ciencias Administrativas y Económicas, Escuela de Economía. (Ibarra-Ecuador). 2022. pp. 1-83. [Consulta: 2022-07-20]. Disponible en: <http://repositorio.utn.edu.ec/handle/123456789/12521>

VALBUENA, Ana, & RODRÍGUEZ, Laura. "Análisis espacial en epidemiología: Revisión de métodos." *Revista de la Universidad Industrial de Santander* [en línea], 2018, (Colombia) 50(4), pp. 358-365. [Consulta: 20 julio 2022]. ISSN 2145-8464. Disponible en: <https://doi.org/10.18273/revsal.v50n4-2018009>

VIDAL, Rene., & FAVARO, Paolo. "Low rank subspace clustering (LRSC)". *Pattern Recognition Letters* [en línea], 2014, (Holanda) 43, pp. 47-61. [Consulta: 20 julio 2022]. ISSN 0167-8655. Disponible en: <https://doi.org/10.1016/j.patrec.2013.08.006>

VILALTA, Carlos. "¿Se pueden predecir geográficamente los resultados electorales? Una aplicación del análisis de clusters y outliers espaciales". *Estudios demográficos y urbanos* [en línea], 2008, (México) 23(3), pp. 571-613. [Consulta: 20 julio 2022]. ISSN 0186-7210. Disponible en: <https://doi.org/10.24201/edu.v23i3.1322>

YASELGA, E. *Clustering y descubrimiento de patrones. YouTube* [vídeo en línea]. 5 de Octubre 2020. [Consulta: 27 de Julio 2022]. Disponible en: https://www.youtube.com/playlist?list=PL2PpISw8vp_rUvqhZcbiMYcXPBQSHjPN_

ZAVALA, Mayumi. Segmentación demográfica y cultura de ahorro en la población del distrito de Huancayo-2018 [en línea]. (Trabajo de Titulación) (Licenciatura). Universidad Continental, Perú, Facultad de Ciencias de la Empresa, Escuela Académica Profesional de Contabilidad. (Huancayo-Perú). 2019. pp. 1-74. [Consulta: 20 de Julio 2022]. Disponible en: https://repositorio.continental.edu.pe/bitstream/20.500.12394/7118/3/IV_FCE_310_TE_Zavala_Lino_2019.pdf

ANEXOS

ANEXO A: CÓDIGO EN R DE CLÚSTER

```
Código R Cluster

##CLUSTER JERARQUICO AGLOMERATIVO
library(factoextra)
datos<-read_xlsx("C:/Users/Lenovo/Desktop/OCTAVOSEMESTRE/TITULACIÓNCURRICULAR/
CLUSTERING/DEF MAPAS/BE5.xlsx", sheet = "AE2020")

Datoss=datos[-1]
Dat<- scale(Datoss)
row.names(Dat)=datos$ACTIVIDAD_ECONÓMICA

pca2<- PCA (X = Dat, scale.unit = TRUE, graph = F)
pca2$eig
fviz_screeplot(pca2, addlabels = TRUE, ylim = c(0, 70))

fviz_pca_ind(
  pca2, col.var="contrib",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE,
  axes = c (1, 2))

scatterplot3d (x = d$coord. Dim.1,
  y = d$coord.Dim.2,
  z = d$coord.Dim.3,
  pch = 20, cex.lab = 0.8,
  grid = TRUE, box = TRUE,
  color = "2")
text(d[1:3], labels =(datos$ACTIVIDAD_ECONÓMICA),
  cex= 0.4, col = "steelblue")

aa=dist(dat, method = "euclidean")
a=hclust(aa,method = "ward.D2")
set.seed(101)

#DENDROGRAMA
fviz_dend(a, cex = 0.3, k = 2, horiz = FALSE, main="INGRESOS POR ACTIVIDAD ECONÓMICA DE LA
CACECH 2020")

#ASIGNACION PARA SABER A QUE CLUSTER PERTENECE
clust <- cutree(a, k = 2)

#CLUSTER
fviz_cluster(list(data = dat, cluster = clust),ellipse.type = "convex", repel = TRUE, show.clust.cent = FALSE,
  labelsize = 8)+
  labs(title = "Hierarchical clustering + Proyección PCA",
  subtitle = "Distancia euclídea, Método Ward, K=2") +
  theme_bw() +
  theme (legend.position = "bottom")

# Estadístico H
hopkins(data = datos1, n = nrow(datos1)-1)
```



epoch

Dirección de Bibliotecas y
Recursos del Aprendizaje

UNIDAD DE PROCESOS TÉCNICOS Y ANÁLISIS BIBLIOGRÁFICO Y
DOCUMENTAL

REVISIÓN DE NORMAS TÉCNICAS, RESUMEN Y BIBLIOGRAFÍA

Fecha de entrega: 28 / 10 / 2022

INFORMACIÓN DEL AUTOR/A (S)
Nombres – Apellidos: Magdalena Estefanía Montero Arias Cristina Elizabeth López Guilcapi
INFORMACIÓN INSTITUCIONAL
Facultad: Ciencias
Carrera: Estadística
Título a optar: Ingeniera en Estadística Informática
f. Analista de Biblioteca responsable: Ing. Leonardo Medina Ñuste MSc.



2034-DBRA-UTP-2022