



MODELOS LINEALES MIXTOS PARA ANALIZAR EL EFECTO DEL
TIEMPO EN LAS VENTAS POR TIENDAS DE UNA GRAN SUPERFICIE

MÓNICA VIVIANA CELY SALAZAR
INGRID LORENA GALEANO ARDILA

FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES
DEPARTAMENTO DE CIENCIAS BÁSICAS
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA
BOGOTÁ, D.C., 2019



MODELOS LINEALES MIXTOS PARA ANALIZAR EL EFECTO DEL
TIEMPO EN LAS VENTAS POR TIENDAS DE UNA GRAN SUPERFICIE

MÓNICA VIVIANA CELY SALAZAR
INGRID LORENA GALEANO ARDILA

Trabajo de grado para optar al título de Especialistas en Estadística
Aplicada

JUAN CARLOS RUBRICHE CÁRDENAS
Magister en Ciencias Matemáticas
Director de trabajo de grado

FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES
DEPARTAMENTO DE CIENCIAS BÁSICAS
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA
BOGOTÁ, D.C., 2019

Nota de Aceptación

Presidente del Jurado

Jurado

Jurado

Bogotá, D.C., Junio de 2019

Las directivas de la Fundación Universitaria los Libertadores, los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documento, estos corresponden únicamente a las autoras.

ÍNDICE

1. INTRODUCCIÓN.....	11
2. PLANTEAMIENTO DEL PROBLEMA.....	13
3. JUSTIFICACIÓN.....	14
4. OBJETIVOS.....	15
4.1. Objetivo general	15
4.2. Objetivos específicos	15
5. MARCO DE REFERENCIA	16
5.1. Grandes superficies	16
5.2. Ventas	17
5.3. Análisis clúster K-means para agrupación de datos.....	18
5.3.1. Criterio Mojena	18
5.4. Modelización estadística para datos longitudinales.....	18
6. METODOLOGÍA	20
6.1. Variables	20
6.2. Fases	20
6.3. Datos.....	20
7. ANÁLISIS DE RESULTADOS	22
7.1. Análisis de clúster para agrupación de datos.....	22
7.2. Análisis descriptivo.....	24
7.3. Descriptivos de datos longitudinales	26
7.4. Modelo de Laird -Ware para el análisis de las ventas de una gran superficie.	29
7.4.1. Modelo 1	30

7.4.2. Modelo 2.....	33
7.4.3. Modelo 3.....	36
7.4.4. Modelo 4.....	39
8. CONCLUSIONES	44
9. REFERENCIAS	45

ÍNDICE DE GRÁFICAS

Gráfica 1. Análisis clúster de las tiendas por medio de criterio Mojena	22
Gráfica 2. Resultados clustering K-means.....	23
Gráfica 3. Boxplot de las ventas según clasificación de grupo	24
Gráfica 4. Boxplot de grupo 1	25
Gráfica 5. Boxplot de grupo 2	25
Gráfica 6. Boxplot de grupo 3	26
Gráfica 7. Modelización estadística para datos longitudinales de las ventas totales por tienda.....	27
Gráfica 8. Modelización estadística para datos longitudinales del promedio de las ventas totales por grupo	27
Gráfica 9. Modelización estadística para datos longitudinales de las ventas del grupo 1	28
Gráfica 10. Modelización estadística para datos longitudinales de las ventas del grupo 2	29
Gráfica 11. Residuales estandarizados del modelo 1	30
Gráfica 12. Normalidad del modelo 1	31
Gráfica 13. Boxplot de residuales modelos 1.....	31
Gráfica 14. Residuales estandarizados modelos 1 por tienda	32
Gráfica 15. Perfiles individuales con modelo 1 ajustado.....	32
Gráfica 16. Residuales estandarizados del modelo 2.....	33
Gráfica 17. Normalidad de modelo 2	34
Gráfica 18. Boxplot residuales modelo 2	34
Gráfica 19. Perfiles individuales modelo 2.....	35
Gráfica 20. Perfiles individuales con modelo 2 ajustado.....	35
Gráfica 21. Residuales estandarizados del modelo 3.....	36
Gráfica 22. Normalidad del modelo 3	37
Gráfica 23. Boxplot residuales modelo 3	37
Gráfica 24. Perfiles individuales modelo 3.....	38
Gráfica 25. Perfiles individuales con modelo 3 ajustado.....	38

Gráfica 26. Residuales estandarizados del modelo 4.....	39
Gráfica 27. Normalidad del modelo 4	40
Gráfica 28. Boxplot de residuales del modelo 4.....	40
Gráfica 29. Perfiles individuales modelo 4	41
Gráfica 30. Perfiles individuales modelo 4 ajustado	41

ÍNDICE DE TABLAS

Tabla 1. Tipo de variables de estudio	20
Tabla 2. Base de datos de ventas de la gran superficie	21
Tabla 3. Clasificación de grupos de acuerdo al criterio Mojena.....	23

RESUMEN

En este trabajo se analiza el efecto del tiempo, el grupo de tiendas y la interacción entre el tiempo y el grupo de tiendas en las ventas de una gran superficie, distribuidas por tienda en Colombia, a partir de los datos obtenidos en el periodo comprendido entre el año 2013 y 2018. Las variables de estudio fueron las ventas, el tiempo y el tipo de tienda. Dado el número de tiendas de la gran superficie fue necesario realizar clasificación de datos por medio de la técnica de análisis clúster k-means, de esta manera se generaron tres grupos a través de la tendencia de ventas de la gran superficie; así mismo se generaron 4 modelos lineales mixtos para determinar el efecto de las variables grupo y tiempo en las ventas. Los resultados obtenidos muestran que las tiendas pueden ser clasificadas en 3 grupos, de acuerdo a su margen de ventas; en cuanto al modelo lineal mixto que mejor explica las ventas de la gran superficie es el número 4.

Palabras clave: *Datos longitudinales, Análisis clúster, Grandes superficies, Ventas, Modelo lineal mixto.*

ABSTRACT

This paper analyzes the effect of time, the group of stores and the interaction between time and the group of stores in the sales of a large area, distributed by the store in Colombia, from the corresponding data in the period covered between 2013 and 2018. The study variables were sales, time and type of store. Given the number of stores in the large area, it was necessary to perform data classification using the k-means clustering technique, in this way three groups were generated through the sales trend of the large area; likewise, 4 mixed linear models were generated to determine the effect of group and time variables on sales.

The results obtained show that stores can be classified into 3 groups, according to their sales margin; as for the mixed linear model that best explains the sales of the large area is the number 4.

Keywords: Longitudinal data, Cluster analysis, Supermarkets, Sales

1. INTRODUCCIÓN

Las grandes superficies hacen referencia a las tiendas, almacenes y mercados grandes, distribuidores de diversidad de productos, caracterizadas por su capacidad de almacenamiento y distribución, así como por su condición económica, cuentan con una superficie superior a 2.500 m² (Superintendencia de Industria y Comercio, 2012) y se ubican principalmente en sectores urbanos (Castro, 2017).

Dado que distribuyen productos de “alimentos, aseo y limpieza, automotores, cuidado personal, electrodomésticos y tecnología, ferretería, juguetería y artículos deportivos, muebles y lencería, papelería, entre otros” (Colombia compra eficiente, 2019, párrafo 2) se cuenta con la estrategia de compra de autoservicio o pague por llevar.

Las ventas en una gran superficie se realizan a través de crédito y contado, y se cuenta con variedad de clientes, entre los que se encuentran personas naturales y jurídicas; el acceso del cliente a los productos puede darse a través de la omnicanalidad, es decir a través de venta en piso o telefónica, lo que indica que el comprador puede acceder a un punto de venta, denominado tienda y seleccionar los artículos que desea adquirir.

El análisis estadístico de ventas por tienda es un insumo para la organización, con este es posible planear estrategias transversales que le faciliten seguir manteniendo su competitividad en el mercado colombiano.

Teniendo en cuenta lo anterior, en este trabajo se analiza el efecto del tiempo, el grupo de tiendas y la interacción entre el tiempo y el grupo de tiendas en las ventas de una gran superficie a través del modelo Laird-Ware, dado que los análisis que incluyen estas variables lograrían aportar a la compañía información más precisa en relación con las ventas; se elige aquel modelo que explique de mejor manera las ventas, esto, a partir de los datos obtenidos en el periodo comprendido entre el año 2013 a 2018, para ello, en el documento se presenta la segmentación de grupos de tiendas realizada por medio de análisis clúster o de conglomerados, la cual permite agrupar los datos y obtener dos grupos homogéneos en función de la mayor diferencia en ventas.

Por otra parte, se presenta la descripción de las ventas de manera individual, según el promedio del grupo y las ventas totales, esto a partir de Boxplot y gráfica de datos longitudinales.

Finalmente, se muestran 4 modelos para los datos, en los cuales se pueden identificar la importancia de cada una de las variables en las ventas de la gran superficie.

2. PLANTEAMIENTO DEL PROBLEMA

Las grandes superficies en Colombia han tenido cada vez mayor participación en el mercado, para el año 2013 en el país el 40% de las compras de la canasta familiar se realizaron en este tipo de centros de distribución (Uribe, 2013), según el DANE para el año 2018 el aporte de este sector al PIB (Producto Interno Bruto) fue de 2,7%, así mismo, el crecimiento para el mismo año fue de 8% en los primeros 11 meses del 2018, en comparación con el año 2017 (El Tiempo, 2018).

En el sector de grandes superficies en Colombia se ha observado la incursión de mercados e hipermercados que ofrecen a los consumidores confianza en los productos, precios que no son bajos pero accesibles y planes de fidelización como el sistema de puntos o el sistema de identificación única por cliente, aspectos que se convierten en beneficios para el comprador; de igual manera, a fin de ganar competitividad en el mercado, las grandes superficies han ofrecido beneficios adicionales acordes con los avances tecnológicos, como lo son compras online y apps que facilitan acceder a descuentos y realizar compras de manera más rápida, que a través de la venta en piso.

El análisis estadístico de ventas por tienda es un insumo para la organización, con este es posible planear estrategias transversales que le faciliten seguir manteniendo su competitividad en el mercado colombiano.

El abordaje actual de los datos relacionados con las ventas en la gran superficie de la cual se obtiene información está basado en los reportes mensuales y anuales, pueden ser abordados por diferentes técnicas y una de ellas es la realizada en este documento.

Es así como el estudio, busca responder la pregunta

¿Cuál es el modelo lineal mixto que mejor explica la evolución de las ventas del periodo 2013 a 2018 de las tiendas de una gran superficie?

3. JUSTIFICACIÓN

La caracterización del sector retail, en el cual se encuentran las grandes superficies fue realizado por Fenalco (2016), con miras a presentar la actualización y análisis del contexto, esto fue realizado por la agremiación partiendo de datos macroeconómicos, hasta llegar al análisis del consumidor individual, permitiendo hacer una proyección de las ventas para el año 2017, considerando variables económicas, sociales y políticas.

El análisis de Fenalco (2016) fue realizado a través de técnicas estadísticas de corte descriptivo, a través de las cuales se presenta un panorama general respecto a los hábitos de consumo, preferencia, canal de compra (de acuerdo a la omnicanalidad) y evaluación de calidad de servicio; este informe en particular carece del análisis de la variación de la venta a través del tiempo y ubicación geográfica.

Por otra parte en el estudio de Pacheco (2017) se realizó muestreo aleatorio simple a estratos, derivados de los canales de distribución de un producto de consumo masivo, con miras a comprender las ventas en un periodo de tiempo determinado, se aplicó el diseño de medidas repetidas, lo cual permitió concluir que el tiempo es determinante en el planteamiento de modelos de predicción en las ventas en una organización.

Como se puede observar el estudio de las ventas de grandes superficies se ha realizado a través de diferentes técnicas estadísticas, sin embargo, no se ha elaborado un análisis estadístico a partir de modelos lineales mixtos.

4. OBJETIVOS

4.1. Objetivo general

Determinar la evolución de las ventas del periodo 2013 a 2018 de las tiendas de una gran superficie a partir del modelo lineal mixto.

4.2. Objetivos específicos

Agrupar las tiendas que integran una gran superficie de acuerdo a la tendencia de ventas del periodo 2013 a 2018 por medio de análisis clúster k-means.

Describir la tendencia en ventas de cada grupo de tiendas de acuerdo al tiempo, y las características del grupo.

Identificar el modelo lineal mixto que más se ajusta a las ventas del periodo 2013 a 2018 de las tiendas de una gran superficie.

5. MARCO DE REFERENCIA

5.1. Grandes superficies

Las grandes superficies son antagónicas a las tiendas de barrio y han sido entendidas como distribuidores de diversos productos, como lo son alimentos, no alimentos, electrodomésticos, vestuario, electrónicos, mobiliario, decoración, textil, entre otros; caracterizadas por su capacidad de almacenamiento y distribución, así como por su condición económica; cuentan con una superficie superior a 2.500 m² (Superintendencia de Industria y Comercio, 2012), en las que se incorporan grandes áreas para estacionamiento de vehículos y servicios completos (Marketing, n.d.) y se ubican principalmente en sectores urbanos (Castro, 2017).

Dentro de las características de las grandes superficies se encuentra el proceso logístico por el que se da la distribución y almacenamiento de productos, las estrategias de comercialización y la forma de compra.

La logística de las grandes superficies corresponde al sistema de distribución moderna (Castro, 2017; Montagut, Vivas, & Montagut, 2008) propio del siglo XX, el cual implica la disponibilidad de productos en estanterías, son de fácil acceso para los clientes; esta colocación hace que la elección del cliente sea más autónoma, provocando la sensación de libertad en la compra. Así mismo, la distribución en las grandes superficies busca aumentar el margen de ganancia por medio de la disminución de costos, como producto de las alianzas y la disminución de la competencia, consecuencia de los beneficios adicionales que ofrece la compra al consumidor. En las grandes superficies se cuenta con franquicias, que ofrecen al cliente servicios adicionales.

Respecto a la forma en la cual comercializan los productos Montagut et al., (2008) plantea que la distribución moderna implementa la “*generalización del libre servicio como sistema de venta*” (p.12) y las estrategias comerciales varían en cada empresa, pasando por la calidad, hasta conceptos más específicos como la venta de productos ecológicos, la especialización o la diversificación.

La Superintendencia de Industria y Comercio (2012) presenta que una de los atributos distintivos de las grandes superficies se da en el nivel de compra, estos se surten a través de manufactura, generando alianzas con proveedores.

Castro (2017) plantea que, en el último siglo, las familias cada vez compran más sus alimentos e insumos en las grandes superficies, promoviendo que las tiendas y el comercio de barrio existentes desaparezcan gradualmente, pues los beneficios económicos y de diversidad a productos que connotan el acceso a las grandes superficies, satisfacen al consumidor, ya sea este entidades jurídicas o clientes individuales, como lo son las familias.

5.2. Ventas

La venta en grandes superficies hace referencia al término todo en uno, el cual implica la comercialización de varias categorías de productos, provenientes de diferentes proveedores, “convirtiéndose en un canal de distribución de: Alimentos, Aseo y limpieza, Automotores, Cuidado personal, Electrodomésticos y tecnología, Ferretería, Juguetería y artículos deportivos, Muebles y lencería, Papelería, Otros” (Colombia compra eficiente, 2019, párrafo 2).

La venta es la acción de vender, traspasar a otro una propiedad de acuerdo a un precio que ha sido convenio, las ventas buscan vender el producto que la empresa produce o distribuye y el marketing pretende que la empresa tenga lo que el cliente quiere; en el caso de las grandes superficies, las ventas implican contar con la disponibilidad de productos en un mismo lugar, es decir en tiendas (Vega, 2012).

5.3. Análisis clúster K-means para agrupación de datos

El análisis clúster k-means es un método comúnmente utilizado, que permite crear una clasificación, descubrir la estructura de las categorías en la que se encajan las observaciones, implica “particionar un conjunto de datos dado en un conjunto de k grupos (es decir, k agrupamientos), donde k representa el número de grupos preespecificados clasifica los objetos en varios grupos (es decir, grupos), de manera que los objetos dentro de los mismos los grupos son lo más similares posible” (Kassambara, 2017 p.36)

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2$$

x_i Diseñar un punto de datos perteneciente al clúster.

μ_k Es el valor medio de los puntos asignados al clúster

5.3.1. Criterio Mojena

Para determinar el número de grupos a elegir en el análisis de conglomerados se utiliza el criterio de Mojena, el cual está relacionado con los saltos relativos en los valores de fusión. En el “método se compara el valor de fusión de cada etapa con el promedio de los valores de fusión sumado con el producto de una cierta constante por la cuasidesviación típica de los valores de fusión. Cuando un valor de fusión supera dicha cantidad se concluye que el nivel precedente es el que origina la solución óptima” (Gutiérrez et al., 1994, p.111).

5.4. Modelización estadística para datos longitudinales

El análisis de datos longitudinales es una técnica gráfica para modelar las medidas repetidas de las unidades de observación en el tiempo, permite analizar y observar de manera secuenciada la evolución de un fenómeno (Arnau y Bono, 2008), constituyendo un conjunto estructurado de procedimientos que facilitan obtener información relacionada con el proceso de cambio.

El proceso de análisis es “una representación simplista de la media de un vector de datos, (...). Las unidades son asignadas al azar en $q \geq 1$ grupos de tratamiento”, es así como los grupos en el estudio se determinaron a través del análisis clúster.

Así mismo para realizar el análisis se tuvo en cuenta que existe una variación aleatoria en las ventas por año, por ejemplo, la variación en ventas existe en cada una de las tiendas, de ahí que considere que las ventas son contextuales, es decir que están afectadas posiblemente por el lugar de ubicación de las tiendas.

El modelo de datos de datos longitudinales de acuerdo con León (2015) es el siguiente:

$$Y_{hlj} = \mu + \pi_l + b_{hl} + \gamma_j + (\tau\gamma)_{lj} + e_{hlj}$$

Donde

μ es la media

π_l es la desviación sobre la media muestral producida al encontrarse una unidad en el l-ésimo grupo.

b_{hl} es un efecto aleatorio, con esperanza nula, que representa la desviación producida al medir la variable Y (ventas) en una h-ésima unidad particular en el l-ésimo grupo. Es una variación producida entre unidades, por su distribución mediante una distribución de probabilidad.

γ_j es la desviación asociada al j-ésimo instante

$(\tau\gamma)_{lj}$ es el efecto de interacción para el l-ésimo grupo en el j-ésimo instante, que conlleva a una desviación adicional.

6. METODOLOGÍA

6.1. Variables

El estudio considera las siguientes variables categóricas o cualitativas de tipo nominal y numéricas de tipo continuo.

Tipo de variable	Variable
Categórica	Tienda
Cuantitativa	Venta por año
	Tiempo

Tabla 1. Tipo de variables de estudio

6.2. Fases

El estudio se desarrolla a través de los siguientes momentos o fases:

Fase 1: Análisis clúster k-means.

Fase 2: Descripción de datos.

Fase 3: Modelo de Laird - Ware

6.3. Datos

La información de la base de datos cuenta con el reporte de ventas total anual de 6 años consecutivos en 17 tiendas, y el reporte de 2 años en 4 tiendas, con un total de 160 datos y una identificación por cada tienda, como se muestra en la tabla 2.

	Venta_2013	Venta_2014	Venta_2015	Venta_2016	Venta_2017	Venta_2018
1	60880,78	68810,31	78831,85	83928,52	76118,65	79984,73
2	98363,92	107274,44	117654,86	126447,88	113558,58	111139,41
3	45541,25	48961,64	57180,9	54826,81	49955,83	55913,66
4	60340,29	61075,75	74758,67	79292,85	70866,21	75832,41
5	114936,89	122432,1	136851,71	126694,72	121904,93	132614,3
6	40147,89	44884,91	49259	54058,25	57724,37	63166,32
7	74445,94	83008,81	88906,79	94256,6	90497,04	90946,24
8	39726	45785,69	55023,32	57102,05	54231,2	56662,64
9	53098,9	60438,19	70499,98	62187,31	59365,98	68777,02
10	26303347	30315,31	35972,22	37802,17	36325,26	41602,65

11	39544,2	48064,03	56681,43	60152,33	60978,58	65497,32
12	31843,6	34416,52	40612,02	42356,7	44791,25	54373,99
13	40560,68	44490,49	51712,98	58363,94	63680,89	66564,48
14	25560,68	32781,88	49150,7	64866,73	62554,33	65627,82
15	31019,22	34815,67	44173,15	43759,17	42221,76	46363,17
16	26847,68	31154,82	37736,92	41517,18	41984,44	44304,31
17	0	0	0	26748,6	38251,71	44435,81
18	0	0	0	0	5574,47	43997,07
19	0	0	0	0	4996,7	42925,37
20	0	0	0	0	6668,92	31467,93

Tabla 2. Base de datos de ventas de la gran superficie

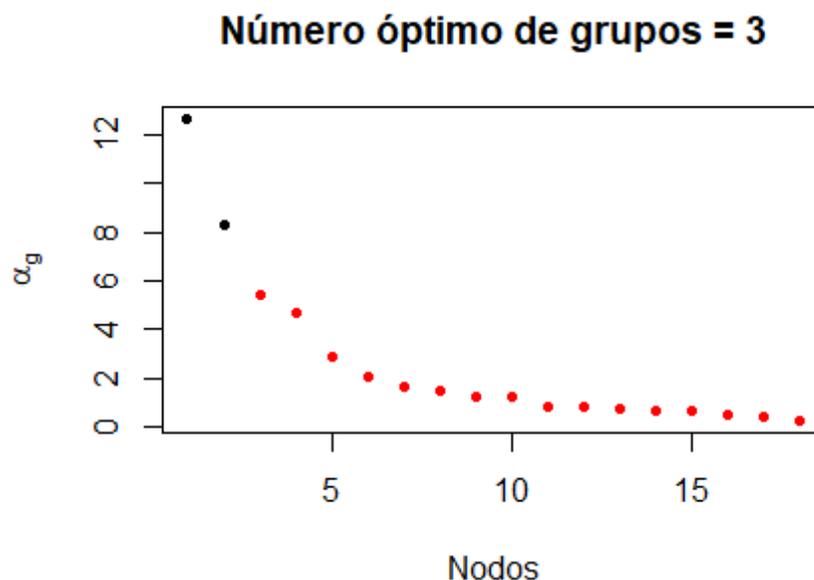
*Datos expresados en miles de millones

7. ANÁLISIS DE RESULTADOS

La base de datos analizada contiene las ventas totales del periodo comprendido entre 2013 a 2018 de 20 tiendas pertenecientes a una gran superficie, ubicadas en diferentes regiones de Colombia. Para el análisis se aplican las técnicas estadísticas clúster, prueba de normalidad y modelación de datos longitudinales.

7.1. Análisis de clúster para agrupación de datos

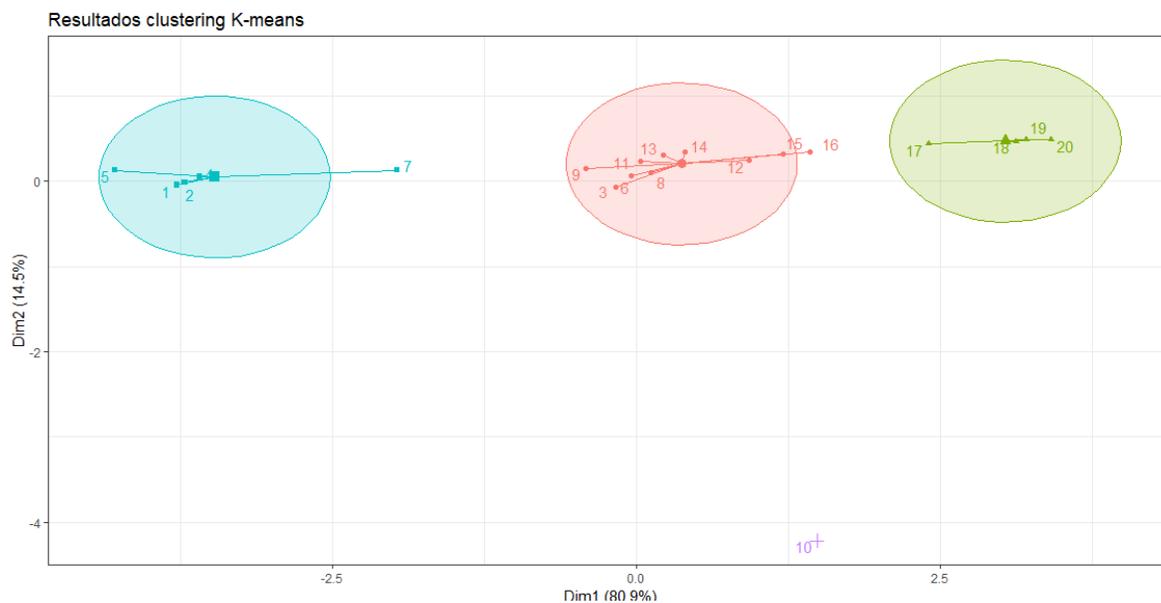
A través del análisis de clúster se identifica la clasificación de las tiendas bajo el criterio de tendencia de venta en el periodo comprendido entre 2013 a 2018, los casos a agrupar son las tiendas codificadas con números, en una escala de 1 a 20, esto a fin de garantizar la confidencialidad de los datos de la gran superficie; por su parte, el criterio de tendencia de ventas hace referencia a las ventas totales por año de cada tienda.



Gráfica 1. Análisis clúster de las tiendas por medio de criterio Mojena

A partir de los resultados obtenidos en la gráfica 1 se puede observar que las tiendas pueden clasificarse en 3 grupos según el criterio Mojena, el cual indica que debe tomarse como punto de referencia la distancia entre cada uno de los grupos y

sus observaciones. Los grupos estarían constituidos por determinadas tiendas, como se presenta en la tabla 2.



Gráfica 2. Resultados clustering K-means

De acuerdo con la gráfica 2 se identifica que las tiendas de la gran superficie se clasifican en tres grupos.

Grupo	Tienda	Característica
1	1, 2, 4, 5, 7	Ubicadas en ciudades con más de 2'500.000 habitantes
2	3, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16	Ubicadas en ciudades con población entre 362.000 a 2'400.000 de habitantes
3	17, 18, 19, 20	Ubicadas en ciudades población menor a 200.000 habitantes

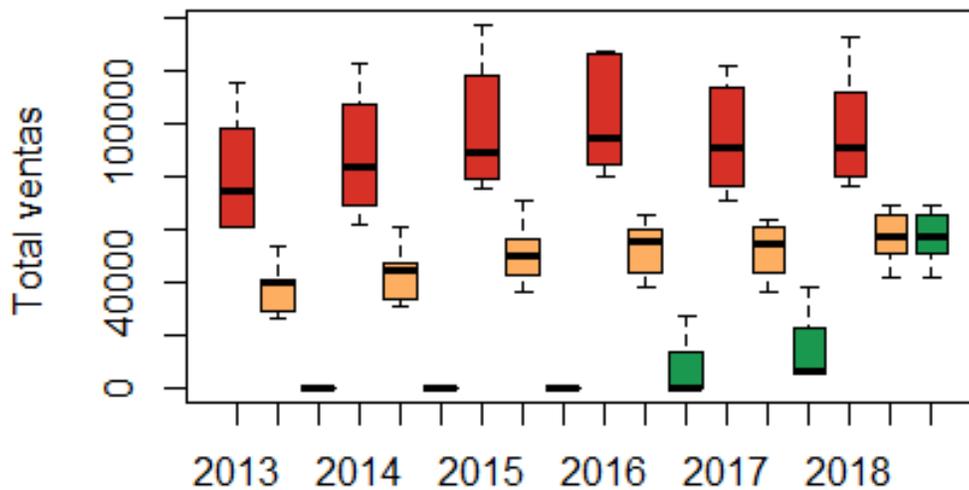
Tabla 3. Clasificación de grupos de acuerdo al criterio Mojena

De acuerdo con la tabla 2 se pueden considerar las siguientes características similares en los grupos; para el grupo 1, las tiendas están ubicadas en ciudades principales que cuentan con más de 2'500.000 de habitantes y por lo tanto las ventas tienen un mayor valor nominal en ventas. El grupo 2 lo integran las ciudades intermedias con población entre 362.000 a 2'400.000 de habitantes; y el grupo 3,

está constituido por tiendas ubicadas en lugares aledaños a ciudades intermedias o principales, con población menor a 200.000 habitantes (DANE, 2017).

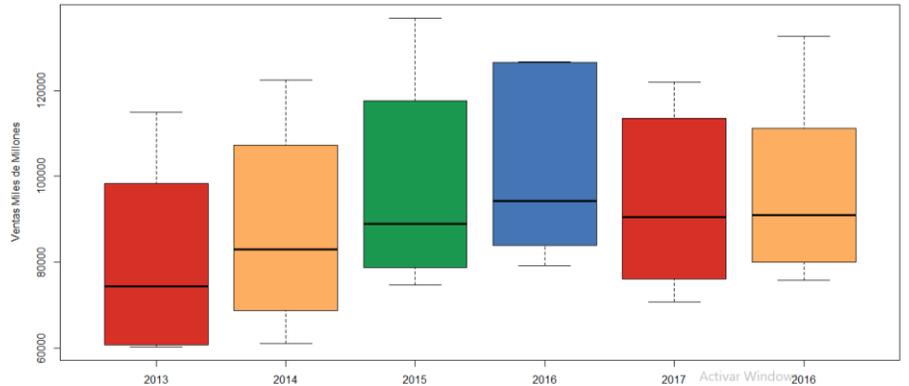
7.2. Análisis descriptivo

A través del análisis descriptivo de los datos se realiza una exploración de las ventas por año.



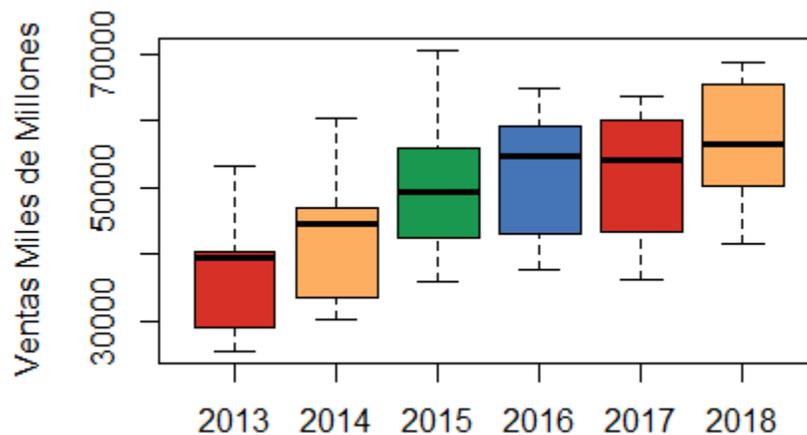
Gráfica 3. Boxplot de las ventas según clasificación de grupo

En la gráfica 3 se observa que el grupo 1 (representado en color rojo) tiene el mayor nivel de ventas para todos años analizados, esto como producto de la densidad poblacional de las ciudades en las cuales están ubicadas las tiendas. Por otra parte, para el año 2018, se identifica que los grupos 2 (representado en el color amarillo) y 3 (representado en el color verde), aunque están conformadas con una densidad poblacional diferente, contaron con el mismo nivel de ventas.



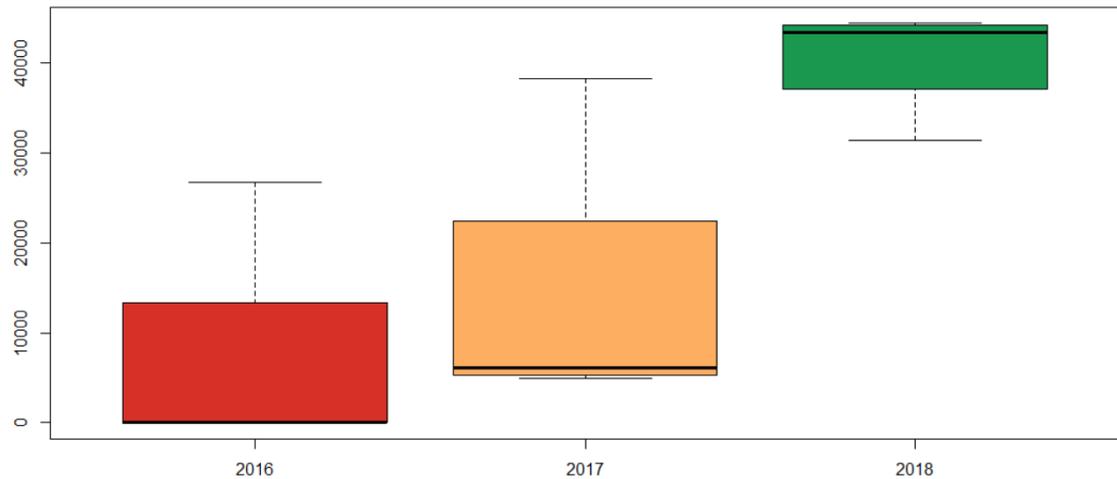
Gráfica 4. Boxplot de grupo 1

En la gráfica 4 se presentan los datos del grupo 1, se identifica una tendencia de crecimiento en las ventas desde el año 2013 hasta el 2018, con una alta dispersión en todos los años analizados, exceptuando el 2016; es de anotar que para el grupo 1 no se encuentran datos atípicos.



Gráfica 5. Boxplot de grupo 2

La gráfica 5 refleja que las tiendas que se encuentran en ciudades intermedias cuentan con un incremento de las ventas significativo desde el 2013 a 2015; a partir del año 2016 y hasta el 2018 las ventas son constantes, con pocos incrementos. En cuanto a la dispersión de los datos, los años 2013, 2014 y 2015 presentan una mayor variación en comparación con los demás años.



Gráfica 6. Boxplot de grupo 3

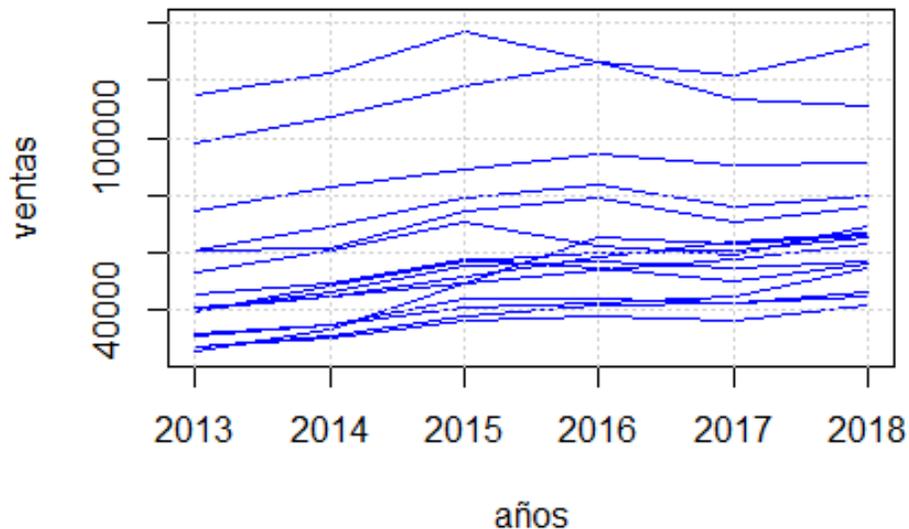
La gráfica 6 muestra un crecimiento significativo en ventas desde el año 2016 al 2018, con una asimetría hacia la derecha y mayor dispersión de los datos en los años 2016 y 2017. En cuanto al año 2018, la asimetría es hacia la izquierda, lo que indica que hubo mayor cantidad de ventas por debajo de la mediana. Dada la ausencia de datos en la conformación del grupo 3, se realizarán los siguientes análisis descriptivos sin este grupo, puesto que no permite la aplicación de técnicas longitudinales.

7.3. Descriptivos de datos longitudinales

Para identificar la tendencia en ventas con base en los análisis de datos obtenidos del periodo comprendido entre 2013 a 2018, así como para obtener predicciones de la tendencia a futuro de las ventas se realiza la modelización estadística para datos longitudinales, esta se presenta en las gráficas 6, 7, 8, 9 y 10 las cuales representan la tendencia por sujeto específico, promediados de la población y de acuerdo a la clasificación de grupos.

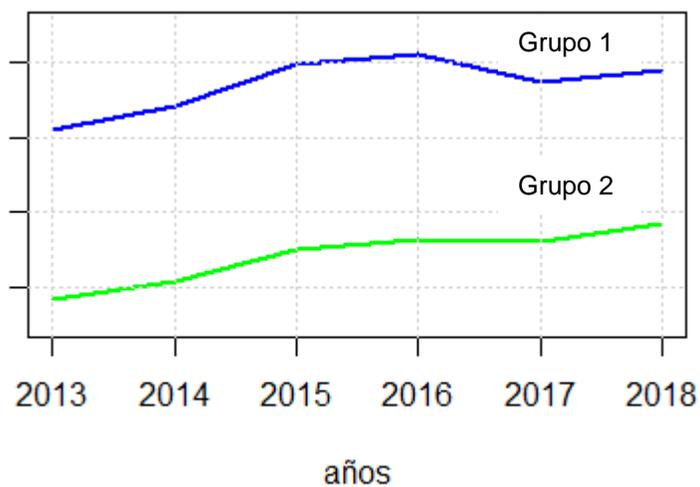
Se presenta a continuación los análisis descriptivos para datos longitudinales, dado que la base de datos está balanceada, es decir que todas las unidades en estudio poseen mediciones repetidas a lo largo de 6 años consecutivos

y los dos grupos a tratar (grupo 1 y 2) presentan homogeneidad en las ventas totales por año. De ahí que se haya eliminado el grupo número 3.



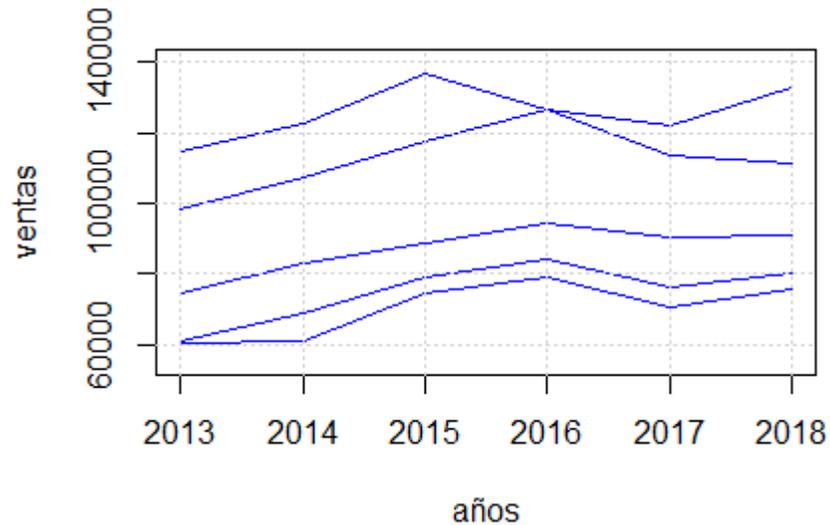
Gráfica 7. Modelización estadística para datos longitudinales de las ventas totales por tienda

En la gráfica 7 puede verse el desarrollo del nivel de ventas de una gran superficie a lo largo 5 años (2013 a 2018), la tendencia general muestra un aumento poco significativo en la mayoría de las tiendas durante el periodo de análisis. Los perfiles para cada una de las tiendas son planos, con crecimiento poco significativo, lo que denota la poca fluctuación en las ventas de la gran superficie.



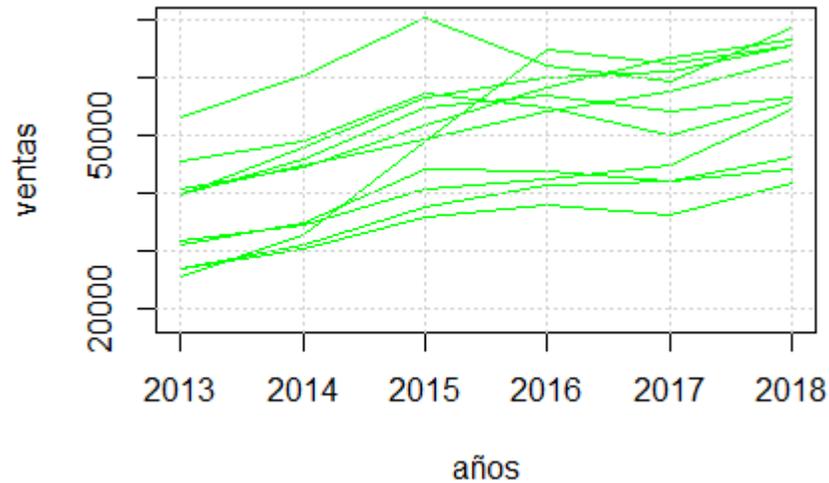
Gráfica 8. Modelización estadística para datos longitudinales del promedio de las ventas totales por grupo

La gráfica 8 muestra que la mayor parte de las trayectorias grupales aumenta de manera general durante el periodo de estudio, para cada grupo las trayectorias de ventas son diferentes, con una tendencia de crecimiento muy poco significativa. Así mismo es de observar que los patrones de ventas a lo largo del tiempo para los dos grupos de tiendas no varían.



Gráfica 9. Modelización estadística para datos longitudinales de las ventas del grupo 1

La gráfica 9 representa la tendencia de ventas de las tiendas que están ubicadas en las ciudades principales, en ella se observa que la distancia entre las trayectorias es equivalente, lo que indica que las ventas son constantes para todos los periodos de tiempo analizados; la tendencia general en el grupo es el aumento de las ventas con el paso de los años de manera muy poco significativa.



Gráfica 10. Modelización estadística para datos longitudinales de las ventas del grupo 2

La gráfica 10 representa la tendencia de ventas de las tiendas que están ubicadas en las ciudades intermedias, en ella se observa que la distancia entre las trayectorias es similar, lo que indica que las ventas son equitativas para todas las tiendas que integran el grupo; sin embargo, en cuanto a la tendencia general se identifica un crecimiento poco significativo en las ventas.

7.4. Modelo de Laird -Ware para el análisis de las ventas de una gran superficie.

Se generaron 4 modelos lineales mixtos para analizar los datos, a fin de determinar cuál de ellos se explica de mejor manera las ventas de la gran superficie.

7.4.1. Modelo 1

Linear mixed-effects model fit by REML

Data: ventas.n

AIC	BIC	logLik
1820.58	1832.91	-905.29

Random effects:

Formula: ~1 | Tienda

(Intercept) Residual

stdDev: 14815.11 5600.54

Fixed effects: Ventas ~ Año + Grupo

	Value	Std. Error	DF	t-value	p-value
(Intercept)	-7192951	696735.8	74	-10.323785	0e+00
Año	3615	345.7	74	10.459311	0e+00
GrupoB	-44695	8210.6	13	-5.443488	1e-04

Correlation:

(Intr) Año

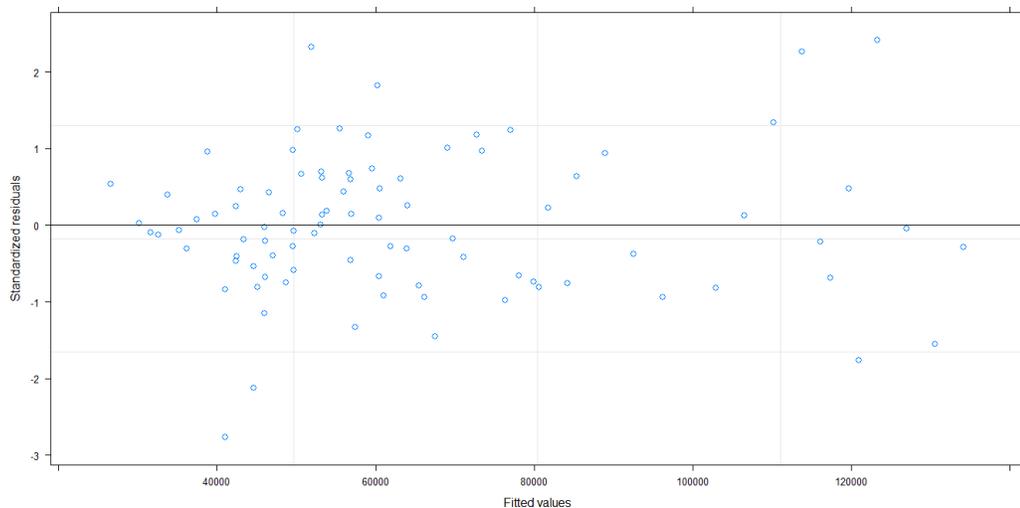
Año -1.000

GrupoB -0.008 0.000

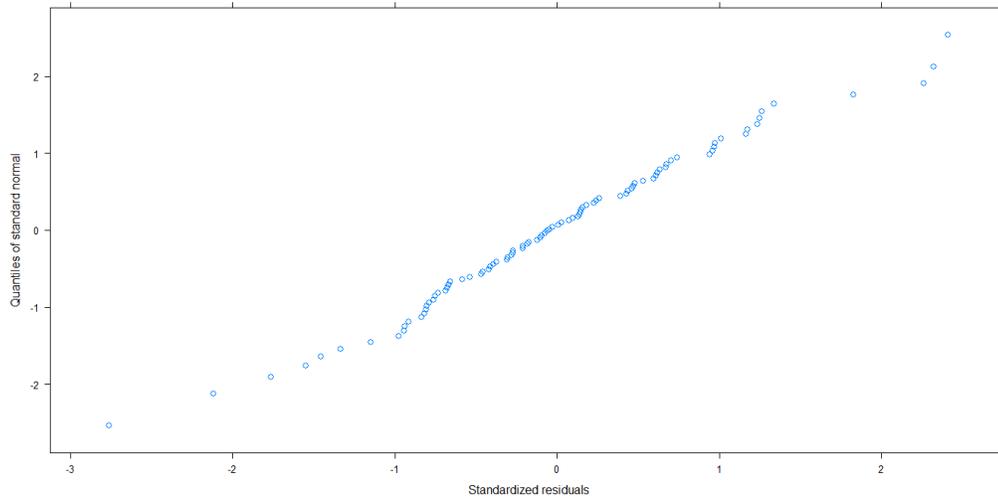
Standardized within-Group Residuals:

Min	Q1	Med	Q3	Max
-2.76307793	-0.64045890	-0.05723961	0.57827750	2.40938674

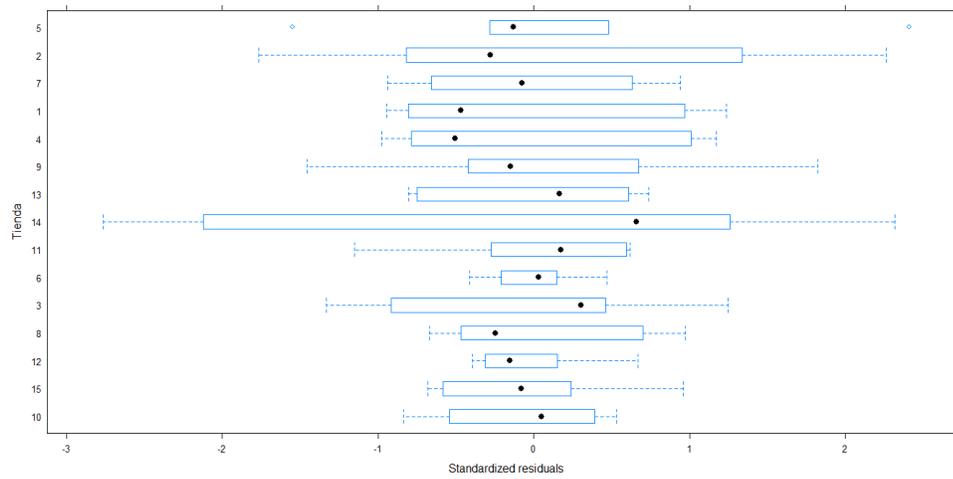
El modelo 1 presenta las covariables, tiempo (año), grupo y el intercepto como efecto aleatorio, se puede observar en él que la intersección entre el año y el grupo 2 es muy significativo en las ventas.



Gráfica 11. Residuales estandarizados del modelo 1

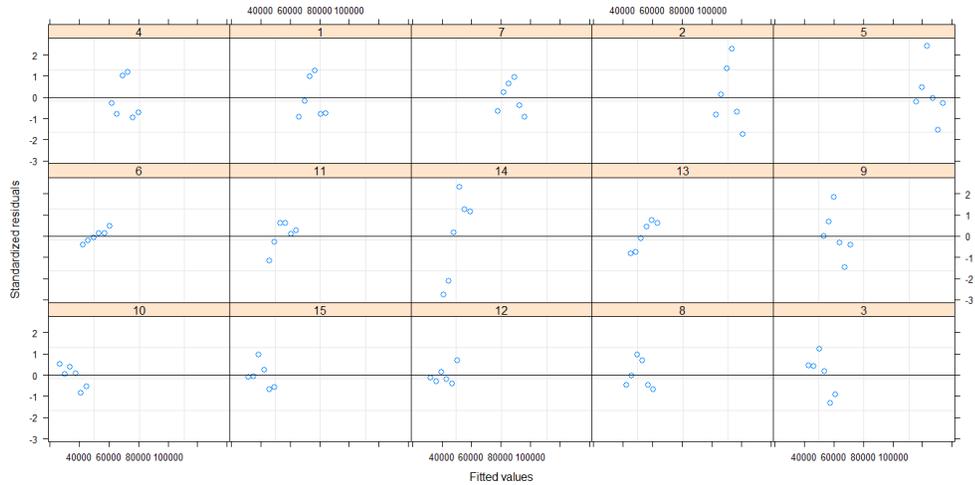


Gráfica 12. Normalidad del modelo 1



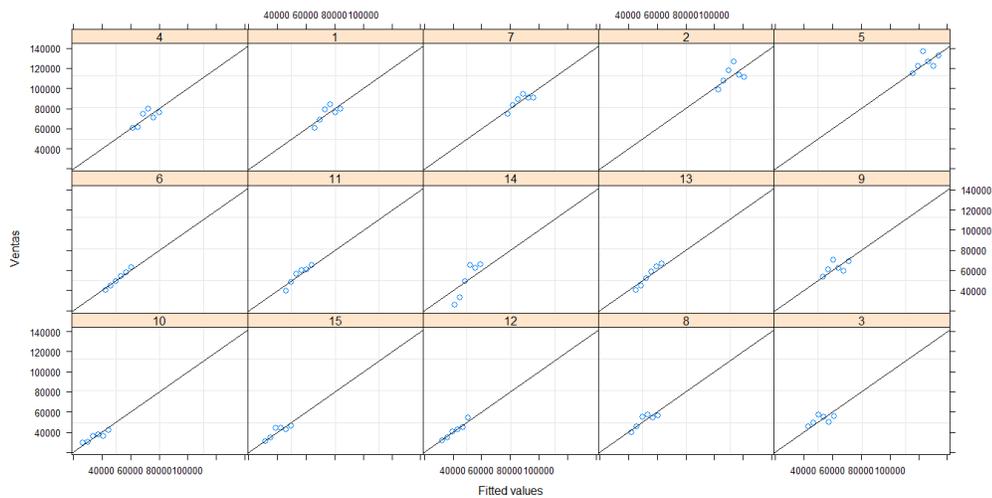
Gráfica 13. Boxplot de residuales modelos 1.

En la gráfica 13 se observa una alta dispersión en los residuales de las tiendas 2, 9 y 14.



Gráfica 14. Residuales estandarizados modelos 1 por tienda

Se observa una tendencia individual muy similar para cada una de las tiendas.



Gráfica 15. Perfiles individuales con modelo 1 ajustado

Se puede apreciar en la grafica 15 un ajuste de todas las tiendas para el modelo 1. Está gráfica es realizada para evaluar el desempeño del modelo 1 en las ventas de la gran superficie.

7.4.2. Modelo 2

Linear mixed-effects model fit by REML

Data: ventas.n

AIC	BIC	logLik
1817.26	1834.521	-901.6299

Random effects:

Formula: ~1 + Grupo | Tienda

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	24050.45	(Intr)
GrupoB	17581.43	-0.979
Residual	5600.54	

Fixed effects: Ventas ~ Año + Grupo

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-7192951	696787.4	74	-10.323022	0.0000
Año	3615	345.7	74	10.459311	0.0000
GrupoB	-44695	11101.8	13	-4.025893	0.0014

Correlation:

	(Intr)	Año
Año	-1.000	
GrupoB	-0.015	0.000

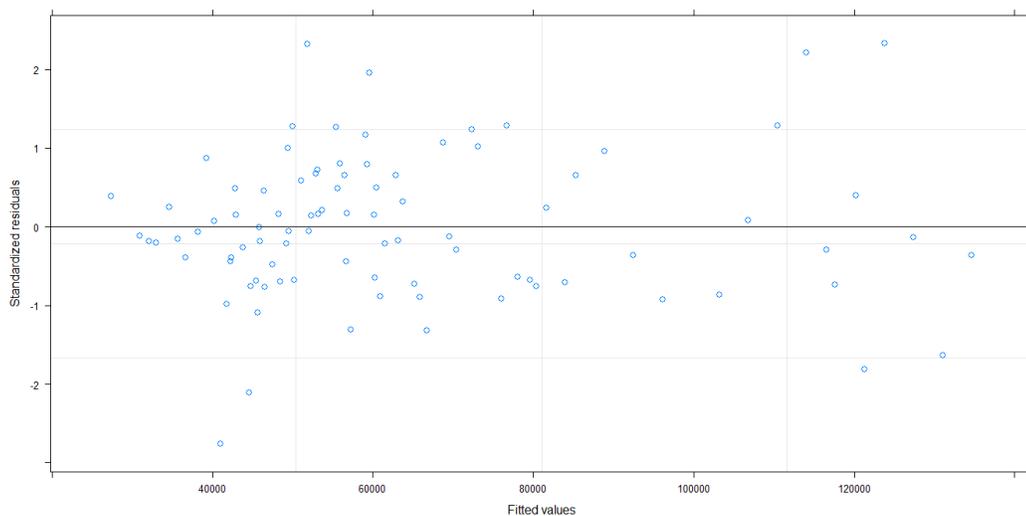
Standardized within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-2.75599964	-0.66846180	-0.09362298	0.56444465	2.32810887

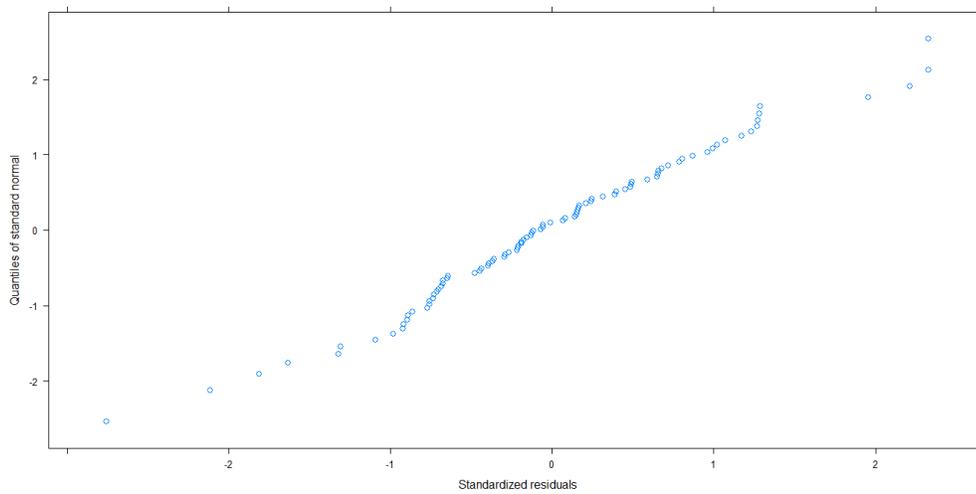
Number of observations: 90

Number of Groups: 15

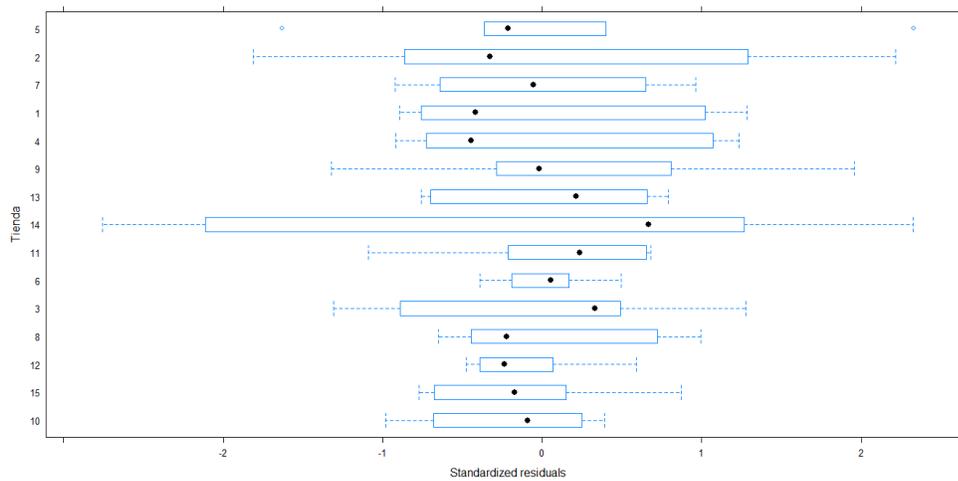
El modelo 2 presenta las covariables, tiempo (año), grupo y el intercepto como efecto aleatorio en el tiempo, se puede observar en él que la intersección entre el año y el grupo 2 es significativo en las ventas.



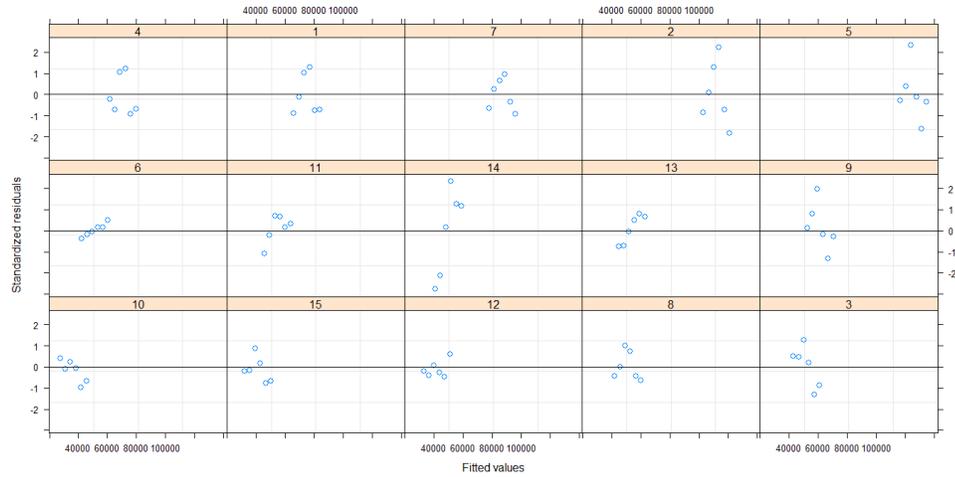
Gráfica 16. Residuales estandarizados del modelo 2



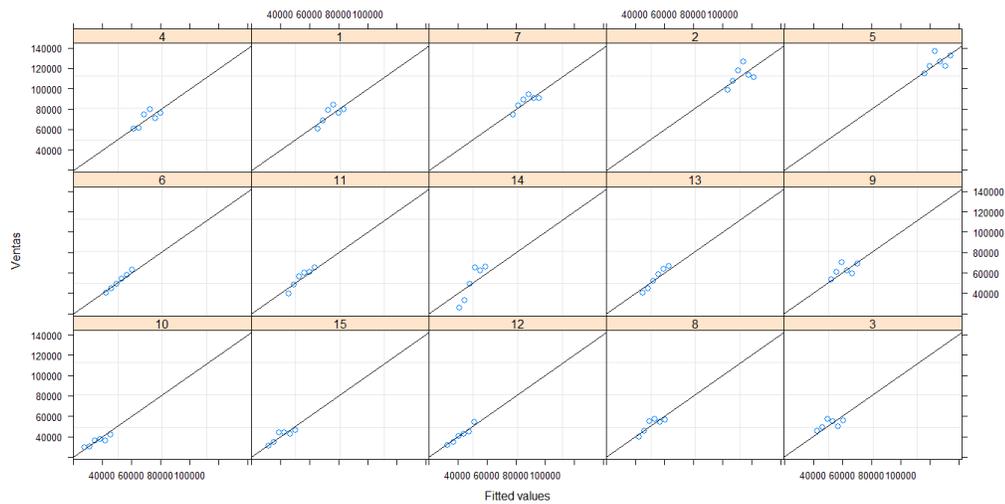
Gráfica 17. Normalidad de modelo 2



Gráfica 18. Boxplot residuales modelo 2



Gráfica 19. Perfiles individuales modelo 2



Gráfica 20. Perfiles individuales con modelo 2 ajustado

Se puede apreciar en la gráfica 20 un ajuste de todas las tiendas para el modelo 2. Esta gráfica es realizada para evaluar el desempeño del modelo 2 en las ventas de la gran superficie.

7.4.3. Modelo 3

Linear mixed-effects model fit by REML

Data: ventas.n

AIC	BIC	logLik
1855.034	1864.943	-923.5169

Random effects:

Formula: ~1 | Tienda

(Intercept) Residual

StdDev: 26058.69 5600.54

Fixed effects: Ventas ~ Año

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-7222748	696736.3	74	-10.36654	0
Año	3615	345.7	74	10.45931	0

Correlation:

(Intr)

Año -1

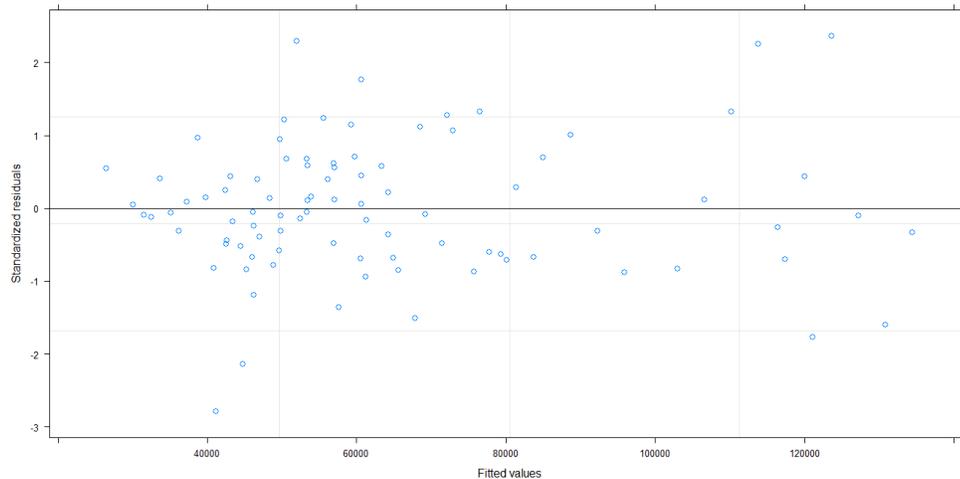
Standardized within-Group Residuals:

Min	Q1	Med	Q3	Max
-2.78534293	-0.59396776	-0.06975302	0.55576516	2.36127060

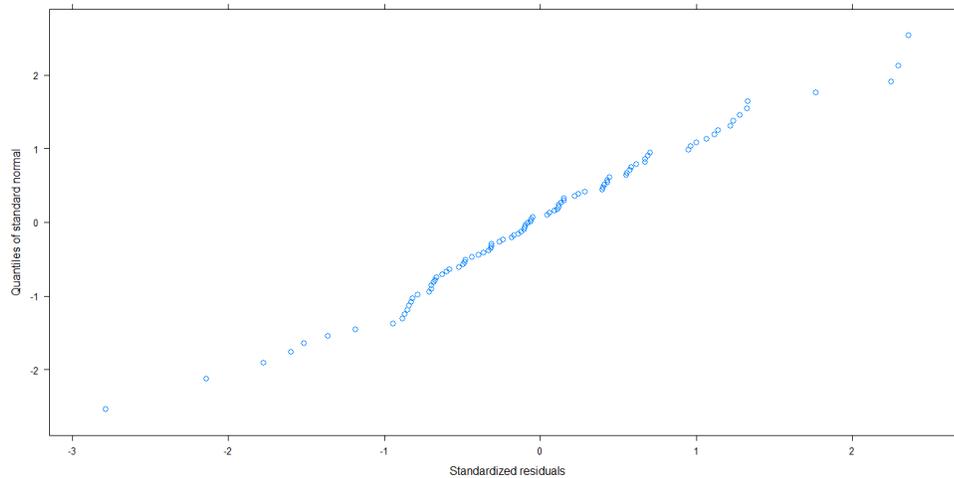
Number of Observations: 90

Number of Groups: 15

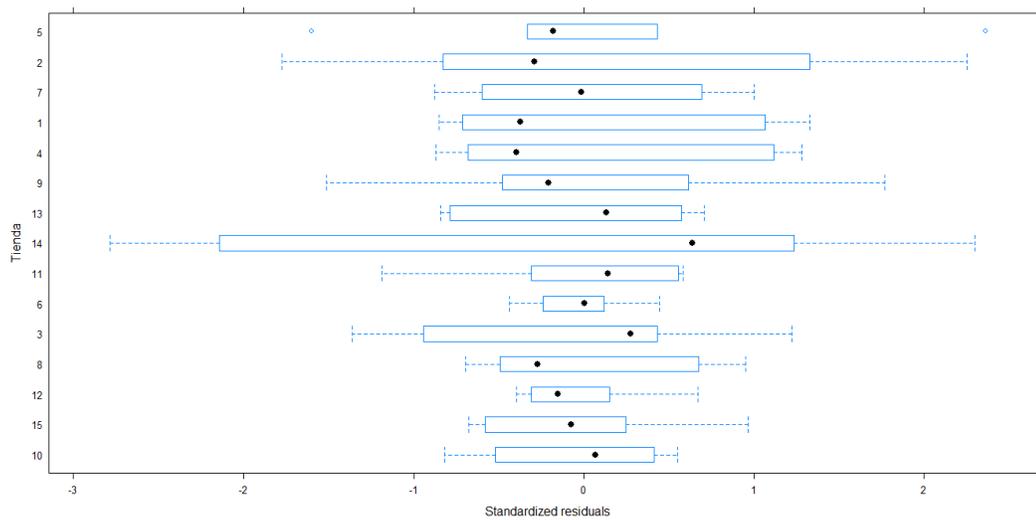
El modelo 3 presenta la covariable, tiempo (año) y el intercepto como efecto aleatorio, se puede observar en él la interacción entre el año y las ventas.



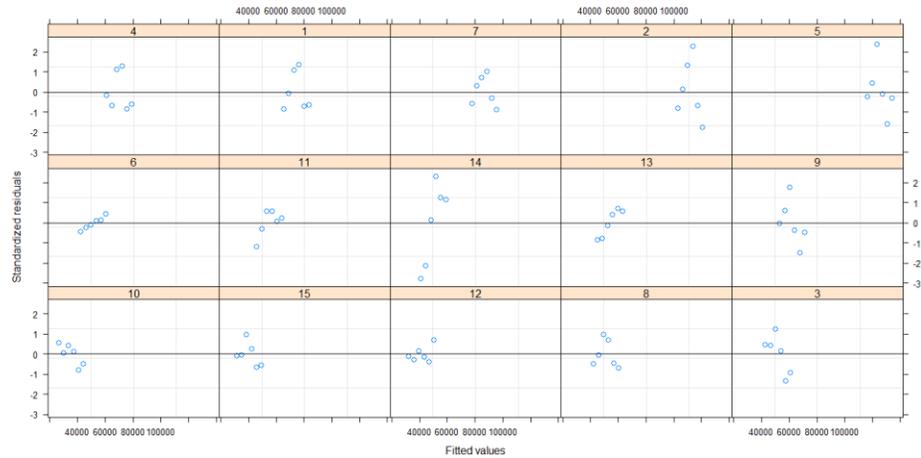
Gráfica 21. Residuales estandarizados del modelo 3



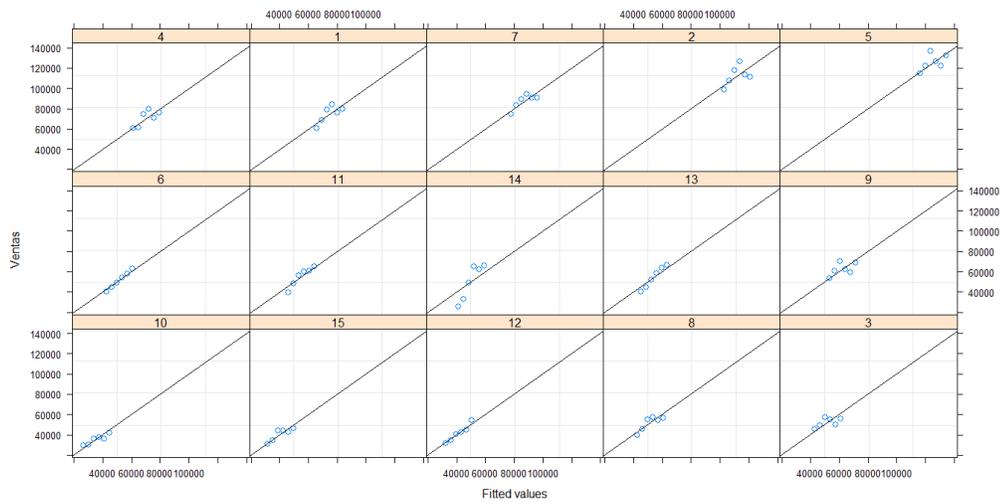
Gráfica 22. Normalidad del modelo 3



Gráfica 23. Boxplot residuales modelo 3



Gráfica 24. Perfiles individuales modelo 3



Gráfica 25. Perfiles individuales con modelo 3 ajustado

Se puede apreciar en la gráfica 25 un ajuste de todas las tiendas para el modelo 3. Esta gráfica es realizada para evaluar el desempeño del modelo 3 en las ventas de la gran superficie.

7.4.4. Modelo 4

```

Linear mixed-effects model fit by REML
Data: ventas.n
      AIC      BIC    logLik
1859.034 1873.898 -923.5169

Random effects:
Formula: ~1 + Año | Tienda
Structure: General positive-definite, Log-Cholesky parametrization
              StdDev      Corr
(Intercept) 2.605872e+04 (Intr)
Año          9.555486e-03 0.002
Residual     5.600537e+03

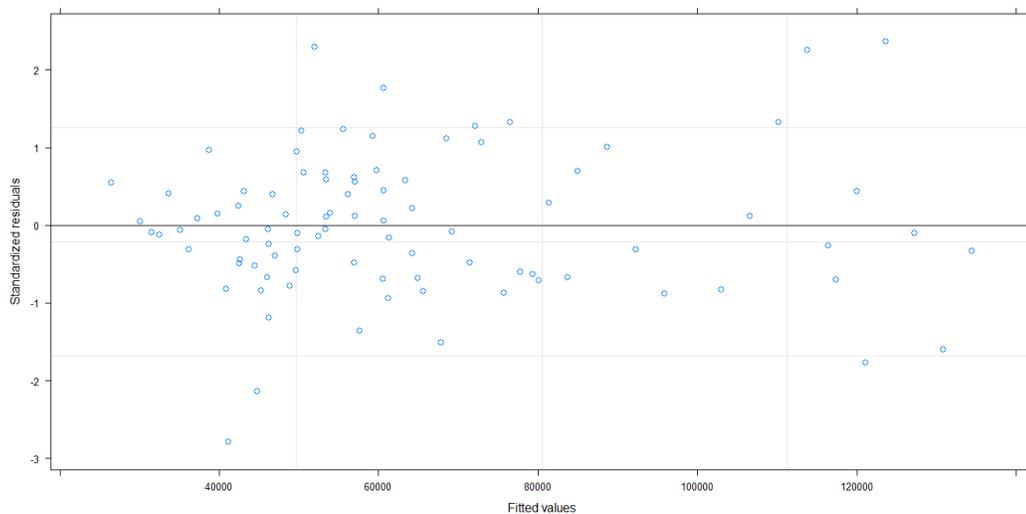
Fixed effects: Ventas ~ Año
              Value Std.Error DF   t-value p-value
(Intercept) -7222748  696736.0 74  -10.36655    0
Año           3615     345.7 74   10.45932    0
Correlation:
(Intr)
Año -1

Standardized within-Group Residuals:
      Min      Q1      Med      Q3      Max
-2.7853442 -0.5939681 -0.0697530  0.5557656  2.3612713

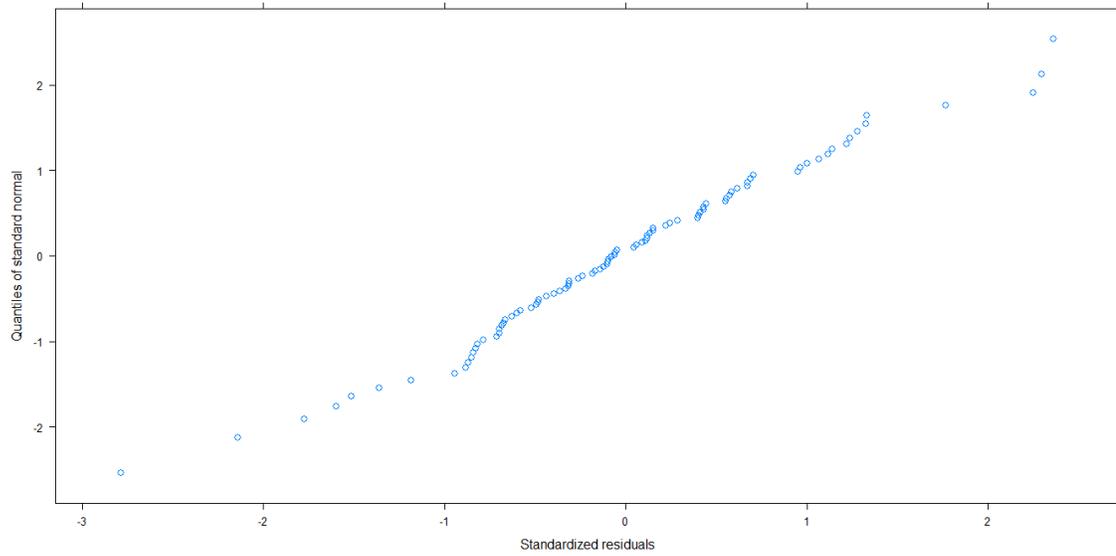
Number of Observations: 90
Number of Groups: 15

```

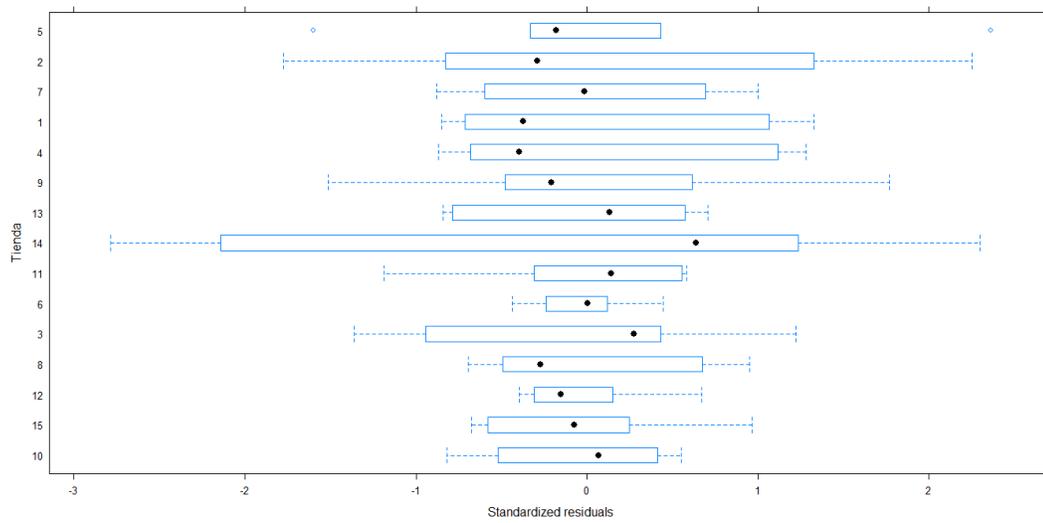
El modelo 4 presenta la covariable, tiempo (año) y el intercepto como efecto aleatorio en el grupo, se puede observar en él la interacción entre el año y las ventas.



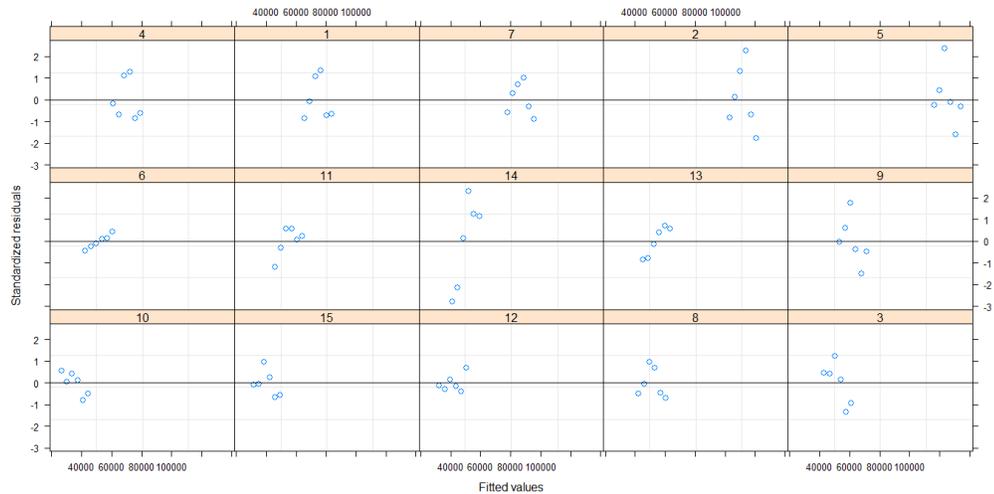
Gráfica 26. Residuales estandarizados del modelo 4



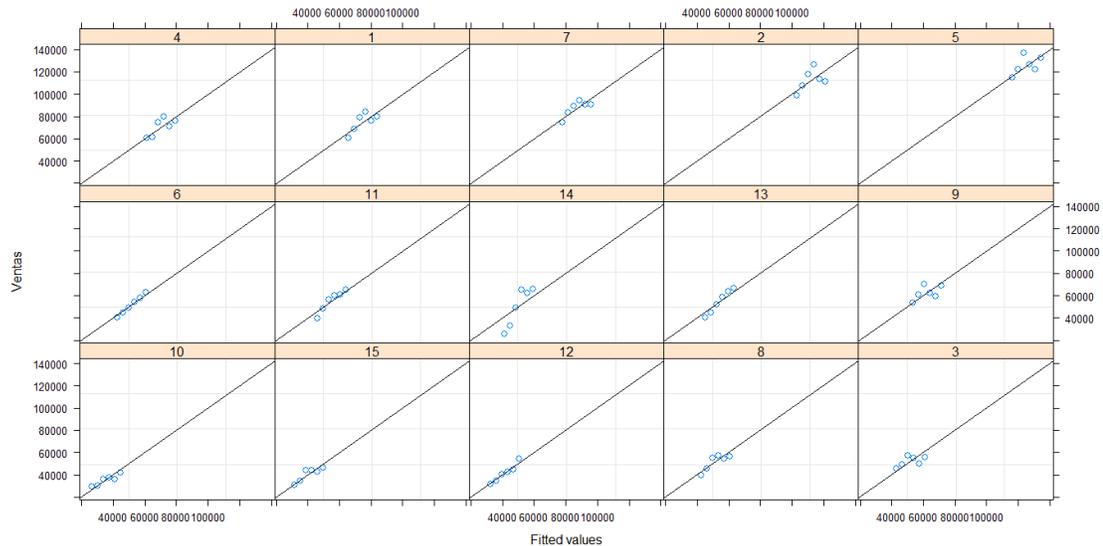
Gráfica 27. Normalidad del modelo 4



Gráfica 28. Boxplot de residuales del modelo 4



Gráfica 29. Perfiles individuales modelo 4



Gráfica 30. Perfiles individuales modelo 4 ajustado

Se puede apreciar en la gráfica 30 un ajuste de todas las tiendas para el modelo 4. Esta gráfica es realizada para evaluar el desempeño del modelo 4 en las ventas de la gran superficie.

Después de comparar los AIC de los cuatro modelos, modelo 1: 1820.58, modelo 2: 1817.26, modelo 3: 1855.03, y, modelo 4: 1859.04, se selecciona el que tiene un menor valor AIC, en este caso particular el modelo 2, el cual cuenta con las covariables tiempo (año) y grupo, con efecto aleatorio del grupo; lo que demuestra que el grupo de tiendas incide en las ventas, es decir, el hecho de que una tienda

pertenezca a un grupo hace que las ventas sean diferentes, así se puede concluir que el grupo ayuda a discriminar las ventas. De igual manera, el tiempo es significativo en las ventas, dado que hay variación en ellas para cada periodo. Por su parte, la interacción entre el grupo y el tiempo presenta una influencia significativa en las ventas.

Los resultados muestran que la matriz de covarianza no se acerca a una matriz identidad y para corregirlo se genera una nueva matriz identidad, la cual no se corrige, puesto que sigue teniendo cercanía a cero. El test de Mauchly nos da evidencia suficiente para rechazar la estructura de tipo H para Σ .

Test de paralelismo de perfiles

$$\begin{cases} H_0 : & (\tau\gamma)_{lj} = 0 \quad \forall l, j \\ H_1 : & \text{Al menos un } (\tau\gamma)_{lj} \neq 0 \end{cases}$$

Se rechaza la hipótesis nula, dado que el resultado es cero.

Test de efecto principal de grupos

$$\begin{cases} H_0 : & \tau_1 = \dots = \tau_q = 0 \\ H_1 : & \text{Al menos un } \tau_l \neq 0 \end{cases}$$

Se rechaza la hipótesis nula, dado que el resultado es cero.

Test de efecto principal de tiempo

$$\begin{cases} H_0 : & \gamma_1 = \dots = \gamma_n = 0 \\ H_1 : & \text{Al menos un } \gamma_j \neq 0 \end{cases}$$

Se rechaza la hipótesis nula, dado que el resultado es cero.

8. DISCUSIÓN DE RESULTADOS

Los cuatro modelos presentan de manera clara un efecto de las variables tiempo y grupo sobre las ventas, sin embargo, como criterio de decisión en torno al modelo que mejor explica el efecto de las covariables en las ventas, se elige aquel que cuente con el menor AIC. El modelo 1, cuenta con AIC= 1820.58, modelo 2, AIC= 1817.26, modelo 3, AIC= 1855.03 ,y, modelo 4, AIC=1859.04.

Es así como el modelo 2 se presenta el menor AIC, se puede comprender entonces que las covariables año y tienda con efecto aleatorio en el grupo inciden de manera significativa en las ventas, es decir que la ciudad en la cual este ubicada la tienda, así como el año analizado pueden generar resultados diferenciales de la tendencia en ventas. Está información resulta de valor para la gran superficie, dado que permite tomar decisiones en torno a la apertura de nuevas tiendas.

En relación con los análisis estadísticos que se han realizado sobre las ventas, se encuentra que los modelos lineales mixtos facilitan el estudio riguroso de variables sobre las ventas, permitiendo determinar el efecto concreto que estas variables pueden tener sobre el valor nominal de la venta en una gran superficie. Como se mencionó previamente, el estudio de Pacheco (2017) presenta el efecto del tiempo, sin embargo queda limitado en relación con la incidencia de la tienda en las ventas, aspecto que fue identificado en el presente estudio.

9. CONCLUSIONES

Se puede comprender la evolución de las ventas del periodo 2013 a 2018 de las tiendas de una gran superficie a partir del modelo lineal mixto de Laird-Ware, puesto que este muestra el efecto y la significancia de las variables tiempo (año), grupo de tienda e interacción grupo de tienda y tiempo, en el periodo estudiado.

El análisis de las tiendas por tendencia de ventas que integran una gran superficie a partir del periodo de tiempo 2013 a 2018 es posible realizarlo a través de análisis clúster k-means, dado que permite clasificar las tiendas observadas en grupos de acuerdo a su valor nominal en ventas y caracterizar los grupos de acuerdo con la ubicación geográfica de la tienda y el número poblacional por ciudad.

Se puede concluir que existen variaciones significativas en las ventas de los años analizados, por lo cual hay un efecto del tiempo en las ventas, es de aclarar que la trayectoria de los grupos 1 y 2 refleja perfiles paralelos en cuanto a la tendencia de la venta.

Por otra parte, pertenecer a un grupo genera efecto en la variabilidad de las ventas, así el grupo 1, el cual está ubicado en ciudades principales presenta el nivel más representativo del valor nominal de las ventas de la gran superficie, para el grupo 2, el cual se ubica en ciudades intermedias se observa un comportamiento de ventas con crecimiento significativo y perfiles constantes desde el primer año observado.

De igual manera se encuentra que el tiempo tiene un efecto en las ventas, es así como se observan cambios en los periodos analizados para cada una de las tiendas. Respecto a la concomitancia entre tiempo y grupo, se encuentra en el modelo que existe un efecto entre estas dos variables sobre las ventas.

De acuerdo con el modelo lineal mixto de Laird-Ware, se observa que el modelo que mejor se explica las ventas de la gran superficie en el periodo 2013 a 2018 es aquel que integra las covariables, tiempo (año) y el efecto aleatorio en el grupo, con un error de 0.05 y un AIC de 1817.26.

10. REFERENCIAS

- Arnau, J., & Bono, R. (2008). Estudios longitudinales modelos de diseño y análisis. *Analysis*, 32–41.
- Castro, N. (2017). *La dictadura de los supermercados*. Ediciones Akal.
- Colombia compra eficiente. (2019). Grandes superficies | Colombia Compra Eficiente. Retrieved May 9, 2019, from <https://www.colombiacompra.gov.co/tienda-virtual-del-estado-colombiano/grandes-superficies>
- DANE. (2017). Proyecciones de población por Departamento Colombia | Datos Abiertos Colombia. Retrieved October 11, 2018, from <https://www.datos.gov.co/Estadisticas-Nacionales/Proyecciones-de-poblacion-por-Departamento-Colombi/5b7v-4tvp/data>
- DANE. (2018). Grandes Almacenes e Hipermercados. Retrieved May 11, 2019, from <https://www.dane.gov.co/index.php/139-espanol/noticias/ultimas-noticias/2370-grandes-almacenes-e-hipermercados>
- El Tiempo. (2018). Las ventas del comercio colombiano en 2018 - Sectores - Economía - ELTIEMPO.COM. Retrieved from <https://www.eltiempo.com/economia/sectores/las-ventas-del-comercio-colombiano-en-2018-316492>
- Fenalco. (2016). *Informe económico y de gestión sector nacional grandes superficies y almacenes de cadena*.
- Gutiérrez, González, Torres, & Gallardo. (1994). *Técnicas de Análisis de datos multivariable: tratamiento computacional*. Granada: Universidad de Granada. Retrieved from <http://www.ugr.es/~gallardo/pdf/cluster-g.pdf>
- León, S. (n.d.). *Modelización estadística de datos longitudinales*. Universidad de Sevilla.
- Marketing, F. F. (n.d.). Gran superficie. Retrieved May 26, 2019, from <https://www.foromarketing.com/diccionario/gran-superficie>
- Montagut, X., Vivas, E., & Montagut, X. (2008). *Supermercados, no gracias: grandes cadenas de distribución: impactos y alternativas*. Barcelona: Icaria Editorial.
- Pacheco, L. (2017). Aplicación de un diseño de medidas repetidas para explicar si el tiempo es un factor determinante en las ventas de productos de consumo masivo, durante el periodo comprendido entre los años 2013 y 2014. *Tesis de Grado*, 6, 5–9.
- Superintendencia de Industria y Comercio. (2012). Estudio económico del sector Retail en Colombia (2010-2012) Estudio, 1–93. Retrieved from http://www.sic.gov.co/recursos_user/documentos/promocion_competencia/Estudios_Economicos/Retail2012.pdf

- Uribe, E. (2013). Propuesta para las grandes superficies. *Model of Quality Management in Customer Service*, 333–354.
- Vega, D. (2012). *Gestión estratégica del departamento de ventas aplicada en una empresa comercial*. Retrieved from http://sisbib.unmsm.edu.pe/bibvirtualdata/tesis/ingenie/vega_sd/cap2.pdf
- Villa Sánchez, A., Arnau, E., Cabezas, C., Cancino, R., Fernández-Lamarra, N., Gresing, C., ... Lopez, A. (2013). Un modelo de evaluación de Innovación Social Universitaria Responsable (ISUR), 430.